Derivatives

1. $\dfrac{d}{dx}\left(x^n\right) = nx^{n-1}$
2. $\dfrac{d}{dx}\left(\ln|x|\right) = \dfrac{1}{x}$
3. $\dfrac{d}{dx}\left(\sin(x)\right) = \cos(x)$
4. $\dfrac{d}{dx}\left(\cos(x)\right) = -\sin(x)$
5. $\dfrac{d}{dx}\left(\tan(x)\right) = \sec^2(x)$
6. $\dfrac{d}{dx}\left(\sec(x)\right) = \sec(x)\tan(x)$
7. $\dfrac{d}{dx}\left(\cot(x)\right) = -\csc^2(x)$
8. $\dfrac{d}{dx}\left(\csc(x)\right) = -\csc(x)\cot(x)$
9. $\dfrac{d}{dx}\left(\arcsin(x)\right) = \dfrac{1}{\sqrt{1-x^2}}$
10. $\dfrac{d}{dx}\left(\arctan(x)\right) = \dfrac{1}{1+x^2}$
11. $\dfrac{d}{dx}\left(\operatorname{arcsec}(x)\right) = \dfrac{1}{|x|\sqrt{x^2-1}}$
12. $\dfrac{d}{dx}\left(e^x\right) = e^x$
13. $\dfrac{d}{dx}\left(a^x\right) = a^x(\ln(a))$
14. $\dfrac{d}{dx}\left(\sinh(x)\right) = \cosh(x)$
15. $\dfrac{d}{dx}\left(\cosh(x)\right) = \sinh(x)$

Antiderivatives

1. $\displaystyle\int x^n\ dx = \frac{1}{n+1}x^{n+1} \qquad n \neq -1$
   $\displaystyle\int \frac{dx}{x} = \ln(x),\ x > 0 \quad \text{or} \quad \ln|x|,\ x \neq 0$
2. $\displaystyle\int e^x\ dx = e^x$
3. $\displaystyle\int \sin(x)\ dx = -\cos(x)$
4. $\displaystyle\int \cos(x)\ dx = \sin(x)$
5. $\displaystyle\int \tan(x)\ dx = \ln|\sec(x)| = -\ln|\cos(x)|$

6. $\int \cot(x)\ dx = \ln|\sin(x)| = -\ln|\csc(x)|$

7. $\int \sec(x)\ dx = \ln|\sec(x)+\tan(x)| = \ln\left|\tan\left(\frac{x}{2}+\frac{\pi}{4}\right)\right|$

8. $\int \csc(x)\ dx = \ln|\csc(x)-\cot(x)| = \ln\left|\tan\left(\frac{x}{2}\right)\right|$

9. $\int \frac{dx}{x^2+a^2} = \frac{1}{a}\arctan\left(\frac{x}{a}\right)$

10. $\int \frac{dx}{\sqrt{a^2-x^2}} = \frac{1}{a}\arcsin\left(\frac{x}{a}\right),\ a>0$

11. $\int \frac{dx}{|x|\sqrt{x^2-a^2}} = \frac{1}{a}\operatorname{arcsec}\left(\frac{x}{a}\right)$

Expressions Containing $ax+b$

12. $\int (ax+b)^n\ dx = \frac{1}{a(n+1)}(ax+b)^{n+1}$

13. $\int \frac{dx}{ax+b} = \frac{1}{a}\ln|ax+b|$

14. $\int \frac{dx}{(ax+b)^2} = \frac{-1}{a(ax+b)}$

15. $\int \frac{x\ dx}{(ax+b)^2} = \frac{b}{a^2(ax+b)} + \frac{1}{a^2}\ln|ax+b|$

16. $\int \frac{dx}{x(ax+b)} = \frac{1}{b}\ln\left|\frac{x}{ax+b}\right|$

17. $\int \frac{dx}{x^2(ax+b)} = \frac{-1}{bx} + \frac{a}{b^2}\ln\left|\frac{ax+b}{x}\right|$

18. $\int \sqrt{ax+b}\ dx = \frac{2}{3a}\sqrt{(ax+b)^3}$

19. $\int x\sqrt{ax+b}\ dx = \frac{2(3ax-2b)}{15a^2}\sqrt{(ax+b)^3}$

20. $\int \frac{dx}{\sqrt{ax+b}} = \frac{2}{a}\sqrt{ax+b}$

21. $\int \frac{\sqrt{ax+b}}{x}\ dx = 2\sqrt{ax+b} + b\int \frac{dx}{x\sqrt{ax+b}}$

22. $\int \frac{dx}{x\sqrt{ax+b}} = \frac{1}{\sqrt{b}}\ln\left|\frac{\sqrt{ax+b}-\sqrt{b}}{\sqrt{ax+b}+\sqrt{b}}\right|,\ b>0$

23. $\int \frac{dx}{x\sqrt{ax+b}} = \frac{2}{\sqrt{-b}}\arctan\sqrt{\frac{ax+b}{-b}},\ b<0$

24. $\int \frac{dx}{x^2\sqrt{ax+b}} = \frac{-\sqrt{ax+b}}{bx} - \frac{a}{2b}\int \frac{dx}{x\sqrt{ax+b}}$

25. $\int \sqrt{\frac{cx+d}{ax+b}}\ dx = \frac{\sqrt{ax+b}\sqrt{cx+d}}{a} + \frac{ad-bc}{2a}\int \frac{dx}{\sqrt{ax+b}\sqrt{cx+d}}$

Expressions Containing ax^2+c, $x^2 \pm p^2$, and p^2-x^2, $p>0$

26. $$\int \frac{dx}{p^2-x^2} = \frac{1}{2p}\ln\left|\frac{p+x}{p-x}\right|$$

27. $$\int \frac{dx}{ax^2+c} = \begin{cases} \frac{1}{\sqrt{ac}}\arctan\left(x\sqrt{\frac{a}{c}}\right) & a>0,\ c>0 \\ \frac{1}{2\sqrt{-ac}}\ln\left|\frac{x\sqrt{a}-\sqrt{-c}}{x\sqrt{a}+\sqrt{-c}}\right| & a>0,\ c<0 \\ \frac{1}{2\sqrt{-ac}}\ln\left|\frac{\sqrt{c}+x\sqrt{-a}}{\sqrt{c}-x\sqrt{-a}}\right| & a<0,\ c>0 \end{cases}$$

28. $$\int \frac{dx}{(ax^2+c)^n} = \frac{1}{2(n-1)c}\frac{x}{(ax^2+c)^{n-1}} + \frac{2n-3}{2(n-1)c}\int\frac{dx}{(ax^2+c)^{n-1}} \qquad n>1$$

29. $$\int x(ax^2+c)^n\ dx = \frac{1}{2a}\frac{(ax^2+c)^{n+1}}{n+1} \qquad n\neq 1$$

30. $$\int \frac{x}{ax^2+c}\ dx = \frac{1}{2a}\ln\left|ax^2+c\right|$$

31. $$\int \sqrt{x^2\pm p^2}\ dx = \frac{1}{2}\left(x\sqrt{x^2\pm p^2} \pm p^2\ln\left|x+\sqrt{x^2\pm p^2}\right|\right)$$

32. $$\int \sqrt{p^2-x^2}\ dx = \frac{1}{2}\left(x\sqrt{p^2-x^2} + p^2\arcsin\left(\frac{x}{p}\right)\right)$$

33. $$\int \frac{dx}{\sqrt{x^2\pm p^2}} = \ln\left|x+\sqrt{x^2\pm p^2}\right|$$

34. $$\int (p^2-x^2)^{3/2}\ dx = \frac{x}{4}\left(p^2-x^2)^{3/2}\right) + \frac{3p^2x}{8}\sqrt{p^2-x^2} + \frac{3p^4}{8}\arcsin\left(\frac{x}{p}\right)$$

Expressions Containing ax^2+bx+c

35. $$\int \frac{dx}{ax^2+bx+c} = \begin{cases} \frac{1}{\sqrt{b^2-4ac}}\ln\left|\frac{2ax+b-\sqrt{b^2-4ac}}{2ax+b+\sqrt{b^2-4ac}}\right| & b^2>4ac \\ \frac{2}{\sqrt{4ac-b^2}}\arctan\left(\frac{2ax+b}{\sqrt{4ac-b^2}}\right) & b^2<4ac \\ \frac{-2}{2ax+b} & b^2=4ac \end{cases}$$

36. $$\int \frac{dx}{(ax^2+bx+c)^{n+1}} = \frac{2ax+b}{n(4ac-b^2)(ax^2+bx+c)^n} + \frac{2(2n-1)a}{n(4ac-b^2)}\int\frac{dx}{(ax^2+bx+c)^n}$$

37. $$\int \frac{x\ dx}{ax^2+bx+c} = \frac{1}{2a}\ln\left|ax^2+bx+c\right| - \frac{b}{2a}\int\frac{dx}{ax^2+bx+c}$$

38. $$\int \frac{dx}{\sqrt{ax^2+bx+c}} = \begin{cases} \frac{1}{\sqrt{a}}\ln\left|2ax+b+2\sqrt{a}\sqrt{ax^2+bx+c}\right| & a>0 \\ \frac{1}{\sqrt{-a}}\arcsin\left(\frac{-2ax-b}{\sqrt{b^2-4ac}}\right) & a<0 \end{cases}$$

39. $$\int \frac{x\ dx}{\sqrt{ax^2+bx+c}} = \frac{\sqrt{ax^2+bx+c}}{a} - \frac{b}{2a}\int\frac{dx}{\sqrt{ax^2+bx+c}}$$

40. $$\int \sqrt{ax^2+bx+c}\ dx = \frac{2ax+b}{4a}\sqrt{ax^2+bx+c} + \frac{4ac-b^2}{8a}\int\frac{dx}{\sqrt{ax^2+bx+c}}$$

Expressions Containing Powers of Trigonometric Functions

41. $\displaystyle\int \sin^2(ax)\ dx = \frac{x}{2} - \frac{\sin(2ax)}{4a}$

42. $\displaystyle\int \sin^3(ax)\ dx = \frac{-1}{a}\cos(ax) + \frac{1}{3a}\cos^3(ax)$

43. $\displaystyle\int \sin^n(ax)\ dx = -\frac{\sin^{(n-1)}(ax)\cos(ax)}{na} + \frac{n-1}{n}\int \sin^{(n-2)}(ax)\ dx$, $n \geq 2$ positive integer

44. $\displaystyle\int \cos^2(ax)\ dx = \frac{x}{2} + \frac{\sin(2ax)}{4a}$

45. $\displaystyle\int \cos^3(ax)\ dx = \frac{1}{a}\sin(ax) - \frac{1}{3a}\sin^3(ax)$

46. $\displaystyle\int \cos^n(ax)\ dx = \frac{\cos^{(n-1)}(ax)\sin(ax)}{na} + \frac{n-1}{n}\int \cos^{(n-2)}(ax)\ dx$, $n \geq$ positive integer

47. $\displaystyle\int \tan^2(ax)\ dx = \frac{1}{a}\tan(ax) - x$

48. $\displaystyle\int \tan^3(ax)\ dx = \frac{1}{2a}\tan^2(ax) + \frac{1}{a}\ln|\cos(ax)|$

49. $\displaystyle\int \tan^n(ax)\ dx = \frac{\tan^{(n-1)}(ax)}{a(n-1)} - \int \tan^{(n-2)}(ax)\ dx$, $n \neq 1$

50. $\displaystyle\int \sec^2(ax)\ dx = \frac{1}{a}\tan(ax)$

51. $\displaystyle\int \sec^3(ax)\ dx = \frac{1}{2a}\sec(ax)\tan(ax) + \frac{1}{2a}\ln|\sec(ax) + \tan(ax)|$

52. $\displaystyle\int \sec^n(ax)\ dx = \frac{\sec^{(n-2)}(ax)\tan(ax)}{a(n-1)} - \frac{n-2}{n-1}\int \sec^{(n-2)}(ax)\ dx$, $n \neq 1$

53. $\displaystyle\int \frac{dx}{1 \pm \sin(ax)} = \mp\frac{1}{a}\tan\left(\frac{\pi}{4} \mp \frac{ax}{2}\right)$

Expressions Containing Algebraic and Trigonometric Functions

54. $\displaystyle\int x\sin(ax)\ dx = \frac{1}{a^2}\sin(ax) - \frac{x}{a}\cos(ax)$

55. $\displaystyle\int x\cos(ax)\ dx = \frac{1}{a^2}\cos(ax) + \frac{x}{a}\sin(ax)$

56. $\displaystyle\int x^n\sin(ax)\ dx = \frac{-1}{a}x^n\cos(ax) + \frac{n}{a}\int x^{n-1}\cos(ax)\ dx \quad n \text{ positive}$

57. $\displaystyle\int x^n\cos(ax)\ dx = \frac{1}{a}x^n\sin(ax) - \frac{n}{a}\int x^{n-1}\sin(ax)\ dx \quad n \text{ positive}$

58. $\displaystyle\int \sin(ax)\cos(bx)\ dx = \frac{-\cos((a-b)x)}{2(a-b)} - \frac{\cos((a+b)x)}{2(a+b))} \quad a^2 \neq b^2$

Expressions Containing Exponential and Logarithmic Functions

59. $\displaystyle\int xe^{ax}\ dx = \frac{1}{a^2}e^{ax}(ax-1)$

60. $\displaystyle\int xb^{ax}\ dx = \frac{1}{a^2}\frac{b^{ax}}{(\ln(b))^2}\left(a\ln(b)x-1\right)$

61. $\displaystyle\int x^n e^{ax}\ dx = \frac{1}{a}x^n e^{ax} - \frac{n}{a}\int x^{n-1}e^{ax}\ dx$

62. $\displaystyle\int e^{ax}\sin(bx)\ dx = \frac{e^{ax}}{a^2+b^2}(a\sin(bx)-b\cos(bx))$

63. $\displaystyle\int e^{ax}\cos(bx)\ dx = \frac{e^{ax}}{a^2+b^2}(a\cos(bx)+b\sin(bx))$

64. $\displaystyle\int \ln(ax)\ dx = x\left(\ln(ax)-1\right)$

65. $\displaystyle\int x^n\ln(ax)\ dx = x^{n+1}\left(\frac{\ln(ax)}{n+1}-\frac{1}{(n+1)^2}\right) \qquad n=0,\ 1,\ 2,\ \ldots$

66. $\displaystyle\int (\ln(ax))^2\ dx = x^2\left((\ln(ax))^2-2\ln(ax)+2\right)$

67. $\displaystyle\int \frac{\ln(ax)}{x}\ dx = \frac{a}{2}\left(\ln(ax)\right)^2$

Expressions Containing Inverse Trigonometric Functions

68. $\displaystyle\int \arcsin(ax)\ dx = x\arcsin(ax)+\frac{1}{a}\sqrt{1-a^2x^2}$

69. $\displaystyle\int \arccos(ax)\ dx = x\arccos(ax)-\frac{1}{a}\sqrt{1-a^2x^2}$

70. $\displaystyle\int \operatorname{arcsec}(ax)\ dx = x\operatorname{arcsec}(ax)-\frac{1}{a}\ln\left|ax+\sqrt{a^2x^2-1}\right|$

71. $\displaystyle\int \operatorname{arccsc}(ax)\ dx = x\operatorname{arccsc}(ax)+\frac{1}{a}\ln\left|ax+\sqrt{a^2x^2-1}\right|$

72. $\displaystyle\int \arctan(ax)\ dx = x\arctan(ax)-\frac{1}{2a}\ln\left(1+a^2x^2\right)$

73. $\displaystyle\int \operatorname{arccot}(ax)\ dx = x\operatorname{arccot}(ax)+\frac{1}{2a}\ln\left(1+a^2x^2\right)$

Some Special Integrals

74. $$\int_0^{\pi/2} \sin^n(x)\ dx = \int_0^{\pi/2} \cos^n(x)\ dx = \begin{cases} \frac{1\cdot3\cdot5\cdot7\cdots(n-1)}{2\cdot4\cdot6\cdot8\cdots(n)}\frac{\pi}{2} & n \text{ even} \\ \frac{2\cdot4\cdot6\cdot8\cdots(n-1)}{1\cdot3\cdot5\cdot7\cdots(n)} & n \text{ odd} \end{cases}$$

75. $$\int_{-\infty}^{\infty} e^{-x^2}\ dx = \sqrt{\pi}$$

Calculus

January 31, 2012

Contents

To the Users of This Text

As we wrote each section of this book, we kept in our minds an image of the student who will be using it. The student will most likely be majoring in a STEM discipline. As such, they will be busy, taking other demanding classes in addition to calculus. The degree programs that require the full three semesters of calculus probably expect their students to have a foundational understanding of vector analysis. That chapter, the last in the book, represents the culmination of the theory and applications within the covers of this book.

That image shaped both the exposition and the exercises in each section.

A section begins with a brief introduction. Then it quickly moves to an informal presentation of the central idea of the section, followed by examples. Formal proofs are given only after the student has a feel for the core of the section.

Those proofs are what hold the course together and serve also as a constant review. For this reason we chose student-friendly proofs, adequately motivated. For instance, instead of the elegant, short proof that absolute convergence of a series implies convergence, we employed a longer, but more revealing proof. We avoid pulling tricks out of thin air; hence our new motivation of the cross product. Where one proof will do, we do not use two. Also, rather than proving the theorem in complete generality, we may treat only a special case, if that case conveys the flow of the general proof, which may be left as an exercise.

As we assembled the exercises we labeled them R (routine), M (medium), and C (challenging), to make sure we had enough of each type. The R- exercises focus on definitions and algorithmic calculations. The M-type require more thought. The C-type either demand a deeper understanding or offer an alternative view of the material.

In order to keep the sections as short as feasible, we concentrated on the mathematics. We avoided the temptation to bring into the text too many applications. Not only would this make the sections too long to be read by a busy student, but we could not do justice to the applications by presenting just little snippets. However, because applications are the reason most students study the subject, each chapter concludes with a thorough treatment of an application in a section called "Calculus is Everywhere" (CIE). Because each CIE stands alone, students and instructors are free to deal with it as they please, depending on time available and interest: skip it, glance at it, browse through it, or read it carefully. The presence of the Calculus is Everywhere sections allowed us to replace exercises that start with a long description of an application and end with a trivial bit of calculus. Our guiding theme is do one thing at a time, whether it's exposition, an example, or an exercise.

As we worked on each section we asked ourselves several questions: Is it the right length? Does it get to the point quickly? Does it focus on just one idea and correspond to one lecture? Are there enough examples? Are there enough exercises, with the right balance of routine to medium to challenging? Only sections that serve mainly as a reference may be longer.

Curvature is treated twice, first in the plane, without vectors, and later, in space, with vectors. We do this for two reasons. First, it provides the student background for appreciating the vector approach. Second, it reduces the vector treatment section to a reasonable length.

Many students will use vector analysis in engineering and physics courses. One of us sat in on a sophomore-level electromagnetic course in order to find out how the concepts were applied and what was expected of the students. That inspired a major revision of that important chapter.

In addition, we reach limits and derivatives as early as possible, and as simply as possible. Also, we introduce the Permanence Property, which asserts that a continuous function that is positive at a number remains positive nearby. This is referred to several times; hence we gave it a name.

The controversy about what to do about $\epsilon - \delta$ proofs will never end. Therefore in our text the instructor is free to choose what to do about such proofs. To make our treatment student-friendly, we broke it into two sections. The first section treats limits at infinity because the diagrams are easier and the concept is more accessible. The second deals with limits at a number. A rigorous proof is given there of the Permanence Property, illustrating the power of the $\epsilon - \delta$ approach to demonstrate something that is not intuitively obvious. Later in the book the rigorous approach appears only in a few exercises, giving the instructor and student an opportunity to reinforce that approach if they so choose.

Throughout the book we include exercises that ask only for computing a derivative or an integral. These exercises are intended to keep those skills sharp. We do not want to assign exercises that explore a new concept the additional responsibility of offering extensive practice in calculations. This is another example of our general principle: do only one thing at a time, and do it clearly.

Another of our objectives is to help our students to develop their mathematical maturity to a level that allows the student to understand the vector analysis in the final chapter. For instance, we often include an exercise which asks the student to state a theorem in their own words without mathematical symbols. We had found while doing some pro-bono tutoring that students do not read theorems carefully, if they read them at all. No wonder they didn't know what to do when a supposedly routine exercise asked them to verify a theorem in a particular case.

We enjoyed writing this book, though the process took much longer than either of us had anticipated. We hope it is an effective tool in your teaching or learning of calculus.

Notes to the Instructor

§1.1 A review and a reference. It gets right to the point. The examples provide background for later work. Exercises 35 to 39 bring in the transcendental functions early.

§1.2 Reinforces the exponential and logarithmic functions early and its summary emphasizes the most difficult functions, logarithms. We save "modeling" for later, abiding by our principle, "one section, one main idea." Exercise 52 asks students to think on their own, to be ready for the last third of the book.

§1.3 Quickly builds all the functions needed. We do this for two reasons: to give the students more time to deal with them and to have them available for examples and exercises.

Following our policy of doing just one thing at a time, we develop limits in Chapter 2, separating them from their application in Chapter 3, which introduces the derivative.

§2.2 Focuses on the basic limits needed in Chapter 3. The binomial theorem is not used because many students are not familiar or comfortable with it.

§2.5 Introduces the Permanence Property, which is used several times in later chapters. Hence, we give it a name.

§2.6 Chapter summaries offer an overall perspective and emphasis not possible in an individual section.

§3.1 Introduces the derivative in the traditional way, by velocity and the tangent line. Because of the earlier development of the key limits, this section can be kept short.

§3.3 By using the Δ-notation, we obtain the derivatives of $f + g$, fg, and f/g without using any "student unfriendly" tricks, such as adding and subtracting $f(x)g(x)$.

§3.4 The rigorous proof of the chain rule is left as an exercise with detailed sketch. That enables the student reading the text to concentrate on learning how to apply the chain rule. After all, the proof with the slight hole of division by zero is valid if $g'(x)$ is not 0.

§3.5 Obtains the derivatives of the inverse functions, using the chain rule. There is no need to wait until implicit differentiation is discussed. That way the chapter can focus on obtaining the differentiation formulas. Exercises 76 and 85 are two of the "Sam and Jane" exercises that add a light touch and invite the students to think on their own.

§3.6 Introduces antiderivatives well before the definite integral appears in Chapter 6, so that the two concepts are adequately separated in time. Slope fields will be used later.

§3.7 Note that the higher derivatives will be put to work as early as Section 5.5, which concerns Taylor polynomials.

§3.8 and 3.9 We delayed the precise definitions of limits in order to give the students more time to work with limits before facing these definitions. These sections are optional. Section 3.8 is easier. One may separate the two sections by several days to let the first one sink in. Note that Example 2 in Section 3.9 shows how useful a precise definition is, as it justifies the Permanence Principle.

§3.9 Emphasizes the essentials and invites more practice in differentiation. Throughout the remaining chapters we include exercises on straightforward differentiation.

Chapter 4 Concentrates on just one theme: using f' and f'' to graph a function. This provides a strong foundation for Chapter 5, which includes optimization.

§5.4 Shows how a higher derivative influences the growth of a function and sets the stage for Section 5.5, Taylor polynomials and their errors. The growth theorem of Section 5.4 is used in exercises in Chapter 6 to obtain the error in approximating a definite integral by the trapezoidal or Simpson's methods.

§5.7 Exercise 39 raises interesting questions about exponential growth.

§6.1 This section keeps to a readable length by avoiding involvement with a formula for the sum $1^2 + 2^2 + \cdots + n^2$.

§6.2 Anticipates the formula $F(b) - F(a)$ for evaluating a definite integral.

§6.5 Exercises such as 44 and 45 are not as hard as one would expect, because the steps are outlined. Such exercises review several important concepts.

Chapter 1

Pre-Calculus Review

This chapter reviews precalculus concepts needed in all subsequent chapters.

Because calculus is the study of functions, Section 1.1 begins with a review of the terminology and notation used when talking about them. In Section 1.2 fundamental types of functions are reviewed: power functions, exponentials, logarithms, and the trigonometric functions. Section 1.3 describes how functions can be combined to create new functions.

The final two sections review two important topics that will be used often, geometric series in Section 1.4 and logarithms in Section 1.5.

1.1 Functions

This section reviews several ideas related to functions: piecewise-defined functions, one-to-one functions, inverse functions, and increasing or decreasing functions.

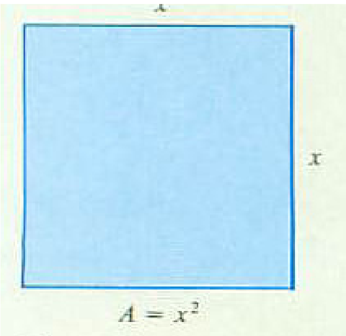


Figure 1.1.1

Definition of a Function

The area A of a square depends on the length of its side x and is given by the formula $A = x^2$. (See Figure 1.1.1.)

Similarly, the distance s (in feet) that a freely falling object drops in the first t seconds is described by the formula $s = 16t^2$. Each choice of t determines a specific value for s. For instance, when $t = 3$ seconds, $s = 16 \cdot 3^2 = 144$ feet.

Both of these formulas illustrate the notion of a function.

> **DEFINITION** *(Function.)* Let X and Y be sets. A **function** from X to Y assigns one (and only one) member in Y to each member in X. A function may assign different values in Y to different members in X.

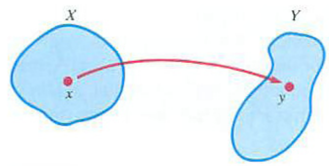


Figure 1.1.2

The notion of a function is illustrated in Figure 1.1.2, where the member y in Y is assigned to the member x in X. Usually X and Y will be sets of numbers.

A function is often denoted by the symbol f. The member that the function assigns to the member x is denoted $f(x)$ (read "f of x"). In practice, though, almost everyone speaks interchangeably of the function f or the function $f(x)$.

If $f(x) = y$, x is called the **input** or **argument** and y is called the **output** or **value** of the function at x. Also, x is called the **independent variable** and y the **dependent variable**.

A function may be given by a formula, as in the function $A = x^2$. Because A depends on x, we say that "A is a function of x." Because A depends on only one number, x, it is called a function of a single variable. The area A of a rectangle depends on its length l and width w; it is a function of *two* variables, $A = lw$.

Intervals

Most of the sets we will be dealing with in calculus are intervals. The following notations are standard.

$[a, b]$ the **closed interval** consisting of all numbers between a and b including both a and b, in short $\{x : a \leq x \leq b\}$.

(a, b) the **open interval** consisting of all numbers between a and b including neither a nor b, in short $\{x : a < x < b\}$.

$[a, b)$ the **half-open interval** consisting of all numbers between a and b including a but not b, in short $\{x : a \leq x < b\}$.

$(a, b]$ the **half-open interval** consisting of all numbers between a and b including b but not a, in short $\{x : a < x \leq b\}$.

$[a, \infty)$ the unbounded interval consisting of all numbers larger than a including the a, in short $\{x : x \geq a\}$.

(a, ∞) the unbounded interval consisting of all numbers larger than a not including the a, in short $\{x : x > a\}$.

$(-\infty, a)$ the unbounded interval consisting of all numbers smaller than a including a, in short $\{x : x \leq a\}$.

$(-\infty, a)$ the unbounded interval consisting of all numbers smaller than a not including a, in short $\{x : x \leq a\}$.

$(-\infty, \infty)$ the set of all numbers, in short $\{x : -\infty < x < \infty\{$.

Ways to write and talk about a function

There are several ways to describe the function that assigns to each argument x the value x^2.You may write $x \mapsto x^2$ (and say "x goes to x^2" or "x is mapped to x^2"). Or you may say simply, "the formula x^2", "the function x^2", or, sometimes, just "x^2." Using this abbreviation, we might say, "How does x^2 behave when x is large?" Some people object to "x^2" because they fear that it might be misinterpreted as the number x^2, with no sense of a general assignment. In practice, the context will make it clear whether x^2 refers to a number or to a function.

A circle is a curve and a **disk** is the flat region inside a

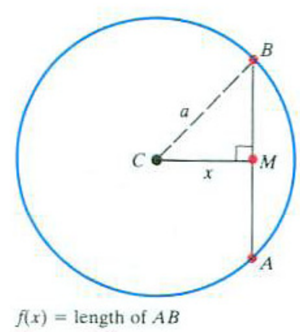


Figure 1.1.3

EXAMPLE 1 In the circle of radius a shown in Figure 1.1.3 let $f(x)$ be the length of chord AB of the circle at a distance x from its center. Find a formula for $f(x)$.

SOLUTION We are trying to find how the length $\overline{AB}$ varies as x varies. That is, we are looking for a formula for $\overline{AB}$, the length of AB, in terms of x.

Before searching for the formula, it is a good idea to calculate $f(x)$ for some easy inputs. They can serve as a check on the formula we work out.

In this case $f(0)$ and $f(a)$ can be read at a glance at Figure 1.1.3: $f(0) = 2a$ and $f(a) = 0$. (Why?) Now let us find $f(x)$ for all x in $[0, a]$.

Let M be the midpoint of the chord AB and let C be the center of the circle. Because $\overline{CM} = x$ and $\overline{CB} = a$, the Pythagorean theorem gives $\overline{BM} = \sqrt{a^2 - x^2}$. Hence $\overline{AB} = 2\sqrt{a^2 - x^2}$. Thus

$$f(x) = 2\sqrt{a^2 - x^2}.$$

Does the formula give the correct values at $x = 0$ and $x = a$? ◇

Domain and Range

The set of permissible inputs and the set of possible outputs of a function are essential parts of the definition of a function. They have special names, which we now introduce.

> **DEFINITION** *(Domain and range)* Let X and Y be sets and let f be a function from X to Y. The set X is called the **domain** of the function. The set of all outputs of the function is called the **range** of the function. (The range is part or all of Y.)

When the function is given by a formula, the domain is usually understood to consist of all the numbers for which the formula is defined.

Try it. What does your calculator do? Some advanced calculators go into "complex number" mode to handle square roots of negative numbers. *No* calculator, however advanced, can permit division by zero.

In Example 1 the domain is the closed interval $[0, a]$ and the range is the closed interval $[0, 2a]$.

When using a calculator you must pay attention to the domain corresponding to a function key or command. If you enter a negative number as x and press the $\sqrt{x}$-key to calculate the square root of x you will get no result. It might display an E for "error" or start flashing. Your error was entering a number not in the domain of the square root function.

You can also get into trouble if you enter 0 and press the $1/x$-key. The domain of $1/x$, the **reciprocal function**, consists of all numbers except 0.

Graph of a Function

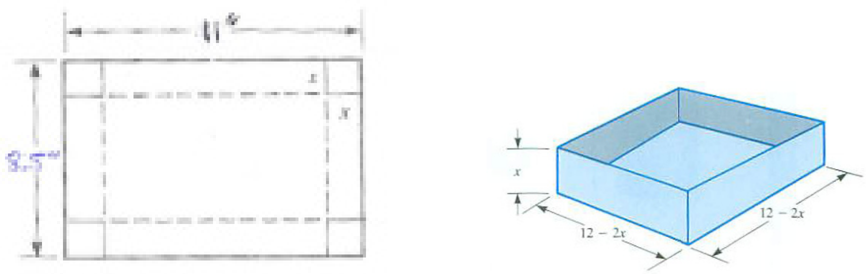
If both the inputs and the outputs of a function are numbers, we can draw a picture of the function, called its **graph**.

> **DEFINITION** *(Graph of a function)* Let f be a function whose inputs and output are numbers. The **graph** of f consists of those points (x, y) in the xy-plane such that $y = f(x)$.

The next example illustrates the usefulness of a graph.

EXAMPLE 2 A tray is to be made from a rectangular piece of paper by cutting congruent squares from each corner and folding up the flaps. The size of the rectangle is $8\frac{1}{2}'' \times 11''$. Find how the volume of the tray depends on the size of the squares.

SOLUTION Let the side of each cut out square be x inches, as shown in Figure 1.1.4(a). The resulting tray is shown in Figure 1.1.4(b).



(a) (b)

Figure 1.1.4 (a) A rectangular sheet with a square cut out from each corner. (b) The tray formed when the sides are folded up.

The volume $V(x)$ of the tray is the height, x, times the area of the base $(11-2x)(8.5-2x)$,

$$V(x) = x(11-2x)(8.5-2x). \tag{1.1.1}$$

The domain of V contains all values of x that lead to an actual tray. This means that x cannot be negative, nor can x cannot be more than half of the shortest side. Thus, the largest corners that can be cut out have sides of length 4.25 inches. So, the domain of interest is only the interval $[0, 4.25]$. The trays obtained when $x=0$ or $x=4.25$ are peculiar. What are their volumes?

Of course we are free to graph (1.1.1) viewed as a polynomial whose domain is $(-\infty, \infty)$.

A short table of inputs and corresponding outputs will help sketch the graph. Figure 1.1.5 displays the graph of $V(x)$.

x (in)	-1	0	1	2	3	4	4.25	5	6
$V(x)$ (in^3)	-136.5	0	5.85	63	37.5	6	0	-7.5	21

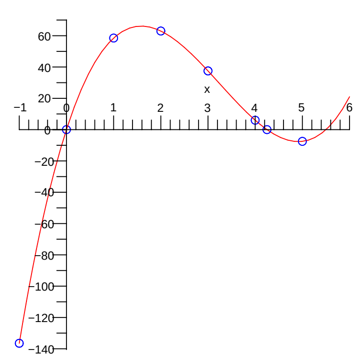


Figure 1.1.5

When $11-2x=0$, that is, when $x = \frac{11}{2} = 5.5$, $V(x)=0$. When x is greater than $\frac{11}{2}$ all three factors in the formula for $V(x)$ are positive, and $V(x)$ becomes very large for large values of x.

For negative x, two factors in (1.1.1) are positive and one is negative. (Which factor is negative?) Thus $V(x)$ is negative and has large absolute value for negative inputs of large absolute value.

Only the part of the graph above the interval $[0, 4.25]$ is meaningful in the tray problem. Other values of x have nothing to do with trays. $\diamond$

To test whether some curve drawn in the xy-plane is the graph of a function, check that each vertical line meets the curve no more than once. If the vertical line $x = a$ meets the curve twice, say at (a, b) and (a, c), there would be the two outputs b and c for the single input a.

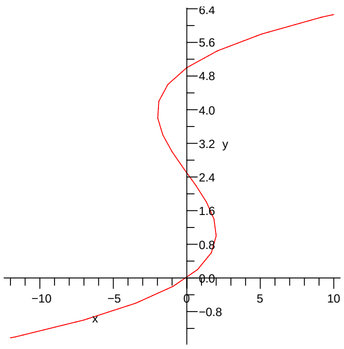


Figure 1.1.6

Vertical Line Test

The input a is in the domain of f if and only if the vertical line $x = a$ intersects the graph of $y = f(x)$ exactly once. Otherwise, a is not in the domain of f.

Figure 1.1.6 shows a graph that does not pass the vertical line test. The corresponding input-output table would have three entries for each input x between -2 and 2, two entries for $x = -2$ and $x = 2$ and exactly one entry for each input $x < -2$ or $x > 2$.

In Example 2 the function is described by a single formula, $V(x) = x(11 - 2x)(8.5 - 2x)$. But a function may be described by different formulas for different intervals or points in its domain, as in the next example.

A **sphere** is a surface and a **ball** is the sphere together with its interior.

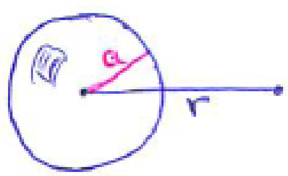


Figure 1.1.7 SHERMAN: Is this (new) figure in your notes?

EXAMPLE 3 A hollow sphere of radius a has mass M, distributed uniformly throughout its surface. Describe the gravitational force it exerts on a particle of mass m at a distance r from the center of the sphere.

SOLUTION Let $f(r)$ be the force at a distance r from the center of the sphere. In an introductory physics course it is shown that the sphere exerts no force at all on objects in its interior. Thus for $0 \le r < a$, $f(r) = 0$.

The sphere attracts an external particle as though all its mass is at its center. Thus, for $r > a$, $f(r) = G\frac{Mm}{r^2}$, where G is a constant, whose value depends on the units used for measuring length, time, mass, and force.

It can be shown by calculus that for a particle on the surface, that is, for $r = a$, the force is $G\frac{Mm}{2a^2}$. The graph of f is shown in Figure 1.1.8. $\diamond$

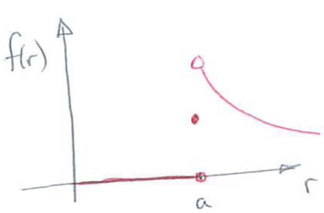


Figure 1.1.8

The formula describing the function in Example 3 changes for different parts of its domain.

$$f(r) = \begin{cases} 0 & \text{if } 0 \le r < a \\ \frac{GMm}{2a^2} & \text{if } r = a \\ \frac{GMm}{r^2} & \text{if } r > a \end{cases}$$

Such a function is called a **piecewise-defined function**.

In a graph that consists of several different pieces, such as Figure 1.1.8, the presence of a point on the graph of a function is indicated by a solid dot ($\bullet$) and the absence of a point by a hollow dot ($\circ$).

Inverse Functions

If you know a particular output of the function $f(x) = x^3$ you can figure out what the input must be. For instance, if $x^3 = 8$, then $x = 2$ – you can go backwards from output to input. This is not possible with the function $f(x) = x^2$. If you are told that $x^2 = 25$, you do not know what x is. It can be 5 or -5. However, if you are told that $x^2 = 25$ and that x is positive, then you know that x is 5.

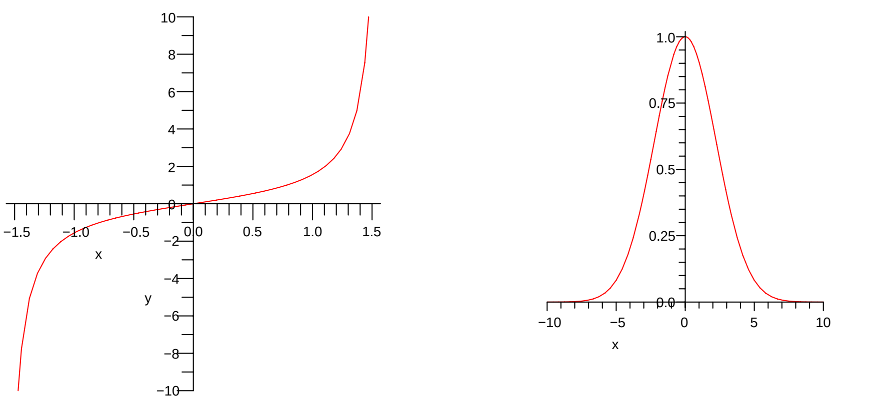
This brings us to the notion of a one-to-one function.

> **DEFINITION** *(One-to-One Function)* A function f that does not assign the same output to two different inputs is **one-to-one**. That is, if $f(x_1) = f(x_2)$, then $x_1 = x_2$.

> **DEFINITION** *(Inverse Function)* If f is a one-to-one function, the **inverse function** is the function g that assigns to each output of f the corresponding input. That is, if $f(x) = y$ then $g(y) = x$.

Horizontal Line Test

The graph of a one-to-one function never meets a horizontal line more than once. (See Figure 1.1.9.)



(a) (b)

Figure 1.1.9 The function in (a) is one-to-one as it passes the horizontal line test. The function in (b) does not pass the horizontal line test, so it is not one-to-one.

The function $f(x) = x^3$ is one-to-one on the entire real line. A few entries in the tables for $f(x)$ and its inverse function are shown in Table 1.1.1(a) and (b), respectively.

input	1	2	$\frac{1}{2}$	3	-2
output	1	8	$\frac{1}{8}$	27	-8

(a)

input	1	8	$\frac{1}{8}$	27	-8
output	1	2	$\frac{1}{2}$	3	-2

(b)

Table 1.1.1 (a) Table of input and output values for $f(x) = x^3$. (b) Table of input and output values for the inverse of $f(x) = x^3$.

In this case an explicit formula for the inverse function can be found algebraically: if $y = x^3$ then $y^{1/3} = (x^3)^{1/3} = x$. Then $x = y^{1/3}$. Since it is customary to use the x-axis for the input and the y-axis for the output, it is convenient to rewrite $x = y^{1/3}$ as $y = x^{1/3}$. (Both say the same thing: "The output is the cube root of the input.")

By the way, an inverse of a one-to-one function may not be given by a nice formula. As will be easily shown in Chapter 4 $f(x) = 2x + \cos(x)$ is one-to-one. However, the inverse function is not described by a convenient formula. Happily, we do not need to deal with an explicit formula for it.

The inverse function of the one-to-one function f is denoted $\mathsf{inv}f$ or f^{-1}.

Notation: The use of $\mathsf{inv}f$ to denote the inverse function of f is based on the fact that many calculators have a button marked inv to indicate the inverse of a function. The mathematical notation for the inverse function of f is f^{-1} or $\mathsf{inv}f$. The -1 is not an exponent, and in general the inverse and reciprocal functions are different: f^{-1} is not equal to $\frac{1}{f}$.

The Graph of an Inverse Function

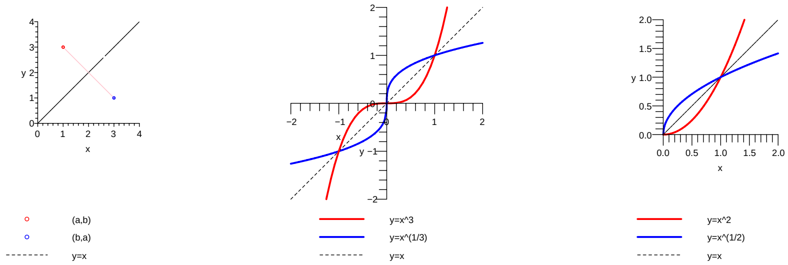
When you know the graph of a one-to-one function, it is easy to draw the graph of the inverse function.

If (a, b) is a point on the graph of the function f, that is, $b = f(a)$, then (b, a) is a point on the graph of $\mathsf{inv}f$, shown in Figure 1.1.10(a).

EXAMPLE 4 Draw the graphs of (a) the inverse of the cubing function given by $f(x) = x^3$, and (b) the squaring function $g(x) = x^2$ restricted to $x \geq 0$.

SOLUTION See Figure 1.1.10(b) and (c). ◇

EXAMPLE 5 Let $m \neq 0$ and b be constants and $f(x) = mx + b$. Show that f is one-to-one and describe its inverse function.



(a) (b) (c)

Figure 1.1.10 (a) The point (b, a) is obtained by reflecting (a, b) around the line $y = x$. (b) Plots of $f(x) = x^3$ and $\mathsf{inv}f(x) = x^{1/3}$. (c) Plots of $g(x) = x^2$ $(x \geq 0)$ and $\mathsf{inv}\ g(x) = \sqrt{x}$.

SOLUTION If $f(x_1) = f(x_2)$ we have

$$\begin{aligned} mx_1 + b &= mx_2 + b \\ mx_1 &= mx_2 \qquad \text{(subtract } b \text{ from both sides)} \\ x_1 &= x_2 \qquad \text{(divide both sides by } m \neq 0\text{)} \end{aligned}$$

Because $f(x_1) = f(x_2)$ only when $x_1 = x_2$, f is one-to-one.

This problem can also be analyzed graphically. The graph of $y = f(x)$ is the line with slope m and y-intercept b. (See Figure 1.1.11.) It passes the horizontal line test.

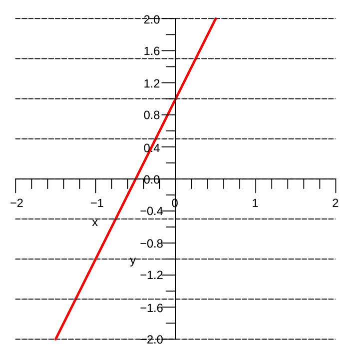


Figure 1.1.11

To find the inverse function, solve the equation $y = f(x)$ to express x in terms of y:

$$\begin{aligned} y &= mx + b \\ y - b &= mx \qquad \text{(subtract } b \text{ from both sides} \\ \tfrac{y-b}{m} &= x \qquad \text{(divide by } m \neq 0 \\ x &= \tfrac{y}{m} - \tfrac{b}{m} \qquad \text{(move } x \text{ to left-hand side)} \\ y &= \tfrac{x}{m} - \tfrac{b}{m}. \qquad \text{(interchange } x \text{ and } y\text{)} \end{aligned}$$

Reversing the roles of x and y in the final step is done only to present the inverse function in a form where the input is called x and the output is called y. The inverse function has the formula

$$f^{-1}(x) = \frac{x}{m} - \frac{b}{m}.$$

The graph of the inverse function is also a line; its slope is $1/m$ the reciprocal of the slope of the original line, and its y-intercept is $-b/m$. (See Figure 1.1.12.)

◇

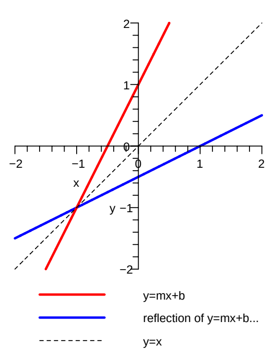


Figure 1.1.12

Decreasing and Increasing Functions

A function is **increasing** on an interval if whenever x_1 and x_2 are in the interval and x_2 is greater than x_1, then $f(x_2)$ is greater than $f(x_1)$. As a pencil moves along the graph of f from left to right, it goes up. This is shown in Figure 1.1.13(a).

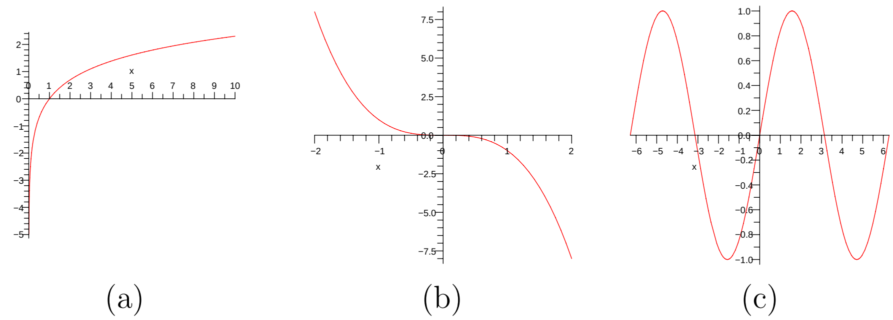


Figure 1.1.13 Graphs of (a) an increasing function, (b) a decreasing function, and (c) a non-monotonic function.

In the case of a **decreasing** function, outputs decrease as the input increases: if $x_2 > x_1$ then $f(x_2) < f(x_1)$. (See Figure 1.1.13(b).)

The graph of $f(x) = \sin(x)$ is shown in Figure 1.1.13(c). On the interval $[-\pi/2, \pi/2]$ the values of $\sin(x)$ increase. On the interval $[\pi/2, 3\pi/2]$ the values of $\sin(x)$ decrease.

A **monotonic** function is either only increasing or only decreasing. It always passes the horizontal line test, as the next example illustrates.

Monotone is from the Greek, *mono*=single, *tonos*=tone, which also gives us the word 'monotonous').

For $k \neq 0$ and $x > 0$, x^k is a monotonic function. For $k < 0$, x^k is monotone decreasing for $x > 0$; for $k > 0$ it is monotone increasing for $x > 0$. The inverse of x^k is $x^{1/k}$. If $k = 0$, we have a constant function, $x^0 = 1$. It does not pass the horizontal line test, so has no inverse.

Because strict inequalities are used in the definitions of increasing and decreasing, we sometimes say these functions are *strictly increasing* or *strictly decreasing* on an interval. A function f is said to be **non-decreasing** on an interval if whenever x_1 and x_2 are in the interval and x_2 is greater than x_1, then $f(x_2) \geq f(x_1)$. The graph of a non-decreasing function is increasing except on intervals where it is constant. Likewise, f is **non-increasing** on an interval if whenever x_1 and x_2 are in the interval and $x_2 > x_1$, then $f(x_2) \leq f(x_1)$.

The sign of a function's outputs provides another way to describe some functions. A function that has only positive outputs is called a **positive function**; for instance, 2^x. A **negative function** has only negative outputs; for instance, $\frac{-1}{1+x^2}$. A **non-negative function** has outputs that are either positive or zero; for instance x^2. The outputs of a **non-positive function** are either negative or zero, for instance, $\sin(x) - 1$.

Summary

This section introduced concepts that will be used throughout the coming chapters: intervals, functions, domains, ranges, graphs, piecewise-defined functions, one-to-one functions, inverse functions, increasing functions, decreasing functions, monotonic functions, non-decreasing functions, non-increasing functions, positive functions, negative functions, non-negative functions, and non-positive functions.

Every monotonic function has an inverse function and the graph of the inverse function is the reflection across the line $y = x$ of the graph of the original function.

A function can be described in several ways: by a formula, such as $V(x) = x(11 - 2x)(8.5 - 2x)$, by a table of values, or by words, such as "the volume of a tray depends on the size of the cut-out squares."

EXERCISES for Section 1.1

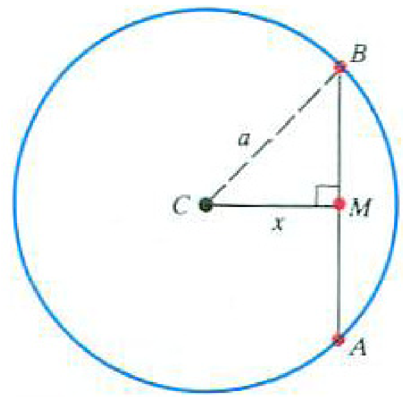


Figure 1.1.14 Exercises 1 to 4.

Exercises 1 to 4 refer to Figure 1.1.14.

1. Express the area of triangle ABC as a function of $x = \overline{CM}$

2. Express the perimeter of triangle ABC as a function of x.

3. Express the area of triangle ABC as a function of θ.

4. Express the perimeter of triangle ABC as a function of θ.

In Example 2 a tray was formed from an $8\frac{1}{2}$" by 11" rectangle by removing squares from the corners. Find and graph the corresponding volume function for trays formed from sheets of sizes given in Exercises 5 to 8.

5. 4" by 13"

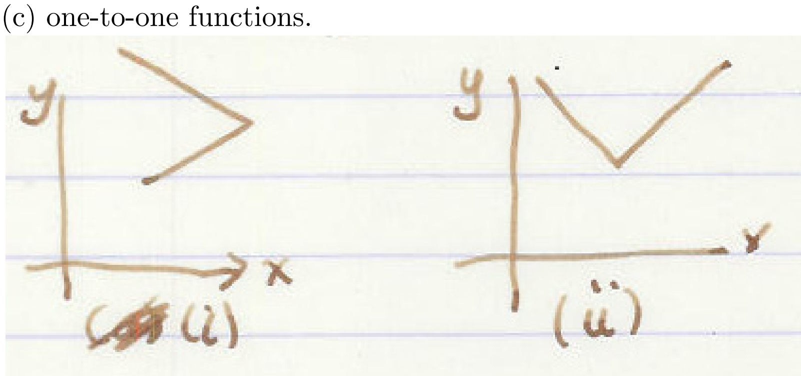
6. 5" by 7"

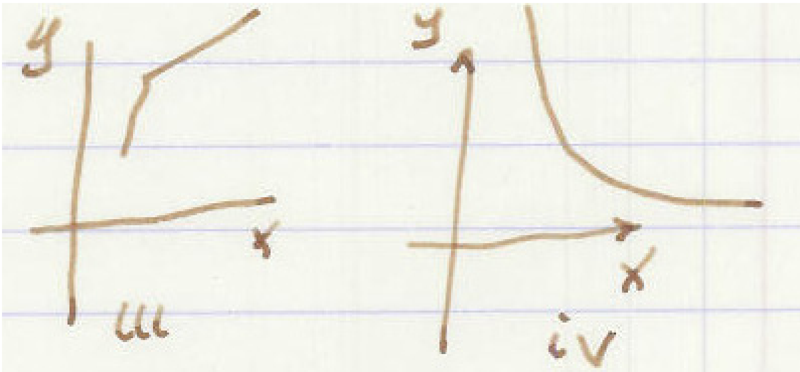
7. 6" by 6"

8. 5" by 5"

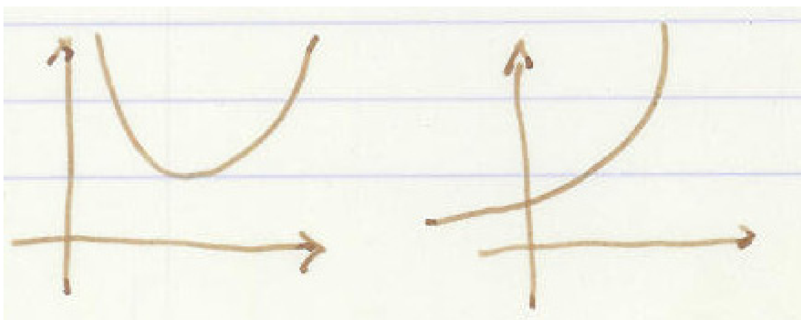
In Exercises 9 and 10 decide which curves are graphs of (a) functions, (b) increasing functions, and (c) one-to-one functions.

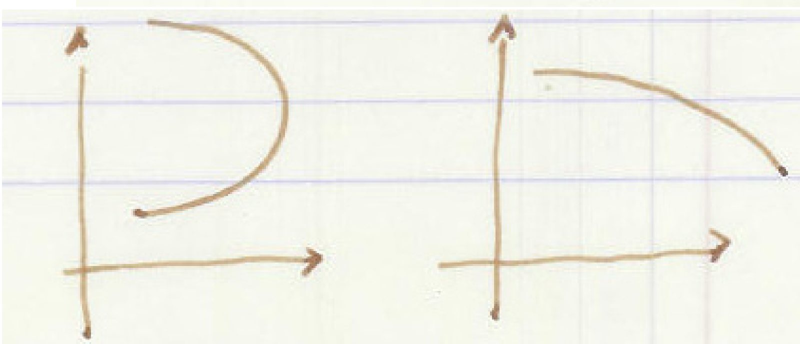
9.





10.





11. Let $f(x) = x^3$.

(a) Fill in this table

x	0	1/4	1/2	$-1/4$	$-1/2$	1	2
x^3							

(b) Graph f.

(c) Use the table in (a) to find seven points on the graph of f^{-1}.

(d) Graph f^{-1} (use the same axes as in (b)).

12. Let $f(x) = \cos(x)$, $0 \le x \le \pi$ (angles in radians).

(a) Fill in this table

x	0	$\pi/6$	$\pi/4$	$2\pi/3$	$\pi/2$	$3\pi/4$	π
$\cos(x)$							

(b) Graph f.

(c) Use the table in (a) to find seven points on the graph of $\mathsf{inv}\cos$.

(d) Graph $\mathsf{inv}\cos$ (use the same axes as in (b)).

In Exercises 13 to 18 the functions are one-to-one. Find the formula for each inverse function, expressed in the form $y = g(x)$, so that the independent variable is labeled x. If you have trouble with the use of logarithms in Exercise 17 or Exercise 18, read Section 1.5.

13. $y = 3x - 2$

14. $y = x/2 + 7$

15. $y = x^5$

16. $y = 3\sqrt{x}$

17. $y = 3^x$

18. $y = 5(2^x)$

In Exercises 19 to 23 the slope of line L is given. Let L' be the reflection of L across the line $y = x$. What is the slope of the reflected line, L'? In each case sketch a possible L and its reflection, L'.

19. L has slope 2.

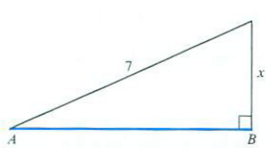
20. L has slope 1.

21. L has slope 1/10.

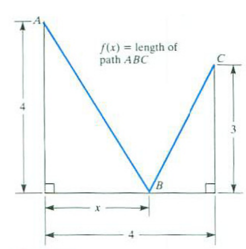
22. L has slope $-1/3$.

23. L has slope -2.

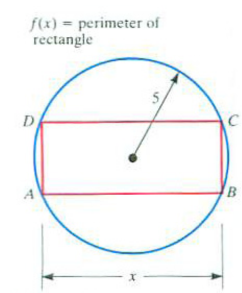
In Exercises 24 to 33 state the formula for the function f and give its domainn.



(a)



(b)



(c)

Figure 1.1.15

24. $f(x)$ is the perimeter of a circle of radius x.

25. $f(x)$ is the area of a disk of radius x.

26. $f(x)$ is the perimeter of a square of side x.

27. $f(x)$ is the volume of a cube of side x.

28. $f(x)$ is the total surface area of a cube of side x.

29. $f(x)$ is the length of the hypotenuse of the right triangle whose legs have lengths 3 and x.

30. $f(x)$ is the length of the side AB in the triangle in Figure 1.1.15(a).

31. For $0 \leq x \leq 4$, $f(x)$ is the length of the path from A to B to C in Figure 1.1.15(b).

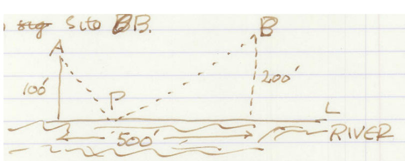
32. For $0 \leq x \leq 10$, $f(x)$ is the perimeter of the rectangle $ABCD$, one side of which has length x, inscribed in the circle of radius 5 shown in Figure 1.1.15(c).

33. A person at point A, two miles from shore in a lake, is going to swim to the shore ST and then walk to point B, five miles from the shore. She swims at 1.5 miles per hour and walks at 4 miles per hour. If she reaches the shore at point P, x miles from S, let $f(x)$ denote the time for her combined swim and walk. Obtain a formula for $f(x)$. (See Figure 1.1.18(a).)

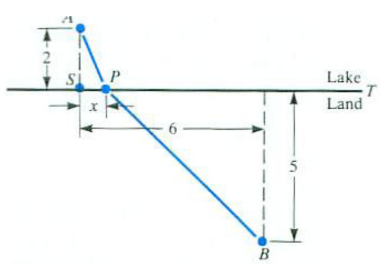
34. A camper at A will walk to the river, put some water in a pail at P, and take it to the campsite at B.

(a) Express the distance $\overline{AP} + \overline{PB}$ as a function of x.

(b) Where should P be located to minimize the length of the walk, $\overline{AP} + \overline{PB}$? (Reflect B across the line L. See Figure 1.1.16.)



(a)



(b)

Figure 1.1.16 Sketches for situations in Exercises 33 and 34.

A geometric trick solved (b). Chapter 4 develops a general procedure for finding the maximum or minimum of a function.

In Exercises 35 to 39 give (a) three functions that satisfy the equation for all positive x and y and (b) one function that does not.

35. $f(x+y) = f(x) + f(y)$

36. $f(x+y) = f(x)f(y)$

37. $f(xy) = f(x) + f(y)$

38. $f(xy) = f(x)f(y)$

39. $f(x) = f(y)$

40. The cost of life insurance depends on whether the person is a smoker or a non-smoker. The following chart lists the annual cost for a male for a million-dollar life insurance policy.

age (yrs)	20	30	40	50	60	70	80
cost for smoker ($)	1150	1164	1944	4344	9864	26500	104600
cost for non-smoker ($)	396	396	600	1490	3684	10900	41600

(A smoker is a person who has used tobacco during the previous three years.)

(a) Plot the data and sketch the graphs on the same axes for both groups of males.

(b) A smoker at age 20 pays as much as a non-smoker of about what age?

(c) A smoker pays about how many times as much as a non-smoker of the same age?

41. Let $f(x)$ be the diameter of the largest circle that fits in a $1 \times x$ rectangle

(a) Graph $y = f(x)$ for $x > 0$.

(b) Give a formula for $f(x)$. (This will be a piecewise-defined function.)

42. If f is an increasing function, what, if anything, can be said about f^{-1}?

43. On a typical summer day in the Sacramento Valley the temperature is at a minimum of $60°$ at 7 A.M. and a maximum of $95°$ at 4 P.M..

(a) Sketch a graph that shows how the temperature may vary during the twenty-four hours from midnight to midnight.

(b) A closed shed with little insulation is in the middle of a treeless field. Sketch a graph that shows how the temperature inside the shed may vary during the same period.

(c) Sketch a graph that shows how the temperature in a well-insulated house may vary. Assume that in the evening all the windows and skylights are opened when the outdoor temperature equals the indoor temperature, and closed in the morning when the two temperatures are again equal.

Use the same set of axes for all three graphs.

44. The monthly average air and water temperatures in Myrtle Beach, SC, are shown in Table 1.1.2.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Air Temp (°)	56	60	68	76	83	88	91	89	85	77	69	60
Water Temp (°)	51	52	57	62	69	77	81	83	80	73	65	55

Table 1.1.2 Source: `http://www.myrtle-beach-resort.com/weather.htm`
Assume, for convenience, that the temperatures in the table are the temperatures on the first day of each month.

(a) Sketch a graph that shows how the water temperature may vary during one calendar year, that is, from January 1 through December 31.

(b) Sketch a graph that shows how the difference between the air and water temperatures may vary during one calendar year. During what month is the temperature difference greatest? least?

(c) During February, the water temperature increases 5° in 28 days so the average daily change is $5/28 \approx 0.1786°/\text{day}$. For each month, estimate the average daily change in the water temperature from one day to the next. During which month is this daily change greatest? least?

(d) Repeat (b) and (c) for the air temperature data.

45. This problem grew out of a question raised by Rebecca Stein-Wexler, the daughter of one of the authors, when cutting cloth for a dress. She wanted to cut out two congruent semicircles from a long strip of fabric 44 inches wide, as shown in Figure 1.1.17. The radius of the semicircles determines d, the length of fabric used, $d = f(r)$.

(a) Draw a picture to show that $f(22) = 44$.

(b) For $0 \le r \le 22$, determine d as a function of r, $d = f(r)$.

(c) For $22 \le r \le 44$, determine d as a function of r, $d = f(r)$.

(d) Obtain an equation expressing r as a function of d.

(e) She had 104 inches of fabric, and guesed that the largest semicircle she could cut has a radius of about 30 inches. Use (c) to see how good her guess is.

(f) Graph f.

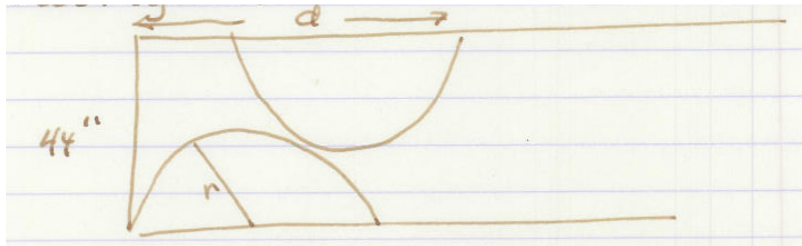


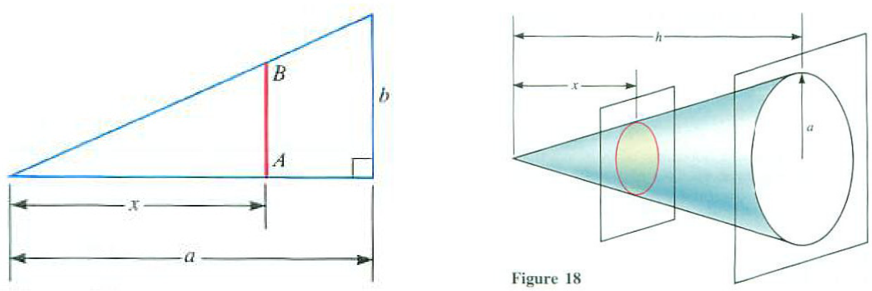
Figure 1.1.17 Exercise 45.

46. Let $f(x)$ be the length of the segment AB in Figure 1.1.18(b).

(a) What are $f(0)$ and $f(a)$?

(b) What is $f(a/2)$?

(c) Find the formula for $f(x)$ and explain your solution.



(a) (b)

Figure 1.1.18

47. Let $f(x)$ be the area of the cross section of a right circular cone shown in Figure 1.1.18(c).

(a) What are $f(0)$ and $f(h)$?

(b) Find a formula for $f(x)$ and explain your solution.

48. This is how the cost of a ride in a New York city taxi is calculated. At the start the meter reads \$2.50. For every fifth of a mile, 40 cents is added. Graph the cost as a function of distance travelled. The cost also depends on other factors.

For every two minutes stopped in traffic, 40 cents is added. During the evening rush, 4–8 p.m., there is a surcharge of one dollar. Between 8 p.m. and 6 a.m. there is a surcharge of 50 cents. So the cost, which depends on distance travelled, time stopped, and time of day, is actually a function of three variables.)

49.

(a) Find all functions of the form $f(x) = a + bx$, where a and b are constants, such that $f = \mathsf{inv}(f)$.

(b) Sketch the graph of one of the functions found in (a).

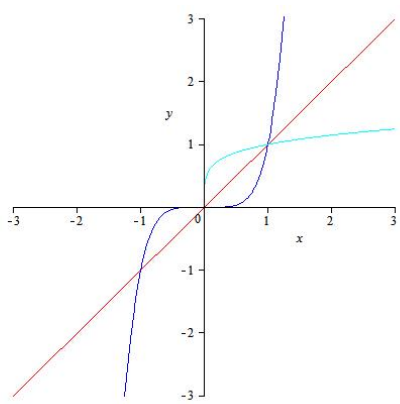
1.2 The Basic Functions in Calculus

This section describes the basic functions in calculus. In the next section they are used build more complicated functions.

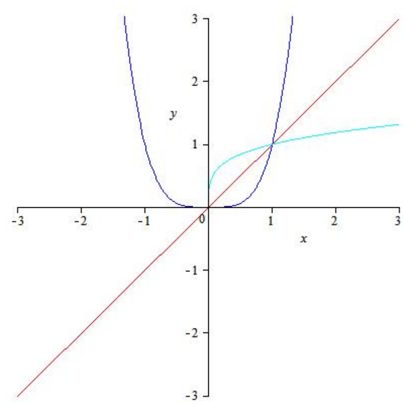
The Power Functions

The first group of functions consists of the **power functions** x^k where the exponent k is a fixed non-zero number and the base x is the input. When the domain of x^k includes all positive numbers, it is one-to-one, and has an inverse, $x^{1/k}$, with, again, a domain consisting of all positive numbers.

In Section 1.1 it was shown that the inverse of $f(x) = x^3$ is $f^{-1}(x) = x^{1/3}$ for all x. However, $g(x) = x^4$ does not pass the horizontal line test unless the domain is restricted to, say, nonnegative inputs ($[0, \infty)$). Thus, the inverse of $g(x) = x^4$ is $g^{-1}(x) = x^{1/4}$ only for $x \geq 0$.



(a)



(b)

Figure 1.2.1 Graphs of power functions. (a) x^k for $k = 1$ (red), 5 (blue), and 1/5 (cyan), (b) x^k for $k = 1$ (red), 4 (blue), and 1/4 (cyan), The pairs of blue and cyan graphs are inverses in both (a) and (b). But in (b), the inverse of x^4 exists only for $x \geq 0$

An important property of power functions is that their inverse functions are also power functions. When the exponent k is an even integer or a rational number (in lowest terms) whose numerator is even (2/3, 4/7, etc.) the graph of $y = x^k$ does not pass the horizontal line test unless the domain is reduced, usually to $[0, \infty)$. And, when $k = 0$ we obtain the function x^0, which is constant (with all outputs equal to 1), the very opposite of being one-to-one.

The Exponential and Logarithm Functions

Next we have the **exponential functions** b^x where the base b is fixed and the exponent x is the input. The inverses of exponential functions are not exponential functions. The inverses are called **logarithms** and are the next class of functions that we will consider.

To be specific, lets look at the case $b = 2$ and study $f(x) = 2^x$.

As x increases, so does 2^x. So the function 2^x has an inverse function. (See Figure 1.2.2.) In other words, if $y = 2^x$, then if we know the output y we can determine the input x, the exponent, uniquely. For instance, if $2^x = 8$ then $x = 3$. This is expressed as $3 = \log_2 8$ and is read as "the logarithm of 8, base 2, is 3." If $y = b^x$, then we write $x = \log_b y$.

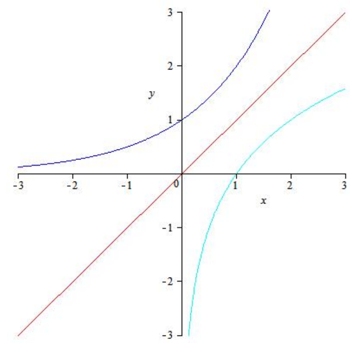


Figure 1.2.2 Graph of $y = 2^x$ and its inverse.

Since we usually denote the independent variable (the input or argument) by x, and the dependent variable (the output, or value) by y, we will rewrite this as $y = \log_b(x)$.

The table of values of $\log_2(x)$ in Table 1.2.1 helps us graph $y = \log_2(x)$. Putting a smooth curve through the seven points in Table 1.2.1 yields the graph in Figure 1.2.3.

x	1	2	4	8	1/2	1/4	1/8
$\log_2(x)$	0	1	2	3	-1	-2	-3

Table 1.2.1 Table of values of $y = \log_2(x)$.

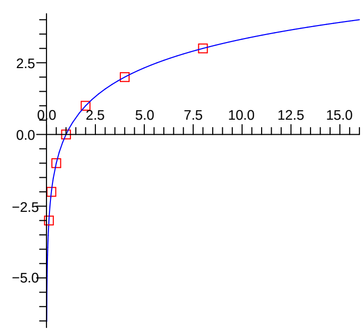


Figure 1.2.3 Plot of $y = \log_2(x)$ based on data in Table 1.2.1.

As x increases, $\log_2(x)$ grows slowly. For instance $\log_2 1024 = 10$, as every computer scientist knows. For x between 0 and 1, $\log_2(x)$ is negative. As x moves from 1 towards 0, $|\log_2(x)|$ grows very large. For instance, $\log_2 \frac{1}{1024} = -10$.

Because $\log_2(x)$ is the inverse of the function 2^x, we can sketch the graph of $y = \log_2(x)$ by first sketching the graph of $y = 2^x$ and reflecting it around the line $y = x$.

For any positive base b, $\log_b(x)$ is defined similarly. For x and b both positive numbers, the logarithm of x to the base b, denoted $\log_b(x)$, is the power to which we must raise b to obtain x. By the definition of the logarithm

$$b^{\log_b(x)} = x.$$

The Trigonometric Functions and Their Inverses

So far we have the power functions, x^k, the exponential functions, b^x, and the logarithm functions, $\log_b(x)$. The last major group of important functions consists of the trigonometric functions, $\sin(x)$, $\cos(x)$, $\tan(x)$, and their inverses (after we shrink their domains to make them one-to-one).

In calculus we generally measure angles in radians.

The trigonometric functions are periodic. A function f is **periodic** if there is a non-zero constant k such that $f(x + k) = f(x)$ for x and $x + k$ in the domain of f. Note that if k is a period, so are $2k$, $3k$, ... and $-k$, $-2k$, $-3k$, ... The smallest positive period is often singled out as "the period" of f. For example, $\sin(x)$ is periodic with period 2π.

$\sin(x)$ and its inverse

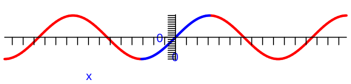


Figure 1.2.4

The graph of the **sine function** $\sin(x)$ has period 2π and is shown in Figure 1.2.4. Its range is $[-1, 1]$. On the domain $[-\pi/2, \pi/2]$, $\sin(x)$ is increasing and its values for these inputs already sweep out the full range.

When we restrict the domain of the function $\sin(x)$ to $[-\pi/2, \pi/2]$ it is one-to-one with range $[-1, 1]$. This means the sine function has an inverse with domain $[-1, 1]$ and range $[-\pi/2, \pi/2]$. The inverse sine function is denoted by $\arcsin(x)$, $\sin^{-1}(x)$, or $\mathsf{inv}\sin(x)$.

Let us stop to summarize our findings: For x in $[-1, 1]$, $\arcsin(x)$ is the angle in $[-\pi/2, \pi/2]$ whose sine is x. In equations:

$$y = \arcsin(x) \iff \sin(y) = x.$$

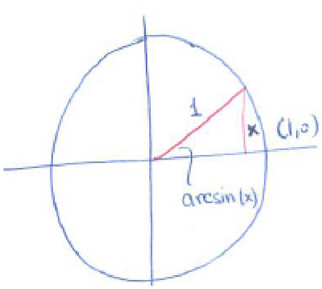
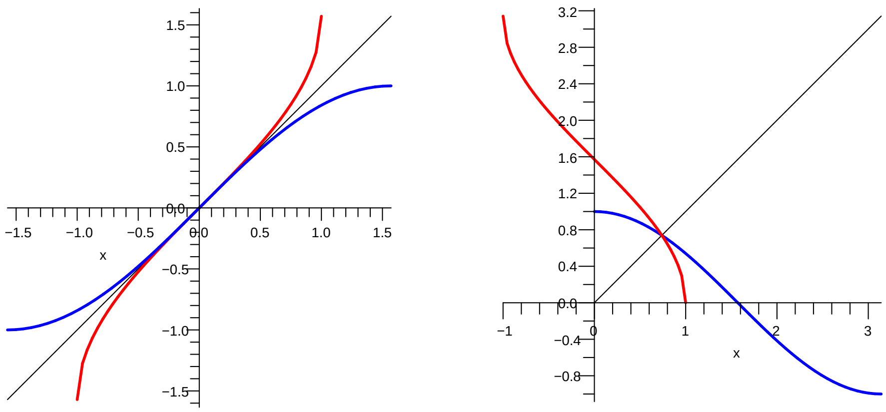


Figure 1.2.5

For instance, $\arcsin(1) = \pi/2$ because the angle in $[-\pi/2, \pi/2]$ whose sine is 1 is $\pi/2$. Similarly, $\sin^{-1}(1/2) = \pi/6$, $\mathsf{inv}\sin(0) = 0$, $\arcsin(-1/2) = -\pi/6$, and $\sin^{-1}(-1) = -\pi/2$. A unit circle helps display these facts, as Figure 1.2.5 illustrates.

We could graph $y = \arcsin(x)$ with the aid of those five values. However, it is easier to reflect the graph of $y = \sin(x)$ around the line $y = x$. (See Figure 1.2.6(a).)



(a) (b)

Figure 1.2.6 (a) The graph of $y = \arcsin(x)$ (red) is the graph of $y = \sin(x)$ (blue), with domain restricted to $[-\pi/2, \pi/2]$, reflected around the line $y = x$. (b) The graph of $y = \arccos(x)$ (red) is the graph of $y = \cos(x)$ (blue), with domain restricted to $[0, \pi]$, around the line $y = x$.

$\cos(x)$ **and its inverse**

The graph of the **cosine function** $\cos(x)$ is shown in Figure 1.2.7.

It is not one-to-one, even if we restrict the domain to the domain used for $\sin(x)$, namely $[-\pi/2, \pi/2]$. Because $\cos(x)$ is decreasing on $[0, \pi]$. It is one-to-one on $[0, \pi]$. Moreover, the values of $\cos(x)$ for x in $[0, \pi]$ sweep out all possible values of the cosine function, namely $[-1, 1]$.

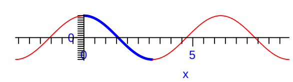


Figure 1.2.7

Because $\cos(x)$ is a one-to-one function on the domain $[0, \pi]$, it has an inverse function, called $\arccos(x)$, $\mathsf{inv}\cos(x)$, or $\cos^{-1}(x)$. Each of these is short for "the angle in $[0, \pi]$ whose cosine is x". For instance, $\arccos(0) = \pi/2$, $\cos^{-1}(1) = 0$, and $\mathsf{inv}\cos(-1) = \pi$. Moreover, because the range of the cosine function is $[-1, 1]$, the domain of arccos is $[-1, 1]$. Figure 1.2.6(b) shows that the graph of $\arccos(x)$ is obtained by reflecting the graph of $\cos(x)$, with domain $[0, \pi]$, around the line $y = x$.

$\tan(x)$ **and its inverse**

The range of the function $\tan(x) = \frac{\sin(x)}{\cos(x)}$ is $(-\infty, \infty)$. (See Figure 1.2.8.)

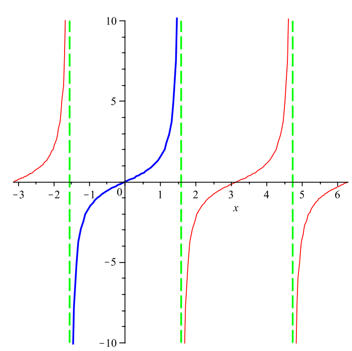


Figure 1.2.8

When the inputs are restricted to $(-\pi/2, \pi/2)$, $\tan(x)$ is one-to-one, and therefore has an inverse function, denoted $\arctan(x)$, $\tan^{-1}(x)$, or $\mathsf{inv}\tan(x)$. The domain of the inverse tangent function is $(-\infty, \infty)$ and its range is $(-\pi/2, \pi/2)$.

For instance, $\tan^{-1}(0) = 0$, $\mathsf{inv}\tan(1) = \pi/4$, and as x increases, $\arctan(x)$

approaches $\pi/2$. Also, $\arctan(-1) = -\pi/4$, and when x is negative and becomes ever more negative (that is, $|x|$ becomes bigger and bigger) $\arctan(x)$ approaches $-\pi/2$. Figure 1.2.9 is the graph of $\arctan(x)$. It is the reflection of the blue part of the graph in Figure 1.2.8 across the line $y = x$. (See Figure 1.2.9.)

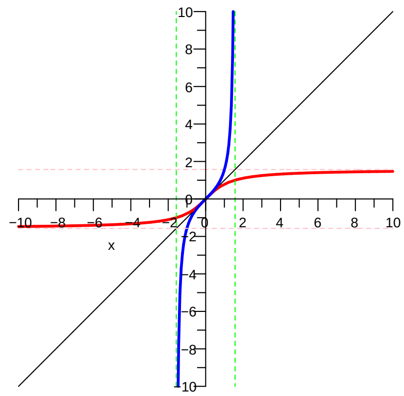


Figure 1.2.9 The graphs of $y = \tan(x)$ and $y = \arctan(x)$

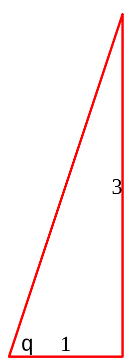


Figure 1.2.10 The traditional symbol for angles is the Greek letter θ (pronounced "theta").

EXAMPLE 1 Evaluate

$$\text{(a) } \sin(\sin^{-1}(0.3)), \quad \text{(b) } \sin(\tan^{-1}(3)), \quad \text{and (c) } \tan(\cos^{-1}(0.4)).$$

SOLUTION

(a) The expression $\sin^{-1}(0.3)$ is short for the angle in the interval $[-\pi/2, \pi/2]$ whose sine is 0.3. So, the sine of $\sin^{-1}(0.3)$ is 0.3.

(b) To find $\sin(\tan^{-1}(3))$, first draw the angle θ whose tangent is 3 (and lies in the interval $(-\pi/2, \pi/2)$. Figure 1.2.10 shows a simple way to draw this angle. To find the sine of θ, recall that sine equals "$\frac{\text{opposite}}{\text{hypotenuse}}$." By the Pythagorean Theorem, the hypotenuse is $\sqrt{3^2 + 1^2} = \sqrt{10}$. Thus, $\sin(\tan^{-1} 3) = 3/\sqrt{10}$.

(c) To evaluate $\tan(\cos^{-1}(0.4))$, first draw an angle whose cosine is $0.4 = \frac{2}{5}$, as in Figure 1.2.11, which is based on the fact that cosine equals "$\frac{\text{adjacent}}{\text{hypotenuse}}$." To find the tangent of this angle, we need the length of the other leg in Figure 1.2.11. By the Pythagorean Theorem it is $\sqrt{5^2 - 2^2} = \sqrt{21}$.

From the relation $\tan(\theta) = \text{opposite}/\text{adjacent}$, we conclude that

$$\tan(\cos^{-1}(0.4)) = \sqrt{21}/2 \approx 2.291.$$

◇

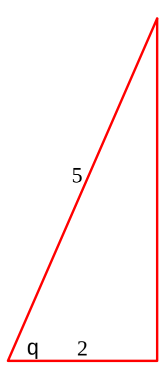


Figure 1.2.11

$\csc(x)$, $\sec(x)$, and $\cot(x)$ and their inverses

The **cosecant**, **secant**, and **cotangent** functions are defined in terms of the sine and cosine functions:

$$\csc(x) = \frac{1}{\sin(x)}, \quad \sec(x) = \frac{1}{\cos(x)}, \quad \text{and} \quad \cot(x) = \frac{\cos(x)}{\sin(x)}.$$

Each is defined only when the denominator is not zero. Figure 1.2.12 shows their graphs.

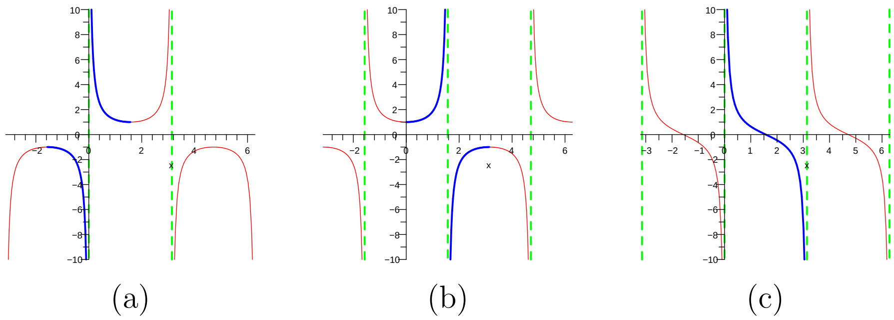


Figure 1.2.12 The graphs of (a) the cosecant, (b) the secant, and (c) the cotangent functions. The restrictions to an interval where the funciton is one-to-one is shown in bold.

Note that $|\sec(x)| \geq 1$ and $|\csc(x)| \geq 1$. In each case the range consists of two separate intervals: $[1, \infty)$ and $(-\infty, -1]$.

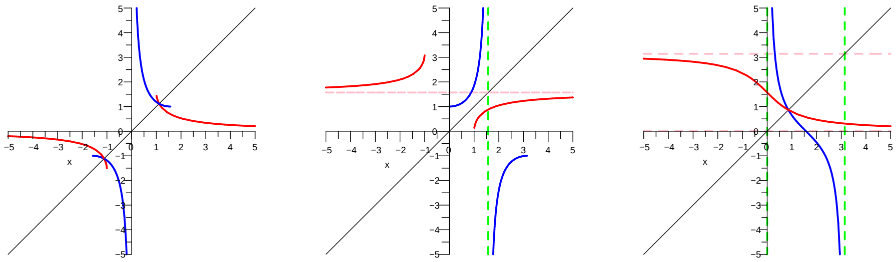
These three functions have inverses when restricted to appropriate intervals. Table 1.2.2 contains a summary of the three inverse functions, $\operatorname{arccsc}(x)$, $\operatorname{arcsec}(x)$, and $\arctan(x)$. Figure 1.2.13 shows the graphs of (a) csc, (b) sec, and (c) cot and their inverses.

function	domain (input)	range (output)
$\operatorname{arccsc}(x)$	$(-\infty, -1]$ and $[1, \infty)$	all of $[-\pi/2, \pi/2]$ except 0
$\operatorname{arcsec}(x)(x)$	$(-\infty, -1]$ and $[1, \infty)$	all of $[0, \pi]$ except 0, that is $(0, \pi]$
$\operatorname{arccot}(x)x$	$(-\infty, \infty)$	the open interval $(0, \pi)$

Table 1.2.2 Summary of the inverse cosecant, inverse secant, and inverse cotangent functions.

Summary

This section reviewed the basic functions in calculus, x^k, b^x, $\sin(x)$, $\cos(x)$, $\tan(x)$, and their inverses. $\log_b(x)$, $\arcsin(x)$, $\arccos(x)$, and $\arctan(x)$. (The inverse of x^k, $k \neq 0$, is just another power function $x^{1/k}$).



(a) (b) (c)

Figure 1.2.13 Graphs of (a) $y = \csc(x)$ and $y = \operatorname{arccsc}(x)$, (b) $y = \sec(x)$ and $y = \operatorname{arcsec}(x)$, and (c) $y = \cot(x)$ and $y = \operatorname{arccot}(x)$. The inverse function is the reflection of the original function across the line $y = x$.

The functions that may be hardest to have a feel for are the logarithms. Now, $\log_2(x)$ is typical of $\log_b(x)$, $b > 1$. These are its key features:

- Its graph crosses the x-axis at $(1, 0)$ because $\log_2(1) = 0$ $(2^0 = 1)$,
- It is defined only for positive inputs, that is, the domain of $\log_2$ is $(0, \infty)$, because only positive numbers can be expressed in the form 2^x,
- It is an increasing function,
- It grows very slowly as the argument increases: $\log_2(8) = 3$, $\log_2(16) = 4$, $\log_2(32) = 5$, $\log_2(64) = 6$, and $\log_2(1024) = 10$,
- For values of x in $(0, 1)$, $\log_2(x)$ is negative ($2^x < 1$ then x is negative),
- For x near 0 (and positive), $|\log_2(x)|$ is large.

The case when the base b is less than 1 is treated in Exercise 54.

EXERCISES for Section 1.2

1. Graph the power function $x^{3/2}$, $x \geq 0$, and its inverse.

2. Graph the power function $\sqrt{x}$ and its inverse.

3. What is the period of $\tan(x)$?

4. What is the period of $\sin(x) + \cos(2x)$?

5. Explain your calculator's response when you try to calculate $\log_{10}(-3)$.

6. Explain your calculator's response when you try to calculate $\arcsin(2)$.

7.

(a) Graph 2^x and $(1/2)^x$ on the same axes.

(b) How could the second graph be obtained from the first?

8.

(a) Graph 3^x and $(1/3)^x$ on the same axes.

(b) How could the second graph be obtained from the first?

9. For any base b, $b^0 = 1$. What is the corresponding property of logarithms? Explain.

10. For any base b, $b^{x+y} = b^x b^y$. What is the corresponding property of logarithms? Explain. If you have trouble with this exercise, study Section 1.5.

11. Explain why $\log_b(1/x) = -\log_b(x)$. ("The log of the reciprocal of x is the negative of the log of x.")

12. Explain why $\log_b(c^x) = x\log_b(c)$. ("The log of a number raised to a power x is x times the log of the number.")

13.

(a) Evaluate $\log_2(x)$ and $\log_4(x)$ at $x = 1$, 2, 4, 8, 16, and 1/16.

(b) Graph $\log_2(x)$ and $\log_4(x)$ on the same axes (clearly label each point found in (a)).

(c) Compute $\frac{\log_4(x)}{\log_2(x)}$ for the six values of x in (a).

(d) Explain the phenomenon observed in (c).

(e) How would the graph of $y = \log_4(x)$ be obtained from that for $y = \log_2(x)$?

14.

(a) Evaluate $\log_2(x)$ and $\log_8(x)$ at x=1, 2, 4, 8, 16, and 1/8.

(b) Graph $\log_2(x)$ and $\log_8(x)$ on the same axes. Clearly label each point found in (a)

(c) Compute $\frac{\log_8(x)}{\log_2(x)}$ for the six values of x in (a).

(d) Explain the phenomenon observed in (c).

(e) How would you obtain the graph of $\log_8(x)$ from that for $\log_2(x)$?

15. Evaluate

(a) $\log_{10}(1000)$

(b) $\log_{100}(10)$

(c) $\log_{10}(0.01)$

(d) $\log_{10}(\sqrt{10})$

(e) $\log_{10}(10)$

16. Evaluate

(a) $\log_3(3^{17})$

(b) $\log_3(1/9)$

(c) $\log_3(1)$

(d) $\log_3(\sqrt{3})$

(e) $\log_3(81)$

17. Evaluate $5^{\log_5(17)}$.

18. Evaluate $3^{-\log_3(21)}$.

19. For positive x near 0, what happens to the functions 2^x, x^2, and $\log_2(x)$?

20. For very large values of x what happens to the quotent $x^2/2^x$? Illustrate by using specific values for x.

21. What happens to $(\log_2(x))/x$ for large values of x? Illustrate by citing specific x.

22. Draw the graphs of $\cos(x)$ for x in $[0, \pi]$, and $\arccos(x)$ on the same axes.
23. Draw the graphs of $\tan(x)$ for x in $(-\pi/2, \pi/2)$, and $\arctan(x)$ on the same axes.

24. Which of these equations is correct?

(a) $\csc(x) = \sin^{-1}(x)$

(b) $\csc(x) = (\sin(x))^{-1}$

(c) $\csc(x) = \mathsf{inv}\sin(x)$

(d) $\csc(x) = 1/\sin(x)$

In Exercises 25 to 41 evaluate the given expressions.
25.

(a) $\sin^{-1}(1/2)$

(b) $\arcsin(1)$

(c) $\mathsf{inv}\sin(-\sqrt{3}/2)$

(d) $\arcsin(\sqrt{2}/2)$

26.

(a) $\cos^{-1}(0)$

(b) $\mathsf{inv}\cos(-1)$

(c) $\arccos(1/2)$

(d) $\arccos(-1/\sqrt{2})$

27.

(a) $\arctan(1)$

(b) $\mathsf{inv}\tan(-1)$

(c) $\arctan(\sqrt{3})$

(d) $\arctan(1000)$ (approximately)

28.

(a) $\operatorname{arcsec}(2)$

(b) $\mathsf{inv}\sec(-2)$

(c) $\operatorname{arcsec}(\sqrt{2})$

(d) $\sec^{-1}(1000)$ (approximately)

29.

(a) $\arcsin(0.3)$

(b) $\arccos(0.3)$

(c) $\arctan(0.3)$

(d) $\frac{\arcsin(0.3)}{\arccos(0.3)}$

Observe that (c) and (d) are different.

30. $\sin(\tan^{-1}(2))$
31. $\sin(\cos^{-1}(0.4))$
32. $\tan(\tan^{-1}(3))$
33. $\tan(\sin^{-1}(0.7))$

34. $\tan(\sec^{-1}(3))$
35. $\sec(\tan^{-1}(0.3))$
36. $\sin(\sec^{-1}(5))$
37. $\sec(\cos^{-1}(0.2))$

38. $\arctan(\tan(\frac{\pi}{3}))$
39. $\arcsin(\sin(\frac{-3\pi}{4}))$
40. $\arccos(\cos(\frac{5\pi}{2}))$
41. $\operatorname{arcsec}(\sec(\frac{-\pi}{3}))$

42. Let k be a period of a function f. Show that $2k$ and $-k$ are also periods of f.

In Exercises 43 to 46, use properties of logarithms to express $\log_{10} f(x)$ as simply as possible.

43. $f(x) = \frac{(\cos(x))^7\sqrt{(x^2+5)^3}}{4+(\tan(x))^2}$

44. $f(x) = \sqrt{(1+x^2)^5(3+x)^4\sqrt{1+2x}}$

45. $f(x) = (x\sqrt{2+\cos(x)})^{x^2}$

46. $f(x) = \sqrt{\frac{x(1+x)}{\sqrt{1+2x^3}}}$

47. Imagine that your calculator fell on the floor and its multiplication and division keys stopped working. However, all the other keys, including the trigonometric, arithmetic, logarithmic, and exponential keys, still functioned. Show how you could use it to calculate the product and quotient of two positive numbers, a and b.

48. (Richter Scale) In 1989, San Francisco and vicinity was struck by an earthquake that measured 7.1 on the **Richter scale**. The strongest earthquake in recent years had a Richter measure of 9.0 (Fukushima, Japan in 2011). There have also been two earthquakes with Richter measures of 8.9 (Colombia-Equador in 1906 and Japan in 1933). A "major earthquake" typically has a measure of at least 7.5.
In his *Introduction to the Theory of Seismology*, Cambridge, 1965, pp. 271–272, K. E. Bullen explains the Richter scale as follows:
"Gutenburg and Richter sought to connect the magnitude M with the energy E of an earthquake by the formula

$$aM = \log_{10}\left(\frac{E}{E_0}\right)$$

and after several revisions arrived in 1956 at the result $a = 1.5$, $E_0 = 2.5\times 10^{11}$ ergs." Energy E is measured in ergs. M is the number assigned to the earthquake on the Richter scale. E_0 is the energy of the smallest instrumentally recorded earthquake.

(a) Deduce that $\log_{10}(E) \approx 11.4 + 1.5M$.

(b) What is the ratio between the energy of the earthquake that struck Japan in 1933 ($M = 8.9$) and the San Francisco earthquake of 1989 ($M = 7.1$)?

(c) What is the ratio between the energy of the San Francisco earthquake of 1906 ($M = 8.3$) and that of the San Francisco earthquake of 1989 ($M = 7.1$)?

(d) Find a formula for E in terms of M.

(e) If one earthquake has a Richter measure 1 larger than that of another earthquake, what is the ratio of their energies?

(f) What is the Richter measure of a 10-megaton H-bomb, that is, of an H-bomb whose energy is equivalent to that of 10 millon tons of TNT?

(One ton of TNT releases an energy of 4.2×10^6 ergs.)

49. Translate the sentence, "She has a five-figure annual income" into logarithms. How small can the income be? How large?

50. As of 2011 the largest known prime was $2^{43112609} - 1$.

(a) When written in decimal notation, how many digits will it have?

(b) How many pages of this book would be needed to print it? (One page can hold 50 70-character lines, a total of 3,500 digits per page).)

There is a prize of \$250,000 for the discovery of the first billion-digit prime. A web search for "largest prime" will find the largest known prime.

51. Newton computed the logarithms of 0.8, 0.9, 1.1, and 1.2 to 57 decimal places by hand using a method developed in Section 10.4. Show how to compute

(a) $\log(2)$, using $\log(1.2)$, $\log(0.8)$ and $\log(0.9)$

(b) $\log(3)$, using $\log(2)$, $\log(1.2)$ and $\log(0.8)$

(c) $\log(4)$, using $\log(2)$

(d) $\log(5)$, using $\log(2)$ and $\log(0.8)$

(e) $\log(6)$, $\log(8)$, $\log(9)$, and $\log(10)$

(f) $\log(11)$

You do not need to know the base. Why?

52. The graph of $y = \log_2(x)$ consists of the part to the right of $(1, 0)$ and the part to the left of $(1, 0)$. Are the two parts congruent?

53. Say that you have drawn the graph of $y = \log_2(x)$. Jane says that to get the graph of $y = \log_2(4x)$, you just raise that graph 2 units parallel to the y-axis. Sam says, "No, just shrink the x-coordinate of each point on the graph by a factor of 4." Who is right?

54. Let b be a positive number less than 1.

(a) Sketch the graphs of $y = b^x$ and $y = \log_b(x)$ on the same set of axes.

(b) What is the domain of $\log_b$?

(c) What is the x-intercept? That is, solve $\log_b(x) = 0$.

(d) For what values of x is $\log_b(x)$ positive? negative?

(e) Is the graph of $y = \log_b(x)$ an increasing or decreasing function?

(f) What can you say about the values of $\log_b(x)$ when x is close to zero (and in the domain)?

(g) What can you say about the values of $\log_b(x)$ when x is a large positive number?

(h) What can you say about the values of $\log_b(x)$ when x is a large negative number?

55. Prove that $\log_3(2)$ is irrational, that is, not rational. (Assume that it is rational, that is, equal to m/n for some integers m and n, and obtain a contradiction.)

1.3 Building More Functions from Basic Functions

This section completes the list of functions needed for calculus. Starting with the basic functions introduced in Section 1.1, we will see how to obtain a function as complicated as

$$f(x) = \frac{\sin(2x) + 3 + 4x + 5x^2}{\log_2(x) + 3^{-5x} + \sqrt{1 + x^3}}. \tag{1.3.1}$$

Before we go see how to construct new functions from old ones, we introduce one more type of basic function. These functions are so simple that they did not deserve to appear with the functions in the preceding section. They are the constant functions, whose graphs are horizontal lines. (See Figure 1.3.1.)

The Constant Functions

DEFINITION *(Constant Function)* A function $f(x)$ is **constant** if there is a number C such that $f(x) = C$ for all x in its domain. A special constant function is the **zero function**: $f(x) = 0$.

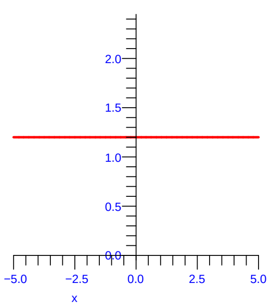


Figure 1.3.1

Using the Four Arithmetic Operations: $+$, $-$, $\times$, $/$

Given functions f and g, we can produce other functions from them by using the four operations of arithmetic:

$f + g$: For an input x, the output is $f(x) + g(x)$.

$f - g$: For an input x, the output is $f(x) - g(x)$.

fg: For an input x, the output is $f(x)g(x)$.

f/g: For an input x with $g(x) \neq 0$, the output is $f(x)/g(x)$.

The domains of $f + g$, $f - g$, and fg consist of the numbers that belong to both the domain of f and the domain of g. The function f/g is defined for all numbers x that belong to the domain of f and the domain of g with the extra condition that $g(x) \neq 0$.

With the aid of these constructions any polynomial or rational function are built from the simple function $f(x) = x$, called the **identity function**, and the constant functions.

A **polynomial** is a function of the form $p(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$, where the coefficients a_0, a_1, a_2, $\ldots$, a_n are numbers. If a_n is not zero, the **degree** of the polynomial is n. A **rational function** is the quotient of two polynomials. The domain of a polynomial is the set of all real numbers. The domain of a rational function is all real numbers except those for which the denominator is zero.

EXAMPLE 1 Use addition, subtraction, and multiplication to form the polynomial $F(x) = x^3 + 3x - 7$.

SOLUTION We first build each of the three terms: x^3, $3x$, and 7. The last is just a constant function. Multiplying the identity function x and the constant function 3 gives $3x$. The first term is obtained by first multiplying x and x to obtain x^2. Then multiplying x^2 and x yields x^3. Adding x^3 and $3x$ gives $x^3 + 3x$. Lastly, subtract the constant function 7 to obtain $x^3 + 3x - 7$.

Each of the three functions involved in forming F is defined for all real numbers. The domain of F is $(-\infty, \infty)$. ◇

Example 1 shows how to build any polynomial using $+$, $-$, and $\times$. Constructing rational functions also requires a use of division.

But how would we build a function like $\sqrt{1 + 3x}$? This leads us to the most important technique for combining functions to build more complicated ones.

Composite Functions

Given two functions f and g we can use the output of g as the input for f. That is, we can find $f(g(x))$. For instance, if $g(x) = 1 + 3x$ and f is the square root function, $f(x) = \sqrt{x}$, then $f(g(x)) = f(1 + 3x) = \sqrt{1 + 3x}$. This brings us to the definition of a composite function.

> **DEFINITION** *(Composition of functions)* Let X, Y, and Z be sets. Let g be a function from X to Y and let f be a function from Y to Z. Then the function that assigns to each element x in X the element $f(g(x))$ in Z is called the **composition of f with g**. It is denoted $f \circ g$, which is read as "f circle g" or as "f composed with g". The function $f \circ g$ is called a **composite function**.

Thinking of f and g as input-output machines we may consider $f \circ g$ as the machine built by hooking up the machine for f to process the outputs of the machine for g (see Figure 1.3.2).

Through Chapter 13 the sets X, Y, and Z will consist of numbers.

Most functions we encounter are composite. For instance, $\sin(2x)$ is the composition of $g(x) = 2x$ and $f(x) = \sin(x)$. We can hook up three or more

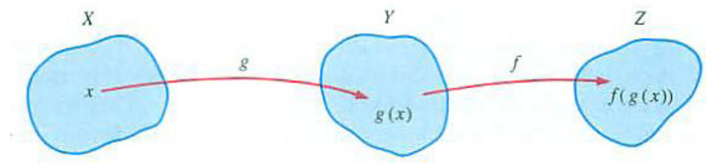


Figure 1.3.2 The output of the g-machine, $g(x)$, becomes the input for the f-machine. The result is the composition of f with g, $(f \circ g)(x) = f(g(x))$.

functions to make even more complicated functions. The function $\sin^3(2x) = (\sin(2x))^3$ is built up in three steps:

$$x \longrightarrow 2x \longrightarrow \sin(2x) \longrightarrow (\sin(2x))^3. \tag{1.3.2}$$

The first doubles the input, the second takes the sine of its input, and the third cubes its input.

The order matters. If, instead, cubing is done first, then the sine, and then end by doubling the input, the result is:

$$x \longrightarrow x^3 \longrightarrow \sin(x^3) \longrightarrow 2\sin(x^3). \tag{1.3.3}$$

Before pressing the sin key, be sure to check that your calculator is in radians mode.

When you enter a function on your calculator or on a computer, you have to be careful of the order in which the functions are applied as you evaluate a composite function. The way to evaluate $\sin(\log_{10}(240))$ depends on your calculator. On a traditional scientific calculator you enter 240, press log10 key, and then the sin key. On many of the newer graphing calculators you press the sin key, then the log10 key, then enter 240, followed by two right parentheses[0],)), and, finally, press the Enter key. The two approaches are different. If you press the sin key before log10, you will get $\log_{10}(\sin 240)$. For most computer software it is necessary to use parentheses to indicate inputs to functions. In this case you might enter sin(log10(240)).

To describe the build-up of a composite function it is convenient to use various letters, not just x, to denote the variables. This is illustrated in Examples 2 to 4.

EXAMPLE 2 Show how the function $\sqrt{4-x^2}$ is built up by the composition of functions. Find its domain.

SOLUTION The function $\sqrt{4-x^2}$ is obtained by applying the square-root function to the function $4-x^2$. To be specific, let

$$g(x) = 4 - x^2 \qquad \text{and} \qquad f(u) = \sqrt{u}\ (u \geq 0). \tag{1.3.4}$$

[0]The left parentheses are automatically included when the sin and log10 keys are pressed.

Then

$$f(g(x)) = f(4 - x^2) = \sqrt{4 - x^2}. \tag{1.3.5}$$

The square-root function is defined for all $u \geq 0$ and the polynomial $g(x)$ is defined for all numbers. So $f(g(x))$ is defined only when $g(x) \geq 0$:

$$\begin{aligned} g(x) &\geq 0 \\ 4 - x^2 &\geq 0 \\ 4 &\geq x^2 \\ 2 &\geq |x|. \end{aligned}$$

Thus, the domain of $\sqrt{4 - x^2}$ is the closed interval $[-2, 2]$. ◇

EXAMPLE 3 Express $1/\sqrt{1 + x^2}$ as a composition of three functions. Find its domain.

SOLUTION Call the input x. First, we compute $1 + x^2$. Second, we take the square root of that output, getting $\sqrt{1 + x^2}$. Third, we take the reciprocal of that result, getting $1/\sqrt{1 + x^2}$. In summary, form

$$u = 1 + x^2, \qquad \text{then } v = \sqrt{u} \qquad \text{then } y = \frac{1}{v}. \tag{1.3.6}$$

Given x, we first evaluate the polynomial $1 + x^2$, then apply the square-root function, then the reciprocal function.

The domain of $1+x^2$ consists of all real numbers; the domain of the square-root function is $[0, \infty)$; and the domain of the reciprocal function is all numbers except zero. Because $u = 1 + x^2 \geq 1$, $v = \sqrt{u} = \sqrt{1 + x^2}$ is defined for all x. Moreover, $v = \sqrt{1 + x^2} \geq 1$, so that $y = \frac{1}{v} = 1/\sqrt{1 + x^2}$ is defined for all real numbers x. ◇

The function in Example 3 can also be written as the composition of two functions: $x \longrightarrow 1 + x^2 \longrightarrow (1 + x^2)^{-1/2}$.

EXAMPLE 4 Let f be the cubing function and g the cube-root function. Compute $(f \circ g)(x)$, $(f \circ f)(x)$, and $(g \circ f)(x)$.

SOLUTION In terms of formulas, $f(x) = x^3$ and $g(x) = \sqrt[3]{x}$.

$$\begin{aligned} (f \circ g)(x) &= f(g(x)) = f(\sqrt[3]{x}) = \left(\sqrt[3]{x}\right)^3 = x. \\ (f \circ f)(x) &= f(f(x)) = f(x^3) = \left(x^3\right)^3 = x^9. \\ (g \circ f)(x) &= g(f(x)) = g(x^3) = \sqrt[3]{x^3} = x. \end{aligned}$$

The domains of f and g are $(-\infty, \infty)$. so $f \circ g$, $f \circ f$, and $g \circ f$ are defined for all real numbers.

Both $f \circ g$ and $g \circ f$ are the identity function. Whenever g is the inverse of f, $f \circ g$ and $g \circ f$ are the identity function. Each function undoes the action of the other. ◇

EXAMPLE 5 Show that every power function x^k, $x > 0$, can be constructed as a composition using exponential and logarithmic functions.

SOLUTION The first step is to write $x = 2^{\log_2(x)}$. Then, $x^k = \left(2^{\log_2(x)}\right)^k$ or, by properties of exponentials, $x^k = 2^{k \log_2(x)}$. So x^k is the composition of three functions: First, find $\log_2(x)$, then multiply by the constant function k, and then raise 2 to the resulting power. ◇

That a power function can be expressed in terms of an exponential and a logarithm will be used in Chapter 4.

An additional consequence of Example 5 is that it provides a meaning to functions like $x^{\sqrt{2}}$ and $x^{-\pi}$ for $x > 0$. We could even go so far as to remove the power functions from our list of basic functions. We chose not to do so because power functions with integer exponents are common and are defined also at negative inputs. Lastly, while it might seem surprising that the power functions can be expressed in terms of exponentials and logarithms, it is more astonishing that trigonometric functions, such as $\sin(x)$, can be expressed in terms of exponentials, as shown in Chapter 12.

Summary

This section showed how to build more complicated functions from power, exponential, and trigonometric functions and their inverses, and the constant functions. One method is to add, multiply, subtract, or divide outputs. The other is to compose functions so that one function operates on the output of a second function. Composite functions are extremely important in our upcoming study of calculus.

Warning about notation

Be careful when composing functions when one of them is a trigonometric function. For instance, what is meant by $\sin x^3$? Is it $\sin(x^3)$ or $(\sin(x))^3$? Do we first cube x, then take the sine, or the other way around? There is a general agreement that $\sin x^3$ stands for $\sin(x^3)$; cube first, then take the sine.

Spoken aloud, $\sin x^3$ is usually "the sine of x cubed," which is ambiguous. We can either insert a brief pause – "sine of (pause) x cubed" – to emphasize that x is cubed rather than $\sin(x)$, or rephrase it as "sine of the quantity x cubed."

On the other hand $(\sin(x))^3$, which is by convention usually written as $\sin^3(x)$, is spoken aloud as "the cube of $\sin(x)$" or "sine cubed of x."

Similar warnings apply to other trigonometric functions and logarithmic functions.

EXERCISES for Section 1.3

The function $y = \sqrt{1+x^2}$ is the composition of $s = 1 + x^2$ and $y = \sqrt{s}$. In Exercises 1 to 12 use a similar format to build the given functions as the composition of two or more functions.

1. $\sin(2x)$

2. $\sin^3(x)$

3. $\sin(3x)$

4. $\sin(x^3)$

5. $\sin^2(x^3)$

6. 2^{x^2}

7. $(x^2+3)^{10}$

8. $\log_{10}(1+x^2)$

9. $1/(x^2+1)$

10. $\cos^3(2x+3)$

11. $\left(\frac{2}{3x+5}\right)^3$

12. $\arcsin(\sqrt{x})$

13. These tables show some of the values of functions f and g:

x	1	2	3	4	5
$f(x)$	6	8	9	7	10

x	6	7	8	9	10
$g(x)$	4	3	2	5	1

(a) Find $f(g(7))$.

(b) Find $g(f(3))$.

14. Figure 1.3.3 shows the graphs of functions f and g. Estimate

(a) $f(g(0.6))$,

(b) $f(g(0.3))$,

(c) $f(f(0.5))$.

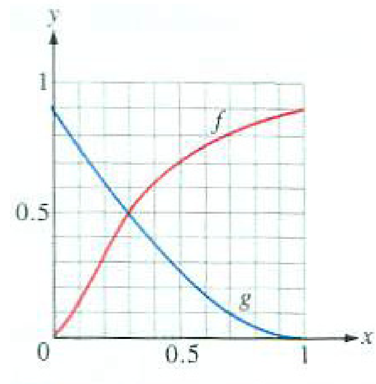


Figure 1.3.3

In Exercises 15 to 24 write y as a function of x. Simplify when possible.

15. $u = \sin(x)$, $y = u^2$

16. $u = x^3$, $y = 1/u$

17. $u = 2x^2 - 3$, $y = 1/u$

18. $u = \sqrt{x}$, $y = u^2$

19. $u = \sqrt{x}$, $y = \sin(u)$

20. $u = \log_3(x)$, $y = 3^u$

21. $v = 2x$, $u = v^2 - 1$, $y = u^2$

22. $v = \sqrt{x}$, $u = 1 + v$, $y = u^2$

23. $v = x + x^2$, $u = \sin(v)$, $y = u^3$

24. $v = \tan(x)$, $u = 1 + v^2$, $y = \cos(u)$

25. Figure 1.3.4 shows the graph of a function f whose domain is $[0, 1]$. Let $g(x) = f(2x)$.

(a) What is the domain of g?

(b) Graph $y = g(x)$

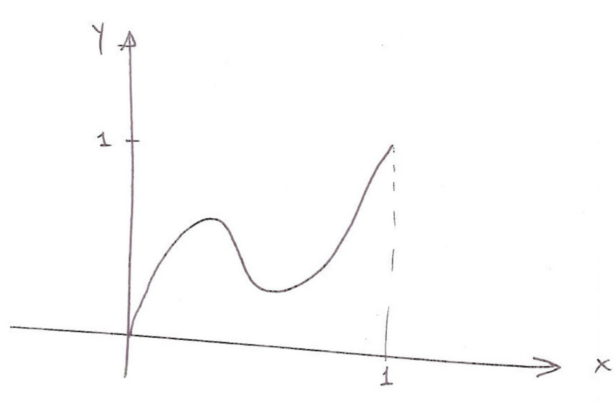


Figure 1.3.4

26. Let $f(x) = x^3$. Is there a function $g(x)$ such that $(f \circ g)(x) = x$ for all numbers x? If so, how many such functions are there?

27. Let $f(x) = x^4$. Is there a function $g(x)$ such that $(f \circ g)(x) = x$ for all negative numbers x? If so, how many such functions are there?

28. Let $f(x) = x^4$. Is there a function $g(x)$ such that $(f \circ g)(x) = x$ for all positive numbers x? If so, how many such functions are there?

29. Let $f(x) = 1/(1-x)$. What is the domain of f? of $f \circ f$? of $f \circ f \circ f$? Show that $(f \circ f \circ f)(x) = x$ for all x in the domain of $f \circ f \circ f$.

30. Find all functions of the form $f(x) = 1/(ax+b)$, where $a \neq 0$, such that $(f \circ f \circ f)(x) = x$ for all x in the domain of $f \circ f \circ f$.

31. Show that there is a function $u(x)$ such that $\cos x = \sin u(x)$. This shows that we did not need to include $\cos x$ among our basic functions.

32. Find a function $u(x)$ such that $3^x = 2^{u(x)}$.

33. If f and g are one-to-one, must $f \circ g$ be one-to-one?

34. If $f \circ g$ is one-to-one, must f be one-to-one? Must g be one-to-one?

35. If f has an inverse, $\mathsf{inv}f$, compute $(f \circ \mathsf{inv}f)(x)$ and $((\mathsf{inv}f) \circ f)(x)$.

36. Let $f(x) = 2x^2 - 1$ and $g(x) = 4x^3 - 3x$.

(a) Find $(f \circ g)(x)$.

(b) Find $(g \circ f)(x)$.

(c) Show that $(f \circ g)(x) = (g \circ f)(x)$.

While any two powers, such as x^3 and x^4, commute under composition, their composition in either order being x^{12}. Pairs of polynomials that commute with each other under composition are rare. To convince yourself of this, try to find more examples. Exercises 37 to 40 consider some specific cases. Exercise 41 shows a way to generate many such examples.

37. Let $g(x) = x^2$. Find all first degree polynomials $f(x) = ax + b$, $a \neq 0$, such that $f \circ g = g \circ f$, that is, $f(g(x)) = g(f(x))$.

38. Let $g(x) = x^2$. Find all second-degree polynomials $f(x) = ax^2 + bx + c$, where $a \neq 0$, such that $f \circ g = g \circ f$, that is, $f(g(x)) = g(f(x))$.

39. Let $f(x) = 2x + 3$. Find all functions of the form $g(x) = ax + b$, where a and b are constants, such that $f \circ g = g \circ f$.

40. Let $f(x) = 2x + 3$. Find all functions of the form $g(x) = ax^2 + bx + c$, where a, b, and c are constants, such that $f \circ g = g \circ f$.

41. This exercise rests on the identifies $\sin(x + y) = \sin(x)\cos(y) + \cos(x)\sin(y)$, $\cos(x + y) = \cos(x)\cos(y) - \sin(x)\sin(y)$, and $\cos^2 x + \sin^2 x = 1$.

(a) Show that $\sin(2x) = 2\sin(x)\cos(x)$ and $\cos(2x) = 2\cos^2(x) - 1$.

(b) Show that $\sin(3x) = 3\sin(x) - 4\sin^3(x)$ and $\cos(3x) = 4\cos^3(x) - 3\cos(x)$.

(c) Show that $\sin(4x) = \cos(x)(4\sin(x) - 8\sin^3(x))$ and $\cos(4x) = 8\cos^4(x) - 8\cos^2(x) + 1$.

(d) Show that for each positive integer n, $\cos(nx)$ is a polynomial in $\cos(x)$, that is, there is a polynomial P_n such that $\cos(nx) = P_n(\cos(x))$. You will have to consider separately the forms of $\sin(nx)$ when n is odd or even.

(e) Explain why $(P_n \circ P_m)(x) = (P_m \circ P_n)(x)$ for x in $[-1, 1]$. Because P_n and P_m are polynomials, it follows that $P_n \circ P_m = P_m \circ P_n$.

1.4 Geometric Series

Let a and r be numbers. The (infinite) sequence of numbers

$$a, ar, ar^2, ar^3, \ldots$$

is called a **geometric sequence**. Its first term is a. Each term after the first term is obtained by multiplying its predecessor by r, which is called the **ratio**. The n^{th} term is ar^{n-1}.

A finite collection of consecutive terms from a geometric sequence is called a **geometric progression**.

Let S_n be the sum of the first n terms of the geometric sequence:

S_n is also called a **partial sum** of the geometric sequence.

$$S_n = a + ar + ar^2 + \cdots + ar^{n-1}.$$

There is a short formula for it, which we will use several times.

To find the formula, subtract rS_n from S_n:

$$\begin{array}{lcll} S_n & = & a+ & ar + ar^2 + \cdots + ar^{n-1} \\ rS_n & = & & ar + ar^2 + \cdots + ar^{n-1} + ar^n. \end{array}$$

The many cancellations give

$$S_n - rS_n = a - ar^n.$$

If r is not 1, we can divide by $1-r$ to obtain:

Short Formula for the Partial Sum of a Geometric Series

$$S_n = \frac{a(1-r^n)}{1-r} \qquad r \neq 1 \tag{1.4.1}$$

EXAMPLE 1 Find (a) $3 + \frac{3}{2} + \frac{3}{4} + \frac{3}{8} + \frac{3}{16} + \frac{3}{32}$ and (b) $1 - \frac{1}{3} + \frac{1}{9} - \frac{1}{27} + \frac{1}{81}$.

SOLUTION (a) Here $a = 3$, $r = \frac{1}{2}$, and $n = 6$. The sum is

$$S_6 = \frac{3\left(1 - \left(\frac{1}{2}\right)^6\right)}{1 - \frac{1}{2}} = 6\left(1 - \left(\frac{1}{2}\right)^6\right) = \frac{378}{64} = \frac{189}{32}.$$

(b) In this case $a = 1$, $r = \frac{-1}{3}$, and $n = 5$. So the sum is

$$S_5 = \frac{1\left(1 - \left(\frac{-1}{3}\right)^5\right)}{1 - \frac{-1}{3}} = \frac{1 - \left(\frac{-1}{3}\right)^5}{\frac{4}{3}} = \frac{3}{4}\left(1 + \left(\frac{1}{3}\right)^5\right) = \frac{61}{81}.$$

$\diamond$

We now use the general formula, (1.4.1), to develop the partial sum of another geometric progression that will be very useful.

Let x and a be two numbers and consider the sequence

$$x^{n-1}, ax^{n-2}, a^2x^{n-3}, a^3x^{n-4}, \ldots, a^{n-3}x^2, a^{n-2}x, a^{n-1}. \tag{1.4.2}$$

The exponent of x decreases from $n-1$ to 0 while the exponent of a increases from 0 to $n-1$. While it might not look like it at first, (1.4.2) is the first n terms of a geometric sequence. The first term is x^{n-1} and the ratio is a/x. Thus, assuming x is not 0 or a,

$$\begin{aligned} & x^{n-1} + ax^{n-2} + a^2x^{n-3} + a^3x^{n-4} + \cdots + a^{n-3}x^2 + a^{n-2}x + a^{n-1} \\ = \; & x^{n-1}\left(\frac{\left(1-\left(\frac{a}{x}\right)^n\right)}{1-\frac{a}{x}}\right) = \frac{x^{n-1}\left(\frac{x^n-a^n}{x^n}\right)}{\frac{x-a}{x}} = \frac{x^n-a^n}{x-a}. \end{aligned}$$

This leads us to conclude that

$$\boxed{x^{n-1}+ax^{n-2}+a^2x^{n-3}+a^3x^{n-4}+\cdots+a^{n-3}x^2+a^{n-2}x+a^{n-1} = \frac{x^n-a^n}{x-a} \qquad x \neq a} \tag{1.4.3}$$

In Chapter 2 we will use (1.4.3) in the reverse order, to express the quotient $\frac{x^n-a^n}{x-a}$ as a sum of n terms.

Equation (1.4.3) can also be established from the factorizations of $x^n - a^n$:

$$\begin{aligned} x^2 - a^2 &= (x-a)(x+a) \\ x^3 - a^3 &= (x-a)(x^2+ax+a^2) \\ x^4 - a^4 &= (x-a)(x^3+ax^2+a^2x+a^3) \end{aligned}$$

and so on. To establish the last, for instance, multiply out its right-hand side:

$$\begin{aligned} & (x-a)(x^3+ax^2+a^2x+a^3) \\ = \; & (x^4+ax^3+a^2x^2+a^3x) - (ax^3+a^2x^2+a^3x+a^4) \\ = \; & x^4 - a^4. \end{aligned}$$

Summary

The key idea of this section is that the sum of the n numbers $a + ar + ar^2 + \cdots + ar^{n-1}$ equals $a\frac{1-r^n}{1-r}$ as long as r is not 1. If r is 1, then the sum is na, because each summand is a. As a particularly useful case, we have

$$\frac{x^n-a^n}{x-a} = x^{n-1} + ax^{n-2} + a^2x^{n-3} + \cdots + a^{n-2}x + z^{n-1}.$$

EXERCISES for Section 1.4

In Exercises 1 to 6 calculate the sum using the formula for the sum of a geometric progression.

1. $1 + 3 + 9 + 27 + 81 + 243$

2. $1 - 3 + 9 - 27 - 81 + 243$

3. $2 + 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16}$

4. $0.5 - 0.05 + 0.005 - 0.0005 + 0.00005 - 0.000005 + 0.0000005$

5. $a^4 + a^3b + a^2b^2 + ab^3 + b^4$

6. $1 - \frac{1}{2} + \frac{1}{4} - \frac{1}{8} + \frac{1}{16} - \frac{1}{32}$

In Exercises 7 and 8 write the given polynomial in two different ways as a product of two polynomials. Don't just change the order of the factors.

7. $x^6 - a^6$

8. $x^9 - a^9$

9. Show that $x^4 - 16 = (x^3 + 2x^2 + 4x + 8)(x - 2)$.

10. Show that $x^5 - 32 = (x^4 + 2x^3 + 4x^2 + 8x + 16)(x - 2)$.

11. This exercise obtains the sum of a geometric progression geometrically. Let r be a a positive number less than 1 and n a positive integer .

(a) In the interval $[0, 1]$ indicate the numbers r, r^2, $\dots$, r^{n+1}.

(b) The numbers in (a) break the interval $[r^{n+1}, 1]$ of length $1 - r^{n+1}$ into $n + 1$ intervals. By adding their lengths show that $1 + r + \cdots + r^n = \frac{1-r^{n+1}}{1-r}$.

12.

(a) Why is $(1 - r) + (r - r^2) + (r^2 - r^3) + \cdots + (r^{n-1} - r^n)$ equal to $1 - r^n$ for any number r?

(b) From (a) deduce the formula for the sum of the geometric series, $1 + r + r^2 + ... + r^{n-1}$ when r is not 1.

13. What happens to $\frac{x^3-1}{x^2-1}$ when you choose x nearer and nearer 2? Nearer and nearer 1?

(a) Base your answers on exponents.

(b) Base your answers on geometric series.

14. What happens to $\frac{x^5+32}{x+2}$ as x approaches 2? as x approaches -2?

(a) Base your answers on exponents.

(b) Base your answers on geometric series.

In Exercises 15 and 16 Sam and Jane discuss the term r^n that appears in the short formula, (1.4.1), for the sum of a geometric progression.

15.

Sam: I just computed 1.001^n for really large values of n.

Jane: What did you find?

Sam: Well, 1.001^{500} is about 1.65 and 1.001^{1000} is about 2.72.

Jane: So?

Sam: So I think that as n grows, 1.001^n is getting near 3 maybe, or maybe π.

Jane: Well, I just computed 1.001^{2000} and got about 7.38. I think it's getting nearer and nearer 20.

After computing some values of 1.001^n for some larger values of n, offer your own opinion on what happens to 1.001^n as n increases. How do you think 1.000001^n behaves?

16.

Sam: When I graph 0.5^n I see a sequence of numbers getting very near 0, as in Figure 1.4.1.

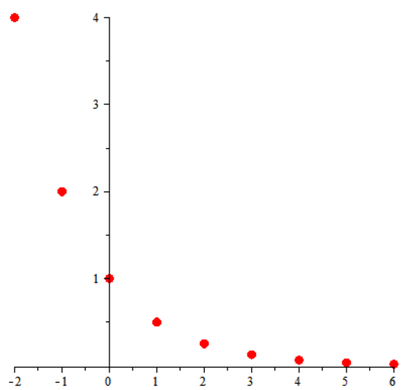


Figure 1.4.1

Jane: Maybe you're right. I computed $(0.999)^{1000}$ and got about 0.37.

Sam: So it looks like those numbers are getting real close to 1/3.

Jane: Why 1/3?

Sam: It's the only number I know near 0.37.

Jane: That's not much of a reason.

Based on your calculations, make a conjecture about what happens to 0.999^n as n gets larger and larger.

17. Express 5.14141414 as a fraction. Use the short formula for the sum of a geometric progression.

18.

(a) Using your calculator, evaluate the product $2 \cdot \sqrt{2} \cdot \sqrt[4]{2} \cdot \sqrt[8]{2}$.

(b) Each factor in (a) except the first is the square root of its predecessor. Continue the pattern with more factors. Evaluate the product in each case.

(c) **Sam:** I think the products will get arbitrarily large.

Jane: Why?

Sam: You're multiplying numbers bigger than 1. So the products keep growing.

Jane: But the factors are getting closer and closer to 1.

Sam: So?

Jane: So maybe the products don't get arbitrarily large.

Decide who is right.

1.5 Logarithms

How many 2s must be multiplied to get 32? Whatever the answer is, it is called "the logarithm of 32 to the base 2." Because $2^5 = 32$, the logarithm of 32 to the base 2 is 5. More generally, a logarithm is defined in terms of an exponential function.

Definition of Logarithm to the Base b

Let b and c be positive numbers, $b \neq 1$. There is a number d such that

$$b^d = c.$$

The exponent d is called the **logarithm of c to the base b**. It is denoted

$$\log_b(c).$$

By the definition of a logarithm,

$$b^{\log_b(c)} = c.$$

The word "logarithm" comes from the Greek. In a Greek restaurant, to get the bill, you ask the waiter for the "logarismo".

EXAMPLE 1 Find (a) $\log_{10}(1000)$, (b) $\log_2(1024)$, (c) $\log_9(3)$, and (d) $\log_{16}\left(\frac{1}{4}\right)$.

SOLUTION (a) Because $10^3 = 1000$, $\log_{10}(1000) = 3$.

(b) Because $2^{10} = 1024$, $\log_2(1024) = 10$.

(c) Because $9^{1/2} = 3$, $\log_9(3) = \frac{1}{2}$.

(d) Because $16^{-1/2} = \frac{1}{4}$, $\log_{16}\left(\frac{1}{4}\right) = \frac{-1}{2}$. $\diamond$

Every property of an exponential function translates into a property of logarithms. For instance, here is how we write the equation $b^{x+y} = b^x b^y$ with the language of logarithms.

Let $c = b^x$ and $d = b^y$. We have

$$x = \log_b(c) \qquad \text{and} \qquad y = \log_b(d). \tag{1.5.1}$$

Because

$$cd = b^x b^y = b^{x+y}$$

we know

$$\log_b(cd) = x + y.$$

Using (1.5.1), we conclude that

$$\log_b(cd) = \log_b(c) + \log_b(d).$$

This generalizes to the logarithm of the product of several numbers. In words:

The logarithm of a product of two or more numbers is the sum of their logarithms.

Table 1.5.1 lists the properties of exponential functions and the corresponding properties of logarithms.

Exponential Language			Logarithm Language		
b^{x+y}	$=$	$b^x b^y$	$\log_b cd$	$=$	$\log_b c + \log_b d$
b^{x-y}	$=$	b^x/b^y	$\log_b c/d$	$=$	$\log_b c - \log_b d$
b^0	$=$	1	$\log_b 1$	$=$	0
b^1	$=$	b	$\log_b b$	$=$	1
b^{-x}	$=$	$1/b^x$	$\log_b(1/c)$	$=$	$-\log_b c$
$(b^x)^y$	$=$	b^{xy}	$\log_b c^d$	$=$	$d \log_b c$

Table 1.5.1

Figure 1.5.1 is the graph of $y = \log_2(x)$. Notice that as x increases, so does

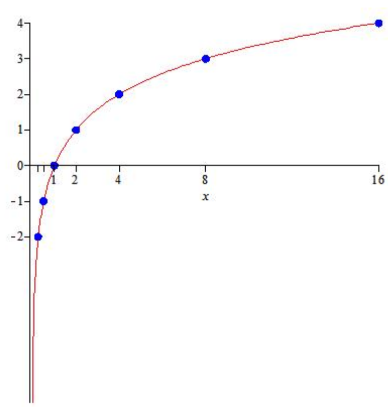


Figure 1.5.1

$\log_2(x)$, but very slowly. Also, when x is near 0, $\log_2(x)$ is negative but has large absolute values.

Logarithms can be used to simplify products, quotients, and powers:

$$\begin{aligned}\log_b\left(\frac{\sqrt{x}(2+x)^3}{(1+x^2)^5}\right) &= \log_b(\sqrt{x}) + \log_b\left((2+x)^3\right) - \log_b\left(\left(1+x^2\right)^5\right) \\ &= \frac{1}{2}\log_b(x) + 3\log_b(2+x) - 5\log_b\left(1+x^2\right).\end{aligned}$$

In the final expression, most of the exponents and radical sign no longer appear. There is no way to simplify $\log_b(2+x)$ and $\log_b(1+x^2)$.

Summary

This section reviews logarithms, which are a different way of talking about exponents. The two key properties of logarithms for a positive base b are $\log_b(xy) = \log_b(x) + \log_b(y)$ and $\log_b(x^y) = y\log_b(x)$.

EXERCISES for Section 1.5

1. Why is $\log_b(c)$ defined only for positive values of c?

2. How is $\log_{b^2}(c)$ related to $\log_b(c)$?

3. Evaluate (a) $\log_b\left(\sqrt{b}\right)$, (b) $\log_b\left(\frac{b^3}{\sqrt{b}}\right)$, (c) $\log_b\left(\sqrt{b}\sqrt[3]{b}b^4\right)$

4. Simplify $\log_2\left(\frac{\left(x^3\right)^5\sqrt[3]{x+2}\left(1+x^2\right)^{15}}{x^5+7}\right)$.

5. Show that $\frac{\log_b(x)-\log_b(y)}{c} = \log_b\left(\left(\frac{x}{y}\right)^{1/c}\right)$.

6. What happens to $\log_{10}(x)/x$ for large values of x? (Experiment and form a conjecture.)

7. Translate "She has a five-figure income" into logarithms.

In Exercises 8 to 12 establish the given property of logarithms by using a property of exponentials. (Assume $b > 0$.)

8. $\log_b(1) = 0$

9. $\log_b(b) = 1$

10. $\log_b(1/c) = -\log_b(c)$ $(c > 0)$

11. $\log_b(c^d) = d\log_b(c)$ $(c > 0)$

12. $\log_b(c/d) = \log_b(c) - \log_b(d)$ $(c, d > 0)$

13.

(a) Graph $\log_{1/2}(x)$ and $\log_2(x)$.

(b) How is $\log_{1/b}(c)$ related to $\log_b(c)$?

14. Show that

$$\frac{\log_b(a+h)-\log_b(a)}{h} = \left(\log_b\left(\left(1+\frac{h}{a}\right)^{a/h}\right)\right)^{1/a}.$$

15. How would you find $\log_5(3^7)$ if your calculator has only a key for logarithms to the base ten? (Start with the equation $5^x = 3^7$.)

16. Until the appearance of calculators, slide rules were commonly used for multiplication and division. Now, the International Slide Rule Museum (`http://www.sliderulemuseum.com/` is the world's largest repository of slide rule information. To see how a slide rule multiplies two numbers, mark two pieces of paper with the numbers 1, 2, 4, 8, 16, and 32 placed at equal distances apart, as shown in Figure 1.5.2. To multiply, say, 4 times 8, slide the lower paper so its 1 is under the

1	2	4	8	16	32

1	2	4	8	16	32

Figure 1.5.2 Slide rule scales for multiplication and division.

4. Then the product of 4 and 8 appears above the 8.

(a) Why does the slide rule work?

(b) How would you make a slide rule for multiplying that has all the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 on both scales?

17.

(a) Show that for positive numbers b and c, neither equal to 1, $\log_b(x)/\log_c(x)$ equals $\log_b(c)$, independent of x ($x > 0$). (Start with $b^{\log_b(x)} = x$.)

(b) What does (a) imply about the graphs of $y = \log_b(x)$ and $y = \log_c(x)$?

18. Rarely is $\log_b(x + y)$ equal to $\log_b(x) + \log_b(y)$.

(a) Show that if $\log_b(x + y) = \log_b(x) + \log_b(y)$, then $y = x/(1 - x)$.

(b) Give an example of x and y that satisfy the equation in (a).

The point of this Exercise is to show that while there is an identify for $\log_b(xy)$, there is no identity involving $\log_b(x + y)$.

19. One way to compute b^4 is to start with b and multiply by b three times, obtaining b^2, b^3, and, finally, b^4. But b^4 can be computed with only two multiplications. First compute b^2, then compute $b^2 \cdot b^2$. This raises a question encountered when programming a computer. What is the fewest number of multiplications needed to compute b^n? Call that numer $m(n)$. For instance, $m(4) = 2$.

(a) Show that $m(2) = 1$, $m(3) = 2$, $m(5) = 3$, $m(6) = 3$, $m(7) = 3$, $m(8) = 3$, and $m(9) = 4$.

(b) Show that $m(n) \geq \log_2(n)$. (Think of the final multiplication.)

(c) Show that, when n is a power of 2, then $m(n) = \log_2(n)$. (n is a power of 2 when $n = 2^k$, k a positive integer.)

20. Jane says to Sam, "I'm thinking of a whole number in the interval from 1 to 32. You have to find what it is. I'll answer each question 'yes' or 'no'."

(a) What five questions, in order, should Sam ask to be sure he will guess the number?

(b) If, instead, the interval is from 1 to 50, how should Sam modify his questions to be guaranteed to guess the number in the fewest number of questions?

(c) How is this Exercise related to logarithms?

1.S Chapter Summary

This chapter reviewed precalculus material concerning functions. Calculus begins in the next chapter when we answer questions such as "What happens to $(2^x - 1)/x$ as x gets very small?". The answers are used in Chapter 3 to settle questions such as "How rapidly does 2^x change for a slight change in x?"

Section 1.1 introduced the terminology of functions: input (argument), output (value), domain, range, independent variable, dependent variable, piecewise-defined function, inverse of a function, graph of a function, decreasing, increasing, non-increasing, non-decreasing, positive, and monotonic.

Section 1.2 reviewed x^k and its inverse $x^{1/k}$ (constant exponent, variable base), b^x (constant base, variable exponent) and its inverse $\log_b(x)$, and the six trigonometric functions and their inverses. All angles are measured in radians, unless otherwise stated.

Section 1.3 described five ways of getting new functions from functions f and g, namely $f+g$, $f-g$, fg, f/g, and the composition $f \circ g$.

Section 1.4 developed a short formula for a finite geometric sum with first term a and ratio r, $r \neq 1$: $a + ar + ar^2 + \cdots + ar^{n-1} = \frac{a(1-r^n)}{1-r}$.

Section 1.5 reviews the logarithm function to base b, b positive and b not 1.

EXERCISES for 1.S

Exercises 1 to 10 concern logarithms.

1. Evaluate (a) $\log_3 \sqrt{3}$, (b) $\log_3(3^5)$, and (c) $\log_3\left(\frac{1}{27}\right)$.

2. If $\log_4 A = 2.1$, evaluate (a) $\log_4(A^2)$, (b) $\log_4(1/A)$, and (c) $\log_4(16A)$.

3. If $\log_3 5 = a$, what is $\log_5 3$?

4. Find x if $5 \cdot 3^x \cdot 7^{2x} = 2$.

5. Solve for x: (a) $2 \cdot 3^x = 7$, (b) $3^{5x} = 2^{7x}$, (c) $3 \cdot 5^x = 6^x$, and (d) $10^{2x}3^{2x} = 5$.

6. Why do only positive numbers have logarithms? (Chapter 12 shows that negative numbers have logarithms also, but they are provided with the aid of complex numbers.)

7. Evaluate (a) $\log_2(2^{43})$, (b) $\log_2(32)$, and (c) $\log_2(1/4)$.

Exercises 8 to 10 concern the relation between logarithms to different bases.

8. Suppose that you want to obtain $\log_2(17)$ in terms of $\log_3(17)$.

(a) Which would be larger, $\log_2(17)$ or $\log_3(17)$?

(b) Show that $\log_2(17) = (\log_2(3)) \log_3(17)$. (Take logarithms to the base 2 of both sides of the equation $3^{\log_3(17)} = 17$.)

9.

(a) Using only the log key (and +, -, ×, and /), compute $\log_2(6)$ and $\log_6(2)$.

(b) Compute the product of $\log_2(6)$ and $\log_6(2)$.

(c) Compute the product of $\log_7(11)$ and $\log_{11}(7)$.

(d) Make a conjecture about $\log_a(b) \cdot \log_b(a)$

(e) Show that the conjecture made in (d) is correct. (Start with $a^{\log_a(b)} = b$ and take logarithms to the base b.)

10. You can use your calculator with a key for base-ten logarithms to compute logarithms to any base.

(a) Show why $\log_b(x) = \frac{\log_{10}(x)}{\log_{10}(b)}$.

(b) Compute $\log_2(3)$.

When using the formula in (a) it is easy to forget whether to multiply or divide by $\log_{10}(b)$. As a memory device keep in mind that when b is large, $\log_b(x)$ is small, so you want to divide by $\log_{10}(b)$.

11. If your scientific calculator lacks a key to display a decimal approximation to π, how could you use other keys to display it?

12. Drawing pictures, find (a) $\tan(\arcsin(1/2))$, (b) $\tan(\arctan(-1/2))$, and (c) $\sin(\arctan(3))$.

13. If f and g are decreasing functions, what (if anything) can be said about (a) $f + g$, (b) $f - g$, (c) f/g, (d) f^2, and (e) $-f$?

14. What type of function is $f \circ g$ if (a) f and g are increasing, (b) f and g are decreasing, and (c) f is increasing and g is decreasing? Explain.

15. If f is increasing, what (if anything), can be said about $g = \mathsf{inv}(f)$?

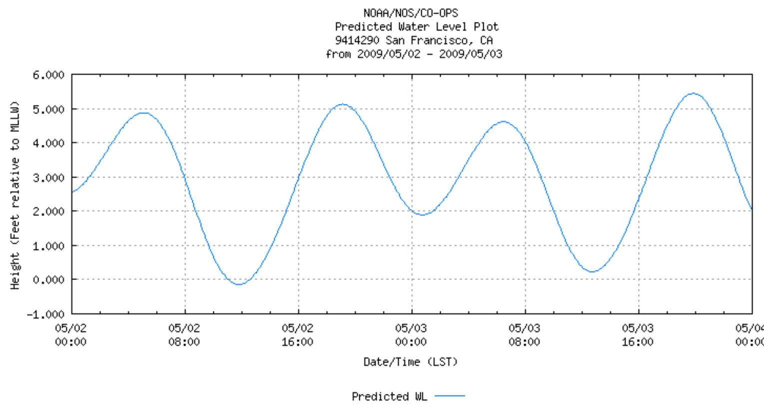


Figure 1.S.1 Source: `http://tidesandcurrents.noaa.gov/gmap3/`

16. The predicted height of the tide at San Francisco for May 3, 2009 is shown in Figure 1.S.1.

(a) At what time(s) was the tide falling the fastest?

(b) At what time(s) was it rising the fastest?

(c) At what time(s) was it changing most slowly?

(d) How high was the highest tide? The lowest?

(e) At what rate was the tide going down at 2 p.m.? Express the answer in feet per hour.

17. Evaluate as simply as possible.

(a) $\log_3\left(3^{17.21}\right)$,

(b) $\log_5\left(5^{\sqrt{2}}/25^{\sqrt{3}}\right)$,

(c) $\log_2\left(4^{123}\right)$,

(d) $\log_2\left((4^5)^6\right)$,

(e) $\tan(\arctan(3))$.

18. Give an example of (a) an increasing function f defined for positive x such that $f(f(x)) = x^9$ and (b) a decreasing function g such that $g(g(x)) = x^9$.

19. Graph

(a) $\sin(x)$, x in $[0, 2\pi]$,

(b) $\sin(3x)$, x in $[0, \pi/2]$,

(c) $\sin(x - \pi)$, x in $[0, 2\pi]$,

(d) $\sin(3x - \pi/6)$, x in $[0, \pi/2]$.

20. Imagine that the exponential key, x^y, on your calculator is broken. How would you compute $(2.73)^{3.09}$?

21. Using only multiplication, exponential, and logarithm keys, how could you compute $0.37 + 1.75$? Carry out the computation.

22.

(a) In many calculators the log key refers to base-ten logarithms. It can be used to find logarithms to any base $b > 0$. To see why, start with the equation $b^{\log_b(x)} = x$ and then take $\log_{10}$ of both sides. This gives the formula

$$\log_b(x) = \frac{\log_{10}(x)}{\log_{10}(b)}.$$

(b) Use (a) to find $\log_3(7)$. (Why should the result be between 1 and 2?)

(Semi-log graphs) In most graphs the scale on the y-axis is the same as the scale on the x-axis, or a constant multiple of it. However, to graph a rapidly increasing function, such as 10^x, it is convenient to distort the y-axis. Instead of plotting the point (x, y) at a height of, say, y inches, plot it at a height of $\log_{10} y$ inches. So the datum $(x, 1)$ is drawn at height zero, the datum $(x, 10)$ at height 1 inch, and the datum $(x, 100)$ at 2 inches. Instead of graphing $y = f(x)$, you graph $Y = \log_{10} f(x)$. In particular, if $f(x) = 10^x$, $y = \log_{10} 10^x = x$: the graph would be a straight line. To avoid having to calculate logarithms, it is convenient to use semi-log graph paper, shown in Figure 1.S.2.

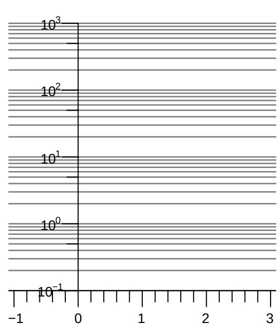


Figure 1.S.2

23. Using semi-log paper, graph $y = 2 \cdot 3^x$.

24. Using semi-log paper, graph $y = \frac{2}{3^x}$.

25. Let a, b, c, d be constants such that $ad - bc \neq 0$.

(a) Show that $y = (ax + b)/(cx + d)$ is one-to-one.

(b) For which a, b, c, d does the function in (a) equal its inverse function?

26. Show that for x in $(0, \pi/2)$, $x - \sin(x)$ is an increasing function. (Display x and $\sin(x)$ using a unit circle, for two values of x, a and b.)

27. Figure 1.S.3 shows a circle of radius 1 and a point P at a distance h from it. An arc of the circle is visible from P. That arc subtends an angle.

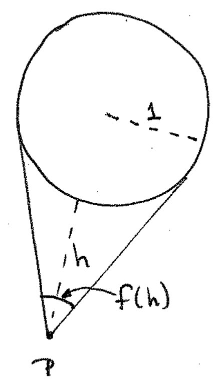


Figure 1.S.3

(a) Express the angle (in radians) as a function of h, $f(h)$.

(b) As P is chosen farther and farther from the circle what happens to $f(h)$?

(c) As P is chosen closer and closer to the circle what happens to $f(h)$?

28. The equation $y = x - e\sin(x)$, known as **Kepler's equation**, is important in the study of the motion of planets. Here e is the eccentricity of an elliptical orbit, y is related to time, and x is related to an angle. For more information, do a web search for `Kepler equation`.
The function $f(x) = x - \sin(x)$ is increasing for all numbers x. (See Exercise 26.)

(a) Graph f.

(b) Explain why, even though it cannot be solved explicitly, you know the equation $y = x - \sin(x)$ can be solved for x as a function of y ($x = g(y)$).

(c) How are the graphs of $y = x - \sin(x)$ and $y = g(x)$ related?

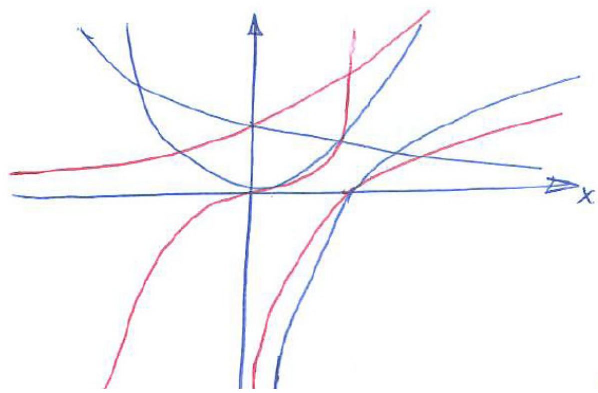


Figure 1.S.4

29. Label the curves in Figure 1.S.4 with their equations.

(a) $y = x^2$,

(b) $y = x^3$,

(c) $y = 2^x$,

(d) $y = \log_2(x)$,

(e) $y = \log_3(x)$, and

(f) $f(x) = \left(\frac{1}{2}\right)^x$.

30. The equation $\log_a(b)\cdot\log_b(a) = 1$ makes one wonder, "Is $\log_a(b)\cdot\log_b(c)\log_c(a) = 1$?" What is the answer? Either exhibit positive a, b, and c for which the equation does not hold or else prove it always holds.

31. Find all numbers a and b such that $\log_a(b)$ equals $\log_b(a)$.

32.

(a) Are the graphs of $y = x^2$ and $y = 4x^2$ congruent?

(b) Are the graphs of $y = x^2$ and $y = 4x^2$ similar? (One figure is similar to another if one is the other magnified by the same factor in all directions.)

33. A solar cooker can be made in the shape of part of a sphere. The one in Figure 1.S.5 spans only $\pi/3$ (60°) at the center O. For simplicity, take the radius to be 1.

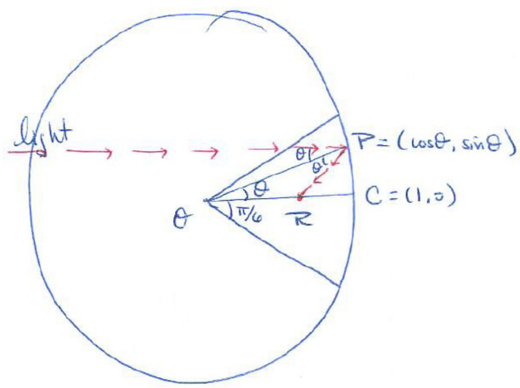


Figure 1.S.5

Light parallel to OC strikes the cooker at $P = (\cos(\theta), \sin(\theta))$ and is reflected to a point R on the radius OC.

(a) There are two angles of measure θ at P. Why is the top one equal to θ?

(b) Why is the bottom angle at P also θ?

(c) Show that $\overline{OR} = 1/(2\cos(\theta))$.

(d) Show that the heated part of the x-axis has length $(1/\sqrt{3}) - (1/2) \approx 0.077$, or about 1/13th of the radius.

The Calculus is Everywhere section at the end of Chapter 3 describes a parabolic reflector, which reflects all the light to a single point.

Calculus is Everywhere # 1

Graphs Tell It All

The graph of a function conveys a great deal of information quickly. Here are four examples, all based on numerical data.

The Hybrid Car

A friend of ours bought a hybrid car that runs on a fuel cell at low speeds, on gasoline at higher speeds, and on a combination of the two in between. He also purchased the gadget that exhibits "miles-per-gallon" at any instant. With the driver glancing at the speedometer and the passenger watching the gadget, we collected data on fuel consumption (miles-per-gallon) as a function of speed. Figure C.1.1 displays what we observed.

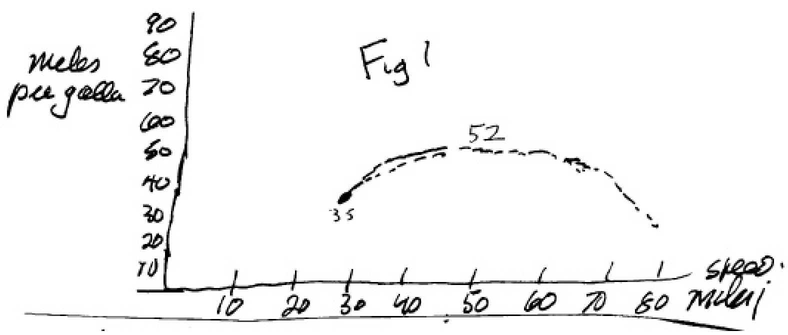


Figure C.1.1 A graph of fuel efficiency (miles per gallon) as a function of speed (miles per hour).

The straight-line part is misleading, for at low speeds no gasoline is used. So 100 plays the role of infinity. The "sweet spot," the speed that maximizes fuel efficiency (as determined by miles-per-gallon), is about 55 mph, while speeds in the range from 40 mph to 70 mph are almost as efficient. However, at 80 mph the car gets only about 30 mpg.

To avoid having to use 100 to represent infinity, we also graph gallons-per-mile, the reciprocal of miles-per-gallon, as shown in Figure C.1.2. In this graph the minimum occurs at 55 mph. The straight part of the graph on the speed axis (horizontal) records zero gallons per mile.

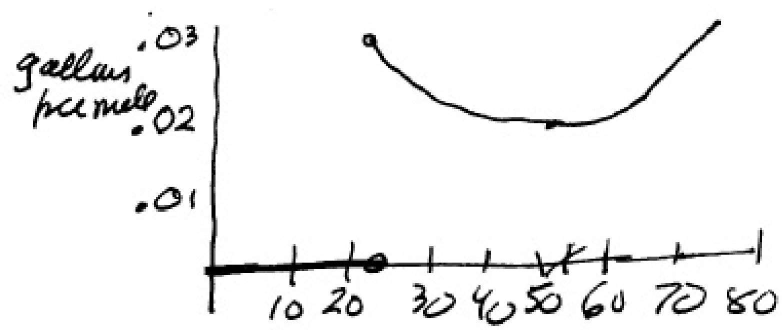


Figure C.1.2

Traffic and Accidents

Figure C.1.3 appears in S.K. Stein's, Risk factors of sober and drunk drivers by time of day, *Alcohol, Drugs, and Driving* 5 (1989), pp. 215–227. The vertical scale is described in the paper.

Glancing at the graph labelled "Traffic" in Figure C.1.3 we see that there are peaks at the morning and afternoon rush hours, with minimum traffic around 3 a.m. However, the number of accidents is fairly high at that hour. "Risk" is measured by the quotient, accidents divided by traffic. It reaches a peak at 1 a.m. This cannot be explained by the darkness at that hour, for the risk rapidly decreases the rest of the night. It turns out that the risk has the same shape as the graph that records the number of drunk drivers.

It is a sobering thought that at any time of day a drunk's risk of being involved in an accident is on the order of one hundred times that of an alcohol-free driver.

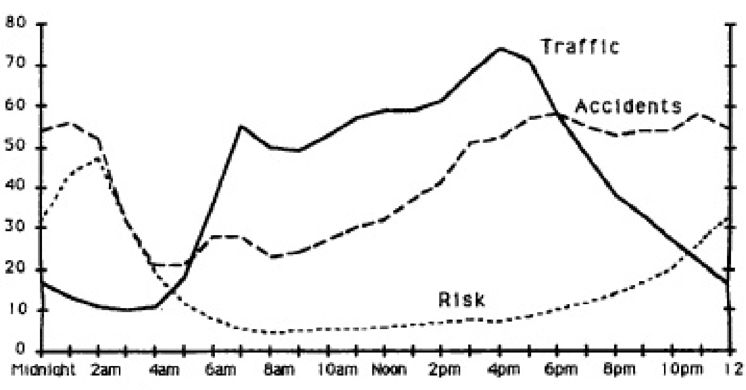


Figure C.1.3

Petroleum

The three graphs in Figure C.1.4 show the rate of crude oil production in the United States, the rate at which it was imported, and their sum, the rate of consumption. They are expressed in millions of barrels per day, as a function of time. A barrel contains 42 gallons. (For a few years after the discovery of oil in Pennsylvania in 1859 it was transported in barrels.)

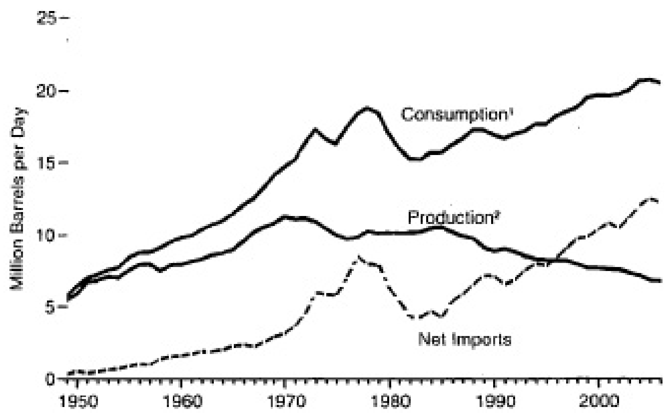


Figure C.1.4 Source: Energy Information Administration (Annual Energy Review, 2006)

The graphs convey a good deal of history and a warning. In 1950 the United States produced almost enough petroleum to meet its needs, but by 2006 it had to import most of the petroleum it consumed. Moreover, domestic production peaked in 1970.

The imbalance between production and consumption raises serious questions, especially as exporting countries need more oil to fuel their own growing economies, and developing nations, such as India and China, place increasing demands on world production. Also, since the total amount of petroleum in the earth is finite, it will run out, and the Age of Oil will end. Geologists, having gone over the globe with a fine-tooth comb, believe they have already found all the major oil deposits. No wonder that the development of alternative sources of energy has become a high priority.

Calculus is Everywhere # 2

Where Does All That Money Come From?

As of 2007 there were over 7 trillion dollars in the form of currency, deposits in banks, in money market mutual funds, and so on. Where did they all come from? How is money created?

Banks create some of the supply, and this is how they do it.

When someone makes a deposit at a bank, the bank lends most of it. It cannot lend all of it, for it must keep a reserve to meet the needs of depositors who may withdraw money from their accounts. The government stipulates what this reserve must be, usually between 10 and 20 percent of the deposit Let's use 20 percent.

If a person deposits $1,000, the bank can lend $800. Assume that the borrower deposits that amount in another bank; the second bank can lend 80 percent of the $800, or $640. The recipient of the $640 can then deposit it at a third bank, which must retain 20 percent, but is free to lend 80 percent, which is $512. At this point there are now

$$1000 + 800 + 640 + 512 \text{ dollars in circulation.} \tag{C.2.1}$$

Each summand is 0.8 times the preceding summand. The sum (C.2.1) can be written as

$$1000\left(1 + 0.8 + 0.8^2 + 0.8^3\right). \tag{C.2.2}$$

The process goes on indefinitely, through a fifth person, a sixth, and so on. A good approximation of the impact of the initial deposit of $1000 after n stages is 1000 times the sum

$$1 + 0.8 + 0.8^2 + 0.8^3 + \cdots + 0.8^n. \tag{C.2.3}$$

Being the sum of a geometric progression, the sum (C.2.3) equals $(1 - 0.8^{n+1})/(1 - 0.8)$ and that quotient approaches $1/0.2 = 5$. Thus the original $1000 could create an amount approaching $5000. Economists say that the **multiplier** is 5, and the total impact is five times the initial deposit. There are now magically 4000 more dollars than at the start. This happens because a bank can lend money it does not have. The sequence of deposits and lends involve having faith in the future. If it is destroyed, then there may be a run on the bank as depositors rush to take their money out. If that disaster can be avoided, then banking is a delightful business, for bankers can lend money they do not have.

The concept of the multiplier also appears in measuring economic activity. Assume that the government spends a million dollars on a new road. That amount goes to firms and individuals who build the road. In turn, those firms and individuals spend a certain fraction of that income. This process of earn and spend continues to trickle through the economy. The total impact may be much more than the initial amount the government spent. Again, the ratio between the total impact and the initial expenditure is called the multiplier.

The mathematics behind the multiplier is the theory of the geometric series, summing the successive powers of a fixed number.

EXERCISES

1. If the amount a bank must keep on reserve is cut in half, what effect does this have on the multiplier?

Chapter 2

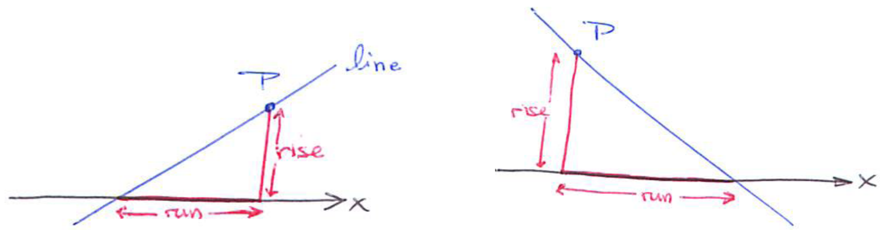
Introduction to Calculus

There are two main concepts in calculus: the derivative and the integral. Underlying both is the concept of a limit. This chapter introduces limits, with an emphasis on developing both your understanding of limits and techniques for finding them.

We start the journey in Section 2.1 where our knowledge about the slope of a line is used to define the slope at a point on a curve. The four limits introduced in Section 2.2 provide the foundation for computing many other limits, particularly the ones needed in Chapter 3. The next few sections present a definition of the limit that pertains to cases other than finding the slope of a tangent line, explores continuous functions (Section 2.4) and three fundamental properties of continuous functions (Section 2.5). We conclude, in Section 2.6, with a first look at graphing functions by hand using intercepts, symmetry, and asymptotes and with the use of technology.

2.1 Slope at a Point on a Curve

The slope of a (straight) line is simply the quotient of "rise over run", as shown in Figure 2.1.1(a).



(a) (b)

Figure 2.1.1 slope $= \frac{\text{rise}}{\text{run}}$; (a) positive slope, (b) negative slope.

It does not matter which point P is chosen on the line. If the line goes down as you move from left to right the "rise" is considered to be negative and the slope is negative. This is the case in Figure 2.1.1(b).

The slope of US Interstates never exceeds 6%=0.06. This means the road can rise (or fall) at most 6 feet in 100 (horizontal) feet, see Figure 2.1.2(a). On the other hand the steepest street in San Francisco is Filbert Street, with a slope of 0.315, see Figure 2.1.2(b).

(a) (b)

Figure 2.1.2 (a) Steepest US interstate has slope 0.06 and (b) Filbert Street has slope 0.315. [EDITOR: Replace with annotated pictures.]

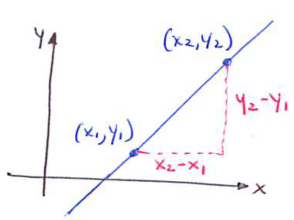


Figure 2.1.3

Now consider a line L placed in an xy-coordinate system, as in Figure 2.1.3. Since two points determine the line, they also determine its slope.

To find that slope pick any two distinct points on the line, (x_1, y_1) and (x_2, y_2). As Figure 2.1.3 shows, they determine a rise of $y_2 - y_1$ and a run of $x_2 - x_1$, hence

$$\text{slope} = \frac{\text{rise}}{\text{run}} = \frac{y_2 - y_1}{x_2 - x_1}.$$

Note that the run could be negative too; that occurs if $x_2 - x_1$ and $y_2 - y_1$ have opposite signs.

EXAMPLE 1 Find the slope of the line through $(4, -1)$ and $(1, 3)$.
SOLUTION Figure 2.1.4 shows two points and the line they determine. Let $(4, -1)$ be (x_1, y_1) and let $(1, 3)$ be (x_2, y_2). So the slope is

$$\frac{3 - (-1)}{1 - 4} = \frac{4}{-3} = -\frac{4}{3}.$$

That the slope is negative is consistent with Figure 2.1.4 which shows that the line descends as you go from left to right. ◇

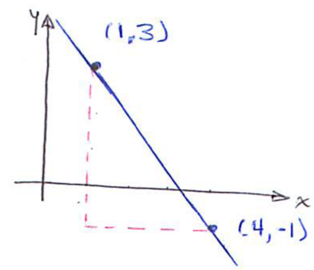


Figure 2.1.4

Note that the slope in Example 1 does not change if $(4, -1)$ is called (x_2, y_2) and $(1, 3)$ is called (x_1, y_1).

If we knew a point on a line and its slope we can draw the line. For instance, say we know a line goes through $(1, 2)$ and has slope 1.4, which is 7/5. We draw a triangle with a vertex at $(1, 2)$ and legs parallel to the axes, as in Figure 2.1.5. The rise and run of the triangle could be 7 and 5, or 1.4 and 1, or any two numbers in the ratio 1.4 : 1.

If we know a point on a line, say (a, b), and the slope of the line, m, we can draw the line and also write its equation. Any point (x, y) on the line, other than (a, b), together with the point (a, b) determine, the slope of the line:

$$\text{slope} = \frac{y - b}{x - a} = m.$$

The equation can be written as

$$y - b = m(x - a) \qquad \text{or} \qquad y = m(x - a) + b.$$

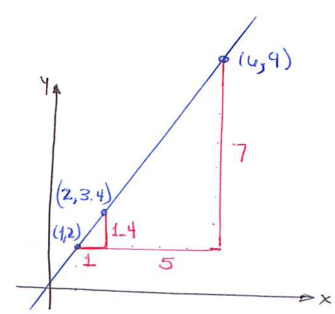


Figure 2.1.5

The slope of a line will be useful when we consider tangents to curves.

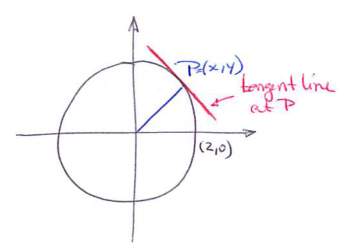


Figure 2.1.6

Slope at Points on a Circle

Consider a circle with radius 2 and center at the origin $(0, 0)$, as shown in Figure 2.1.6. How do we find the tangent line to the circle at $P = (x, y)$? By "tangent line" we mean, informally, the line that most closely resembles the curve near P. The tangent line is perpendicular to the line OP, and the slope of OP is y/x. Thus the slope of the tangent line at (x, y) is $-x/y$. (Exercise 21 shows that the product of the slopes of perpendicular lines is -1 assuming neither line has slope 0.) For instance, at $(0, 2)$ the slope is $-0/2 = 0$, which records that the tangent line at $(0, 2)$, is horizontal, that is, the tangent line at the top of the circle is parallel to the x-axis.

We say that the slope of the circle at (x, y) is $-y/x$ because that is the slope of the tangent line at this point.

For this special curve we could find the tangent line first, and then its slope. If we had been able to find the slope of the tangent line first, we would then be able to draw the tangent line. That is what we will have to do for other curves, like the three considered next.

The Slope at a Point on the Curve $y = x^2$

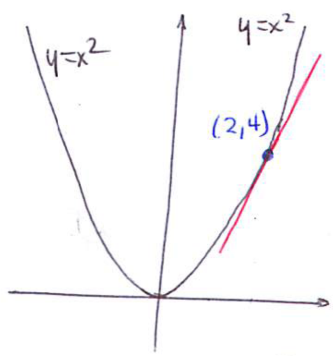
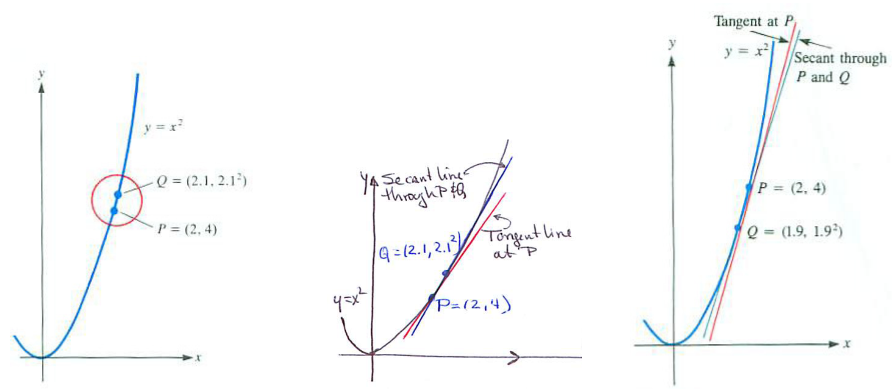


Figure 2.1.7

Figure 2.1.7 shows the graph of $y = x^2$. How can we find the slope of the tangent line at $(2, 4)$? If we knew that slope, we could draw the tangent.

If we know two points on the tangent, we could calculate its slope. But we know only one point on that line, namely $(2, 4)$. To get around this difficulty we will choose a point Q on the parabola $y = x^2$ near P and compute the slope of the line through P and Q. Such a line is called a **secant**. As Figure 2.1.8 suggests that when Q is near P, such a secant line resembles the tangent line at $(2, 4)$. For instance, choose $Q = (2.1, 2.1^2)$ and compute the slope of the



(a) (b) (c)

Figure 2.1.8

line through P and Q shown in Figure 2.1.8(b).

$$\text{Slope of secant} \quad = \quad \frac{\text{Change in } y}{\text{Change in } x} = \frac{2.1^2 - 2^2}{2.1 - 2} = \frac{4.41 - 4}{0.1} = \frac{0.41}{0.1} = 4.1.$$

Thus an estimate of the slope of the tangent line is 4.1. If you look at Figure 2.1.8, you will see that this is an overestimate of the slope of the tangent line. So the slope of the tangent line is less than 4.1.

We can also choose the point Q on the parabola to the left of $P = (2, 4)$. For instance, choose $Q = (1.9, 1.9^2)$. (See Figure 2.1.8(c).) Then

$$\text{slope of secant} \quad = \quad \frac{\text{Change in } y}{\text{Change in } x} = \frac{1.9^2 - 2^2}{1.9 - 2} = \frac{3.61 - 4}{-0.1} = \frac{-0.39}{-0.1} = 3.9.$$

Inspecting Figure 2.1.8(c) shows that this underestimates the slope of the tangent line. So the slope of the tangent line is greater than 3.9.

We have trapped the slope of the tangent line between 3.9 and 4.1. To get closer bounds we choose Q even nearer to $(2, 4)$.

Using $Q = (2.01, 2.01^2)$ leads to the estimate

$$\frac{2.01^2 - 2^2}{2.01 - 2} = \frac{4.0401 - 4}{0.01} = \frac{0.0401}{0.01} = 4.01$$

and using $Q = (1.99, 1.99^2)$ yields the estimate

$$\frac{1.99^2 - 2^2}{1.99 - 2} = \frac{3.9601 - 4}{-0.01} = \frac{-0.0399}{-0.01} = 3.99.$$

Now we know the slope of the tangent at $(2, 4)$ is between 3.99 and 4.01.

To make better estimates we could choose Q even nearer to $(2, 4)$, say $(2.0001, 2.0001^2)$. But, still, the slopes we would get would just be estimates.

What we need to know is what happens to the quotient

$$\frac{x^2 - 2^2}{x - 2} \qquad \text{as } x \text{ gets closer and closer to 2.}$$

This chapter is devoted to answering this and other questions of the type: "What happens to the values of a function as the inputs are chosen nearer and nearer to some fixed number?"

The Slope at a Point on the Curve $y = 1/x$

Figure 2.1.9 shows the graph of $y = 1/x$. Let us estimate the slope of the tangent line to this curve at $(3, 1/3)$.

It's clear that the slope will be negative. We could draw a run-rise triangle on the tangent and get an estimate for the slope. But let's use the nearby point Q method because we can get better estimates that way.

We pick $Q = (3.1, 1/(3.1))$. The points $P = (3, 1/3)$ and Q determine a secant whose slope is

$$\frac{\frac{1}{3} - \frac{1}{3.1}}{3 - 3.1} = \frac{\frac{0.1}{3(3.1)}}{-0.1} = -\frac{1}{3(3.1)} = -\frac{1}{9.3}.$$

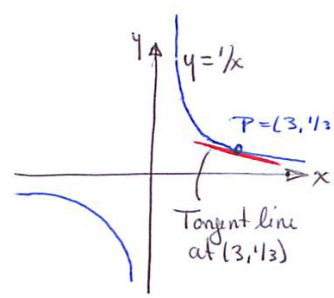


Figure 2.1.9

That's just an estimate of the slope of the tangent line.

Using $Q = (2.9, 1/2.9)$, we get another estimate:

$$\frac{\frac{1}{3} - \frac{1}{2.9}}{3 - 2.9} = \frac{\frac{-0.1}{3(2.9)}}{0.1} = -\frac{1}{3(2.9)} = -\frac{1}{8.7}.$$

By choosing Q nearer $(3, 1/3)$ we could get better estimates.

The Slope at a Point on the Curve $y = \log_2(x)$

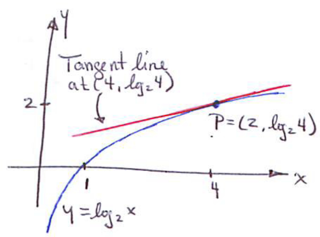


Figure 2.1.10

Figure 2.1.10 shows the graph of $y = \log_2(x)$. Clearly, its slope is positive at all points.

We will make two estimates of the slope at $(4, \log_2(4))$. Before going any further, observe that $(4, \log_2(4)) = (4, 2)$ (because $\log_2(4) = \log_2(2^2) = 2$).

For the nearby point Q, let us use $(4.001, \log_2(4.001))$. The slope of the secant through $P = (4, 2)$ and Q is

$$\frac{\log_2(4.001) - 2}{4.001 - 4} = \frac{\log_2(4.001) - 2}{0.001}.$$

We use a calculator to estimate $\log_2(4.001)$. First, we have, by Exercise 22 in Section 1.2, to five decimal places,

$$\log_2(4.001) = \frac{\log_{10}(4.001)}{\log_{10}(2)} \approx \frac{0.60217}{0.30103} \approx 2.00036.$$

So the estimate of the slope of the tangent to $y = 1/x$ at $(2, 4)$ is

$$\frac{2.00036 - 2}{0,001} = \frac{0.00036}{0.001} = 0.36.$$

The number 0.36 is an estimate of the slope of the graph of $y = \log_2(x)$ at $P = (4, \log_2(4))$. It is not the slope there, but, even so, it could help us draw the tangent at P.

Summary

We introduced the "nearby point Q" method to estimate the slope of the tangent line to a curve at a given point P on the curve. The closer Q is to P, the better the estimate. We applied the techniques to the curves $y = x^2$, $y = 1/x$, and $y = \log_2(x)$. Note that in no case did we have to draw the curve. Nor did we find the slope of the tangent except in the special cases of a line and a circle. We found only estimates. The rest of this chapter develops methods for finding what happens to a function, such as $f(x) = (x^2 - 4)/(x - 2)$, as the argument gets nearer and nearer a given number.

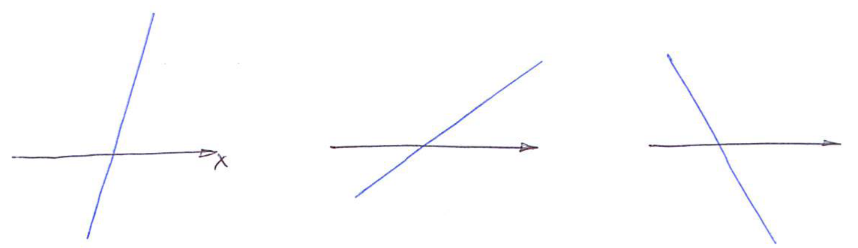
EXERCISES for Section 2.1

1. Draw an x axis and lines of slope $1/2$, 1, 2, 4, 5, -1, and $-1/2$.

2. Draw an x axis and lines of slope $1/3$, 1, 3, -1, and $-2/3$.

In Exercises 3 and 4 copy the figure and estimate the slope of each line as well as you can. In each case draw a "run–rise" triangle and measure the rise and run with a ruler. (A centimeter ruler is more convenient than one marked in inches.)

3.



(a) (b) (c)

4.

(a) (b) (c)

In Exercises 5 to 8 draw the line determined by the given information and give an equation for the line.

5. through $(1, 2)$ with slope -3

6. through $(1, 4)$ and $(4, 1)$

7. through $(-2, -4)$ and $(0, 4)$

8. through $(2, -1)$ with slope 4

9.

(a) Graph the line whose equation is $y = 2x + 3$.

(b) Find the slope of this line.

10.

(a) Graph the line whose equation is $y = -3x + 1$.

(b) Find the slope of this line.

11. Estimate the slope of the tangent line to $y = x^2$ at $(1, 1)$ using the nearby points $(1.001, 1.001^2)$ and $(0.999, 0.999^2)$.

12. Estimate the slope of the tangent line to $y = x^2$ at $(-3, 9)$ using the nearby points $(-3.01, (-3.01)^2)$ and $(-2.99, (-2.99)^2)$.

13. Estimate the slope of the tangent line to $y = 1/x$ at $(1, 1)$

(a) by drawing a tangent line at $(1, 1)$ and a rise-run triangle.

(b) by using the nearby point $(1.01, 1/1.01)$. (Is the slope of the tangent line smaller or larger than this estimate?)

14. Estimate the slope of the tangent line to $y = 1/x$ at $(0.5, 2)$

(a) by drawing a tangent line at $(0.5, 2)$ and a rise-run triangle.

(b) by using the nearby point $(0.49, 1/0.49)$. (Is the slope of the tangent line smaller or larger than this estimate?)

15. Estimate the slope of the tangent line to $y = \log_2(x)$ at $(2, \log_2(2))$

(a) by drawing a tangent line at $(2, \log_2(2))$ and a rise-run triangle.

(b) by using the nearby point $(2.01, \log_2(2.01))$. (Is the slope of the tangent line smaller or larger than this estimate?)

16. Estimate the slope of the tangent line to $y = \log_2(x)$ at $(4, 2)$

(a) by drawing a tangent line at $(4, 2)$ and a rise-run triangle.

(b) by using the nearby point $(3.99, \log_2(3.99))$. (Is the slope of the tangent line smaller or larger than this estimate?)

17.

(a) Graph $y = x^2$ carefully for x in $[-2, 3]$.

(b) Draw the tangent line to $y = x^2$ at $(1, 1)$ as well as you can and estimate its slope.

(c) Using the nearby points $(1.1, 1.1^2)$ and $(0.9, 0.9^2)$, estimate the slope of the tangent line at $(1, 1)$. (Is the slope of the tangent line smaller or larger than this estimate?)

18.

(a) Graph $y = 2^x$ carefully for x in $[0, 2]$.

(b) Draw the tangent line to $y = 2^x$ at $(1, 2)$ as well as you can and estimate its slope.

(c) Using the nearby point $(1.03, 2^{1.03})$, estimate the slope of the tangent line at $(1, 2)$. (Is the slope of the tangent line smaller or larger than this estimate?)

19.

(a) Show that if you compute the slope of the line through $P = (1, 2)$ and $Q = (5, 3)$, you will get the same answer with either choice of labeling.

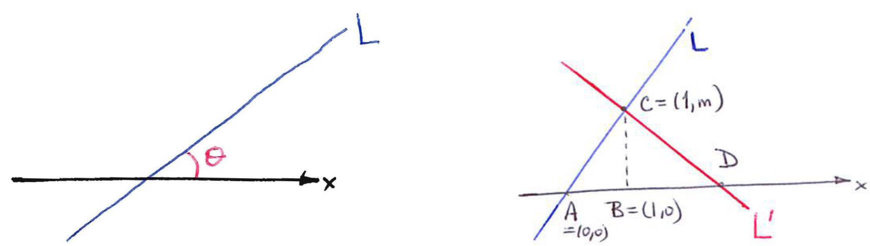
(b) Show that in general both ways of labeling the points P and Q give the same slope.

20. The angle between a line L that crosses the x axis and the x axis is called its **angle of inclination**. It is measured counterclockwise from the positive x axis to the line, as shown in Figure 2.1.11(b). The symbol θ denotes both the angle and its measure, $0 < \theta < \pi$. For a line parallel to the x axis, θ is defined to be 0. Show that $\tan(\theta)$ equals the slope of the line.

21. (This exercise shows that the product of the slopes of perpendicular lines is -1.) Let one line, L, have the positive slope m. Let L' be a line through $(1, m)$ perpendicular to L of slope m'. For convenience, we assume L goes through the origin. Note that the point $(1, m)$ also lies on L. (See Figure 2.1.11(b).)

(a) Use similar triangles ABC and BCD to show that L' crosses the x-axis at $(1 + m^2, 0)$.

(b) Show that the slope of L' is $-1/m$. Thus $mm' = -1$.



(a) (b)

Figure 2.1.11

2.2 Four Special Limits

This section develops the notion of a limit of a function, using four examples that play a key role in Chapter 3.

A Limit Involving x^n

Let a and n be fixed numbers, with n a positive integer.

$$\text{What happens to the quotient } \frac{x^n - a^n}{x - a} \text{ as } x \text{ is chosen nearer and nearer to } a? \tag{2.2.1}$$

To keep the reasoning down-to-earth, let's look at a typical concrete case:

$$\text{What happens to } \frac{x^3 - 2^3}{x - 2} \text{ as } x \text{ gets closer and closer to 2?} \tag{2.2.2}$$

As x approaches 2, the numerator approaches $2^3 - 2^3 = 0$. Because 0 divided by anything (other than 0) is 0 we suspect that the quotient may approach 0. But the denominator approaches $2 - 2 = 0$. This is unfortunate because division by zero is not defined.

That $x^3 - 2^3$ approaches 0 as x approaches 2 may make the quotient small. That the denominator approaches 0 as x approaches 2 may make the quotient very large. How these two opposing forces balance determines what happens to the quotient (2.2.2) as x approaches 2.

We have already seen that it is pointless to replace x in (2.2.2) by 2 as this leads to $(2^3 - 2^3)/(2 - 2) = 0/0$, a meaningless expression.

Instead, let's do some experiments and see how the quotient behaves for specific values of x near 2; some less than 2, some more than 2. Table 2.2.1 shows the results as x increases from 1.9 to 2.1. You are invited to fill in the empty squares in the table below and add to the list with values of x even closer to 2.

Math is not a spectator sport. Check some of the calculations reported in Table 2.2.1.

The cases with $x = 1.99$ and 2.01, being closest to 2, should provide the best estimates of the quotient. This suggests that the quotient (2.2.2) approaches a number near 12 as x approaches 2, whether from below or from above.

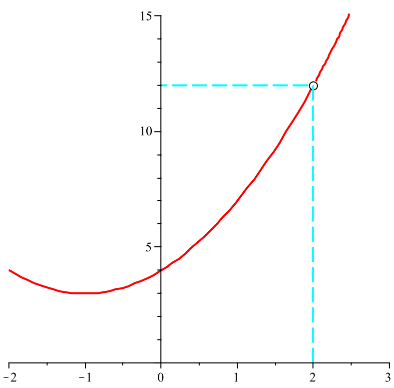
Figure 2.2.1 provides a graphical view of this problem. In (a), the graph of $(x^3 - 2^3)/(x - 2)$ for x between -2 and 3 is a parabola with the point for $x = 2$ deleted. Confirmation that the values of the quotient approach the same value when x approaches 2 from both the left and from the right is obtained by zooming in to a smaller interval around $x = 2$. In (b), the quotient is plotted for x between 1.6 and 2.4. On this scale it seems reasonable that the quotient approaches 12 as x approaches 2.

While the numerical and graphical evidence is suggestive, this question can be answered once-and-for-all with a little bit of algebra. By the formula for

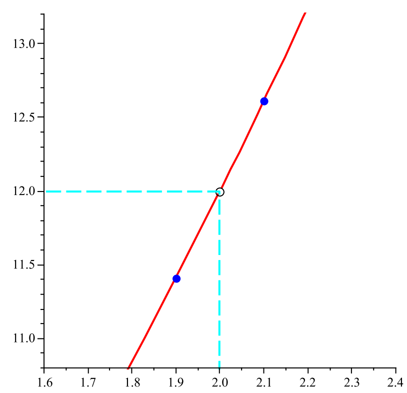
x	x^3	$x^3 - 2^3$	$x - 2$	$\frac{x^3-2^3}{x-2}$
1.90	6.859	-1.141	-0.1	11.41
1.99	7.8806	-0.1194	-0.01	11.94
1.999				
2.00	8.0000	0.0000	0.00	undefined
2.001				
2.01	8.1206	0.1206	0.01	12.06
2.10	9.261	1.261	0.1	12.61

Table 2.2.1 Table showing the steps in the evaluation of $\frac{x^3-2^3}{x-2}$ for four choices of x near 2.

A hollow dot on a graph indicates that that point is NOT on the graph.



(a)



(b)

Figure 2.2.1 The graph of a $y = \frac{x^3-2^3}{x-2}$ suggests that the quotient approaches 12 as x approaches 2. In (b), zooming for x near 2 shows how the data in Table 2.2.1 also suggests the quotient approaches 12 as x approaches 2.

the sum of a geometric series (see (1.4.3) in Section 1.4), $x^3 - 2^3 = (x-2)(x^2 + 2x + 2^2)$. We have

$$\frac{x^3 - 2^3}{x - 2} = \frac{(x-2)(x^2 + 2x + 2^2)}{x - 2} \qquad \text{for all } x \text{ other than } 2. \tag{2.2.3}$$

When x is not 2, (2.2.3) is meaningful, and we can cancel the $(x-2)$, showing that

$$\frac{x^3 - 2^3}{x - 2} = x^2 + 2x + 2^2, \qquad x \neq 2.$$

It is easy to see what happens to $x^2 + 2x + 2^2$ as x gets nearer and nearer to 2: $x^2 + 2x + x^2$ approaches $4 + 4 + 4 = 12$. This agrees with the calculations (see Table 2.2.1).

We say "the limit of $(x^3 - 2^3)/(x - 2)$ as x approaches 2 is 12" and use the shorthand

$$\begin{aligned} \lim_{x\to 2} \frac{x^3 - 2^3}{x - 2} &= \lim_{x\to 2}(x^2 + 2x + 2^2) \qquad (2.2.4) \\ &= 3 \cdot 2^2 = 12. \qquad (2.2.5) \end{aligned}$$

Similar algebra, depending on the formula for the sum of a geometric series, yields

For any positive integer n and fixed number a,

$$\lim_{x\to a} \frac{x^n - a^n}{x - a} = n \cdot a^{n-1} \tag{2.2.6}$$

See also Exercises 42 and 43.

A Limit Involving b^x

What happens to $\frac{2^x - 1}{x}$ and to $\frac{4^x - 1}{x}$ as x approaches 0?

Consider $(2^x - 1)/x$ first: As x approaches 0, $2^x - 1$ approaches $2^0 - 1 = 1 - 1 = 0$. Since the numerator and denominator in $(2^x - 1)/x$ both approach 0 as x approaches 0, we face the same challenge as with $(x^3 - 2^3)/(x - 2)$. There is a battle between two opposing forces.

There are no algebraic tricks to help in this case. Instead, we will rely upon numerical data. While this motivation will be convincing, it is not mathematically rigorous. Later, we will present a way to evaluate these limits that does not depend upon any numerical computations.

Table 2.2.2 records some results (rounded off) for four choices of x. You are invited to fill in the blanks and to add values of x even closer to 0.

You should also take some time to examine the two graphs of $(2^x - 1)/x$ to convince yourself that this quotient approaches a single value as x approaches 0 from the left and from the right. Also, note how the view in Figure 2.2.2(b) provides an better estimate of the y-coordinate of the missing point.

x	2^x	$2^x - 1$	$\frac{2^x-1}{x}$
-0.01	0.993093	-0.006907	0.691
-0.001	0.999307	-0.000693	0.693
-0.0001			
0.0001			
0.001	1.000693	0.000693	0.693
0.01	1.006956	0.006956	0.696

Table 2.2.2 Numerical evaluation of $(2^x - 1)/x$ for four different choices of x. The numbers in the last column are rounded to three decimal places. See also Figure 2.2.2.

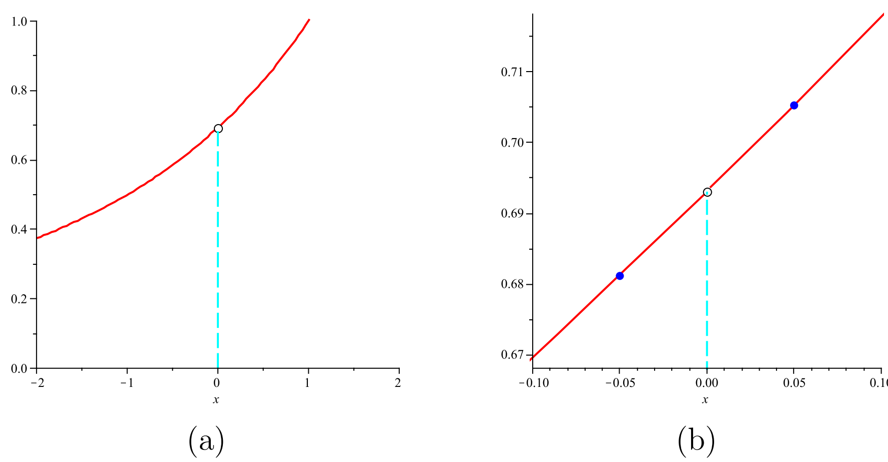


Figure 2.2.2 (a) Graph of $y = (2^x - 1)/x$ for x near 0. (b) View for x nearer to 0, with the data points from Table 2.2.2. Note that there is no point for $x = 0$ since the quotient is not defined when x is 0.

WARNING *(Do not believe your eyes!)* The graphs in Figure 2.2.1(b) and Figure 2.2.2(b) are not graphs of straight lines. They look straight only because the viewing windows are so small.

> Compare the labels on the axes in the two views in each of Figure 2.2.1 and Figure 2.2.2. That the graphs of many common functions look straight as you zoom in on a point will be important in Section 3.1.

It seems that as x approaches 0, $(2^x - 1)/x$ approaches a number whose decimal value begins 0.693. We write

$$\lim_{x \to 0} \frac{2^x - 1}{x} \approx 0.693 \qquad \text{rounded to three decimal places.} \tag{2.2.7}$$

It is then a simple matter to find

$$\lim_{x \to 0} \frac{4^x - 1}{x}.$$

In view of the factoring of the difference of two squares, $a^2 - b^2 = (a-b)(a+b)$, we have $4^x - 1 = (2^x)^2 - 1^2 = (2^x - 1)(2^x + 1)$. Hence

$$\frac{4^x - 1}{x} = \frac{(2^x - 1)(2^x + 1)}{x} = (2^x + 1)\frac{2^x - 1}{x}.$$

As $x \to 0$, $2^x + 1$ approaches $2^0 + 1 = 1 + 1 = 2$ and $(2^x - 1)/x$ approaches (approximately) 0.693. Thus,

$$\lim_{x \to 0} \frac{4^x - 1}{x} \approx 2 \cdot 0.693 \approx 1.386 \qquad \text{rounded to three decimal places.}$$

We now have strong evidence about the values of $\lim_{x \to 0} \dfrac{b^x - 1}{x}$ for $b = 2$ and $b = 4$. They suggest that the larger b is, the larger the limit is. Since $\lim_{x \to 0} \frac{2^x - 1}{x}$ is less than 1 and $\lim_{x \to 0} \frac{4^x - 1}{x}$ is more than 1, it seems reasonable that there should be a value of b such that $\lim_{x \to 0} \frac{b^x - 1}{x} = 1$. This special number is called e, **Euler's number**. We know that e is between 2 and 4 and that $\lim_{x \to 0} \frac{e^x - 1}{x} = 1$. It turns out that e is an irrational number with an endless decimal representation that begins $2.718281828\ldots$. In Chapter 3 we will see that e is as important in calculus as π is in geometry and trigonometry.

Euler named this constant e, but no one knows why he chose this symbol.

In any case we have

Basic Property of e

$$\lim_{x \to 0} \frac{e^x - 1}{x} = 1, \qquad \text{and } e \approx 2.71828.$$

In Section 1.2 it was remarked that the logarithm with base b, $\log_b$, can be defined for any base $b > 0$. The logarithm with base $b = e$ deserves special attention. The $\log_e(x)$ is called the **natural logarithm**, and is typically written as $\ln(x)$ or $\log(x)$. Thus, in particular,

$$y = \ln(x) \quad \text{is equivalent to} \quad x = e^y.$$

Note that, as with any logarithm function, the domain of ln is the set of positive numbers $(0, \infty)$ and the range is the set of all real numbers $(-\infty, \infty)$.

SHERMAN: Where do we show this? Not in this Exercise?

In Exercise 45 it is shown that $\lim_{x\to 0} \frac{b^x-1}{x}$ is $\ln(b)$ for any positive number b.

Often the exponential function with base e is written as exp. This notation is convenient when the input is complicated:

$$\exp\left(\frac{\sin^3(\sqrt{x})}{\cos(x)}\right) \qquad \text{is easier to read than} \qquad e^{\sin^3(\sqrt{x})/\cos(x)}.$$

Many calculators and computer languages use exp to name the exponential function with base e.

Three Important Bases for Logarithms

While logarithms can be defined for any positive base, three numbers have been used most often: 2, 10, and e. Logarithms to the base 2 are used in information theory, for they record the number of "yes – no" questions needed to pinpoint a particular piece of information. Base 10 was used for centuries to assist in computations. Since the decimal system is based on powers of 10, certain convenient numbers had obvious logarithms; for instance, $\log_{10}(1000) = \log_{10}(10^3) = 3$. Tables of logarithms to several decimal places facilitated the calculations of products, quotients, and roots. To multiply two numbers, you looked up their logarithms, and then searched the table for the number whose logarithm was the sum of the two logarithms. The calculator made the tables obsolete, just as it sent the slide rule into museums. However, a Google search for "slide rule" returns a list of more than 15 million websites full of history, instruction, and sentiment. The number e is the most convenient base for logarithms in calculus. Euler, as early as 1728, used e for the base of logarithms.

A Limit Involving $\sin(x)$

What happens to $\frac{\sin(x)}{x}$ as x gets nearer and nearer to 0?

Here x represents an angle, measure in radians. In Chapter 3 we will see that in calculus radians are much more convenient than degrees.

Consider first $x > 0$. Because we are interested in x near 0, we assume that $x < \pi/2$. Figure 2.2.3 identifies both x and $\sin(x)$ on a circle of radius 1, the **unit circle**.

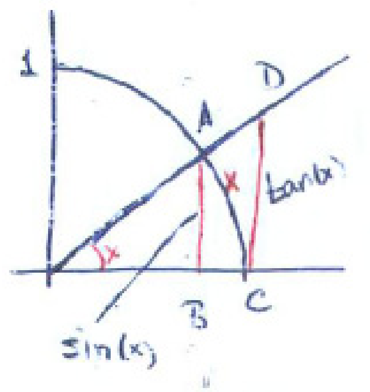


Figure 2.2.3 On the circle with radius 1, (a) x is the arclength subtended by an angle of x radians and $\sin(x) = \overline{AB}$.

To get an idea of the value of this limit, let's try $x = 0.1$. Setting our calculator in the "radian mode", we find

$$\frac{\sin(0.1)}{0.1} \approx \frac{0.099833}{0.1} = 0.99833. \qquad (2.2.8)$$

Likewise, with $x = 0.01$,

$$\frac{\sin(0.1)}{0.01} \approx \frac{0.0099998}{0.01} = 0.99998. \qquad (2.2.9)$$

These results lead us to suspect maybe this limit is 1.

Geometry and a bit of trigonometry show that $\lim_{x\to 0} \frac{\sin(x)}{x}$ is indeed 1. First, using Figure 2.2.3, we show that $\frac{\sin(x)}{x}$ is less than 1 for x between 0 and $\pi/2$. Recall that $\sin(x) = \overline{AB}$. Now, $\overline{AB}$ is shorter than $\overline{AC}$, since a leg of a right triangle is shorter than the hypotenuse. Then $\overline{AC}$ is shorter than the circular arc joining A to C, since the shortest distance between two points is a straight line. Thus,

$$\sin(x) < \overline{AC} < x.$$

So $\sin(x) < x$. Since x is positive, dividing by x preserves the inequality. We have

$$\frac{\sin(x)}{x} < 1. \qquad (2.2.10)$$

Next, we show that $\frac{\sin(x)}{x}$ is greater than something which gets near 1 as x approaches 0. Figure 2.2.3 again helps with this step.

The area of triangle OCD is greater than the area of the sector OCA. (The area of a sector of a disk of radius r subtended by an angle θ is $\theta r^2/2$.) Thus

$$\underbrace{\frac{1}{2} \cdot 1 \cdot \tan(x) >}_{\text{area of } \Delta OCD} \quad \underbrace{\frac{x \cdot 1^2}{2}}_{\text{area of sector } OCA} \quad .$$

Multiplying this inequality by 2 simplifies it to

$$\tan(x) > x.$$

In other words,

$$\frac{\sin(x)}{\cos(x)} > x.$$

Now, multiplying by $\cos(x)$, which is positive, and dividing by x (also positive) gives

$$\frac{\sin(x)}{x} > \cos(x). \tag{2.2.11}$$

Putting (2.2.10) and (2.2.11) together, we have

$$\cos(x) < \frac{\sin(x)}{x} < 1. \tag{2.2.12}$$

Since $\cos(x)$ approaches 1 as x approaches 0 , $\frac{\sin(x)}{x}$ is squeezed between 1 and something that gets closer and closer to 1, $\frac{\sin(x)}{x}$ must itself approach 1.

We still must look at $\frac{\sin(x)}{x}$ for $x < 0$ as x gets nearer and nearer to 0. Define u to be $-x$. Then u is positive, and

$$\frac{\sin(x)}{x} = \frac{\sin(-u)}{-u} = \frac{-\sin u}{-u} = \frac{\sin u}{u}.$$

As x is negative and approaches zero, u is positive and approaches 0. Thus $\frac{\sin(x)}{x}$ approaches 1 as x approaches 0 through positive or negative values.

In short,

$$\boxed{\lim_{x\to 0} \frac{\sin(x)}{x} = 1 \qquad \text{where the angle, } x \text{, is measured in radians.}}$$

A Limit Involving $\cos(x)$

Knowing that $\lim_{x\to 0} \frac{\sin(x)}{x} = 1$, we can show that

$$\boxed{\lim_{x\to 0} \frac{1-\cos(x)}{x} = 0. \qquad (2.2.13)}$$

All we will say about this limit now is that the numerator, $1 - \cos(x)$ is the length of BC in Figure 2.2.3. Exercises 29 and 30 outline how to establish this limit.

The Meaning of $\lim_{x \to \to 0} \frac{\sin(x)}{x} = 1$

When x is near 0, $\sin(x)$ and x are both small. That their quotient is near 1 tells us much more, namely, that x is a "very good approximation of $\sin(x)$."

That means that the difference $\sin(x) - x$ is small, even in comparison to $\sin(x)$. In other words, the "relative error"

$$\frac{\sin(x) - x}{\sin(x)} \tag{2.2.14}$$

approaches 0 as x approaches 0.

To show that this is the case, we compute

$$\lim_{x \to 0} \frac{\sin(x) - x}{\sin(x)}.$$

We have

$$\begin{aligned}
\lim_{x \to 0} \frac{\sin(x) - x}{\sin(x)} &= \lim_{x \to 0} \left(\frac{\sin(x)}{\sin(x)} - \frac{x}{\sin(x)} \right) \\
&= \lim_{x \to 0} \left(1 - \frac{x}{\sin(x)} \right) \\
&= \lim_{x \to 0} \left(1 - \frac{1}{\left(\frac{x}{\sin(x)} \right)} \right) \\
&= 1 - \frac{1}{1} = 0.
\end{aligned}$$

As you may check by graphing, the relative error in (2.2.14) stays less than 1 percent for x less than 0.24 radians, just under 14 degrees.

It is often useful to replace $\sin(x)$ by the much simpler quantity x. For instance, the force tending to return a swinging pendulum is proportional to $\sin(\theta)$, where θ is the angle that the pendulum makes with the vertical. As one physics book says, "If the angle is small, $\sin(\theta)$ is nearly equal to θ"; it then replaces $\sin(\theta)$ by θ.

Summary

This section discussed four important limits:

$$\begin{aligned}
\lim_{x\to a}\frac{x^n-a^n}{x-a} &= na^{n-1} && (n \text{ a positive integer})\\
\lim_{x\to 0}\frac{e^x-1}{x} &= 1 && (e\approx 2.71828)\\
\lim_{x\to 0}\frac{\sin(x)}{x} &= 1 && (\text{angle in radians})\\
\lim_{x\to 0}\frac{1-\cos(x)}{x} &= 0 && (\text{angle in radians}).
\end{aligned}$$

That $\lim_{x\to 0}\frac{e^x-1}{x}=1$ says, informally, that $\frac{\text{exp(a small number)}-1}{\text{same small number}}$ is near 1.

Each of these limits will be needed in Chapter 3, which introduces the derivative of a function.

The next section examines the general notion of a limit. This is the basis for all of calculus.

EXERCISES for Section 2.2

In each of Exercises 1 to 10 describe the two opposing forces involved in the limit. If you can figure out the limit on the basis of results in this section, give it. Otherwise, use a calculator to estimate the limit.

1. $\lim_{x \to 2} \frac{x^4 - 16}{x - 2}$

2. $\lim_{x \to 0} \frac{\sin(x)}{x \cos(x)}$

3. $\lim_{x \to 0} (1 - x)^{1/x}$

4. $\lim_{x \to 0} (\cos(x))^{1/x}$

5. $\lim_{x \to 0} x^x$, $x > 0$

6. $\lim_{x \to 0} \frac{\arcsin(x)}{x}$

7. $\lim_{x \to 0} \frac{\tan(x)}{x}$ (Write $\tan(x) = \sin(x)/\cos(x)$.)

8. $\lim_{x \to 0} \frac{\tan(2x)}{x}$

9. $\lim_{x \to 0} \frac{8^x - 1}{2^x - 1}$ (The numerator is the difference of two cubes; how does $b^3 - a^3$ factor?)

10. $\lim_{x \to 0} \frac{9^x - 1}{3^x - 1}$

Exercises 11 to 15 concern $\lim_{x \to a} \frac{x^n - a^n}{x - a}$.

11. Using the factorization $(x - a)(x + a) = x^2 - a^2$ find $\lim_{x \to a} \frac{x^2 - a^2}{x - a}$.

12. Using Exercise 11,

(a) find $\lim_{x \to 3} \frac{x^2 - 9}{x - 3}$

(b) find $\lim_{x \to \sqrt{3}} \frac{x^2 - 3}{x - \sqrt{3}}$

13.

(a) By multiplying it out, show that $(x - a)(x^2 + ax + a^2) = x^3 - a^3$.

(b) Use (a) to show that $\lim_{x \to a} \frac{x^3 - a^3}{x - a} = 3a^2$.

(c) By multiplying it out, show that $(x - a)(x^3 + ax^2 + a^2x + a^3) = x^4 - a^4$.

(d) Use (c) to show that $\lim_{x\to a} \frac{x^4-a^4}{x-a} = 4a^3$.

14.

(a) What is the domain of $(x^2-9)/(x-3)$?

(b) Graph $(x^2-9)/(x-3)$.

Use a hollow dot to indicate an absent point in the graph.

15.

(a) What is the domain of $(x^3-8)/(x-2)$?

(b) Graph $(x^3-8)/(x-2)$.

Exercises 16 to 19 concern $\lim\limits_{x\to 0} \dfrac{e^x-1}{x}$.

16. What is a definition of the number e?

17. Use a calculator to compute $(2.7^x-1)/x$ and $(2.8^x-1)/x$ for $x = 0.001$. This suggests that e is between 2.7 and 2.8.

18. Use a calculator to estimate $(2.718^x-1)/x$ for $x = 0.1$, 0.01, and 0.001.

19. Graph $y = (e^x-1)/x$ for $x \neq 0$.

Exercises 20 to 31 concern $\lim_{x\to 0} \frac{\sin(x)}{x}$ and $\lim_{x\to 0} \frac{1-\cos(x)}{x}$.

20. Use your calculator to create a graph of $y = \frac{\sin(x)}{x}$.

21. Use your calculator to create a graph of $y = \frac{1-\cos(x)}{x}$.

22. Using the fact that $\lim\limits_{x\to 0} \dfrac{\sin(x)}{x} = 1$, find the limits of the following as x approaches 0.

(a) $\dfrac{\sin(3x)}{3x}$ (Let $u = 3x$.)

(b) $\dfrac{\sin(3x)}{x}$

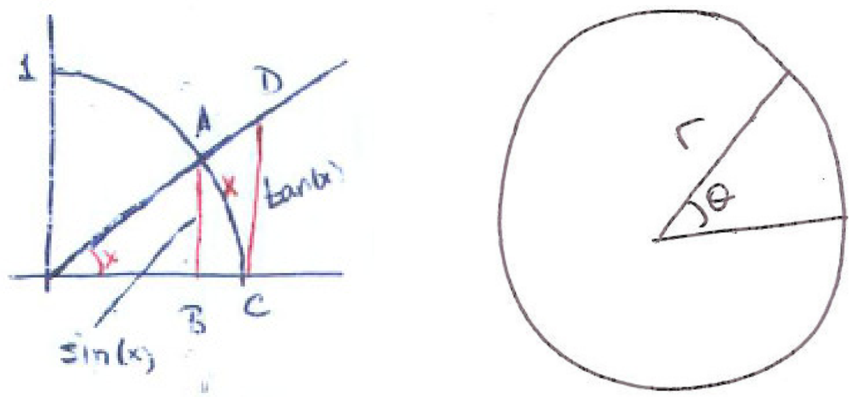
(c) $\dfrac{\sin(3x)}{\sin(x)}$ (Resist the temptation to cancel the sin's. Instead, do a little algebra.)

(d) $\dfrac{\sin^2(x)}{x}$

23. Use $\lim\limits_{x\to 0} \dfrac{e^x-1}{x} = 1$ to find each limit.

(a) $\lim\limits_{x\to 0} \dfrac{e^{10x}-1}{x}$

(b) $\lim\limits_{x\to 0} \dfrac{e^{\pi x}-1}{x}$



(a) (b)

Figure 2.2.4

24. Why is the arc length from A to C in Figure 2.2.4(a) equal to x?

25. Why is the length of CD in Figure 2.2.4(a) equal to $\tan x$?

26. Why is the area of triangle OCD in Figure 2.2.4(a) equal to $(\tan x)/2$?

27. An angle of θ radians in a circle of radius r subtends a sector, as shown in Figure 2.2.4(b). What is the area of this sector? A review of trigonometry is available in an appendix that is available on the web.

28.

(a) Graph $\sin(x)/x$ for x in $[-\pi, 0)$

(b) Graph $\sin(x)/x$ for x in $(0, \pi]$.

(c) How are the graphs in (a) and (b) related?

(d) Graph $\sin(x)/x$ for $x \neq 0$.

29. When $x = 0$, $(1-\cos(x))/x$ is not defined. Estimate $\lim\limits_{x\to 0} \dfrac{1-\cos(x)}{x}$ by evaluating $(1-\cos(x))/x$ at $x = 0.1$ (radians).

30. To find $\lim\limits_{x\to 0} \dfrac{1-\cos(x)}{x}$ first check this algebra and trigonometry:

$$\frac{1-\cos(x)}{x} = \frac{1-\cos(x)}{x}\frac{1+\cos(x)}{1+\cos(x)} = \frac{1-\cos^2(x)}{x(1+\cos(x))} = \frac{\sin^2(x)}{x(1+\cos(x))} = \frac{\sin(x)}{x}\frac{\sin(x)}{1+\cos(x)}.$$

Then show that

$$\lim_{x\to 0}\frac{\sin(x)}{x}\frac{\sin(x)}{1+\cos(x)} = 0.$$

31. Show that

$$\lim_{x\to 0}\frac{1-\cos(x)}{x^2} = \frac{1}{2}.$$

This suggests that, for small values of x, $1-\cos(x)$ is close to $\frac{x^2}{2}$, so that $\cos(x)$ is approximately $1-\frac{x^2}{2}$.

(a) Use a calculator to compare $\cos(x)$ with $1-\frac{x^2}{2}$ for $x = 0.2$ and 0.1 radians. 0.2 radians is about 11°.

(b) Use a graphing calculator to compare the graphs of $\cos(x)$ and $1-\frac{x^2}{2}$ for x in $[-\pi,\pi]$.

(c) What is the largest interval on which the values of $\cos(x)$ and $1-\frac{x^2}{2}$ differ by no more than 0.1? That is, for what values of x is it true that $\left|\cos(x)-(1-\frac{x^2}{2})\right| < 0.1$?

See Exercise 30.

32. The limit $\lim_{\theta\to 0}\frac{\sin(4\theta)}{\sin(\theta)}$ appears in the design of a water sprinkler in the "Calculus is Everywhere" in Chapter 5 Find that limit.

33.

(a) We examined $(2^x-1)/x$ only for x near 0. When x is large and positive 2^x-1 is large. So both the numerator and denominator of $(2^x-1)/x$ are large. The numerator influences the quotient to become large. The large denominator pushes the quotient toward 0. Use a calculator to see how the two forces balance for large values of x.

(b) Sketch the graph of $f(x) = (2^x-1)/x$ for $x > 0$. (Pay special attention to the behavior of the graph for large values of x.)

34.

(a) When x is negative and $|x|$ is large what happens to $(2^x-1)/x$?

(b) Sketch the graph of $f(x) = (2^x-1)/x$ for $x < 0$. (Pay special attention to the behavior of the graph for large negative values of x.)

35.

(a) Using a calculator, explore what happens to $\sqrt{x^2+x}-x$ for large positive values of x.

(b) Show that for $x>0$, $\sqrt{x^2+x}<x+(1/2)$.

(c) Using algebra, find what number $\sqrt{x^2+x}-x$ approaches as x increases. (Multiply $\sqrt{x^2+x}-x$ by $\frac{\sqrt{x^2+x}+x}{\sqrt{x^2+x}+x}$, an operation that removes square roots from the denominator.)

36. Using a calculator, examine the behavior of the quotient $(\theta-\sin(\theta))/\theta^3$ for θ near 0.

37. Using a calculator, examine the behavior of the quotient $\left(\cos(\theta)-1+\frac{\theta^2}{2}\right)/\theta^4$ for θ near 0.

Exercises 38 to 41 concern $f(x)=(1+x)^{1/x}$, x in $(-1,0)$ and $(0,\infty)$.

38.

(a) Why is $(1+x)^{1/x}$ not defined when $x=-3/2$ but is defined when $x=-5/3$. Give an infinite number of $x<-1$ for which it is not defined.

(b) For x near 0, $x>0$, $1+x$ is near 1. So we might expect $(1+x)^{1/x}$ to be near 1 then. However, the exponent $1/x$ is very large. So perhaps $(1+x)^{1/x}$ is also large. To see what happens, fill in this table.

x	1	0.5	0.1	0.01	0.001
$1+x$	2				
$1/x$	1				
$(1+x)^{1/x}$	2				

(c) For x near 0 but negative, investigate $(1+x)^{1/x}$ with the use of this table

x	-0.5	-0.1	-0.01	-0.001
$1+x$	0.5			
$1/x$	-2			
$(1+x)^{1/x}$	4			

39. Graph $y=(1+x)^{1/x}$ for x in $(-1,0)$ and $(0,10)$.

Exercises 38 and 39 show that $\lim_{x\to 0}(1+x)^{1/x}$ is about 2.718. This suggests that the number e may equal $\lim_{x\to 0}(1+x)^{1/x}$. In Section 3.2 we show that this is the case. However, the next two exercises give persuasive arguments for this fact. Unfortunately, each argument has a big hole or "unjustified leap," which you are asked to find.

40. Assume that all we know about the number e is that $\lim_{x\to 0} \frac{e^x-1}{x} = 1$. We will write this as

$$\frac{e^x-1}{x} \sim 1,$$

and read this as "$(e^x - 1)/x$ is close to 1 when x is near 0." Multiplying both sides by x gives

$$e^x - 1 \sim x.$$

Adding 1 to both sides of this gives

$$e^x \sim 1 + x.$$

Finally, raising both sides to the power $1/x$ yields

$$(e^x)^{1/x} \sim (1 + x)^{1/x},$$

hence

$$e \sim (1 + x)^{1/x}.$$

This suggests that

$$e = \lim_{x\to 0} (1 + x)^{1/x}.$$

The conclusion is correct. Most of the steps are justified. Which step is the "big leap"?

41. Assume that $b = \lim_{x\to 0} (1 + x)^{1/x}$. We will "show" that

$$\lim_{x\to 0} \frac{b^x-1}{x} = 1.$$

First of all, for x near (but not equal to) 0

$$b \sim (1 + x)^{1/x}.$$

Then

$$b^x \sim 1 + x.$$

Hence

$$b^x - 1 \sim x.$$

Dividing by x gives

$$\frac{b^x-1}{x} \sim 1.$$

Hence

$$\lim_{x\to 0} \frac{b^x-1}{x} = 1.$$

Where is the "suspect step" this time?

42. Let n be a positive integer and define $P_n(x) = x^{n-1} + ax^{n-2} + a^2x^{n-3} + \cdots + a^{n-2}x + a^{n-1}$. As shown in Section 1.4, this polynomial equals the quotient $\frac{x^n-a^n}{x-a}$. That is, $(x - a)P_n(x) = x^n - a^n$.

(a) Verify that $(x-a)P_2(x) = x^2 - a^2$. (Compare with Exercise 11)

(b) Verify that $(x-a)P_3(x) = x^3 - a^3$. (Compare with Exercise 13(a))

(c) Verify that $(x-a)P_4(x) = x^4 - a^4$. (Compare with Exercise 13(c))

(d) Explain why $(x-a)P_n(x) = x^n - a^n$ for all positive integers n.

43. Using the formula for the sum of a geometric progression ((1.4.3) in Section 1.4), show that $\lim\limits_{x\to a} \dfrac{x^n - a^n}{x-a} = na^{n-1}$.

44. An intuitive argument suggested that $\lim\limits_{\theta\to 0}(\sin\theta)/\theta = 1$, which turned out to be correct. Try your intuition on another limit associated with the unit circle shown in Figure 2.2.5.

(a) What do you think happens to the quotient

$$\frac{\text{Area of triangle } ABC}{\text{Area of shaded region}} \qquad \text{as } \theta \to 0?$$

More precisely, what does your intuition suggest is the limit of that quotient as $\theta \to 0$?

(b) Estimate the limit in (a) using $\theta = 0.01$.

This limit, which arose during some research in geometry, is determined in Exercise 54 in Section 5.6. The authors guessed wrong, as has everyone they asked.

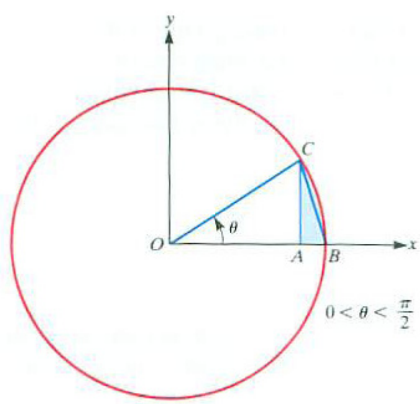


Figure 2.2.5

45. Show that $\lim_{x\to 0} \frac{b^x - 1}{x} = \ln(b)$ for any positive number b. (Write b as $e^{\ln(b)}$ and use $\lim_{x\to 0} \frac{e^x-1}{x} = 1$.)

2.3 The Limit of a Function

Section 2.2 concerned four important limits:

$$\lim_{x\to a}\frac{x^n-a^n}{x-a}=na^{n-1},\quad \lim_{x\to 0}\frac{e^x-1}{x}=1,\quad \lim_{x\to 0}\frac{\sin(x)}{x}=1,\quad \lim_{x\to 0}\frac{1-\cos(x)}{x}=0.$$

These are all of the form $\lim_{x\to a}\frac{f(x)}{g(x)}$, in which $\lim_{x\to a} f(x)=0$ and $\lim_{x\to a} g(x)=0$. However a limit may have a different form, as illustrated in Exercises 40 and 41 in Section 2.2, which concern $\lim_{x\to 0}(1+x)^{1/x}$.

Limits are fundamental to all of calculus. In this section, we pause to discuss the concept of a limit, beginning with the notion of a one-sided limit.

One-Sided Limits

The domain of the function shown in Figure 2.3.1 is $(-\infty,\infty)$. In particular, the function is defined when $x=2$ and $f(2)=1/2$. This fact is conveyed by the solid dot at $(2,1/2)$ in the figure. The hollow dots at $(2,0)$ and $(2,1)$ indicate that these points are not on the graph of this function (but some nearby points are on the graph).

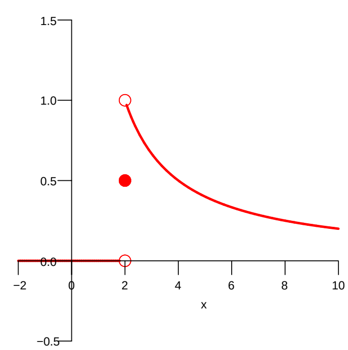


Figure 2.3.1

Consider the part of the graph for inputs $x>2$, that is, for inputs to the right of 2. As x approaches 2 from the right, $f(x)$ approaches 1. This conclusion can be expressed as

$$\lim_{x\to 2^+} f(x)=1$$

and is read "the limit of f of x, as x approaches 2, from the right, is 1." Similarly, looking at the graph of f in Figure 2.3.1 for x to the left of 2, that is, for $x<2$, we see that the values of $f(x)$ approach a different number, namely, 0. This is expressed with the shorthand

$$\lim_{x\to 2^-} f(x)=0.$$

It might sound strange to say the values of $f(x)$ "approach" 0 since the function values are exactly 0 for all inputs $x<2$. But, it is convenient, and customary, to use the word "approach" even for constant functions.

This illustrates the concept of the "right-hand" and "left-hand" limits, the two **one-sided limits**.

> **DEFINITION** *(Right-hand limit of $f(x)$ at a)* Let f be a function and a some fixed number. Assume that the domain of f contains an open interval (a, c). If, as x approaches a from the right, $f(x)$ approaches a specific number L, then L is called the **right-hand limit** of $f(x)$ as x approaches a. This is written
>
> $$\lim_{x \to a^+} f(x) = L$$
>
> or
>
> $$f(x) \to L \quad \text{as} \quad x \to a^+.$$

The assertion that

$$\lim_{x \to a^+} f(x) = L$$

is read "the limit of f of x as x approaches a from the right is L" or "as x approaches a from the right, $f(x)$ approaches L."

> **DEFINITION** *(Left-hand limit of $f(x)$ at a)* Let f be a function and a some fixed number. Assume that the domain of f contains an open interval (b, a). If, as x approaches a from the left, $f(x)$ approaches a specific number L, then L is called the **left-hand limit** of $f(x)$ as x approaches a. This is written
>
> $$\lim_{x \to a^-} f(x) = L$$
>
> or
>
> $$f(x) \to L \quad \text{as} \quad x \to a^-.$$

Notice that the definitions of the one-sided limits do not require that the number a be in the domain of the function f. If f is defined at a, we do not consider $f(a)$ when examining limits as x approaches a.

The Two-Sided Limit

If the two one-sided limits of $f(x)$ at $x = a$, $\lim\limits_{x \to a^-} f(x)$ and $\lim\limits_{x \to a^+} f(x)$, exist and are equal to L then we say the limit of $f(x)$ as x approaches a is L.

$$\lim_{x \to a} f(x) = L \qquad \text{means} \qquad \lim_{x \to a^-} f(x) = L \qquad \text{and} \qquad \lim_{x \to a^+} f(x) = L.$$

For the function graphed in Figure 2.3.1 we found that $\lim_{x \to 2^+} f(x) = 1$ and $\lim_{x \to 2^-} f(x) = 0$. Because they are different, the two-sided limit of $f(x)$ at 2, $\lim_{x \to 2} f(x)$, does not exist.

EXAMPLE 1 Figure 2.3.2 shows the graph of a function f whose domain is the closed interval $[0, 5]$.

(a) Does $\lim_{x\to 1} f(x)$ exist?

(b) Does $\lim_{x\to 2} f(x)$ exist?

(c) Does $\lim_{x\to 3} f(x)$ exist?

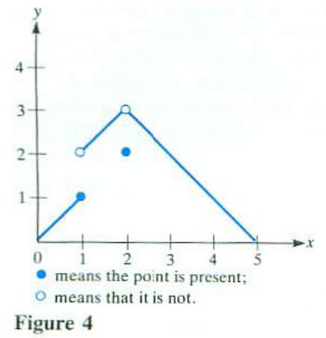


Figure 2.3.2

SOLUTION

(a) Inspection of the graph shows that

$$\lim_{x\to 1^-} f(x) = 1 \qquad \text{and} \qquad \lim_{x\to 1^+} f(x) = 2.$$

Although the two one-sided limits exist, they are not equal. Thus, $\lim_{x\to 1} f(x)$ does not exist. In short, "f does not have a limit as x approaches 1."

(b) Inspection of the graph shows that

$$\lim_{x\to 2^-} f(x) = 3 \qquad \text{and} \qquad \lim_{x\to 2^+} f(x) = 3.$$

Thus $\lim_{x\to 2} f(x)$ exists and is 3. That $f(2) = 2$, as indicated by the solid dot at $(2, 2)$, plays no role in our examination of the limit of $f(x)$ as $x \to 2$ (either one-sided or two-sided).

(c) Inspection, once again, shows that

$$\lim_{x\to 3^-} f(x) = 2 \qquad \text{and} \qquad \lim_{x\to 3^+} f(x) = 2.$$

Thus $\lim_{x\to 3} f(x)$ exists and is 2. Incidentally, the fact that $f(3) = 2$ is irrelevant in determining $\lim_{x\to 3} f(x)$.

◇

We now define the (two-sided) limit without referring to one-sided limits.

DEFINITION *(Limit of $f(x)$ at a.)* Let f be a function and a some fixed number. Assume that the domain of f contains open intervals (b, a) and (a, c), as shown in Figure 2.3.3. If there is a number L such that as x approaches a, from both the right and the left, $f(x)$ approaches L, then L is called the **limit** of $f(x)$ as x approaches a. This is expressed as either

$$\lim_{x\to a} f(x) = L \qquad \text{or} \qquad f(x) \to L \quad \text{as} \quad x \to a.$$

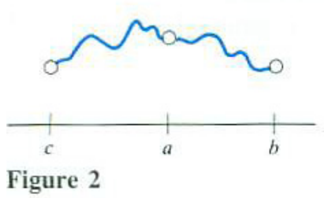


Figure 2.3.3 The function f is defined on open intervals on both sides of a.

EXAMPLE 2 Let f be the function defined by by $f(x) = \dfrac{x^n - a^n}{x - a}$ where n is a positive integer. This function is defined for all x except a. How does it behave for x near a?

SOLUTION In Section 2.2 and its Exercises we found that as x gets closer and closer to a, $f(x)$ gets closer and closer to na^{n-1}. This is summarized with the shorthand

$$\lim_{x\to a} \frac{x^n - a^n}{x - a} = na^{n-1},$$

read as "the limit of $\frac{x^n-a^n}{x-a}$ as x approaches a is na^{n-1}." ◇

EXAMPLE 3 Investigate the one-sided and two-sided limits for the square root function at 0.

SOLUTION The function $\sqrt{x}$ is defined only for x in $[0, \infty)$. We can say that the right-hand limit at 0 exists since $\sqrt{x}$ approaches 0 as $x \to 0$ through positive values of x; that is, $\lim_{x\to 0^+} \sqrt{x} = 0$. Because $\sqrt{x}$ is not defined for any negative values of x, the left-hand limit of $\sqrt{x}$ at 0 does not exist. Consequently, the two-sided limit of $\sqrt{x}$ at 0, $\lim_{x\to 0} \sqrt{x}$, does not exist. ◇

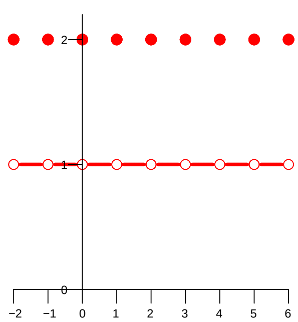


Figure 2.3.4

EXAMPLE 4 Consider the function f defined so that $f(x) = 2$ if x is an integer and $f(x) = 1$ otherwise. For which a does $\lim_{x\to a} f(x)$ exist?
SOLUTION The graph of f, shown in Figure 2.3.4, will help us decide. If a is not an integer, then for all x sufficiently near a, $f(x) = 1$. So $\lim_{x\to a} f(x) = 1$. Thus the limit exists for all a that are not integers.

Now consider the case when a is an integer. In deciding whether $\lim_{x\to a} f(x)$ exists we never consider the value of f at a, namely $f(a) = 2$. For all x sufficiently near an integer a, $f(x) = 1$. Thus, once again, $\lim_{x\to a} f(x) = 1$. The limit exists but is not $f(a)$.

Thus, $\lim_{x\to a} f(x)$ exists and equals 1 for every number a. ◇

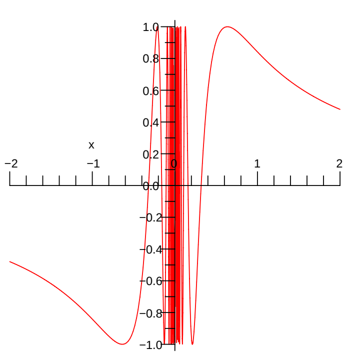


Figure 2.3.5 $y = g(x) = \sin(1/x)$.

EXAMPLE 5 Let $g(x) = \sin(1/x)$. For which a does $\lim_{x\to a} g(x)$ exist?

SOLUTION To begin, graph the function. Notice that the domain of g consists of all x except 0. When x is very large, $1/x$ is very small, so $\sin(1/x)$ is small. As x approaches 0, $1/x$ becomes large. For instance, when $x = \frac{1}{2n\pi}$, for a non-zero integer n, $1/x = 2n\pi$ and therefore $\sin(1/x) = \sin(2n\pi) = 0$. Thus, the graph of $y = g(x)$ for x near 0 crosses the x-axis infinitely often. Similarly, $g(x)$ takes the values 1 and -1 infinitely often for x near 0. The graph is shown in Figure 2.3.5.

Does $\lim_{x\to 0} g(x)$ exist? In other words, does $g(x)$ tend toward one specific number as $x \to 0$? No. The function oscillates, taking on all values from -1 to 1 (repeatedly) for x arbitrarily close to 0. Thus $\lim_{x\to 0} \sin(1/x)$ does not exist.

At all other values of a, $\lim_{x\to a} g(x)$ does exist and equals $g(a) = \sin(1/a)$. ◇

Infinite Limits at a

A function may assume arbitrarily large values as x approaches a fixed number. One important example is the tangent function. As x approaches $\pi/2$ from the left, $\tan(x)$ takes on arbitrarily large positive values. (See Figure 2.3.6.) We write

$$\lim_{x\to \frac{\pi}{2}^-} \tan(x) = +\infty.$$

However, as $x \to \frac{\pi}{2}$ from inputs larger than $\pi/2$, $\tan(x)$ takes on negative values of arbitrarily large absolute value. We write

$$\lim_{x\to \frac{\pi}{2}^+} \tan(x) = -\infty.$$

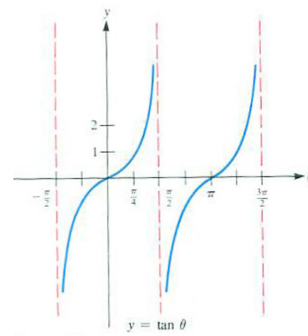


Figure 2.3.6

DEFINITION *(Infinite limit of $f(x)$ at a)* Let f be a function and a some fixed number. Assume that the domain of f contains an open interval (a, c). If, as x approaches a from the right, $f(x)$ becomes *and remains* arbitrarily large and positive, then the limit of $f(x)$ as x approaches a is said to be positive infinity. This is written

$$\lim_{x\to a^+} f(x) = +\infty$$

or sometimes just

$$\lim_{x\to a^+} f(x) = \infty.$$

If, as x approaches a from the left, $f(x)$ becomes and remains arbitrarily large and positive, then we write

$$\lim_{x\to a^-} f(x) = +\infty.$$

Similarly, if $f(x)$ assumes values that are negative and remain arbitrarily large in absolute value, we write either

$$\lim_{x\to a^+} f(x) = -\infty \qquad \text{or} \qquad \lim_{x\to a^-} f(x) = -\infty,$$

depending upon whether x approaches a from the right or from the left.

Limits as $x \to \infty$

Sometimes it is useful to know how $f(x)$ behaves when x is a very large positive number (or a negative number of large absolute value).

EXAMPLE 6 Determine how $f(x) = 1/x$ behaves for

(a) large positive inputs

(b) negative inputs of large absolute value

(c) small positive inputs

(d) negative inputs of small absolute value

SOLUTION

x	$1/x$
10	0.1
100	0.01
1000	0.001

(a) To get started, make a table of values as shown in the margin. As x becomes arbitrarily large, $1/x$ approaches 0: $\lim\limits_{x\to\infty} \dfrac{1}{x} = 0$. This conclusion would be read as "as x approaches ∞, $f(x)$ approaches 0."

(b) This is similar to (a), except that the reciprocal of a negative number with large absolute value is a negative number with a small absolute value. Thus, $\lim\limits_{x\to-\infty} \dfrac{1}{x} = 0$.

(c) For inputs that are positive and approaching 0, the reciprocals are positive and large: $\lim_{x\to 0^+} \frac{1}{x} = +\infty$.

(d) Lastly, the reciprocal of inputs that are negative and approaching 0 from the left are negative and arbitrarily large in absolute value: $\lim\limits_{x\to 0^-} \dfrac{1}{x} = -\infty$.

◇

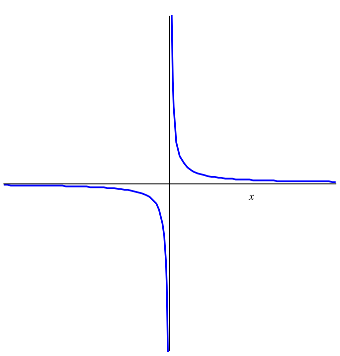


Figure 2.3.7

More generally, for any fixed positive exponent p,

$$\lim_{x\to\infty} \frac{1}{x^p} = 0.$$

Limits of the form $\lim_{x\to\infty} P(x)$ and $\lim_{x\to\infty} \frac{P(x)}{Q(x)}$, where P and Q are polynomials are easy to treat, as the following examples show.

Keep in mind that ∞ is not a number. It is just a symbol that tells us that something — either the inputs or the outputs of a function — become arbitrarily large.

EXAMPLE 7 Find $\lim\limits_{x\to\infty} (2x^3 - 5x^2 + 6x + 5)$.

SOLUTION When x is large, x^3 is much larger than either x^2 or x. With this in mind, we use a little algebra to determine the limit:

$$2x^3 - 5x^2 + 6x + 5 = x^3\left(2 - \frac{5}{x} + \frac{6}{x^2} + \frac{5}{x^3}\right).$$

The expression in parentheses approaches 2, while x^3 gets arbitrarily large. Thus

$$\lim_{x\to\infty} \frac{2x^3 - 5x^2 + 6x + 5}{x^3} = \infty.$$

◇

EXAMPLE 8 Find $\lim_{x\to\infty} \frac{2x^3-5x^2+6x+5}{7x^4+3x+2}$.

SOLUTION We use the same technique as in Example 7.

$$\begin{array}{rrcl} & 2x^3 - 5x^2 + 6x + 5 & = & x^3\left(2 - \frac{5}{x} + \frac{6}{x^2} + \frac{5}{x^3}\right) \\ \text{and} & 7x^4 + 3x + 2 & = & x^4\left(7 + \frac{3}{x^3} + \frac{2}{x^4}\right) \\ \text{so that} & \frac{2x^3-5x^2+6x+5}{7x^4+3x+2} & = & \frac{x^3\left(2-\frac{5}{x}+\frac{6}{x^2}+\frac{5}{x^3}\right)}{x^4\left(7+\frac{3}{x^3}+\frac{2}{x^4}\right)} \\ & & = & \frac{1}{x}\frac{2-\frac{5}{x}+\frac{6}{x^2}+\frac{5}{x^3}}{7+\frac{3}{x^3}+\frac{2}{x^4}}. \end{array}$$

As x gets arbitrarily large, $\frac{1}{x}$ approaches 0, $2 - \frac{5}{x} + \frac{6}{x^2} + \frac{5}{x^3}$ approaches 2, and $7 + \frac{3}{x^3} + \frac{2}{x^4}$ approaches 7. Thus,

$$\lim_{x\to\infty} \frac{2x^3 - 5x^2 + 6x + 5}{7x^4 + 3x + 2} = 0 \cdot \left(\frac{2}{7}\right) = 0.$$

◇

As these two examples suggest, the limit of a quotient of two polynomials, $\frac{P(x)}{Q(x)}$, is completely determined by the limit of the quotient of the highest degree term in $P(x)$ and in $Q(x)$.

Let

$$P(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

and

$$Q(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0,$$

where a_n and b_m are not 0. Then

$$\lim_{x\to\infty} \frac{P(x)}{Q(x)} = \lim_{x\to\infty} \frac{a_n x^n}{b_m x^m}.$$

In particular, if $m = n$, the limit is a_n/b_m. If $m > n$, the limit is 0. If $n > m$, the limit is infinite, either ∞ or $-\infty$, depending on the signs of a_n and b_m.

Summary

This section introduces the concept of a limit and notations for the various types of limits. One-sided limits are the foundation for the two-sided limit as well as for infinite limits and limits at infinity.

It is important to keep in mind that when deciding whether $\lim_{x \to a} f(x)$ exists, you never consider $f(a)$. Perhaps a isn't even in the domain of the function. Even if a is in the domain, the value $f(a)$ plays no role in deciding whether $\lim_{x \to a} f(x)$ exists.

EXERCISES for Section 2.3

In Exercises 1 to 8 the limits exist. Find them.

1. $\lim_{x \to 3} \frac{x^2 - 9}{x - 3}$

2. $\lim_{x \to 4} \frac{x^2 - 9}{x - 3}$

3. $\lim_{x \to 0} \frac{\sin(x)}{x}$

4. $\lim_{x \to \frac{\pi}{2}} \frac{\sin(x)}{x}$

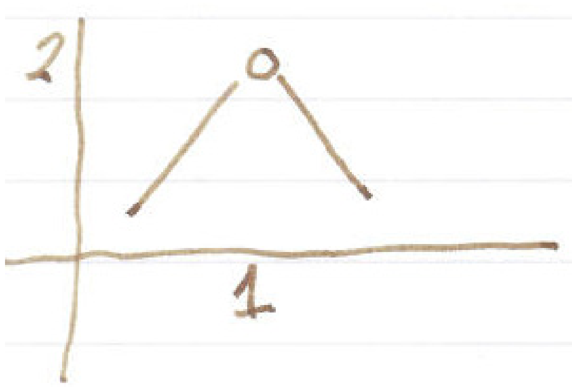
5. $\lim_{x \to 0} \frac{e^x - 1}{2x}$

6. $\lim_{x \to 2} \frac{e^x - 1}{2x}$

7. $\lim_{x \to 0} \frac{1 - \cos(x)}{3x}$

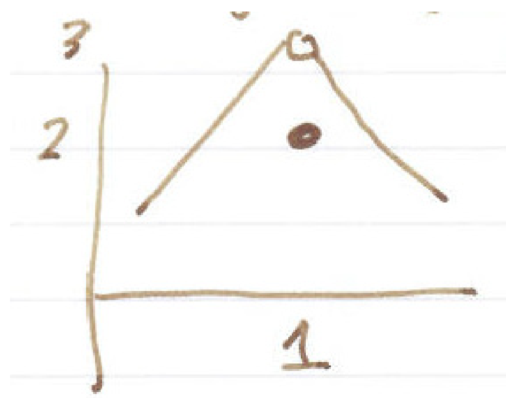
8. $\lim_{x \to \pi} \frac{1 - \cos(x)}{3x}$

In Exercises 9 to 12 the graph of a function $y = f(x)$ is given. Decide whether $\lim_{x \to 1^+} f(x)$, $\lim_{x \to 1^-} f(x)$, and $\lim_{x \to 1} f(x)$ exist. If they do exist, give their values.



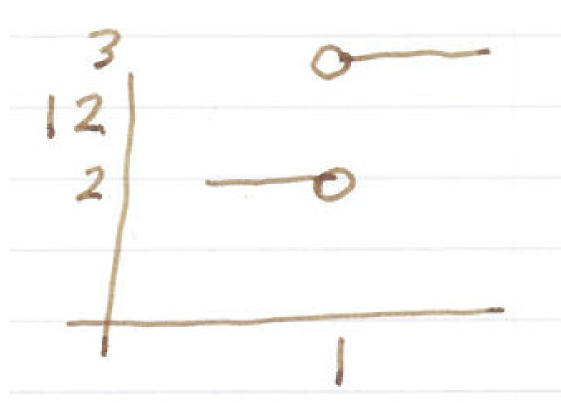
9.

10.



11.

12.



ARTIST: In the above figure, which goes with Exercise 12, please remove the "12" from the vertical axis.

13.

(a) Sketch the graph of $y = \log_2(x)$.

(b) What are $\lim_{x\to\infty} \log_2(x)$, $\lim_{x\to 4} \log_2(x)$, and $\lim_{x\to 0^+} \log_2(x)$?

14.

(a) Sketch the graph of $y = 2^x$.

(b) What are $\lim_{x\to\infty} 2^x$, $\lim_{x\to 4} 2^x$, and $\lim_{x\to-\infty} 2^x$?

15. Find $\lim_{x\to a} \dfrac{x^3-8}{x-2}$ for $a = 1$, 2, and 3.

16. Find $\lim_{x\to a} \dfrac{x^4-16}{x-2}$ for $a = 1$, 2, and 3.

17. Examine $\lim_{x\to a} \dfrac{e^x-1}{x-2}$ for $a = -1$, 0, 1, and 2.

18. Find $\lim_{x\to a} \dfrac{\sin(x)}{x}$ for $a = \frac{\pi}{6}$, $\frac{\pi}{4}$, and 0.

In Exercises 19 to 24, find the given limit (if it exists).

19. $\lim_{x\to\infty} 2^{-x}\sin(x)$

20. $\lim_{x\to\infty} 3^{-x}\cos(2x)$

21. $\lim_{x\to\infty} \dfrac{3x^5+2x^2-1}{6x^5+x^4+2}$

22. $\lim_{x\to\infty} \dfrac{13x^5+2x^2+1}{2x^6+x+5}$

23. $\lim_{x\to\infty} \dfrac{10x^6+x^5+x+1}{x^6}$

24. $\lim_{x\to\infty} \dfrac{25x^5+x^2+1}{x^3+x+2}$

In Exercises 25 to 27, information is given about functions f and g. In each case decide whether the limit asked for can be determined on the basis of that information. If it can, give its value. If it cannot, show by specific choices of f and g that it cannot.

25. Given that $\lim_{x\to\infty} f(x) = 0$ and $\lim_{x\to\infty} g(x) = 1$, discuss

(a) $\lim_{x\to\infty} (f(x)+g(x))$

(b) $\lim_{x\to\infty} (f(x)/g(x))$

(c) $\lim_{x\to\infty} (f(x)g(x))$

(d) $\lim_{x\to\infty} (g(x)/f(x))$

(e) $\lim_{x\to\infty} (g(x)/|f(x)|)$

26. Given that $\lim_{x\to\infty} f(x) = \infty$ and $\lim_{x\to\infty} g(x) = \infty$, discuss

(a) $\lim_{x\to\infty} (f(x) + g(x))$

(b) $\lim_{x\to\infty} (f(x) - g(x))$

(c) $\lim_{x\to\infty} (f(x)g(x))$

(d) $\lim_{x\to\infty} (g(x)/f(x))$

27. Given that $\lim_{x\to\infty} f(x) = 1$ and $\lim_{x\to\infty} g(x) = \infty$, discuss

(a) $\lim_{x\to\infty} (f(x)/g(x))$

(b) $\lim_{x\to\infty} (f(x)g(x))$

(c) $\lim_{x\to\infty} (f(x) - 1)g(x)$

28. Let $f(x) = \cos(1/x)$.

(a) What is the domain of f?

(b) Does $\lim_{x\to 0} \cos(1/x)$ exist?

(c) Graph $f(x) = \cos(1/x)$.

29. Let $f(x) = x \sin(1/x)$.

(a) What is the domain of f?

(b) Graph the lines $y = x$ and $y = -x$.

(c) For which x does $f(x) = x$? When does $f(x) = -x$? (Notice that the graph of $y = f(x)$ goes back and forth between these lines.)

(d) Does $\lim_{x\to 0} f(x)$ exist? If so, what is it?

(e) Does $\lim_{x\to\infty} f(x)$ exist? If so, what is it?

(f) Graph $y = f(x)$.

(g) Does $\lim_{x\to\infty} |f(x)|$ exist? If so, what is it?

30. Let $f(x) = \frac{|x|}{x}$, which is defined except at $x = 0$.

(a) What is $f(3)$?

(b) What is $f(-2)$?

(c) Graph $y = f(x)$.

(d) Does $\lim\limits_{x\to 0^+} f(x)$ exist? If so, what is it?

(e) Does $\lim\limits_{x\to 0^-} f(x)$ exist? If so, what is it?

(f) Does $\lim\limits_{x\to 0} f(x)$ exist? If so, what is it?

In Exercises 31 to 33, find $\lim\limits_{h\to 0} \dfrac{f(3+h) - f(3)}{h}$ for the following functions.

31. $f(x) = 5x$

32. $f(x) = x^2$

33. $f(x) = e^x$

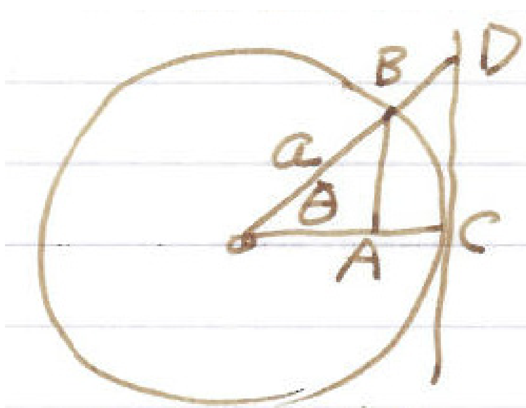


Figure 2.3.8 Exercise 34

34. Figure 2.3.8 shows a circle of radius a. Find

(a) $\lim\limits_{\theta\to 0^+} \dfrac{\overline{AB}}{\overset{\frown}{CB}}$ $\overset{\frown}{CB}$ is the length of the arc of the circle with radius a.

(b) $\lim\limits_{\theta\to 0^+} \dfrac{\overline{AB}}{\overline{CD}}$

(c) $\lim\limits_{\theta\to 0} \dfrac{\text{area of } ABC}{\text{area of } ABCD}$.

35. Let $f(x)$ be the diameter of the largest circle that fits in a 1 by x rectangle.

(a) Find a formula for $f(x)$.

(b) Graph $y = f(x)$.

(c) Does $\lim_{x\to 1} f(x)$ exist?

36. I am thinking of two numbers near 0. What, if anything, can you say about their

(a) product?

(b) quotient?

(c) difference?

(d) sum?

37. I am thinking about two large positive numbers. What, if anything, can you say about their

(a) product?

(b) quotient?

(c) difference?

(d) sum?

38. Find $\lim_{h\to 0} \frac{f(\theta+h)-f(\theta)}{h}$ for $f(x) = \sin(x)$. ($\sin(a+b) = \sin(a)\cos(b) + \cos(a)\sin(b)$.)

39. Find $\lim_{h\to 0} \frac{f(\theta+h)-f(\theta)}{h}$ for $f(x) = \cos(x)$. ($\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$.)

40. Find $\lim_{x\to 0} \frac{e^{2x}-1}{x}$.

41. Sam and Jane are discussing

$$f(x) = \frac{3x^2+2x}{x+5}.$$

Sam: For large x, $2x$ is small in comparison to $3x^2$, and 5 is small in comparison to x. So the quotient $\frac{3x^2+2x}{x+5}$ behaves like $\frac{3x^2}{x} = 3x$. Hence, the graph of $y = f(x)$ is very close to the graph of the line $y = 3x$ when x is large.

Jane: "Nonsense. After all,

$$\frac{3x^2 + 2x}{x + 5} = \frac{3x + 2}{1 + (5/x)}$$

which clearly behaves like $3x+2$ for large x. Thus the graph of $y = f(x)$ stays very close to the line $y = 3x + 2$ when x is large.

Settle the argument.

42. Sam, Jane, and Wilber are arguing about limits in a case where $\lim\limits_{x\to\infty} f(x) = 0$ and $\lim\limits_{x\to\infty} g(x) = \infty$.

Sam: $\lim\limits_{x\to\infty} f(x)g(x) = 0$, since $f(x)$ is going toward 0.

Jane: Rubbish! Since $g(x)$ gets large, it will turn out that $\lim\limits_{x\to\infty} f(x)g(x) = \infty$.

Wilber: You're both wrong. The two influences will balance out and you will see that $\lim_{x\to\infty} f(x)g(x)$ is near 1.

Settle the argument.

43. Sam and Jane are arguing about limits in a case where $f(x) \geq 1$ for $x > 0$, $\lim\limits_{x\to 0^+} f(x) = 1$ and $\lim\limits_{x\to 0} g(x) = \infty$. What can be said about $\lim_{x\to 0^+} f(x)^{g(x)}$?

Sam: That's easy. Multiply a bunch of numbers near 1 and you get a number near 1. So the limit will be 1.

Jane: Rubbish! Since $f(x)$ may be bigger than 1 and you are multiplying it lots of times, you will get a really large number. There's no doubt in my mind: $\lim\limits_{x\to 0} f(x)^{g(x)} = \infty$.

Settle the argument.

44. An urn contains n marbles. One is green and the remaining $n - 1$ are red. When picking one marble at random without looking, the probability is $1/n$ of getting the green marble, and $(n-1)/n$ of getting a red marble. If you do this experiment n times, each time putting the chosen marble back, the probability of *not* getting the green marble on any of the n experiments is $((n-1)/n)^n$.

(a) Let $p(n) = \left(\frac{n-1}{n}\right)^n$. Compute $p(2)$, $p(3)$, and $p(4)$ to at least three decimal digits (to the right of the decimal point).

(b) Show that as $n \to \infty$, $p(n)$ approaches the reciprocal of $\lim\limits_{x\to 0} (1 + x)^{1/x}$.

2.4 Continuous Functions

This section introduces the notion of a continuous function. While almost all functions met in practice are continuous, we must always remain alert that a function might not be continuous. We begin with an informal description and then give a more useful working definition.

An Informal Introduction to Continuous Functions

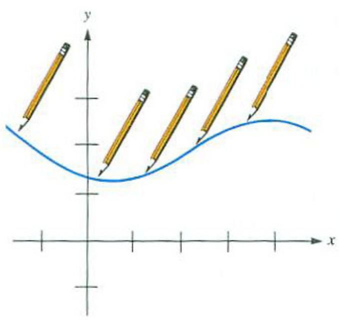


Figure 2.4.1

When we draw the graph of a function defined on some interval, we usually do not have to lift the pencil off the paper. Figure 2.4.1 shows this typical situation.

A function is said to be **continuous** if, when considered on any interval in its domain, its graph can he traced without lifting the pencil off the paper. (The domain may consist of several intervals.) According to this definition any polynomial is continuous. So is each of the basic trigonometric functions, including $y = \tan(x)$, whose graph is shown in Figure 2.3.6 of Section 2.3.

You may be tempted to say "But $\tan(x)$ blows up at $x = \pi/2$ and I have to lift my pencil off the paper to draw the graph." However, $x = \pi/2$ is not in the domain of the tangent function. *On every interval in its domain,* $\tan(x)$ *behaves quite decently;* on such an interval we can sketch its graph without lifting the pencil from the paper. That is why $\tan(x)$ is continuous. The function $1/x$ is also continuous, since it "explodes" only at a number not in its domain, namely at $x = 0$. The function whose graph is shown in Figure 2.4.2 is not continuous. It is defined throughout the interval $[-2, 3]$, but to draw its graph you must lift the pencil from the paper near $x = 1$. However, when you consider the function *only for* x *in* $[1, 3]$*, then it is continuous.* By the way, a formula for this piecewise-defined function given graphically in Figure 2.4.2 is:

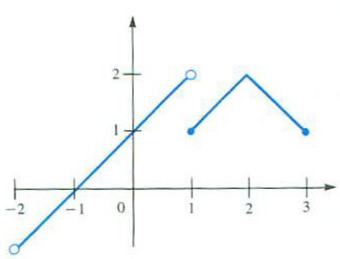


Figure 2.4.2

$$f(x) = \begin{cases} x+1 & \text{for } x \text{ in } [-2,1) \\ x & \text{for } x \text{ in } [1,2) \\ -x+4 & \text{for } x \text{ in } [2,3]. \end{cases}$$

It is pieced together from three different continuous functions.

The Definition of Continuity

Our informal "moving pencil" notion of a continuous function requires drawing a graph of the function. Our working definition does not require such a graph. Moreover, it easily generalizes to functions of more than one variable in later chapters.

To get the feeling of this second definition, imagine that you had the information shown in the table in the margin about some function f. What would you expect the output $f(1)$ to be?

x	$f(x)$
0.9	2.93
0.99	2.9954
0.999	2.9999997

It would be quite a shock to be told that $f(1)$ is, say, 625. A reasonable function should present no such surprise. The expectation is that $f(1)$ will be 3. More generally, we expect the output of a function at the input a to be closely connected with the outputs of the function at inputs near a. The functions of interest in calculus usually behave that way. In short, "What you expect is what you get." With this in mind, we define the notion of *continuity at a number* a. We first assume that the domain of f contains an open interval around a.

> **DEFINITION** *(Continuity at a number a)* Assume that $f(x)$ is defined in some open interval that contains the number a. Then the function f is **continuous at** a if $\lim_{x\to a} f(x) = f(a)$. This means that
>
> 1. $f(a)$ is defined (that is, a is in the domain of f).
> 2. $\lim_{x\to a} f(x)$ exists.
> 3. $\lim_{x\to a} f(x)$ equals $f(a)$.

As Figure 2.4.3 shows, whether a function is continuous at a depends on its behavior both at a and at inputs near a. Being continuous at a is a local matter, involving perhaps very tiny intervals about a.

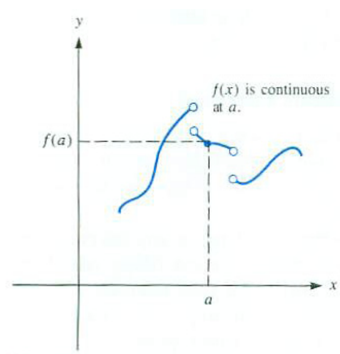


Figure 2.4.3

To check whether a function f is continuous at a number a, we ask three questions:

Question 1: Is a in the domain of f?

Question 2: Does $\lim_{x\to a} f(x)$ exist?

Question 3: Does $f(a)$ equal $\lim_{x\to a} f(x)$?

If the answer is "yes" to each of these questions, we say that f is continuous at a.

If a is in the domain of f and the answer to Question 2 or to Question 3 is "no," then f is said to be **discontinuous at** a. If a is not in the domain of f, we do not speak of it being continuous or discontinuous there.

We are now ready to define a continuous function.

> **DEFINITION** *(Continuous function)* Let f be a function whose domain is the x-axis or is made up of open intervals. Then f is a **continuous function** if it is continuous at each number a in its domain. A function that is not continuous is called a **discontinuous function**.

EXAMPLE 1 Use the definition of continuity to decide whether $f(x) = 1/x$ is continuous.

SOLUTION This function f is continuous at every point a for which the answers to Questions 1, 2, and 3 are all "yes".

If a is not 0, it is in the domain of f. So, for a not 0, the answer to Question 1 is "yes." Since

$$\lim_{x \to a} \frac{1}{x} = \frac{1}{a},$$

the answer to Question 2 is "yes." Because

$$f(a) = \frac{1}{a},$$

the answer to Question 3 is also "yes." Thus $f(x) = 1/x$ is continuous at every number in its domain. Hence f is a continuous function. Note that the conclusion agrees with the "moving pencil" picture of continuity. ◇

Not every important function is continuous. Let $f(x)$ be the greatest integer that is less than or equal to x. For instance, $f(1.8) = 1$, $f(1.9) = 1$, $f(2) = 2$, and $f(2.3) = 2$. This function is often used in number theory and computer science, where it is denoted $[x]$ or $\lfloor x \rfloor$ and called the **floor** of x. People use the floor function every time they answer the question, "How old are you?" The next example examines where the floor function fails to be continuous.

EXAMPLE 2 Let f be the floor function, $f(x) = \lfloor x \rfloor$. Graph f and find where it is continuous. Is f a continuous function?

SOLUTION We begin with the following table to show the behavior of $f(x)$ for x near 1 or 2.

x	0	0.5	0.8	1	1.1	1.99	2	2.01
$\lfloor x \rfloor$	0	0	0	1	1	1	2	2

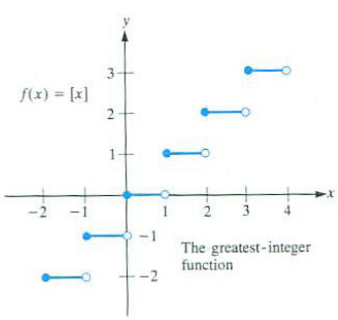


Figure 2.4.4

For $0 \le x < 1$, $\lfloor x \rfloor = 0$. But at the input $x = 1$ the output jumps to 1 since $\lfloor 1 \rfloor = 1$. For $1 \le x < 2$, $\lfloor x \rfloor$ remains at 1. Then at 2 it jumps to 2. More generally, $\lfloor x \rfloor$ has a jump at every integer, as shown in Figure 2.4.4.

Let us show that f is not continuous at $a = 2$ by seeing which of the three conditions in the definition are not satisfied. First of all, Question 1 is answered "yes" since 2 lies in the domain of the function; indeed, $f(2) = 2$.

What is the answer to Question 2? Does $\lim_{x \to 2} f(x)$ exist? We see that

$$\lim_{x \to 2^-} f(x) = 1 \qquad \text{and} \qquad \lim_{x \to 2^+} f(x) = 2.$$

Since the left-hand and right-hand limits are not equal, $\lim_{x\to 2} f(x)$ does not exist. Question 2 is answered "no."

Already we know that the function is not continuous at $a = 2$. Since the limit does not exist there is no point in considering Question 3. Because there is a point in the domain where $\lfloor x \rfloor$ is not continuous, this is a discontinuous function. More specifically, the floor function is discontinuous at a whenever a is an integer.

Is f continuous at a if a is not an integer? Let us take the case $a = 1.5$, for instance.

Question 1 is answered "yes," because $f(1.5)$ is defined.
(In fact, $f(1.5) = 1$.)

Question 2 is answered "yes," since $\lim_{x\to 1.5} f(x) = 1$.

Question 3 is answered "yes," since $\lim_{x\to 1.5} f(x) = f(1.5)$.
(Both values are 1.)

The floor function is continuous at $a = 1.5$. Similarly, f is continuous at every number that is not an integer.

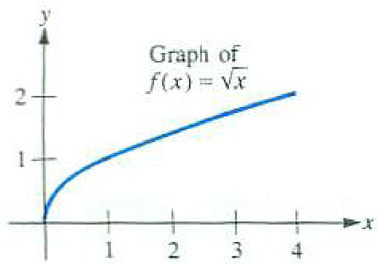
Note that $\lfloor x \rfloor$ is continuous on any interval that does not include an integer. For instance, if we consider the function only on the interval $(1.1, 1.9)$, it is continuous there. ◇

Continuity at an Endpoint

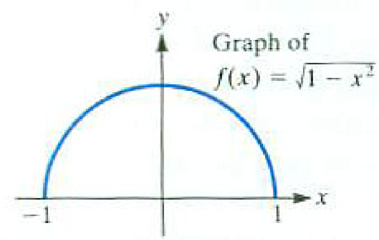
The functions $f(x) = \sqrt{x}$ and $g(x) = \sqrt{1 - x^2}$ are graphed in Figures 2.4.5(a) and (b), respectively. We would like to call both of these functions continuous. However, there is a slight technical problem. The number 0 is in the domain of f, but there is no open interval around 0 that lies completely in the domain, as our definition of continuity requires. Since $f(x) = \sqrt{x}$ is not defined for x to the left of 0, we are not interested in numbers x to the left of 0. Similarly, $g(x) = \sqrt{1 - x^2}$ is defined only when $1 - x^2 \geq 0$, that is, for $-1 \leq x \leq 1$. To cover this type of situation we utilize one-sided limits to define **one-sided continuity**.

DEFINITION *(Continuity from the right at a number.)* Assume that $f(x)$ is defined in some closed interval $[a, c]$. Then the function f **is continuous from the right at** a if

1. $f(a)$ is defined
2. $\lim_{x\to a^+} f(x)$ exists
3. $\lim_{x\to a^+} f(x)$ equals $f(a)$



(a)



(b)

Figure 2.4.5

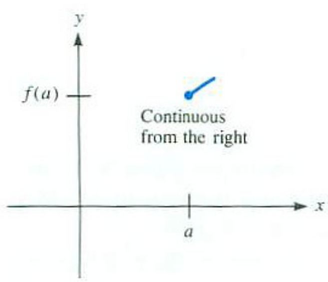


Figure 2.4.6

Figure 2.4.6 illustrates this definition, which also takes care of the continuity of $g(x) = \sqrt{1 - x^2}$ at -1 in Figure 2.4.5(b). The next definition takes care of the right-hand endpoints.

DEFINITION *(Continuity from the left at a number a.)* Assume that $f(x)$ is defined in some closed interval $[b, a]$. Then the function f **is continuous from the left at** a if

1. $f(a)$ is defined
2. $\lim_{x \to a^-} f(x)$ exists
3. $\lim_{x \to a^-} f(x)$ equals $f(a)$

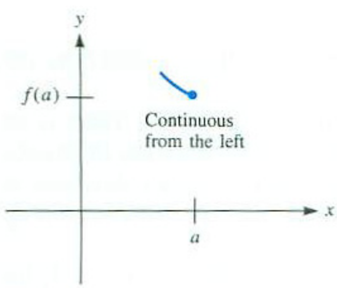


Figure 2.4.7

Figure 2.4.7 illustrates this definition.

With these two extra definitions to cover some special cases in the domain, we can extend the definition of continuous function to include those functions whose domains may contain endpoints. We say, for instance, that $\sqrt{1 - x^2}$ is continuous because it is continuous at any number in $(-1, 1)$, is continuous from the right at -1, and continuous from the left at 1.

These special considerations are minor matters that will little concern us in the future. The key point is that $\sqrt{1 - x^2}$ and $\sqrt{x}$ are both continuous functions. So are practically all the functions studied in calculus.

The following example reviews the notion of continuity.

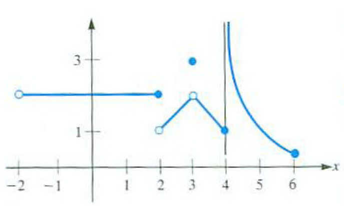


Figure 2.4.8

EXAMPLE 3 Figure 2.4.8 is the graph of a certain (piecewise-defined) function $f(x)$ whose domain is the interval $(-2, 6]$. Discuss the continuity of $f(x)$ at (a) 6, (b) 4, (c) 3, (d) 2, (e) 1, and (f) -2.

SOLUTION

(a) Since $\lim\limits_{x \to 6-} f(x)$ exists and equals $f(6)$, f is continuous from the left at 6.

(b) Since $\lim_{x\to 4} f(x)$ does not exist, f is not continuous at 4.

(c) Inspection of the graph shows that $\lim_{x\to 3} f(x) = 2$. However, Question 3 is answered "no" because $f(3) = 3$, which is *not* equal to $\lim_{x\to 3} f(x)$. Thus f is not continuous at 3.

(d) Though $\lim_{x\to 2-} f(x)$ and $\lim_{x\to 2+} f(x)$ both exist, they are not equal. (The left-hand limit is 2; the right-hand limit is 1.) Thus $\lim_{x\to 2} f(x)$ does not exist, the answer to Question 2 is "no," and f is discontinuous at $x = 2$.

(e) At 1, "yes" is the answer to all three questions: $f(1)$ is defined, $\lim_{x\to 1} f(x)$ exists (it equals 2) and, finally, it equals $f(1)$. f is continuous at $x = 1$.

(f) Since -2 is not even in the domain of this function, we do not speak of continuity or discontinuity of f at -2.

◇

As Example 3 shows, a function can fail to be continuous at a given number a in its domain for either of two reasons:

1. $\lim_{x\to a} f(x)$ might not exist

2. when, $\lim_{x\to a} f(x)$ does exist, $f(a)$ might not be equal to that limit.

Continuity and Limits

Some limits are so easy that you can find them without any work; for instance, $\lim_{x\to 2} 5^x = 5^2 = 25$. Others offer a challenge; for instance, $\lim_{x\to 2} \frac{x^3-2^3}{x-2}$.

If you want to find $\lim_{x\to a} f(x)$, and you know f is a continuous function with a in its domain, then you just calculate $f(a)$. In such a case there is no challenge and the limit is called *determinate*.

The interesting case for finding $\lim_{x\to a} f(x)$ occurs when f is not defined at a. That is when you must consider the influences operating on $f(x)$ when x is near a. You may have to do some algebra or computations. Such limits are called *indeterminate*.

The four limits encountered in Section 2.2, $\lim_{x\to a} \dfrac{x^n - a^n}{x-a}$, $\lim_{x\to 0} \dfrac{b^x - 1}{x}$, $\lim_{x\to 0} \dfrac{\sin(x)}{x}$, and $\lim_{x\to 0} \dfrac{1-\cos(x)}{x}$ are indeterminate. Each required some work to find its value. These types of limits will be discussed in detail in Section 5.6.

We list the properties of limits which are helpful in computing limits.

Each of these properties remains valid when the two-sided limit is replaced with a one-sided limit.

Theorem 2.4.1 (Properties of Limits)**.** *Let g and h be two functions and assume that $\lim_{x\to a} g(x) = A$ and $\lim_{x\to a} h(x) = B$. Then*

Sum $\displaystyle\lim_{x\to a}(g(x)+h(x)) = \lim_{x\to a} g(x) + \lim_{x\to a} h(x) = A + B$
the limit of the sum is the sum of the limits

Difference $\displaystyle\lim_{x\to a}(g(x)-h(x)) = \lim_{x\to a} g(x) - \lim_{x\to a} h(x) = A - B$
the limit of the difference is the difference of the limits

Product $\displaystyle\lim_{x\to a}(g(x)h(x)) = \left(\lim_{x\to a} g(x)\right)\left(\lim_{x\to a} h(x)\right) = AB$
the limit of the product is the product of the limits

Constant Multiple $\displaystyle\lim_{x\to a}(kg(x)) = k\left(\lim_{x\to a} g(x)\right) = kA$*, for any constant* k
this is a special case of the Product property when one factor is a constant

Quotient $\displaystyle\lim_{x\to a}\left(\frac{g(x)}{h(x)}\right) = \frac{(\lim_{x\to a} g(x))}{(\lim_{x\to a} h(x))} = \frac{A}{B}$*, provided the denominator,* B*, is not* 0
the limit of the quotient is the quotient of the limits, provided the denominator is not 0

Power $\displaystyle\lim_{x\to a}\left(g(x)^{h(x)}\right) = \left(\lim_{x\to a} g(x)\right)^{\lim_{x\to a} h(x)} = A^B$*, provided* $A > 0$
the limit of a varying base to a varying power

EXAMPLE 4 Find $\displaystyle\lim_{x\to 0}\frac{(x^4-16)\sin(5x)}{x^2-2x}$.

SOLUTION Notice that the denominator can be factored to obtain

$$\frac{(x^4-16)\sin(5x)}{x^2-2x} = \frac{x^4-2^4}{x-2}\cdot\frac{\sin(5x)}{x}.$$

This allows the limit to be rewritten as

$$\lim_{x\to 0}\frac{x^4-2^4}{x-2}\cdot\lim_{x\to 0}\frac{\sin(5x)}{x}$$

where we have also used $16 = 2^4$. Now, $\lim_{x\to 0}\frac{x^4-2^4}{x-2} = 4\cdot 2^{4-1} = 32$. Also,

$$\lim_{x\to 0}\frac{\sin(5x)}{x} = \lim_{x\to 0} 5\frac{\sin(5x)}{5x} = 5\lim_{x\to 0}\frac{\sin(5x)}{5x} = 5\cdot 1 = 5.$$

We conclude that

$$\lim_{x\to 0}\frac{(x^4-16)\sin(5x)}{x^2-2x} = \lim_{x\to 0}\frac{x^4-2^4}{x-2}\cdot\lim_{x\to 0}\frac{\sin(5x)}{5x} = 32\cdot 5 = 160.$$

◇

Summary

This section opened with an informal view of continuous functions, expressed in terms of a moving pencil. It then gave the definition, phrased in terms of limits, which we will use throughout the text.

The development concludes in the next section, which describes three important properties of continuous functions.

EXERCISES for Section 2.4

In Exercises 1 to 12, which of these limits can be found at a glance and which require some analysis? That is, decide in each case whether the limit is determinate or indeterminate. *Do not evaluate the limit.*

1. $\lim_{x\to 0}(2^x - 1)$

2. $\lim_{x\to\infty}\left(\left(\frac{1}{2}\right)2^x - 1\right)$

3. $\lim_{x\to 1}\frac{3^x - 1}{2^x - 1}$

4. $\lim_{x\to 2}\frac{3^x - 1}{2^x - 1}$

5. $\lim_{x\to\infty}\frac{x}{2^x}$

6. $\lim_{x\to 0}\frac{x}{2^x}$

7. $\lim_{x\to 0^+}\frac{x^2}{e^x - 1}$

8. $\lim_{x\to\frac{\pi}{2}^-}(\sin(x))^{\tan(x)}$

9. $\lim_{x\to 0^+}x\log_2(x)$

10. $\lim_{x\to 0^+}(2+x)^{3/x}$

11. $\lim_{x\to\infty}(2+x)^{3/x}$

12. $\lim_{x\to 0^-}\frac{(2+x)^3}{x}$

In Exercises 13 to 16, evaluate the limit.

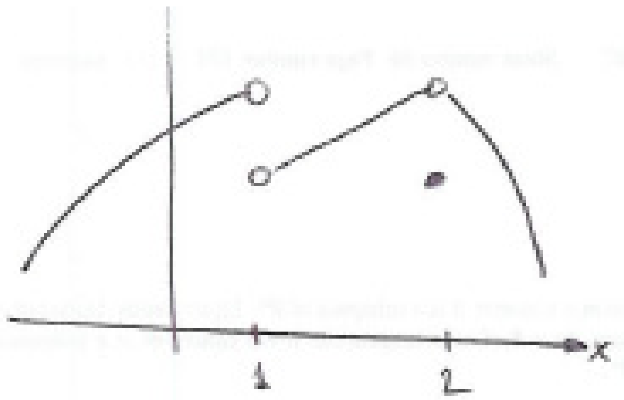
13. $\lim_{x\to\frac{\pi}{2}}\sin(x)\frac{e^x - 1}{x}$

14. $\lim_{x\to 0}\frac{\cos(x)\,(e^x - 1)}{x}$

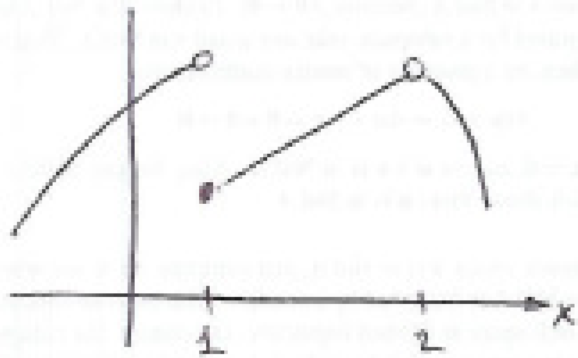
15. $\lim_{x\to 0}\frac{\sin(2x)}{x(\cos(3x))^2}$

16. $\lim_{x\to 1}\frac{(x-1)\cos(x)}{x^3 - 1}$

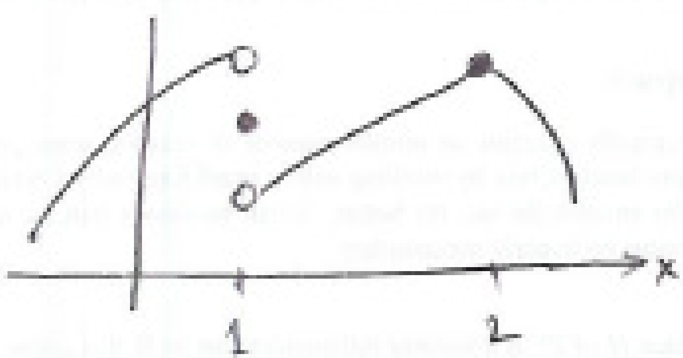
In Exercises 17 to 20 the graph of a function $y = f(x)$ is given. Determine all numbers c for which $\lim_{x\to c}f(x)$ does not exist.



17.

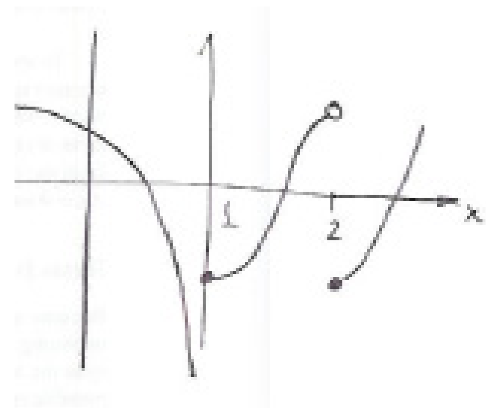


18.



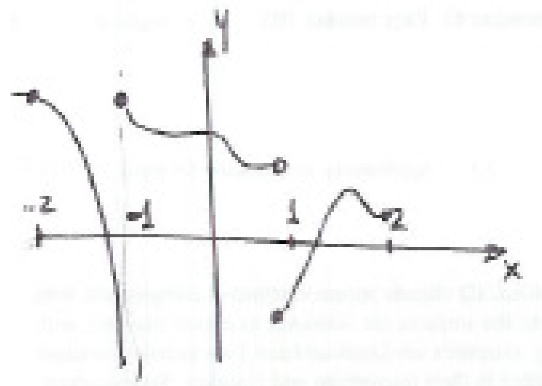
19.

20.



In Exercises 21 and 22 the graph of a function $y = f(x)$ and several intervals are given. For each interval, decide if the function is continuous on that interval.

21.



(a) $[-2, -1]$

(b) $(-2, -1)$

(c) $(-1, 1)$

(d) $[-1, 1)$

(e) $(-1, 1]$

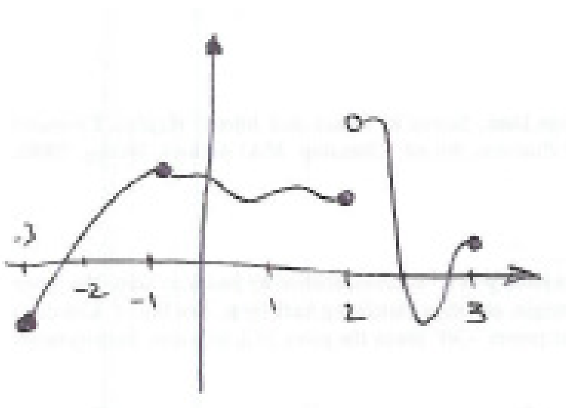
(f) $[-1, 1]$

(g) $(1, 2)$

(h) $[1, 2)$

(i) $(1, 2]$

(j) $[1, 2]$



22.

(a) $[-3, 2]$

(b) $(-1, 3)$

(c) $(-1, 2)$

(d) $[-1, 2)$

(e) $(-1, 2]$

(f) $[-1, 2]$

(g) $(2, 3)$

(h) $[2, 3)$

(i) $(2, 3]$

(j) $[2, 3]$

23. Let $f(x) = x + |x|$.

(a) Graph f.

(b) Is f continuous at -1?

(c) Is f continuous at 0?

24. Let $f(x) = 2^{1/x}$ for $x \neq 0$.

(a) Find $\lim_{x\to\infty} f(x)$.

(b) Find $\lim_{x \to -\infty} f(x)$.

(c) Does $\lim_{x \to 0^+} f(x)$ exist?

(d) Does $\lim_{x \to 0^-} f(x)$ exist?

(e) Graph f, incorporating the information from parts *(a)* to *(d)*.

(f) Is it possible to define $f(0)$ in such a way that f is continuous throughout the x-axis?

25. Let $f(x) = x \sin(1/x)$ for $x \neq 0$.

(a) Find $\lim_{x \to \infty} f(x)$.

(b) Find $\lim_{x \to -\infty} f(x)$.

(c) Find $\lim_{x \to 0} f(x)$.

(d) Is it possible to define $f(0)$ in such a way that f is continuous throughout the x-axis?

(e) Sketch the graph of f.

In Exercises 26 to 28 find equations that the numbers k, p, and/or m must satisfy to make each function continuous.

26. $f(x) = \begin{cases} \frac{\sin(x)}{2x} & x \neq 0 \\ p & x = 0 \end{cases}$

27. $f(x) = \begin{cases} k & x \leq 0 \\ \arcsin(x) & 0 < x \leq \frac{\pi}{2} \\ p & x > \frac{\pi}{2} \end{cases}$

28. $f(x) = \begin{cases} \ln(x) & x > 1 \\ k - m\sqrt{x} & 0 < x \leq 1 \\ pe^{-x} & x \leq 0 \end{cases}$

29.

(a) Let f and g be two functions defined for all numbers. If $f(x) = g(x)$ when x is not 3, must $f(3) = g(3)$?

(b) Let f and g be two continuous functions defined for all numbers. If $f(x) = g(x)$ when x is not 3, must $f(3) = g(3)$?

Explain your answers.

30. *The reason* 0^0 *is not defined.* It might be hoped that if the positive numbers b and x are both close to 0, then b^x might be close to some fixed number. If that were so, it would suggest a definition for 0^0. Experiment with various choices of b and x near 0 and on the basis of your data write a paragraph on the theme, "Why 0^0 is not defined."

2.5 Three Important Properties of Continuous Functions

Continuous functions have three properties important in calculus: the "extreme-value" property, the "intermediate-value" property, and the "permanence" property. All three are quite plausible, and a glance at the graph of a typical continuous function may persuade us that they are obvious. No proofs will be offered: they depend on the precise definitions of limits given in Sections 3.8 and 3.9 and are part of an advanced calculus course.

We will say that a function has a **local or relative maximum** at a point $(c, f(c))$ when $f(c) \geq f(x)$ for x near c. More precisely, there is an open interval I containing c such that if x is in I, and $f(x)$ is defined, then $f(x) \leq f(c)$. Likewise, a function has a **local or relative minimum** at a point $(c, f(c))$ when $f(c) \leq f(x)$ for x near c. Each maximum or minimum is referred to as an **extreme value** or **extremum** of the function.

The plural of extremum is extrema.

Extreme-Value Property

The first property is that a function continuous throughout the closed interval $[a, b]$ takes on a largest value somewhere in the interval. (When we refer to an interval $[a, b]$ it is assumed that a and b are numbers with $a < b$.)

Theorem 2.5.1 (Maximum-Value Property). *Let f be continuous throughout a closed interval $[a, b]$. Then there is at least one number in $[a, b]$ at which f takes on a maximum value. That is, for some number c in $[a, b]$, $f(c) \geq f(x)$ for all x in $[a, b]$.*

To persuade yourself that this is plausible, imagine sketching the graph of a continuous function. (See Figure 2.5.1.)

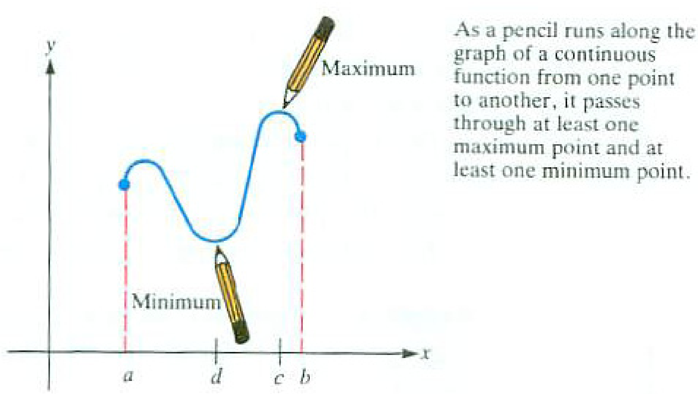


Figure 2.5.1

The maximum-value property guarantees that a maximum value exists, but it does *not tell how* to find it. The problem of finding it is addressed in Chapter 4.

There is also a **minimum-value property** that states that every continuous function on a closed interval takes on a smallest value somewhere in this interval. See Figure 2.5.1 for an illustration of this property. Combining the two properties, we have:

Theorem 2.5.2 (Extreme-Value Property). *Let f be continuous throughout the closed interval $[a, b]$. Then there is at least one number in $[a, b]$ at which f takes on a minimum value and there is at least one number in $[a, b]$ at which f takes on a maximum value. That is, for some numbers c and d in $[a, b]$, $f(d) \leq f(x) \leq f(c)$ for all x in $[a, b]$.*

EXAMPLE 1 Find all numbers in $[0, 3\pi]$ at which the cosine function, $f(x) = \cos(x)$, takes on a maximum value. Also, find all numbers in $[0, 3\pi]$ at which f takes on a minimum value.

SOLUTION Figure 2.5.2 is a graph of $f(x) = \cos(x)$ for x in $[0, 3\pi]$. Inspection of the graph shows that the maximum value of $\cos(x)$ for $0 \leq x \leq 3\pi$ is 1, and it is attained twice: when $x = 0$ and when $x = 2\pi$. The minimum value is -1, which is also attained twice: when $x = \pi$ and when $x = 3\pi$. ◇

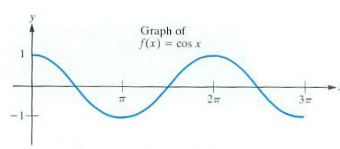


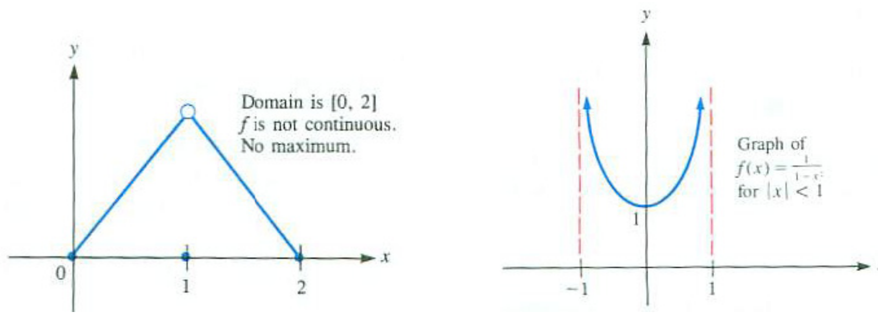
Figure 2.5.2

The Extreme-Value Property has two assumptions: "f is continuous" and "the domain is a closed interval." If either of these conditions is removed, the conclusion need not hold.

Figure 2.5.3(a) shows the graph of a function that is *not* continuous, is defined on a closed interval, but has no maximum value. On the other hand $f(x) = \frac{1}{1-x^2}$ is continuous on $(-1, 1)$. It has no maximum value, as a glance at Figure 2.5.3(b) shows. This does not violate the Extreme-Value Property, since the domain $(-1, 1)$ is not a closed interval.

Intermediate-Value Property

Imagine graphing a continuous function f defined on the closed interval $[a, b]$. As your pencil moves from the point $(a, f(a))$ to the point $(b, f(b))$ the y-coordinate of the pencil point goes through all values between $f(a)$ and $f(b)$. (Similarly, if you hike all day, starting at an altitude of 5,000 feet and ending at 11,000 feet, you must have been, say, at 7,000 feet at least once during the day. In mathematical terms, not in terms of a pencil (or a hike), "a function that is continuous throughout an interval takes on all values between any two of its values".



(a) (b)

Figure 2.5.3

Theorem 2.5.3 (Intermediate-Value Property). *Let f be continuous throughout the closed interval $[a, b]$. Let m be any number such that $f(a) \leq m \leq f(b)$ or $f(a) \geq m \geq f(b)$. Then there is at least one number c in $[a, b]$ such that $f(c) = m$.*

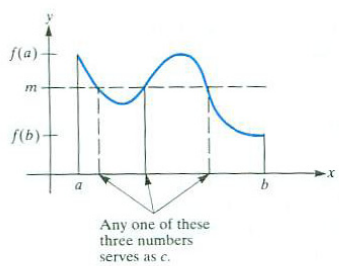


Figure 2.5.4

Pictorially, the Intermediate-Value Property asserts that, if m is between $f(a)$ and $f(b)$, a horizontal line of height m must meet the graph of f at least once, as shown in Figure 2.5.4.

Even though the property guarantees the existence of a certain number c, it does *not tell how* to find it. To find c we must be able to solve an equation, namely, the equation $f(x) = m$.

EXAMPLE 2 Use the Intermediate-Value Property to show that the equation $2x^3 + x^2 - x + 1 = 5$ has a solution in the interval $[1, 2]$.

SOLUTION Let $P(x) = 2x^3 + x^2 - x + 1$. Then

$$\begin{array}{rrcl} & P(1) & = & 2 \cdot 1^3 + 1^2 - 1 + 1 = 3 \\ \text{and} & P(2) & = & 2 \cdot 2^3 + 2^2 - 2 + 1 = 19. \end{array}$$

Since P is continuous (on $[1, 2]$) and $m = 5$ is between $P(1) = 3$ and $P(2) = 19$, the Intermediate-Value Property says there is at least one number c between 1 and 2 such that $P(c) = 5$.

To get a more accurate estimate for a number c such that $P(c) = 5$, find a shorter interval for which the Intermediate-Value Property can be applied. For instance, $P(1.2) = 4.696$ and $P(1.3) = 5.784$. By the Intermediate-Value Property, there is a number c in $[1.2.1.3]$ such that $P(c) = 5$. ◇

EXAMPLE 3 Show that the equation $-x^5 - 3x^2 + 2x + 11 = 0$ has at least one real root. In other words, the graph of $y = -x^5 - 3x^2 + 2x + 11$ crosses the x-axis.

SOLUTION Let $f(x) = -x^5 - 3x^2 + 2x + 11$. We wish to show that there is a number c such that $f(c) = 0$. In order to use the Intermediate-Value Property, we need an interval $[a, b]$ for which 0 is between $f(a)$ and $f(b)$, that is, one of $f(a)$ and $f(b)$ is positive and the other is negative. Then we could apply that property, using $m = 0$.

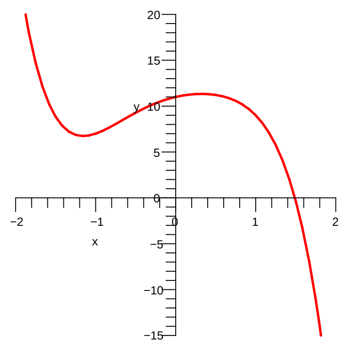


Figure 2.5.5

We show that there are numbers a and b with $a < b$, $f(a) > 0$ and $f(b) < 0$. Because $\lim_{x\to\infty} f(x) = -\infty$, $f(x)$ is negative for x large and positive. Thus, there is a positive number b such that $f(b) < 0$. Similarly, $\lim_{x\to-\infty} f(x) = \infty$, means that when x is negative and of large absolute value, $f(x)$ is positive. So there is a negative number a such that $f(a) > 0$. Thus there are numbers a and b, with $a < b$, such that $f(a) > 0$ and $f(b) < 0$. For instance, $f(-1) = 7$ and $f(2) = -29$.

The number 0 is between $f(a)$ and $f(b)$. Since f is continuous on the interval $[a, b]$, there is a number c in $[a, b]$ such that $f(c) = 0$. (In particular there is a number c in $[-1, 2]$. This number c is a solution to the equation $-x^5 - 3x^2 + 2x + 11 = 0$. ◇

Note that the argument in Example 3 shows that any polynomial of odd degree has a real root. The argument does not hold for polynomials of even degree; the equation $x^2 + 1 = 0$, for instance, has no real solutions.

EXAMPLE 4 Use the Intermediate-Value Property to show that there is a negative number such that $\ln(x + 4) = x^2 - 3$.

SOLUTION We wish to show that there is a negative number c where the function $\ln(x + 4)$ has the same value as the function $x^2 - 3$. The equation $\ln(x+4) = x^2 - 3$ is equivalent to $\ln(x+4) - x^2 + 3 = 0$. The problem reduces to showing that the function $f(x) = \ln(x+4) - x^2 + 3$ has the value 0 for some input c (with $c < 0$).

We will proceed, as we did in the previous example. We want to find numbers a and b (both in $(-\infty, 0)$) such that $f(a)$ and $f(b)$ have opposite signs.

Before beginning the search for a and b, note that $\ln(x + 4)$ is defined only for $x + 4 > 0$, that is, for $x > -4$. To complete the search for a and b, make a table of values of $f(x)$ for some sample arguments in $(-4, 0)$.

x	-3	-2	-1	0
$f(x)$	-6	-0.307	3.099	4.386

We see that $f(-2)$ is negative and $f(-1)$ is positive. Since $m = 0$ lies between $f(-2)$ and $f(-1)$, and f is continuous on $[-2, -1]$, the Intermediate-Value Property asserts that there is a number in $[-2, -1]$ such that $f(c) = 0$. It follows that $\ln(c + 4) = c^2 - 3$. ◇

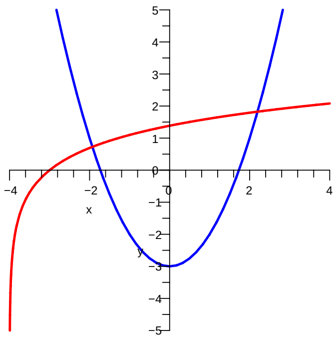


Figure 2.5.6

In Example 4 the Intermediate-Value Property does not tell what c is. The graphs of $\ln(x+1)$ and x^2-3 in Figure 2.5.6 suggest that there are two points of intersection, but only one with a negative input. The graph, and the table of values, suggest that the intersection point occurs when the input is close to -2. Calculations on a calculator or computer show that $c \approx -1.931$.

Permanence Property

The extrema property as well as the intermediate-value property involve the behavior of a continuous function throughout an interval. The next property concerns the "local" behavior of a continuous function.

Consider a continuous function f on an open interval that contains the number a. Assume that $f(a) = p$ is positive. Then it seems plausible that f remains positive in some open interval that contains a. We can say something stronger:

Theorem 2.5.4 (The Permanence Property)**.** *Assume that the domain of a function f contains an open interval that includes the number a. Assume that f is continuous at a and that $f(a) = p$ is positive. Let q be any number less than p. Then there is an open interval including a such that $f(x) \geq q$ for all x in that interval.*

To persuade yourself that the permanence principle is plausible, imagine what the graph of $y = f(x)$ looks like near $(a, f(a))$, as in Figure 2.5.7. That the permanence property is a consequence of the definition of continuity is shown in Example 2 of Section 3.9.

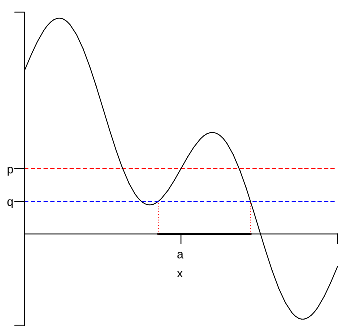


Figure 2.5.7

Summary

This section stated, without proofs, the Extreme-Value Property, the Intermediate-Value Property, and the Permanence Property. Each will be used several times in later chapters.

EXERCISES for Section 2.5

1. For each of the given intervals, find the maximum value of $\cos(x)$ over that interval and the value of x at which it occurs.

(a) $[0, \pi/2]$

(b) $[0, 2\pi]$

2. Does the function $\frac{x^3+x^4}{1+5x^2+x^6}$ have (a) a maximum value for x in $[1,4]$? (b) a minimum value for x in $[1,4]$? If so, use a graphing device to estimate the extreme values.

3. Does the function $2^x - x^3 + x^5$ have (a) a maximum value for x in $[-3,10]$? (b) a minimum value for x in $[-3,10]$? If so, use a graphing device to estimate the extreme values.

4. Does the function x^3 have a maximum value for x in (a) $[2,4]$? (b) $[-3,5]$? (c) $(1,6)$? If so, where does the maximum occur and what is the maximum value?

5. Does the function x^4 have a minimum value for x in (a) $[-5,6]$? (b) $(-2,4)$? (c) $(3,7)$? (d) $(-4,4)$? If so, where does the minimum occur and what is the minimum value?

6. Does the function $2 - x^2$ have (a) a maximum value for x in $(-1,1)$? (b) a minimum value for x in $(-1,1)$? If so, where?

7. Does the function $2 + x^2$ have (a) a maximum value for x in $(-1,1)$? (b) a minimum value for x in $(-1,1)$? If so, where?

8. Show that the equation $x^5 + 3x^4 + x - 2 = 0$ has at least one solution in the interval $[0,1]$.

9. Show that the equation $x^5 - 2x^3 + x^2 - 3x = -1$ has at least one solution in the interval $[1,2]$.

In Exercises 10 to 14 verify the Intermediate-Value Property for the specified function f, the interval $[a,b]$, and the indicated value m. Find all c's in each case.

10. $f(x) = 3x + 5$, $[a,b] = [1,2]$, $m = 10$.

11. $f(x) = x^2 - 2x$, $[a,b] = [-1,4]$, $m = 5$.

12. $f(x) = \sin(x)$, $[a,b] = [\frac{\pi}{2}, \frac{11\pi}{2}]$, $m = -1$.

13. $f(x) = \cos(x)$, $[a, b] = [0, 5\pi]$, $m = \frac{\sqrt{3}}{2}$.
14. $f(x) = x^3 - x$, $[a, b] = [-2, 2]$, $m = 0$.

15. Use the Intermediate-Value Property to show that the equation $3x^3 + 11x^2 - 5x = 2$ has a solution.

16. Show that the equation $2^x = 3x$ has a solution in the interval $[0, 1]$.

17. Does the equation $x + \sin(x) = 1$ have a solution?

18. Does the equation $x^3 = 2^x$ have a solution?

19. Let $f(x) = 1/x$, $a = -1$, $b = 1$, $m = 0$. Note that $f(a) \leq 0 \leq f(b)$. Is there at least one c in $[a, b]$ such that $f(c) = 0$? If so, find c; if not, does this imply the Intermediate-Value Property sometimes does not hold?

20. Use the Intermediate-Value Property to show that there is a positive number such that $\ln(x + 4) = x^2 + 3$.

Exercises 21 and 22 illustrate the Permanence Property.
21. Let $f(x) = 5x$. Then $f(1) = 5$. Find an interval (a, b) containing 1 such that $f(x) \geq 4.9$ for all x in (a, b).

22. Let $f(x) = x^2$. Then $f(2) = 4$. Find an interval (a, b) containing 2 such that $f(x) \geq 3.8$ for all x in (a, b).

23. Let $P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$ be a polynomial of odd degree n and with positive leading coefficient a_n. Show that there is at least one real number r such that $P(r) = 0$.

24. (This continues Exercise 23.) The **factor theorem** from algebra asserts that the number r is a root of a polynomial $P(x)$ if and only if $x - r$ is a factor of $P(x)$. For instance, 2 is a root of the polynomial $x^2 - 3x + 2$ and $x - 2$ is a factor of it: $x^2 - 3x + 2 = (x - 2)(x - 1)$. See also Exercise 47 in Section 8.4.

(a) Use the factor theorem and Exercise 23 to show that every polynomial of odd degree has a factor of degree 1.

(b) Show that none of the polynomials $x^2 + 1$, $x^4 + 1$, or $x^{100} + 1$ has a first-degree factor.

(c) Verify that $x^4 + 1 = \left(x^2 + \sqrt{2}x + 1\right)\left(x^2 - \sqrt{2}x + 1\right)$. (It can be shown using complex numbers that every polynomial with real coefficients is the product of polynomials with real coefficients of degrees at most 2.)

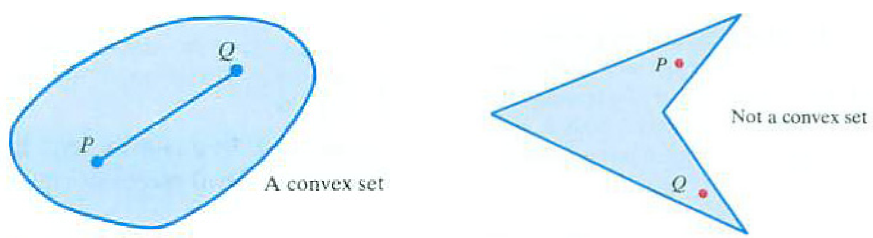
25. Let $f(x) = a_n x^n + \cdots + a_1 x + a_0$ where a_n and a_0 have opposite signs.

(a) Show that the $f(x)$ has a positive root, that is, the equation $f(x) = 0$ has a positive solution.

(b) What can you say about the roots of $f(x)$ if a_n and a_0 have the same sign?

Convex Sets and Curves

A set in the plane bounded by a curve is **convex** if for any two points P and Q in the set the line segment joining them also lies in the set. (See Figure 2.5.8(a).) The boundary of a convex set we will call a **convex curve**. (These ideas generalize to a solid and its boundary surface.) The notion of convexity dates back to Archimedes. Disks, triangles, and parallelograms are convex sets. The quadrilateral shown in Figure 2.5.8(b) is not convex. Convex sets will be referred to in the following exercises and occasionally in the exercises in later chapters.



(a) (b)

Figure 2.5.8 (a) There are no dents in the boundary of a convex set. (b) Not a convex set.

Exercises 26 to 32 concern convex sets and show how the Intermediate-Value Property gives geometric information. In these exercises you will need to define various functions geometrically. *You may assume these functions are continuous.*

26. Let L be a line in the plane and let K be a convex set. Show that there is a line parallel to L that cuts K into two pieces with equal areas.

Follow these steps.

(a) Introduce an x-axis perpendicular to L with its origin on L. Each line parallel to L and meeting K crosses the x-axis at a number x. Label the line L_x. Let a be the smallest and b the largest of these numbers x. (See Figure 2.5.9.) Let the area of K be A.

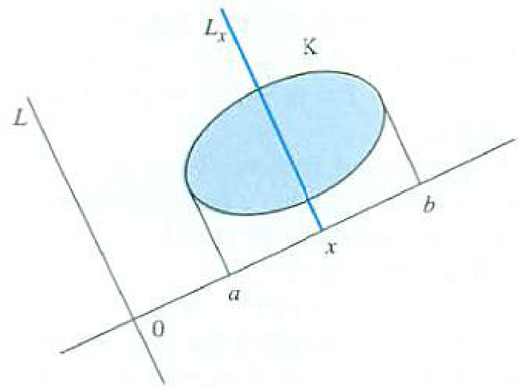


Figure 2.5.9

(b) Let $A(x)$ be the area of K situated to the left of the line L_x corresponding to x. What is $A(a)$? $A(b)$?

(c) Use the Intermediate-Value Property to show that there is an x in $[a, b]$ such that $A(x) = \frac{A}{2}$.

(d) Why does *(c)* show that there is a line parallel to L that cuts K into two pieces of equal areas?

27. Solve the preceding exercise by applying the Intermediate-Value Property to the function $f(x) = A(x) - B(x)$, where $B(x)$ is the area to the right of L_x.

28. Let P be a point in the plane and let K be a convex set. Is there a line through P that cuts K into two pieces of equal areas?

29. Let K_1 and K_2 be two convex sets in the plane. Is there a line that simultaneously cuts K_1 into two pieces of equal areas and cuts K_2 into two pieces of equal areas? This is known as the "two pancakes" question.

30. Let K be a convex set in the plane. Show that there is a line that simultaneously cuts K into two pieces of equal area and cuts the boundary of K into two pieces of equal length.

31. Let K be a convex set in the plane. Show that there are two perpendicular lines that cut K into four pieces of equal areas. (It is not known whether it is always possible to find two perpendicular lines that divide K into four pieces whose areas are $\frac{1}{8}$, $\frac{1}{8}$, $\frac{3}{8}$, and $\frac{3}{8}$ of the area of K, with the parts of equal area sharing an edge, as in Figure 2.5.10.) What if the parts of equal areas are to be opposite each other, instead?

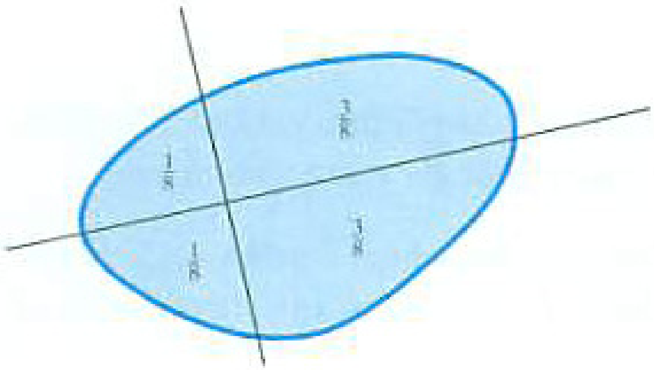


Figure 2.5.10

32. Let K be a convex set in the plane whose boundary contains no line segments. A polygon is said to **circumscribe** K if each edge of the polygon is tangent to the boundary of K.

(a) Is there necessarily a circumscribing equilateral triangle? If so, how many?

(b) Is there necessarily a circumscribing rectangle? If so, how many?

(c) Is there necessarily a circumscribing square?

33. Let f be a continuous function whose domain is the x-axis and has the property that

$$f(x+y) = f(x) + f(y) \qquad \text{for all numbers } x \text{ and } y.$$

For any constant c, $f(x) = cx$ satisfies this equation since $c(x+y) = cx + cy$. This exercise shows that f must be of the form $f(x) = cx$ for some constant c.

(a) Let $f(1) = c$. Show that $f(2) = 2c$.

(b) Show that $f(0) = 0$.

(c) Show that $f(-1) = -c$.

(d) Show that that for any positive integer n, $f(n) = cn$.

(e) Show that that for any negative integer n, $f(n) = cn$.

(f) Show that $f(\frac{1}{2}) = \frac{c}{2}$.

(g) Show that that for any non-zero integer n, $f(\frac{1}{n}) = \frac{c}{n}$.

(h) Show that that for any intger m and any positive integer n, $f(\frac{m}{n}) = \frac{m}{n}c$.

(i) Show that for any irrational number x, $f(x) = cx$. This is where the continuity of f enters. Parts *(h)* and *(i)* together complete the solution.

34.

(a) Let f be a continuous function defined for all real numbers. Is there necessarily a number x such that $f(x) = x$?

(b) Let f be a continuous function with domain $[0, 1]$ such that $f(0) = 1$ and $f(1) = 0$. Is there necessarily a number x such that $f(x) = x$?

35. Let f be a continuous function defined on $(-\infty, \infty)$ such that $f(0) = 1$ and $f(2x) = f(x)$ for all numbers x.

(a) Give an example of such a function f.

(b) Find all functions satisfying these conditions.

Explain your answers.

2.6 Techniques for Graphing

One way to graph a function $f(x)$ is to compute $f(x)$ at several inputs x, plot the points $(x, f(x))$ that you get, and draw a curve through them. This procedure may be tedious and, if you happen to choose inputs that give misleading information, may result in an inaccurate graph.

Another way is to use a calculator that has a graphing routine built in. However, only a portion of the graph is displayed and, if you have no idea what to expect, you may have asked it to display a part of the graph that is misleading or of little interest. At points with large function values, the graph may be distorted by the calculator's choice of scale.

So it pays to be able to get some idea of the general shape of a graph quickly, without having to compute lots of values. This section describes some shortcuts.

Intercepts

The x-coordinates of the points where the graph of a function meets the x-axis are the x-**intercepts** of the function. The y coordinates of the points where a graph meets the y-axis are the y-**intercepts** of the function.

EXAMPLE 1 Find the intercepts of the graph of $y = x^2 - 4x - 5$.
SOLUTION To find the x-intercepts, set $y = 0$, obtaining

$$0 = x^2 - 4x - 5.$$

Fortunately, this quadratic factors nicely:

$$0 = x^2 - 4x - 5 = (x - 5)(x + 1).$$

The equation is satisfied when $x = 5$ or $x = -1$. There are two x-intercepts, 5 and -1. (If the equation did not factor easily, the quadratic formula could be used.)

To find y-intercepts, set $x = 0$, obtaining

$$y = 0^2 - 4 \cdot 0 - 5 = -5.$$

There is only one y-intercept, namely -5.

The intercepts in this case give us three points on the graph. Tabulating a few more points gives the parabola in Figure 2.6.1, where the intercepts are shown as well. ◇

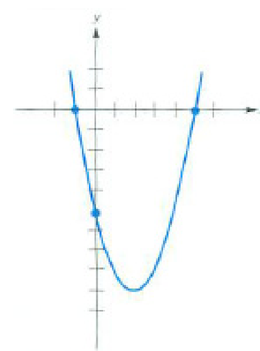


Figure 2.6.1 The graph of $y = x^2 - 4x - 5$, with intercepts.

If $f(x)$ is not defined when $x = 0$, there is no y-intercept. If $f(x)$ is defined when $x = 0$, then it's easy to get the y-intercept; just evaluate $f(0)$. While there is at most one y-intercept, there may be many x-intercepts. To find them, solve the equation $f(x) = 0$. In short,

Finding Intercepts
To find the y-intercept, compute $f(0)$.
To find the x-intercepts, solve the equation $f(x) = 0$.

Symmetry of Odd and Even Functions

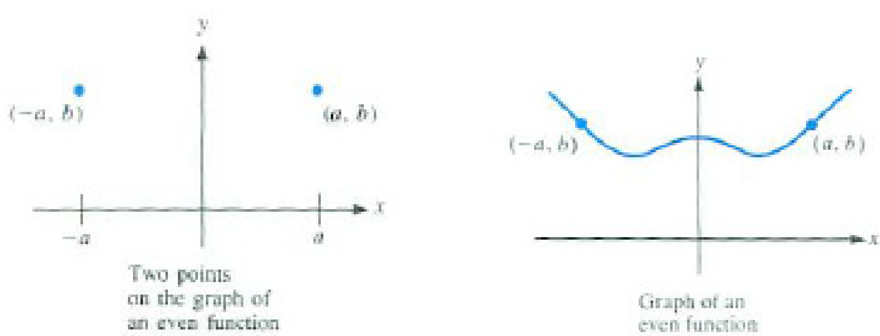
Some functions have the property that when you replace x by $-x$ you get the same value of the function. For instance, the function $f(x) = x^2$ has this property since

$$f(-x) = (-x)^2 = x^2 = f(x).$$

So does the function $f(x) = x^n$ for any *even* integer n. There are fancier functions, such as $3x^4 - 5x^2 + 6x$, $\cos(x)$, and $e^x + e^{-x}$, that also have this property.

DEFINITION *(Even function.)* A function f such that $f(-x) = f(x)$ is called an **even function**.

For an even function f, if $f(a) = b$, then $f(-a) = b$ also. In other words, if the point (a, b) is on the graph of f, so is the point $(-a, b)$, as indicated by Figure 2.6.2(a).



(a) (b)

Figure 2.6.2

This means that the graph of f is symmetric with respect to the y-axis, as shown in Figure 2.6.2(b). So if you notice that a function is even, you can save half the work in finding its graph. First graph it for positive x and then get the part for negative x free of charge by reflecting across the y-axis. If you wanted to graph $y = x^4/(1 - x^2)$, for example, first stick to $x > 0$, then reflect the result.

DEFINITION *(Odd function.)* A function f with $f(-x) = -f(x)$ is called an **odd function**.

The function $f(x) = x^3$ is odd since

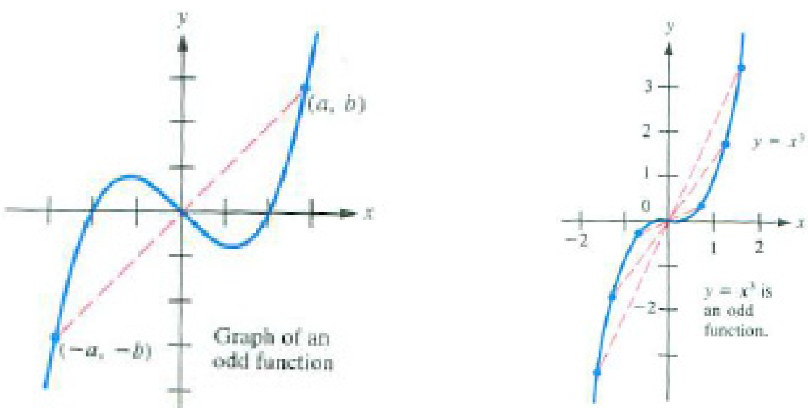
$$f(-x) = (-x)^3 = -(x^3) = -f(x).$$

For any odd integer n, $f(x) = x^n$ is an odd function. The sine function is also odd, since $\sin(-x) = -\sin(x)$.

If the point (a, b) is on the graph of an odd function, so is the point $(-a, -b)$, since

$$f(-a) = -f(a) = -b.$$

(See Figure 2.6.3(a).) Note that the origin $(0, 0)$ is the midpoint of the segment whose ends are (a, b) and $(-a, -b)$. The graph is said to be "symmetric with respect to the origin."



(a) (b)

Figure 2.6.3

If you work out the graph of an odd function for positive x, you can obtain the graph for negative x by reflecting it point by point through the origin. For example, if you graph $y = x^3$ for $x \geq 0$, as in Figure 2.6.3(b), you can complete the graph by reflection with respect to the origin, as indicated by the dashed lines.

Most functions are neither even nor odd. For instance, $x^3 + x^4$ is neither even nor odd since $(-x)^3 + (-x)^4 = -x^3 + x^4$, which is neither $x^3 + x^4$ nor $-(x^3 + x^4)$.

Asymptotes

If $\lim_{x\to\infty} f(x) = L$ where L is a real number, the graph of $y = f(x)$ gets arbitrarily close to the horizontal line $y = L$ as x increases. The line $y = L$ is

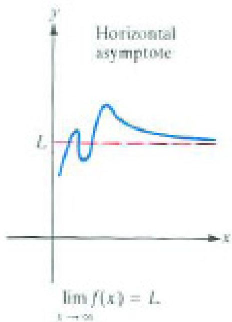


Figure 2.6.4

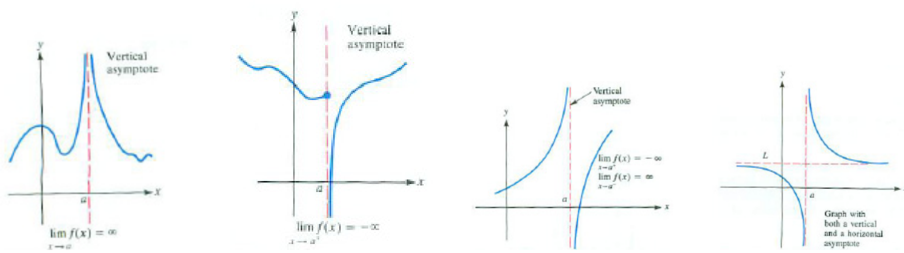
called a **horizontal asymptote** of the graph of f. (See Figure 2.6.4.)

If a graph has an asymptote, we can draw it and use it as a guide in drawing the graph.

If $\lim_{x\to a} f(x) = \infty$, then the graph resembles the vertical line $x = a$ for x near a. The line $x = a$ is called a **vertical asymptote** of the graph of $y = f(x)$. The same term is used if

$$\lim_{x\to a} f(x) = -\infty, \qquad \lim_{x\to a^+} f(x) = \infty \text{ or } -\infty, \quad \text{or} \quad \lim_{x\to a^-} f(x) = \infty \text{ or } -\infty.$$

Figure 2.6.5 illustrates these situations.



(a) (b) (c) (d)

Figure 2.6.5

EXAMPLE 2 Graph $f(x) = 1/(x-1)^2$.

SOLUTION To see if there is any *symmetry*, check whether $f(-x)$ is $f(x)$ of $-f(x)$. We have

$$f(-x) = \frac{1}{(-x-1)^2} = \frac{1}{(x+1)^2}.$$

Since $1/(x+1)^2$ is neither $1/(x-1)^2$ nor $-1/(x-1)^2$, the function $f(x)$ is neither even nor odd. Therefore the graph is *not* symmetric with respect to the y-axis or with respect to the origin.

To determine the y-intercept compute $f(0) = 1/(0-1)^2 = 1$. The y-intercept is 1. To find any x-intercepts, solve the equation $f(x) = 0$, that is,

$$\frac{1}{(x-1)^2} = 0.$$

Since no number has a reciprocal equal to zero, there are no x-intercepts.

To search for a *horizontal asymptote* examine

$$\lim_{x\to\infty} 1/(x-1)^2 \qquad \text{and} \qquad \lim_{x\to-\infty} 1/(x-1)^2.$$

Both limits are 0. The line $y = 0$, that is, the x-axis, is an asymptote both to the right and to the left. Since $1/(x-1)^2$ is positive, the graph lies above the asymptote.

To discover any *vertical asymptotes*, find where the function $1/(x-1)^2$ "blows up" — that is, becomes arbitrarily large (in absolute value). This happens when the denominator $(x-1)^2$ becomes zero. Solving $(x-1)^2 = 0$ we find $x = 1$. The function is not defined for $x = 1$. The line $x = 1$ is a vertical asymptote.

To determine the shape of the graph near the line $x = 1$, we examine the one-sided limits: $\lim_{x\to 1^+} 1/(x-1)^2$ and $\lim_{x\to 1^-} 1/(x-1)^2$. Since the square of a nonzero number is always positive, we see that $\lim_{x\to 1^+} 1/(x-1)^2 = \infty$ and $\lim_{x\to 1^-} 1/(x-1)^2 = \infty$. All this information is displayed in Figure 2.6.6. ◇

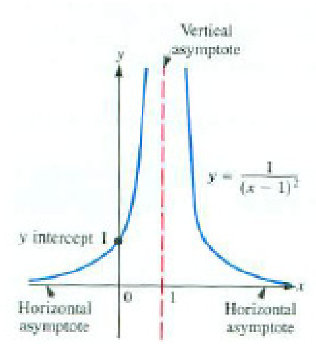


Figure 2.6.6

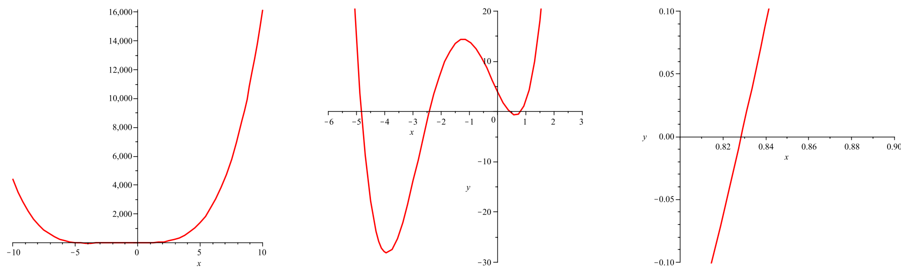
Technology-Assisted Graphing

A graphing utility needs to "know" the function and the viewing window. We will show by three examples some of the obstacles you may run into and how to avoid them. More techniques to help overcome these challenges will be presented in Chapter 4.

The **viewing window** is the portion of the xy-plane to be displayed. We will say the viewing window is $[a, b] \times [c, d]$ when the window extends horizontally from $x = a$ to $x = b$ and vertically from $y = c$ to $y = d$. The graph of a function $y = f(x)$ is created by evaluating $f(x)$ for a sample of numbers x between a and b. The point $(x, f(x))$ is added to the plot. It is customary to connect these points to form the graph of $y = f(x)$. The examples in the remainder of this section demonstrate some of the unpleasant messes that can happen, and how you can avoid them.

EXAMPLE 3 Find a viewing window that shows the general shape of the graph of $y = x^4 + 6x^3 + 3x^2 - 12x + 4$. Use graphs to estimate the location of the rightmost x intercept.

SOLUTION Figure 2.6.7(a) is typical of the first plot of a function. Choose a fairly wide x interval, here $[-10, 10]$, and let the graphing software choose an appropriate vertical range. While this view is useless for estimating any specific x intercept, it is tempting to say that any x intercepts will be between $x = -6$ and $x = 3$. Figure 2.6.7(b) is the graph of this function on the viewing window $[-6, 3] \times [-30, 30]$. Now four x intercepts are visible. The rightmost one occurs around $x = 0.8$. Figure 2.6.7(c) is the result of zooming in on this part of the graph. From this view we estimate that the rightmost x intercept is about 0.83.



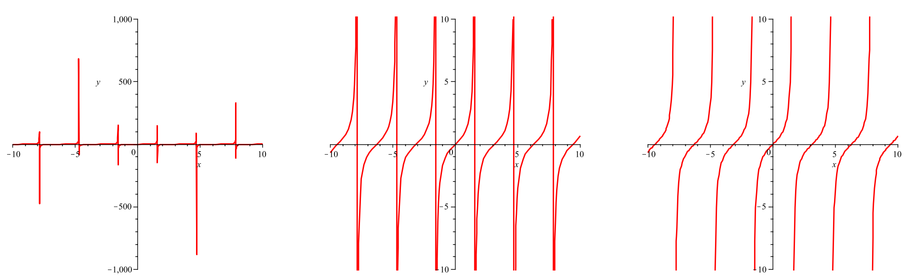
(a) (b) (c)

Figure 2.6.7

In fact, using a CAS, the four x intercepts for this function are found to occur at 0.8284, 0.4142, -2.4142, and -4.8284 (to four decimal places). $\diamond$

Generating a collection of points and connecting the dots can sometimes lead to ridiculous results, as in Example 4.

EXAMPLE 4 Find a viewing window that clearly shows the general shape and periodicity of the graph of $y = \tan(x)$.
SOLUTION A computer-generated plot of $y = \tan(x)$ for x between -10 and 10 with no vertical height of the viewing window is shown in Figure 2.6.8(a). This graph is not periodic; it looks more like an echocardiogram than the graph of one of the trigonometric functions.



(a) (b) (c)

Figure 2.6.8

Notice that the default vertical height is very long: $[-1000, 1000]$. Reducing this by a factor of 100, that is, to $[-10, 10]$, yields Figure 2.6.8(b). This graph is periodic and exhibits the expected behavior.

To understand this plot you must realize that the software selects a sample of input values from the domain, computes the value of tangent of each input, then connects the points in order of the input values. The tangent of the last input smaller than $\pi/2$ is large and positive and the tangent of the first input larger than $\pi/2$ is large but negative. Neither of these points is in the viewing window, but the line segment connecting these points does pass through the viewing window and appears as the "vertical" line at $x = \pi/2$ in Figure 2.6.8(b). Because the tangent is not defined for every odd multiple of $\pi/2$, similar reasoning explains the other "vertical" lines at every odd multiple of $\pi/2$

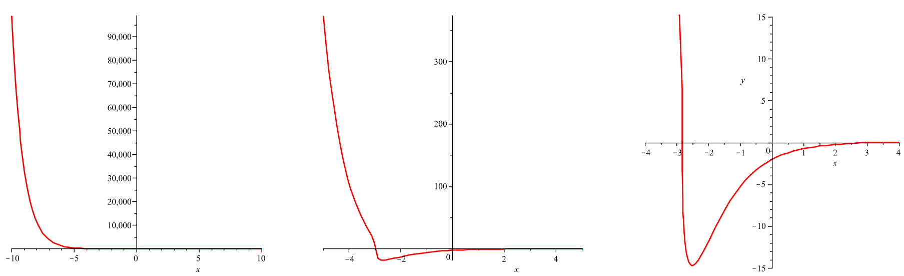
These segments are not really a part of the graph. Figure 2.6.8(c) shows the graph of $y = \tan(x)$ with these extraneous segments removed. ◇

Example 4 illustrates why we must remain alert when using technology. We have to check that the results are consistent with what we already know.

The next example shows that sometimes it is not possible to show all of the important features of a function in a single graph.

EXAMPLE 5 Use one or more graphs to show all major features of the graph $y = e^{-x}\sqrt[3]{x^2 - 8}$.

SOLUTION The graph of this function on the x interval $[-10, 10]$ with the vertical window chosen by the software is shown in Figure 2.6.9(a). In this window, the exponential function dominates the graph.



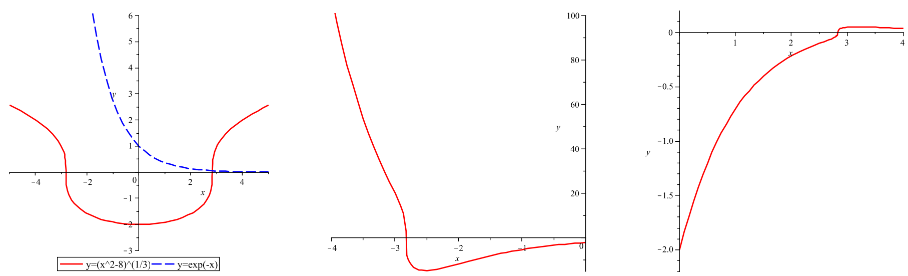
(a) (b) (c)

Figure 2.6.9 Three views of the graph of $y = e^{-x}\sqrt[3]{x^2 - 8}$.

At $x = 0$ the value of the function is $(0 - 8)^{1/3}e^0 = -2$. To get enough detail to see both the positive and negative values of the function, zoom in by reducing the x interval to $[-5, 5]$. The result is Figure 2.6.9(b). Reducing the x interval to $[-4, 4]$ and specifying the y interval as $[-15, 15]$ gives Figure 2.6.9(c).

We could continue to adjust the viewing window until we find suitable views. A more systematic approach is to look at the graphs of $y = \sqrt[3]{x^2 - 8}$ and $y = e^{-x}$ separately, but on the same pair of axes. (See Figure 2.6.10(a), where the solid red curve is $y = \sqrt[3]{x^2 - 8}$ and the dashed blue curve is $y = e^{-x}$.) The exponential growth of e^{-x} for negative values of x stretches (vertically) the graph of $y = \sqrt[3]{x^2 - 8}$ to the left of the y-axis while the exponential decay for $x > 0$ (vertically) compresses the graph of $y = \sqrt[3]{x^2-8}$ to the right of the y-axis.

It is prudent to produce two separate plots to represent the sketch of this function. To the left of the y-axis, with a viewing window of $[-4, 0]\times[-15, 100]$, the graph of the function is shown in Figure 2.6.10(b). To the right of the y-axis, with a much shorter viewing window of $[0, 4] \times [-2.2, 0.2]$, the graph is as shown in Figure 2.6.10(c). ◇



(a) (b) (c)

Figure 2.6.10 Three more views of the graph of $y = e^{-x}\sqrt[3]{x^2-8}$.

Summary

The first half of this section presents three tools for making a quick sketch of the graph of $y = f(x)$ by hand.

1. *Check for intercepts.* Find $f(0)$ to get the y-intercept. Solve $f(x) = 0$ to get the x-intercepts.

2. *Check for symmetry.* Is $f(-x)$ equal to $f(x)$ or $-f(x)$?

3. *Check for asymptotes.* If $\lim_{x\to\infty} f(x) = L$ or $\lim_{x\to-\infty} f(x) = L$ (where L is some real number), then the line $y = L$ is a horizontal asymptote. If $\lim_{x\to a} f(x) = +\infty$ of $-\infty$, then the line $x = a$ is a vertical asymptote. This is also the case whenever $\lim_{x\to a^+} f(x)$ or $\lim_{x\to a^-} f(x)$ is $+\infty$ or $-\infty$.

The second half of the section provides some pointers for using an automatic graphing utility. The key to their use for graphing is to specify an appropriate viewing window.

Computer-Based Mathematics

Graphing calculators provide an easy way to graph a function. Computer algebra systems (CAS) such as Maple, Mathematica, and Derive can perform symbolic operations on mathematical expressions: for example, they can factor a polynomial:

$$x^5 - 2x^4 - 2x^3 + 4x^2 + x - 2 = (x-1)^2(x+1)^2(x-2),$$

express the quotient of two polynomials as the sum of simpler quotients:

$$\frac{36}{x^5 - 2x^4 - 2x^3 + 4x^2 + x - 2} = \frac{-3}{(x+1)^2} - \frac{9}{(x-1)^2} - \frac{4}{x+1} + \frac{4}{x-2},$$

and solve equations, such as

$$\arctan(x^2+1) = \pi/3 \qquad \text{and} \qquad \sin\left(\frac{\pi}{x}\right) - \frac{\pi}{x}\cos\left(\frac{\pi}{x}\right) = 0.$$

Some of these symbolic features are now available on calculators, PDAs, telephones, and other handheld devices.

These tools will continue to develop and you need to be aware that they do exist, and can do much more than graph functions. As they become more widely available and easier to use, they will change the way mathematics is used in the real world. The ability to factor a polynomial or to solve an equation will be less important than the ability to apply basic principles of mathematics and science to set up and to analyze the equations.

EXERCISES for Section 2.6

1. Show that these are even functions.

(a) $x^2 + 2$

(b) $\sqrt{x^4 + 1}$

(c) $1/x^2$

2. Show that these are even functions.

(a) $5x^4 - x^2$

(b) $\cos(2x)$

(c) $7/x^6$

3. Show that these are odd functions.

(a) $x^3 - x$

(b) $x + 1/x$

(c) $\sqrt[3]{x}$

4. Show that these are odd functions.

(a) $2x + \frac{1}{2}x$

(b) $\tan(x)$

(c) $x^{5/3}$

5. Show that these functions are neither odd nor even.

(a) $3 + x$

(b) $(x + 2)^2$

(c) $\frac{x}{x+1}$

6. Show that these functions are neither odd nor even.

(a) $2x - 1$

(b) e^x

(c) $x^2 + 1/x$

7. Label each function as even, odd, or neither.

(a) $x + x^3 + 5x^4$

(b) $7x^4 - 5x^2$

(c) $e^x - e^{-x}$

8. Label each function as even, odd, or neither.

(a) $\frac{1+x}{1-x}$

(b) $\ln(x^2 + 1)$

(c) $\sqrt[3]{x^2 + 1}$

In Exercises 9 to 18 find the x- and y-intercepts, if there are any.

9. $y = 2x + 3$

10. $y = 3x - 7$

11. $y = x^2 + 3x + 2$

12. $y = 2x^2 + 5x + 3$

13. $y = 2x^2 + 1$

14. $y = x^2 + x + 1$

15. $y = \sin(x + 1)$

16. $y = \ln(x^2 + 1)$

17. $y = \frac{x^2-1}{x^2+1}$

18. $y = e^{\cos(x)}$

In Exercises 19 to 24 find all the horizontal and vertical asymptotes.

19. $y = \frac{x+2}{x-2}$

20. $y = \frac{x-2}{x^2-9}$

21. $y = \frac{x}{x^2+1}$

22. $y = \frac{3}{1+e^{-x}}$

23. $y = \frac{\sin(2x)}{x}$

24. $y = \frac{x}{x^2+2x+1}$

In Exercises 25 to 38 graph the function.

25. $y = \frac{1}{x-2}$

26. $y = \frac{1}{x+3}$

27. $y = \frac{1}{x^2-1}$

28. $y = \frac{x}{x^2-2}$

29. $y = \frac{x^2}{1+x^2}$

30. $y = \frac{1}{x^3+x^{-1}}$

31. $y = \frac{3}{1+e^{-x}}$

32. $y = \frac{e^{-x}}{3+e^{-x}}$

33. $y = \frac{e^{-x/2}}{1+e^{-x}}$

34. $y = \frac{2+e^{-x}}{3+e^{-2x}}$

35. $y = \frac{1-e^{x}}{1+e^{x}}$

36. $y = \frac{2-e^{-2x}}{3+e^{-3x}}$

37. $y = \frac{1}{x(x-1)(x+2)}$

38. $y = \frac{x+2}{x^3+x^2}$

Use a graphing utility to sketch a graph of the functions in Exercise 39 to 57. Be sure to indicate the viewing window used to generate your graph.

39. $(x^2 + x - 6)\ln(x + 2)$

40. $(x^2 - x + 6)\ln(x + 2)$

41. $(x^2 + 4)\ln(x + 1)$

42. $(x^2 - 4)\ln(x + 1)$

43. $\frac{x^3}{x^2-4}\arctan\left(\frac{x}{5}\right)$

44. $\frac{(x^2-4)}{x^3}\arctan\left(\frac{x}{5}\right)$

45. $\frac{x^3-3x}{x^2-4}$

46. $\frac{x^3-2x}{x^2-4}$

47. $\frac{\sin(x)}{x}$

48. $\frac{\sin(2x)}{x}$

49. $\frac{\sin(2x)}{3x}$

50. $\frac{\sin(x)}{3x}$

51. $\frac{x-\arctan(x)}{x^3}$

52. $\frac{x-\arctan(x)}{x^3+x}$

53. $\frac{x-\arctan(x)}{x^3-1}$

54. $\frac{x-\arctan(x)}{x^3+1}$

55. $\frac{5x^3+x^2+1}{7x^3+x+4}$

56. $\frac{x^3-3x}{x^2-4}\arctan\left(\frac{x}{4}\right)$

57. $\frac{x^3-2x}{x^2-4}\arctan\left(\frac{x}{4}\right)$

Exercises 58 to 64 concern even and odd functions.

58. If two functions are odd, what can you say about

(a) their sum?

(b) their product?

(c) their quotient?

59. If two functions are even, what can you say about

(a) their sum?

(b) their product?

(c) their quotient?

60. If f is odd and g is even, what can you say about

(a) $f+g$?

(b) fg?

(c) f/g?

61. What, if anything, can you say about $f(0)$ if

(a) f is an even function?

(b) f is an odd function?

Assume 0 is in the domain of f.

62. Which polynomials are even? Explain.

63. Which polynomials are odd? Explain.

64. Is there a function that is both odd and even? Explain.

Exercises 65 to 68 concern tilted asymptotes. Let $A(x)$ and $B(x)$ be polynomials such that the degree of $A(x)$ is equal to 1 more than the degree of $B(x)$. Then when you divide $B(x)$ into $A(x)$, you get a quotient $Q(x)$, which is a polynomial of degree 1, and a remainder $R(x)$, which is a polynomial of degree less than the degree of $B(x)$.

For example, if $A(x) = x^2 + 3x + 4$ and $B(x) = 2x + 2$,

$$\begin{array}{rl} \frac{1}{2}x + 1 & \swarrow Q(x) \\ 2x+2\overline{\big)\,x^2+3x+4} & \\ \underline{x^2+x} & \\ 2x+4 & \\ \underline{2x+2} & \\ 2 & \nwarrow R(x) \end{array}$$

Thus

$$x^2 + 3x + 4 = \left(\frac{1}{2}x + 1\right)(2x + 2) + 2$$

This tells us that

$$\frac{x^2 + 3x + 4}{2x + 2} = \frac{1}{2}x + 1 + \frac{2}{2x + 2}.$$

When x is large, $2/(2x + 2) \to 0$. Thus the graph of $y = \frac{x^2+3x+4}{2x+2}$ is asymptotic to the line $y = \frac{1}{2}x + 1$. (See Figure 2.6.11.)

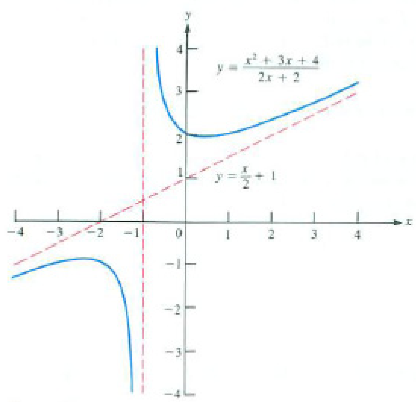


Figure 2.6.11

Whenever the degree of $A(x)$ exceeds the degree of $B(x)$ by exactly 1, the graph of $y = A(x)/B(x)$ has a tilted asymptote. You find it as we did in the example, by dividing $B(x)$ into $A(x)$, obtaining a quotient $Q(x)$ and a remainder $R(x)$. Then

$$\frac{A(x)}{B(x)} = Q(x) + \frac{R(x)}{B(x)}.$$

The asymptote is $y = Q(x)$. In each exercise graph the function, showing all asymptotes.

65. $y = \frac{x^2}{x-1}$

66. $y = \frac{x^3}{x^2-1}$

67. $y = \frac{x^2-4}{x+4}$

68. $y = \frac{x^2+x+1}{x-2}$

Read the directions for your graphing software to learn how to graph a piecewise-defined function. Then use your graphing utility to sketch a graph of the functions in Exercises 69 and 70.

69. $y = \begin{cases} x^2 - x & x < 1 \\ \sqrt{x-1} & x \geq 1 \end{cases}$

70. $y = \begin{cases} \frac{\sin(x)}{x} & x < 0 \\ \sin x & 0 \leq x \geq \pi \\ x - 2 & x > \pi \end{cases}$

Some graphing utilities have trouble plotting functions with fractional exponents. General rules when graphing $y = x^{p/q}$ where p/q is a positive fraction in lowest terms are:

- If p is even and q is odd, then graph $y = |x|^{p/q}$.
- If p and q are both odd, then graph $y = \frac{|x|}{x}|x|^{p/q}$.

Use that advice and a calculator to sketch the graph of each function in Exercises 71 to 74.

71. $y = x^{1/3}$

72. $y = x^{2/3}$

73. $y = x^{4/7}$

74. $y = x^{3/7}$

75. Let $P(x)$ be a polynomial of degree m and $Q(x)$ a polynomial of degree n. For which m and n does the graph of $y = P(x)/Q(x)$ have a horizontal asymptote?

76. Assume you already have drawn the graph of a function $y = f(x)$. How would you obtain the graph of $y = g(x)$ from that graph if

(a) $g(x) = f(x) + 2$?

(b) $g(x) = f(x) - 2$?

(c) $g(x) = f(x - 2)$?

(d) $g(x) = f(x + 2)$?

(e) $g(x) = 2f(x)$?

(f) $g(x) = 3f(x - 2)$?

77. Is there a function f defined for all x such that $f(-x) = 1/f(x)$? If so, how many? If not, explain why there are no such functions.

78. Is there a function f defined for all x such that $f(-x) = 2f(x)$? If so, how many? If not, explain why there are no such functions.

79. Is there a constant k such that the function $f(x) = \dfrac{1}{3^x - 1} + k$ is odd? even?

2.S Chapter Summary

One concept underlies calculus: the limit of a function. For a function defined near a (but not necessarily at a) we ask, "What happens to $f(x)$ as x gets nearer and nearer to a." If the values get nearer and nearer one specific number, we call that number the limit of the function as x approaches a. This concept, which is not met in arithmetic or algebra or trigonometry, distinguishes calculus.

For instance, when $f(x) = (2^x - 1)/x$, which is not defined at $x = 0$, we conjectured on the basis of numerical evidence that $f(x)$ approaches 0.693 (to three decimals). Later we will see that this limit is a certain logarithm. With that information we found that $(4^x - 1)/x$ must approach 2(0.693), which is larger than 1. We then defined e as that number (between 2 and 4) such that $(e^x - 1)/x$ approaches 1 as x approaches 0. The number e is as important in calculus as π is in geometry or trigonometry. The number e is about 2.718 (again to three decimals) and is called Euler's number. That is why a scientific calculator has a key for e^x, the most convenient exponential for calculus, as will become clear in the next chapter.

When angles are measured in radians,

$$\lim_{x\to 0} \frac{\sin(x)}{x} = 1 \qquad \text{and} \qquad \lim_{x\to 0} \frac{1-\cos(x)}{x} = 0.$$

These two limits will serve as the basis of the calculus of trigonometric functions developed in the next chapter. The simplicity of the first limit is one reason that in calculus and its applications angles are measured in radians. If angles were measured in degrees, the first limit would be $\pi/180$, which would complicate computations.

Most of the functions of interest in later chapters are "continuous." The value of such a function at a number a in its domain is the same as the limit of the function as x approaches a. However, we will be interested in a few functions that are not continuous.

A continuous function has three properties, which will be referred to often:

- On a closed interval it attains a maximum value and a minimum value. Extreme-Value Property
- On a closed interval it takes on all values between its values at the end points of the interval. Intermediate -Value Property
- If it is positive at some number and defined at least on an open interval containing that number, then it remains positive at least on some open interval containing that number. More generally, if $f(a) = p > 0$, and q is less than p, then $f(x)$ remains larger than q, at least on some open interval containing a. A similar statement holds when $f(a)$ is negative. Permanence Property

A quick sketch of the graph of a generic continuous function makes the three properties plausible. The permanence property is easily established from the definition of continuity, as shown in Example 2 of Section 3.9. The other two are established, in an advanced calculus course, using only the precise definition of continuity and properties of the real numbers — but no pictures. Such strictness is necessary because there are some pretty wild continuous functions. For instance, there is one such that when you zoom in on its graph at any point, the parts of the graph nearer and nearer the point do not look like straight line segments.

The initial steps in the analysis of a function utilize intercepts, symmetry, and asymptotes. The same ideas are also helpful when selecting an appropriate viewing window when using an electronic graphing utility. Additional techniques will be added in Chapter 4, particularly Section 4.3.

EXERCISES for 2.S

1. Define Euler's constant, e, and give its decimal value to five places.

In Exercises 2 to 4 state the given property in your own words, using as few mathematical symbols as possible.

2. The Maximum-Value Property.

3. The Intermediate-Value Property.

4. The Permanence Property.

5.

(a) Verify that $x^5 - y^5 = (x-y)(x^4 + x^3y + x^2y^2 + xy^3 + y^4)$.

(b) Use (a) to find $\lim_{x\to a} \frac{x^5-a^5}{x-a}$.

6. Let $f(x) = \frac{1}{x+2}$ for x not equal to -2. Is there a continuous function $g(x)$, defined for all x, that equals $f(x)$ when x is not -2? Explain your answer.

7. Let $f(x) = \frac{2^x-1}{x}$ for x not equal to 0. Is there a continuous function $g(x)$, defined for all x, that equals $f(x)$ when x is not 0? Explain your answer.

8. Let $f(x) = \sin(1/(x-1))$ for x not equal to 1. Is there a continuous function $g(x)$, defined for all x, that equals $f(x)$ when x is not 1? Explain your answer.

9. Let $f(x) = x\sin(1/x)$ for x not equal to 0. Is there a continuous function $g(x)$, defined for all x, that equals $f(x)$ when x is not 0? Explain your answer.

10. Show that $\lim_{x\to 1} \frac{x^{1/3}-1}{x-1} = \frac{1}{3}$ by first writing the denominator as $\left(x^{1/3}\right)^3 - 1$ and using the factorization $u^3 - 1 = (u-1)(u^2+u+1)$.

11. Use the factorization in Exercise 5 to find $\lim_{x\to a} \frac{x^{-5}-a^{-5}}{x-a}$.

12. Assume $b > 1$. If $\lim_{x\to 0} \frac{b^x-1}{x} = L$, find $\lim_{x\to 0} \frac{(1/b)^x-1}{x}$

13. By sketching a graph, show that if a function is not continuous it may not

(a) have a maximum even if its domain is a closed interval,

(b) satisfy the Intermediate-Value Theorem, even if its domain is a closed interval,

(c) have the Permanence Property, even if its domain is an open interval.

14. Let g be an increasing function such that $\lim_{x\to a} g(x) = L$.

(a) Sketch the graph of a function f whose domain includes an open interval around L such that

$$f\left(\lim_{x\to a} g(x)\right) \text{ and } \lim_{x\to a} f(g(x))$$

both exist but they are not equal

(b) What property of f would assure us that the two limits in (a) would be equal?

We obtained $\lim_{x\to a} \frac{x^n-a^n}{x-a}$ be exploiting the factorization of $x^n - a^n$. Calling $x - a$ simply h, that limit can be written as $\lim_{h\to 0} \frac{(a+h)^n-a^n}{h}$. This limit can be evaluated, but by different algebra, as Exercises 15 and 16 show.

15.

(a) Show that $(a+h)^2 = a^2 + 2ah + h^2$.

(b) Use (a) to evaluate $\lim_{h\to 0} \frac{(a+h)^2-a^2}{h}$.

16.

(a) Show that $(a+h)^3 = a^3 + 3a^2h + 3ah^2 + h^3$.

(b) Use (a) to evaluate $\lim_{h\to 0} \frac{(a+h)^3-a^3}{h}$.

17. If you are familiar with the Binomial Theorem, use it to show that for any positive integer n, $\lim_{h\to 0}\frac{(a+h)^n-a^n}{h} = na^{n-1}$. The Binomial Theorem expresses $(a+b)^n$, when multiplied out, as the sum of $n+1$ terms. Using calculus, we will develop it in Section 5.5 (Exercise 31).

In Exercises 18 to 21 find each limit.

18. $\lim_{x\to\infty}\frac{\ln(5x)}{\ln(4x^2)}$

19. $\lim_{x\to\infty}\frac{\ln(5x)}{\ln(4x)}$

20. $\lim_{x\to\infty}\frac{\log_2(x^2)}{\log_4(x)}$

21. $\lim_{x\to\infty}\frac{\log_3(x^5)}{\log_9(x)}$

22. Find $\lim_{h\to 0}\frac{(e^2)^h-1}{h}$ by factoring the numerator.

23. Assuming that $\lim_{x\to 0^+} x^x = 1$ and that $\lim_{x\to\infty}\ln(x) = \infty$, deduce each of the following limits:

(a) $\lim_{x\to 0} x\ln(x)$

(b) $\lim_{x\to\infty}\frac{\ln(x)}{x}$ (Use (a).)

(c) $\lim_{x\to\infty} x^{1/x}$

(d) $\lim_{x\to\infty}\frac{\ln(x)}{x^k}$, k a positive constant

(e) $\lim_{x\to\infty}\frac{x}{e^x}$

(f) $\lim_{x\to\infty}\frac{x^n}{e^x}$, n a positive integer

(g) $\lim_{x\to\infty}\frac{(\ln(x))^n}{x}$, n a positive integer

24. Consider the equation $2000^x = 2001^{2000}$.

(a) Without doing any calculations, estimate x.

(b) Use logarithms to estimate x.

25. Consider the equation $x^{2001} = 2001^{2000}$.

(a) Without doing any calculations, estimate x.

(b) Use logarithms to estimate x.

26. Let $f = g + h$, where g is an even function and f is an odd function. Express g and h in terms of f.

27.

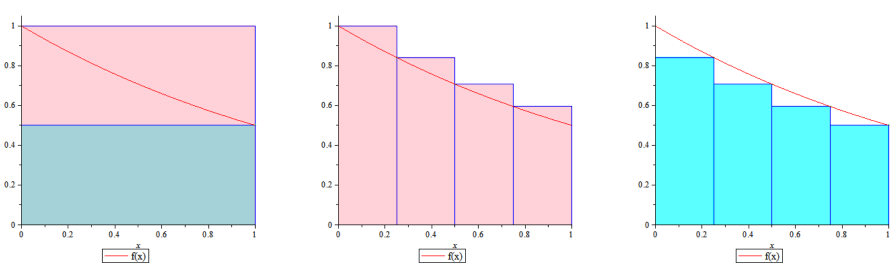
(a) Show that any function f can be written as the sum of an even function and an odd function.

(b) In how many ways can a given function be written that way?

28. If f is an odd function and g is an even function, what, if anything, can be said about (a) fg, (b) f^2, (c) $f + g$, (d) $f + f$, and (e) f/g? Explain.

Exercises 29 to 31 provide an early glimpse of an area problem that is easily addressed with calculus. These exercises require only a basic understanding of geometric series and limits.

29. Figure 2.S.1 shows the region R under the curve $y = 2^{-x}$ and above the interval $[0, 1]$.



(a) (b) (c)

Figure 2.S.1

As shown in Figure 2.S.1(a), the region R lies inside a square of area 1 and contains a rectangle of area 1/2. So its area is between 1/2 and 1.

(a) Use the four rectangles in Figure 2.S.1(b) to estimate the area of R.

(b) Use the four rectangles in Figure 2.S.1(c) to estimate the area of R.

(c) On the basis of (a) and (b), fill in the blanks to complete the following sentence:

The area of R is between _____________ and _____________.

30. Instead of the four rectangles used in the preceding exercise use 100 rectangles, all of the same width, to get even closer estimates of the area of R. (You will run into the sum of 100 numbers. To save work, note that they form a geometric progression.)

31. Let R be the region under $y = 2^{-x}$, to the right of the y-axis, and above the x-axis. Decide if R is finite or infinite. (The technique in Exercises 29 and 30 can help you decide. Assume that $\lim_{n\to\infty} \frac{1}{2^n} = 0$.)

32.

Sam: Did you know that the AB limit comes right out of the $A + B$ limit, at least for positive functions?

Jane: I don't believe it.

Sam: All I do is write $g(x)$ as $e^{\ln g(x)}$ and $h(x)$ as $e^{\ln h(x)}$.

Jane: I see. But you also use the continuity of the exponential and natural logarithm functions.

Sam: So I do. Even so it's neat.

Check that Sam is right.

33. Using an approach like the one in Exercise 32, obtain the A^B limit from the AB limit for two positive functions.

34. The graph of some function f whose domain is $[2, 4]$ and range is $[1, 3]$ is shown in Figure 2.S.2. Sketch the graphs of the following functions and state their domain and range.

(a) $g(x) = -3f(x)$,

(b) $g(x) = f(x + 1)$,

(c) $g(x) = f(x - 1)$,

(d) $g(x) = 3 + f(x)$,

(e) $g(x) = f(2x)$,

(f) $g(x) = f(x/2)$,

(g) $g(x) = f(2x - 1)$.

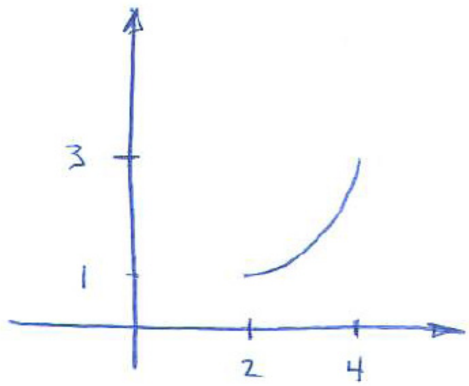


Figure 2.S.2

35. For a constant k, find $\lim_{h\to 0}\frac{(e^k)^h-1}{h}$. (Replace h in the denominator by hk, but do it legally.)

36.

(a) Calculate $(0.99999)^x$ for various large values of x.

(b) Using the evidence gathered in (a), conjecture the value of $\lim_{x\to\infty}(0.99999)^x$.

(c) Why is $\lim_{x\to\infty}(0.99999)^{x+1}$ the same as $\lim_{x\to\infty}(0.99999)^x$?

(d) Denoting the limit in (b) as L, show that $0.99999L = L$.

(e) Using (d), find L.

37. (Contributed by G. D. Chakerian) This exercise obtains $\lim_{\theta\to 0}\frac{\sin(\theta)}{\theta}$ without using areas. Figure 2.S.3 shows a circle C of radius 1 with center at the origin and a circle $C(r)$ of radius $r > 1$ that passes through the center of C. Let $S(r)$ be the part of $C(r)$ that lies within C. Its ends are P and Q. Let θ be the angle subtended by the top half of $S(r)$ at the center of $C(r)$. Note that as $r\to\infty$, $\theta\to 0$. Define $A(\theta)$ to be the length of the arc $S(r)$ as a function of θ.

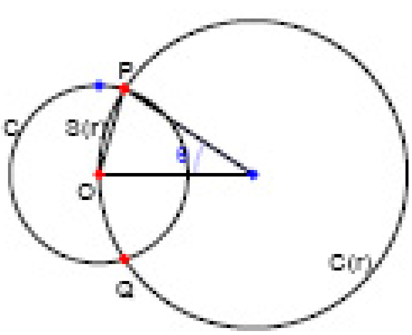


Figure 2.S.3

(a) Looking at Figure 2.S.3, determine $\lim_{\theta \to 0} A(\theta)$. (What happens to P as $r \to \infty$?)

(b) Show that $A(\theta)$ is $\frac{\theta/2}{\sin(\theta/2)}$.

(c) Combining (a) and (b), show that $\lim_{\theta \to 0} \frac{\sin(\theta)}{\theta} = 1$.

Calculus is Everywhere # 3
Bank Interest and the Annual Percentage Yield

The Truth in Savings Act, passed in 1991, requires a bank to post the Annual Percentage Yield (APY) on deposits. That yield depends on how often the bank computes the interest earned, perhaps as often as daily or as seldom as once a year. Imagine that you open an account on January 1 by depositing \$1000. The bank pays interest monthly at the rate of 5 percent a year. How much will there be in your account at the end of the year? For simplicity, assume all the months have the same length. To begin, we find out how much there is in the account at the end of the first month. The account then has the initial amount, \$1000, plus the interest earned during January. Because there are 12 months, the interest rate in each month is 5 percent divided by 12, which is 0.05/12 percent per month. So the interest earned in January is \$1000 times 0.05/12. At the end of January the account then has

$$\$1000 + \$1000(0.05/12) = \$1000(1 + 0.05/12).$$

The initial deposit is "magnified" by the factor $(1 + 0.05/12)$.

The amount in the account at the end of February is found the same way, but the initial amount is $\$1000(1 + 0.05/12)$ instead of \$1000. Again the amount is magnified by the factor $1 + 0.05/12$, to become

$$\$1000(1 + 0.05/12)^2.$$

The amount at the end of March is

$$\$1000(1 + 0.05/12)^3.$$

At the end of the year the account has grown to

$$\$1000(1 + 0.05/12)^{12},$$

which is about \$1051.16.

The deposit earned \$51.16. If instead the bank computed the interest only once, at the end of the year, so-called "simple interest," the deposit would earn only 5 percent of \$1000, which is \$50. The depositor benefits when the interest is computed more than once a year, so-called "compound interest." A competing bank may offer to compute the interest every day. In that case, the account would grow to

$$\$1000(1 + 0.05/365)^{365},$$

which is about \$1051.27, eleven cents more than the first bank offers. More generally, if the initial deposit is A, the annual interest rate is r, and interest is computed n times a year, the amount at the end of the year is

$$A(1 + r/n)^n. \tag{C.3.1}$$

In the examples, A is \$1000, r is 0.05, and n is 12 and then 365. Of special interest is the case when A is 1 and r is a generous 100 percent, that is, $r = 1$. Then (C.3.1) becomes

$$(1 + 1/n)^n. \tag{C.3.2}$$

How does (C.3.2) behave as n increase?

Table C.3.1 shows a few values of (C.3.2), to five decimal places.

n	$(1+1/n)^n$	$(1+1/n)^n$
1	$(1+1/1)^1$	2.00000
2	$(1+1/2)^2$	2.25000
3	$(1+1/3)^3$	2.37037
10	$(1+1/10)^{10}$	2.59374
100	$(1+1/100)^{100}$	2.70481
1000	$(1+1/1000)^{1000}$	2.71692

Table C.3.1

The base, $1 + 1/n$, approaches 1 as n increases, suggesting that (C.3.2) may approach a number near 1. However, the exponent gets large, so we are multiplying lots of numbers, all of them larger than 1. It turns out that as n increases $(1 + 1/n)^n$ approaches the number e defined in Section 2.2. One can write

$$\lim_{x\to 0^+} (1 + x)^{1/x} = e.$$

Note that the exponent, $1/x$, is the reciprocal of the "small number" x.

With that fact at our disposal, we can figure out what happens when an account opens with \$1000, the annual interest rate is 5 percent, and the interest is compounded more and more often. In that case we would be interested in

$$1000 \lim_{n\to\infty} \left(1 + \frac{0.05}{n}\right)^n.$$

Unfortunately, the exponent n is not the reciprocal of the small number $0.05/n$. But a little algebra can overcome that nuisance, for

$$\left(1 + \frac{0.05}{n}\right)^n = \left(\left(1 + \frac{0.05}{n}\right)^{\frac{n}{0.05}}\right)^{0.05}. \tag{C.3.3}$$

The expression in parentheses has the form "(1 + small number) raised to the reciprocal of that small number." Therefore, as n increases, (C.3.3)) approaches $e^{0.05}$, which is about 1.05127. No matter how often interest is compounded, the \$1000 would never grow beyond \$1051.27.

The definition of e given in Section 2.2 has no obvious connection to the fact that $\lim_{x\to 0^+}(1+x)^{1/x}$ equals the number e. It seems "obvious," by thinking in terms of banks, that as n increases, so does $(1+1/n)^n$. However, as n increases, the base decreases and the exponent increases, producing two competing influences. Without thinking about banks, try showing that $(1+1/n)^n$ does increase. (This limit will be evaluated in Section 3.4.)

EXERCISES

1. A dollar is deposited at the beginning of the year in an account that pays an interest rate r of 100% a year. Let $f(t)$, for $0 \le t \le 1$, be the amount in the account at time t. Graph the function if the bank pays

(a) only simple interest, computed only at $t = 1$.

(b) compound interest, twice a year computed at $t = 1/2$ and 1.

(c) compound interest, three times a year computed at $t = 1/3$, $2/3$, and 1.

(d) compound interest, four times a year computed at $t = 1/4$, $1/2$, $3/4$, and 1.

(e) Are the functions in (a), (b), (c), and (d) continuous?

(f) One could expect the account that is compounded more often than another would always have more in it. Is that the case?

Chapter 3

The Derivative

The two main applied concepts of calculus are defined with the aid of limits. In this chapter we meet the first of these: the **derivative of a function**. The derivative tells how rapidly or slowly a function changes. For instance, if the function describes the position of a moving particle, the derivative tells us its velocity.

The definition of a derivative rests on the notion of a limit. The particular limits examined in Chapter 2 are the basis for finding the derivatives of all functions of interest.

The goal of this chapter is twofold: to develop those techniques and to impart an understanding of the meaning of a derivative.

3.1 Velocity and Slope: Two Problems with One Theme

This section discusses two problems that at first glance may seem unrelated. The first concerns the slope of a tangent line to a curve. The second involves velocity. A little arithmetic will show that they are both different versions of one mathematical idea: the *derivative*.

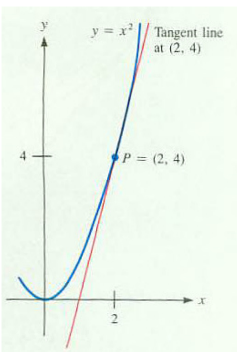


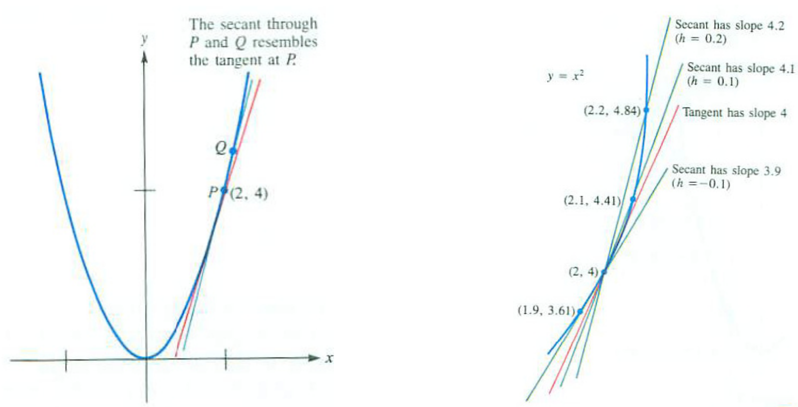
Figure 3.1.1

Slope

Our first problem is important because it is related to finding the straight line that most closely resembles a given graph near a point on the graph.

EXAMPLE 1 What is the slope of the tangent line to the graph of $y = x^2$ at the point $P = (2, 4)$, as shown in Figure 3.1.1?

In Section 2.1 we used a point Q on the curve near P to determine a line that closely resembles the tangent line at $(2, 4)$. Using $Q = (2.01, 2.01^2)$ and also $Q = (1.99, 1.99^2)$, we found that the slope of the tangent line is between 4.01 and 3.99. We did not find the slope of the tangent at $(2, 4)$. Rather than making more estimates by choosing points nearer $(2, 4)$, such as $(2.00001, 2.00001^2)$, it is simpler to consider a typical point.



(a) (b)

Figure 3.1.2

SOLUTION Consider the line through $P = (2, 4)$ and $Q = (x, x^2)$ when x is close to 2, but not equal to 2. (See Figure 3.1.2(a).) It has slope

$$\frac{x^2 - 2^2}{x - 2}.$$

To find out what happens to the quotient as Q moves closer to P (and x moves closer to 2) apply the techniques of limits developed in Chapter 2. We have

$$\lim_{x\to 2}\frac{x^2-2^2}{x-2}=\lim_{x\to 2}\frac{(x+2)(x-2)}{x-2}=\lim_{x\to 2}(x+2)=4.$$

Thus, we expect the tangent line to $y=x^2$ at $(2,4)$ to have slope 4.

Figure 3.1.2(b) shows how secants approximate the tangent line. It suggests a blowup of a small part of the curve $y=x^2$. $\diamond$

We never had to make any estimates with specific choices of the nearby point Q. We did not even have to draw the curve.

Velocity

If an airplane or automobile is moving at a constant velocity, we know that distance traveled equals velocity times time. Thus

$$\text{velocity}=\frac{\text{distance traveled}}{\text{elapsed time}}.$$

If the velocity is not constant, we still may speak of its average velocity, which is defined as

$$\text{average velocity}=\frac{\text{distance traveled}}{\text{elapsed time}}.$$

If a trip from San Francisco to Los Angeles, a distance of 400 miles, takes 8 hours, the average velocity is 400/8 or 50 miles per hour.

Suppose that up to time t_1 you have traveled a distance D_1, while up to time t_2 you have traveled a distance D_2, where $t_2>t_1$. Then during the time interval $[t_1,t_2]$ the distance traveled is D_2-D_1. Thus the average velocity during the time interval $[t_1,t_2]$, which has duration t_2-t_1, is

$$\text{average velocity}=\frac{D_2-D_1}{t_2-t_1}.$$

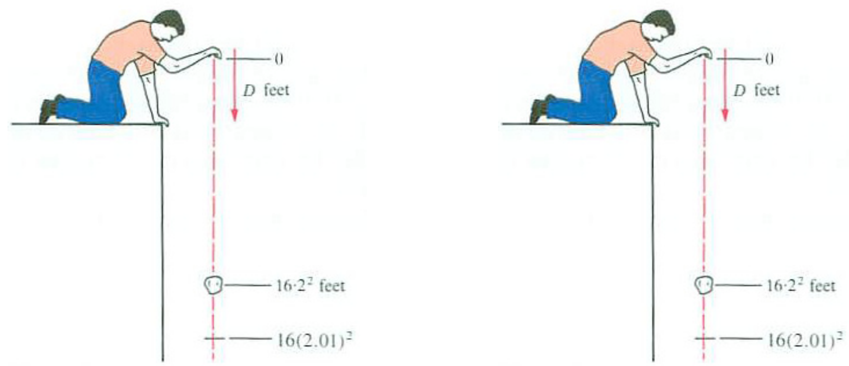
The arithmetic of average velocity is the same as that for the slope of a line.

The next problem shows how to find the velocity at any instant for an object whose velocity is not constant.

EXAMPLE 2 A rock initially at rest falls $16t^2$ feet in t seconds. What is its velocity after 2 seconds? Whatever it turns out to be, it will be called the **instantaneous velocity**.

SOLUTION

To start, make an estimate by finding the average velocity of the rock during a short time interval, say from 2 to 2.01 seconds. At the start of



(a) (b)

Figure 3.1.3 Note: (b) needs to have 2.01 replaced by t.

this interval the rock has fallen $16(2^2) = 64$ feet. By its end it has fallen $16(2.01^2) = 16(4.0401) = 64.6416$ feet. So, during 0.01 seconds the rock fell 0.6416 feet. Its average velocity is

$$\text{average velocity} = \frac{64.6416 - 64}{2.01 - 2} = \frac{0.6416}{0.01} = 64.16 \text{ feet per second.}$$

This is an estimate of the velocity at time $t = 2$ seconds. (See Figure 3.1.3(a).)

Rather than make another estimate with the aid of a shorter interval of time, let us consider the typical time interval from 2 to t seconds, $t > 2$. (Although we will keep $t > 2$, estimates could just as well be made with $t < 2$.) During $t-2$ seconds the rock travels $16(t^2) - 16(2^2) = 16(t^2 - 2^2)$ feet, as shown in Figure 3.1.3(b). The average velocity of the rock is

$$\text{average velocity} = \frac{16t^2 - 16(2^2)}{t-2} = \frac{16(t^2 - 2^2)}{t-2} \text{ feet per second.}$$

When t is close to 2, what happens to the average velocity? It approaches

$$\lim_{t\to 2} \frac{16(t^2 - 2^2)}{t-2} = 16 \lim_{t\to 2} \frac{t^2 - 2^2}{t-2} = 16 \lim_{t\to 2}(t+2) = 16 \cdot 4 = 64 \text{ feet per second.}$$

We say that the (instantaneous) velocity at time $t = 2$ is 64 feet per second. ◇

Even though Examples 1 and 2 seem unrelated, their solutions turn out to be practically identical: the slope in Example 1 is approximated by the quotient

$$\frac{x^2 - 2^2}{x-2}$$

and the velocity in Example 2 is approximated by the quotient

$$\frac{16t^2 - 16(2^2)}{t-2} = 16 \cdot \frac{t^2 - 2^2}{t-2}.$$

The only difference between the solutions is that the second quotient has a factor of 16 and x is replaced with t. This may not be too surprising, since the functions involved, x^2 and $16t^2$ differ by a factor of 16. (That the independent variable is named t in one case and x in the other does not affect the computations.)

The Derivative of a Function

In both the slope and velocity problems we were lead to studying similar limits. For the function x^2 it was

$$\frac{x^2 - 2^2}{x-2} \quad \text{as } x \text{ approaches 2.}$$

For the function $16t^2$ it was

$$\frac{16t^2 - 16(2^2)}{t-2} \quad \text{as } t \text{ approaches 2.}$$

In both cases we formed the change in outputs divided by change in inputs and then found the limit of this quotient as the change in inputs became smaller and smaller. This can be done for other functions, and brings us to one of the two key ideas in calculus, the derivative of a function.

> **DEFINITION** *(Derivative of a function at a number a)* Let f be a function that is defined in an open interval that contains the number a. If
>
> $$\lim_{x\to a} \frac{f(x) - f(a)}{x-a}$$
>
> exists, it is called the **derivative of** f **at** a, and is denoted $f'(a)$. In this case the function f is said to be **differentiable at** a.

Read $f'(a)$ as "f prime at a" or "the derivative of f at a."

EXAMPLE 3 Find the derivative of $f(x) = 16x^2$ at 2.
SOLUTION In this case, $f(x) = 16x^2$ for any input x. By definition, its derivative at 2 is

$$\lim_{x\to 2} \frac{f(x) - f(2)}{x-2} = \lim_{x\to 2} \frac{16x^2 - 16(2^2)}{x-2} = 16 \lim_{x\to 2} \frac{x^2 - 2^2}{x-2} = 16 \lim_{x\to 2}(x+2) = 64.$$

We say that "the derivative of the function $f(x)$ at 2 is 64" and write $f'(2) = 64$. ◇

Now that we have the derivative of f, we can define the slope of its graph at a point $(a, f(a))$ as the value of the derivative, $f'(a)$. Then we define the **tangent line** at $(a, f(a))$ as the line through $(a, f(a))$ whose slope is $f'(a)$.

EXAMPLE 4 Find the derivative of e^x at a.
SOLUTION We must find

$$\lim_{x\to a} \frac{e^x - e^a}{x - a}. \tag{3.1.1}$$

The limit is not obvious. Let us write x as $a + h$ and see what happens as h approaches 0. The denominator $x - a$ is just h. Then (3.1.1) now reads

$$\lim_{h\to 0} \frac{e^{a+h} - e^a}{h}.$$

This form of the limit is more convenient:

$$\begin{aligned}
\lim_{h\to 0} \frac{e^{a+h} - e^a}{h} &= \lim_{h\to 0} \frac{e^a e^h - e^a}{h} && \text{law of exponents} \\
&= e^a \lim_{h\to 0} \frac{e^h - 1}{h} && \text{factor out a constant} \\
&= e^a \cdot 1 && \text{Section 2.2} \\
&= e^a.
\end{aligned}$$

So the limit is e^a. That is, the derivative of e^x is e^x itself. $\diamond$

Differentiability and Continuity

If a function is differentiable at each point in its domain the function is said to be **differentiable**.

A small piece of the graph of a differentiable function at a looks like part of a straight line. You can check this by zooming in on the graph of a function of your choice. Differential calculus can be described as the study of functions whose graphs locally look almost like a line.

It is no surprise that a differentiable function is always continuous. To show that a function is continuous at an argument a in its domain we must show that $\lim_{x\to a} f(x)$ equals $f(a)$, which amounts to showing $\lim_{x\to a}(f(x) - f(a))$ equals 0. To relate this limit to $f'(a)$ we rewrite it:

$$\begin{aligned}
\lim_{x\to a}(f(x) - f(a)) &= \lim_{x\to a} \frac{f(x) - f(a)}{x - a}(x - a) \\
&= \lim_{x\to a} \left(\frac{f(x) - f(a)}{x - a}\right) \lim_{x\to a}(x - a) \\
&= f'(a) \cdot 0 && \text{(definition of the derivative} \\
&= 0.
\end{aligned}$$

So, f is continuous at a.

A function can be continuous yet not differentiable. For instance, $f(x) = |x|$ is continuous but not differentiable at 0, as Figure 3.1.4 suggests.

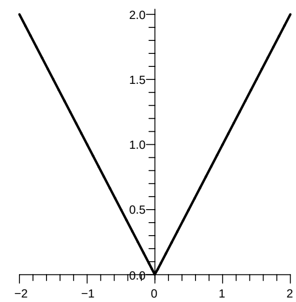


Figure 3.1.4

Summary

From a mathematical point of view, the problems of finding the slope of the tangent line and the velocity of the rock are the same. In each case estimates lead to the same type of quotient, $\frac{f(x)-f(a)}{x-a}$. The behavior of this difference quotient is studied as x approaches a. In each case the answer is a limit, called the derivative of the function at the given number, a. Finding the derivative of a function is called "differentiating" the function.

EXERCISES for Section 3.1

1. Let g be a function and b a number. Define the "derivative of g at b".

2. How is the tangent line to the graph of f at $(a, f(a))$ defined?

3.

(a) Find the slope of the tangent line to $y = x^2$ at $(4, 16)$.

(b) Use it to draw the tangent line to the curve at $(4, 16)$.

4.

(a) Find the slope of the tangent line to $y = x^2$ at $(-1, 1)$.

(b) Use it to draw the tangent line to the curve at $(-1, 1)$.

Exercises 5 to 17 concern slope. Use the technique of Example 1 to find the slope of the tangent line to the curve at the point.

5. $y = x^2$ at the point $(3, 3^2) = (3, 9)$

6. $y = x^2$ at $(\frac{1}{2}, (\frac{1}{2})^2) = (\frac{1}{2}, \frac{1}{4})$

7. $y = x^3$ at $(2, 2^3) = (2, 8)$

8. $y = x^3$ at $(-2, (-2)^3) = (-2, -8)$

9. $y = \sin(x)$ at $(0, \sin(0)) = (0, 0)$

10. $y = \cos(x)$ at $(0, \cos(0)) = (0, 1)$

11. $y = \cos(x)$ at $(\pi/4, \cos(\pi/4)) = (\pi/4, \sqrt{2}/2)$

12. $y = \sin(x)$ at $(\pi/6, \sin(\pi/6)) = (\pi/6, 1/2)$

13. $y = 2^x$ at $(1, 2^1) = (1, 2)$

14. $y = 4^x$ at $(1/2, 4^{1/2}) = (1/2, 2)$

15.

(a) Graph $y = 1/x$ and, by eye, draw the tangent at $(2, 1/2)$.

(b) Using a ruler, measure a rise-run triangle to estimate the slope of the tangent line drawn in (a).

(c) Using no pictures at all, find the slope of the tangent line to the curve $y = 1/x$ at $(2, 1/2)$.

16.

(a) Sketch the graph of $y = x^3$ and the tangent line at $(0, 0)$.

(b) Find the slope of the tangent line to the curve $y = x^3$ at the point $(0, 0)$

Be careful when sketching the graph near $(0, 0)$. In this case the tangent line crosses the curve.

17.

(a) Sketch the graph of $y = x^2$ and the tangent line at $(1, 1)$.

(b) Find the slope of the tangent line to the curve $y = x^2$ at the point $(0, 0)$

In Exercises 18 to 21 use the method of Example 2 to find the velocity of the rock after

18. 3 seconds

19. $\frac{1}{2}$ second

20. 1 second

21. $\frac{1}{4}$ second

22. An object travels t^3 feet in the first t seconds.

(a) How far does it travel during the time interval from 2 to 2.1 seconds?

(b) What is the average velocity during that time interval?

(c) Let h be any positive number. Find the average velocity of the object from time 2 to $2 + h$ seconds. (To find $(2 + h)^3$, multiply out the product $(2 + h)(2 + h)(2 + h)$.)

(d) Find the velocity of the object at 2 seconds by letting h approach 0 in the result found in (c).

23. An object travels t^3 feet in the first t seconds.

(a) Find the average velocity during the time interval from 3 to 3.01 seconds?

(b) Find its average velocity during the time interval from 3 to t seconds, $t > 3$.

(c) By letting t approach 3 in the result found in (b), find the velocity of the object at 3 seconds.

Exercises 24 and 25 illustrate a different notation to find the slope of the tangent.

24. Consider the parabola $y = x^2$.

(a) Find the slope of the line through $P = (2, 4)$ and $Q = (2+h, (2+h)^2)$, where $h \neq 0$.

(b) Show that as h approaches 0, the slope in (a) approaches 4.

25. Consider the curve $y = x^3$.

(a) Find the slope of the line through $P = (2, 8)$ and $Q = (1.9, 1.9^3)$.

(b) Find the slope of the line through $P = (2, 8)$ and $Q = (2.01, 2.01^3)$.

(c) Find the slope of the line through $P = (2, 8)$ and $Q = (2+h, (2+h)^3)$, where $h \neq 0$.

(d) Show that as h approaches 0, the slope in (a) approaches 12.

26. Consider the curve $y = \sin(x)$.

(a) Find the slope of the line through $P = (0, 0)$ and $Q = (-0.1, \sin(-0.1))$.

(b) Find the slope of the line through $P = (0, 0)$ and $Q = (0.01, \sin(0.01))$.

(c) Find the slope of the line through $P = (0, 0)$ and $Q = (h, \sin(h))$, where $h \neq 0$.

(d) Show that as h approaches 0, the slope in (c) approaches 1.

(e) Use (d) to draw the tangent line to $y = \sin(x)$ at $(0, 0)$.

27. Consider the curve $y = \cos(x)$.

(a) Find the slope of the line through $P = (0, 1)$ and $Q = (-0.1, \cos(-0.1))$.

(b) Find the slope of the line through $P = (0, 1)$ and $Q = (0.01, \cos(0.01))$.

(c) Find the slope of the line through $P = (0, 1)$ and $Q = (h, \cos(h))$, where $h \neq 0$.

(d) Show that as h approaches 0, the slope in (c) approaches 0.

(e) Use (d) to draw the tangent line to $y = \cos(x)$ at $(0, 1)$.

28. Consider the curve $y = 2^x$.

(a) Find the slope of the line through $P = (2, 2^2)$ and $Q = (1.9, 2^{1.9})$.

(b) Find the slope of the line through $P = (2, 2^2)$ and $Q = (2.1, 2^{2.1})$.

(c) Find the slope of the line through $P = (2, 2^2)$ and $Q = (2 + h, 2^{2+h})$, where $h \neq 0$.

(d) Show that the slope of the curve $y = 2^x$ at $(2, 2^2)$ is approximately $4(0.693) = 2.772$.

(e) Use (d) to draw the tangent line to $y = 2^x$ at $(2, 4)$.

29. Consider the curve $y = e^x$.

(a) Find the slope of the line through $P = (-0.5, e^{-0.5})$ and $Q = (-0.6, e^{-0.6})$.

(b) Find the slope of the line through $P = (-0.5, e^{-0.5})$ and $Q = (-0.49, e^{-0.49})$.

(c) Find the slope of the line through $P = (-0.5, e^{-0.5})$ and $Q = (-0.5 + h, e^{-0.5+h})$, where $h \neq 0$.

(d) Show that as h approaches 0, the slope in (c) approaches $e^{-0.5}$.

30. Show that the slope of the curve $y = 2^x$ at $(3, 8)$ is approximately $8(0.693) = 5.544$.

31.

(a) Use the method of this section to find the slope of the curve $y = x^3$ at $(1, 1)$.

(b) What does the graph of $y = x^3$ look like near $(1, 1)$?

32.

(a) Use the method of this section to find the slope of the curve $y = x^3$ at $(-1, -1)$.

(b) What does the graph of $y = x^3$ look like near $(-1, -1)$?

33.

(a) Draw the curve $y = e^x$ for x in the interval $[-2, 1]$.

(b) Using a straightedge, draw the tangent line at $(1, e)$ as well as you can.

(c) Estimate the slope of the tangent line by measuring its rise and run.

(d) Using the derivative of e^x, find the slope of the curve at $(1, e)$.

34.

(a) Sketch the curve $y = e^x$ for x in $[-1, 1]$.

(b) Where does the curve cross the y-axis?

(c) What is the (smaller) angle between the graph of $y = e^x$ and the y-axis at the point found in (b)?

The phrase "slope of the graph of $y = f(x)$" is often shortened to "slope of $y = f(x)$," as in Exercises 35 and 36.

35. With the aid of a calculator, estimate the slope of $y = 2^x$ at $x = 1$, using the intervals

(a) $[1, 1.1]$

(b) $[1, 1.01]$

(c) $[0.9, 1]$

(d) $[0.99, 1]$

36. With the aid of a calculator, estimate the slope of $y = \frac{x+1}{x+2}$ at $x = 2$, using the intervals

(a) $[2, 2.1]$

(b) $[2, 2.01]$

(c) $[2, 2.001]$

(d) $[1.999, 2]$

37. Estimate the derivative of $\sin(x)$ at $x = \pi/3$

(a) to two decimal places.

(b) to three decimal places.

38. Estimate the derivative of $\ln(x)$ at $x = 2$

(a) to two decimal places.

(b) to three decimal places.

The ideas common to both slope and velocity also appear in other applications. Exercises 39 to 42 present the same ideas in biology, economics, and physics.

39. A bacterial culture has a mass of t^2 grams after t minutes of growth.

(a) How much does it grow during the time interval $[2, 2.01]$?

(b) What is the average rate of growth during the time interval $[2, 2.01]$?

(c) What is the instantaneous rate of growth when $t = 2$?

40. A thriving business has a profit of t^2 million dollars in its first t years. Thus from time $t = 3$ to time $t = 3.5$ (the first half of its fourth year) it has a profit of $(3.5)^2 - 3^2$ million dollars, giving an annual rate of

$$\frac{(3.5)^2 - 3^2}{0.5} = 6.5 \text{ million dollars per year.}$$

(a) What is its annual rate of profit during the time interval $[3, 3.1]$?

(b) What is its annual rate of profit during the time interval $[3, 3.01]$?

(c) What is its instantaneous rate of profit after 3 years?

Exercises 41 and 42 concern density.

41. The mass of the leftmost x centimeters of a nonhomogeneous string 10 centimeters long is x^2 grams, as shown in Figure 3.1.5. For instance, the string in the interval $[0, 5]$ has a mass of $5^2 = 25$ grams and the string in the interval $[5, 6]$ has mass $6^2 - 5^2 = 11$ grams. The **average density** of any part of the string is its mass divided by its length, $\frac{\text{total mass}}{\text{length}}$ grams per centimeter.

(a) Consider the leftmost 5 centimeters of the string, the middle 2 centimeters of the string, and the rightmost 2 centimeters of the string. Which piece has the largest mass?

(b) Which piece is densest?

(c) What is the mass of the string in the interval $[3, 3.01]$?

(d) Using the interval $[3, 3.01]$, estimate the density at 3.

(e) Using the interval $[2.99, 3]$, estimate the density at 3.

(f) By considering intervals of the form $[3, 3+h]$, h positive, find the density at the point 3 centimeters from the left end.

(g) By considering intervals of the form $[3+h, 3]$, h negative, find the density at the point 3 centimeters from the left end.

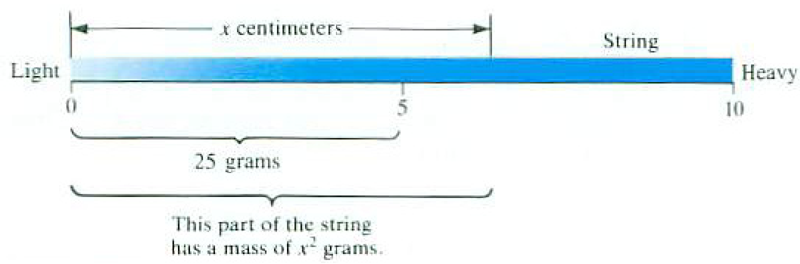


Figure 3.1.5

42. The left x centimeters of a string have a mass of x^2 grams.

(a) What is the mass of the string in the interval $[2, 2.01]$?

(b) Using the interval $[2, 2.01]$, estimate the density at 2.

(c) Using the interval $[1.99, 2]$, estimate the density at 2.

(d) By considering intervals of the form $[2, 2+h]$, h positive, find the density at the point 2 centimeters from the left end.

(e) By considering intervals of the form $[2+h, 2]$, h negative, find the density at the point 2 centimeters from the left end.

43.

(a) Graph the curve $y = 2x^2 + x$.

(b) By eye, draw the tangent line to the curve at the point $(1, 3)$. Using a ruler, estimate its slope.

(c) Sketch the line that passes through the point $(1, 3)$ and the point $(x, 2x^2 + x)$.

(d) Find the slope of the line in (c).

(e) Letting x get closer and closer to 1, find the slope of the tangent line at $(1, 3)$.

(f) How close was your estimate in (b)?

44. An object travels $2t^2 + t$ feet in t seconds.

(a) Find its average velocity during the interval of time $[1, x]$, for $x > 1$.

(b) Letting x get closer and closer to 1, find the velocity at time 1.

45. Find the slope of the tangent line to the curve $y = x^2$ of Example 1 at $P = (x, x^2)$. To do this, consider the slope of the line through P and the nearby point $Q = (x + h, (x + h)^2)$ and let h approach 0.

46. Find the velocity of the falling rock of Example 2 at any time t. To do this, consider the average velocity during the time interval $[t, t + h]$ and then let h approach 0.

47. Does the tangent line to the curve $y = x^2$ at the point $(1, 1)$ pass through the point $(6, 12)$?

48.

(a) Graph the curve $y = 2^x$ as well as you can for $-2 \le x \le 3$.

(b) Using a straightedge, draw as well as you can a tangent to the curve at $(2, 4)$. Estimate its slope by using a ruler to draw and measure a rise-and-run triangle.

(c) Using a secant through $(2, 4)$ and $(x, 2^x)$ for x near 2, estimate the slope of the tangent to the curve at $(2, 4)$. (Choose particular values of x and use your calculator to create a table of the results.)

49. Using your calculator estimate the slope of the tangent line to the graph of $f(x) = \sin(x)$ at $\left(\frac{\pi}{4}, \frac{\sqrt{2}}{2}\right)$ to two decimal places.

50.

(a) Sketch the curve $y = x^3 - x^2$.

(b) Using the method of the nearby point, find the slope of the tangent line to the curve at the point $(a, a^3 - a^2)$.

(c) Find all points on the curve where the tangent line is horizontal.

(d) Find all points on the curve where the tangent line has slope 1.

51. Repeat Exercise 50 for the curve $y = x^3 - x$.

52. An astronaut is traveling from left to right along the curve $y = x^2$. When she shuts off the engine, she will fly off along the line tangent to the curve at the point where she is at the moment the engine turns off. At what point should she shut off the engine in order to reach the point

(a) $(4, 9)$?

(b) $(4, -9)$?

53. See Exercise 52. Where can an astronaut who is traveling from left to right along $y = x^3 - x$ shut off the engine and pass through the point $(2, 2)$?

54.

Sam: I don't like the book's definition of the derivative.

Jane: Why not?

Sam: I can do it without limits, and more easily.

Jane: How?

Sam: Just define the derivative of f at a as the slope of the tangent line at $(a, f(a))$ on the graph of f.

Jane: Something must be wrong with that.

Who is right, Sam or Jane?

3.2 The Derivatives of the Basic Functions

In this section we use the definition of the derivative to find the derivatives of the important functions x^a (a rational), e^x, $\sin x$, and $\cos x$. We also introduce some of the standard notations for the derivative. For convenience, we begin by repeating the definition of the derivative.

> **DEFINITION** *(Derivative of a function at a number)* Assume that the function f is defined in an open interval containing a. If
>
> $$\lim_{x\to a} \frac{f(x) - f(a)}{x - a} \tag{3.2.1}$$
>
> exists, it is called the **derivative of f at a**.

There are several notations for the quotient that appears in (3.2.1) and also for the derivative. Sometimes it is convenient to use $a + h$ instead of x and let h approach 0. Then, (3.2.1) reads

$$\lim_{h\to 0} \frac{f(a+h) - f(a)}{h}. \tag{3.2.2}$$

Expression (3.2.2) says the same thing as (3.2.1): determine how the quotient, change in output divided by change in input, behaves as the change in input gets smaller and smaller.

Sometimes it is useful to call the change in output "Δf" and the change in input "Δx." That is, $\Delta f = f(x) - f(a)$ and $\Delta x = x - a$. Then

$$f'(a) = \lim_{\Delta x\to 0} \frac{\Delta f}{\Delta x}.$$

Δ, pronounced del-t$_{ə}$, is the upper-case Greek letter corresponding to the Latin "D". In mathematics, "Δf," read 'delta eff," generally indicates a difference or change in f.

There is nothing sacred about the letters a, x, and h. One could say

$$f'(x) = \lim_{t\to x} \frac{f(t) - f(x)}{t - x}$$

or

$$f'(x) = \lim_{u\to x} \frac{f(u) - f(x)}{u - x}.$$

The symbol "$f'(a)$" is read aloud as "f prime at a" or "the derivative of f at a." The symbol $f'(x)$ is read similarly. The notation $f'(x)$ reminds us that f', like f, is a function. For each input x the derivative, $f'(x)$, is the output. The derivative of the function f is also written as $D(f)$.

For instance, the derivative of the squaring function, x^2, is

$$\lim_{u\to x} \frac{u^2 - x^2}{u - x} = \lim_{u\to x} \frac{(u-x)(u+x)}{u - x} = \lim_{u\to x} (u + x) = 2x.$$

The derivative of a specific function, in this case x^2, is denoted $(x^2)'$ or $D(x^2)$. Then, $D(x^2) = 2x$ is read aloud as "the derivative of x^2 is $2x$." This is shorthand for "the derivative of the function that assigns x^2 to x is the function that assigns $2x$ to x." Since the value of the derivative depends on x, the derivative is a function.

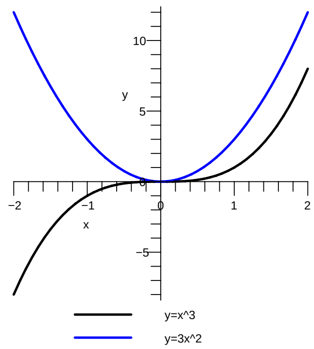


Figure 3.2.1

EXAMPLE 1 Find the derivative of x^3 at a.
SOLUTION

$$(x^3)' = \lim_{x \to a} \frac{x^3 - a^3}{x - a} = 3a^2.$$

This limit was evaluated by noticing that it is one of the four limits in Section 2.2. We can write $(x^3)' = 3x^2$ or $D(x^3) = 3x^2$. ◇

In the same manner, $\lim_{x \to a} \frac{x^n - a^n}{x - a} = n \cdot a^{n-1}$ implies that for any positive integer n, the derivative of x^n is nx^{n-1}. The exponent n becomes the coefficient and the exponent of x shrinks from n to $n-1$:

Derivative of x^n

$$(x^n)' = nx^{n-1} \qquad \text{where } n \text{ is a positive integer.}$$

The next example treats an exponential function with a fixed base.
EXAMPLE 2 Find the derivative of 2^x.

SOLUTION

$$\begin{aligned} D(2^x) &= \lim_{h \to 0} \frac{2^{(x+h)} - 2^x}{h} \\ &= \lim_{h \to 0} \frac{2^x 2^h - 2^x}{h} \\ &= \lim_{h \to 0} 2^x \frac{2^h - 1}{h} \\ &= 2^x \lim_{h \to 0} \frac{2^h - 1}{h}. \end{aligned}$$

In Section 2.2 we found that $\lim_{h \to 0} \frac{2^h - 1}{h} \approx 0.693$. Thus,

$$D(2^x) \approx (0.693)2^x.$$

◇

No one wants to remember the (approximate) constant 0.693, which appears when we use base 2. Recall that in Section 3.1 we found that the derivative of e^x is e^x, with no multiplying constant.

We emphasize this simple and important formula

Derivative of e^x

$$D(e^x) = e^x.$$

The function e^x has the remarkable property that it equals its derivative.

Next, we turn to trigonometric functions.

EXAMPLE 3 Find the derivative of $\sin(x)$.
SOLUTION

Recall that $\sin(a + b) = \sin(a)\cos(b) + \cos(a)\sin(b)$.

$$\begin{aligned} D(\sin x) &= \lim_{h\to 0} \frac{\sin(x+h) - \sin(x)}{h} \\ &= \lim_{h\to 0} \frac{\sin(x)\cos(h) + \cos(x)\sin(h) - \sin(x)}{h} \\ &= \lim_{h\to 0} \frac{\sin(x)(\cos(h) - 1) + \cos(x)\sin(h)}{h} \\ &= \lim_{h\to 0} \sin x \frac{\cos(h) - 1}{h} + \cos(x)\frac{\sin(h)}{h}. \end{aligned}$$

In Section 2.2 we found that $\lim_{h\to 0} \frac{\sin(h)}{h} = 1$ and $\lim_{h\to 0} \frac{1-\cos(h)}{h} = 0$. Thus $\lim_{h\to 0} \frac{\cos(h)-1}{h} = 0$ and

$$D(\sin x) = (\sin x)(0) + (\cos x)(1) = \cos(x).$$

◇

We have the important formula

Derivative of $\sin(x)$

$$D(\sin(x)) = \cos(x).$$

If we graph $y = \sin(x)$ (see Figure 3.2.2), and consider its shape, the formula $D(\sin(x)) = \cos(x)$ is not a surprise. For instance, for x in $(-\pi/2, \pi/2)$ the slope is positive. So is $\cos(x)$. For x in $(\pi/2, 3\pi/2)$ the slope of the sine curve is negative. So is $\cos(x)$. Since $\sin(x)$ has period 2π, we would expect its derivatve also to have period 2π. Indeed, $\cos(x)$ has period 2π.

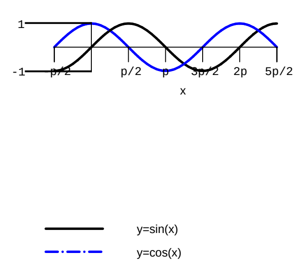


Figure 3.2.2

Similarly, using the definition of the derivative and the identity $\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$, we have the formula

Derivative of $\cos(x)$

$$D(\cos(x)) = -\sin(x).$$

Derivatives of Other Power Functions

We showed that if n is a positive integer, $D(x^n) = nx^{n-1}$. Now let us find the derivative of power functions x^n where n is not a positive integer.

EXAMPLE 4 Find the derivative of $x^{-1} = \frac{1}{x}$.
SOLUTION Before we calculate the necessary limit, let's pause to see how the slope of $y = 1/x$ behaves. Figure 3.2.3 shows that the slope is always negative. For x near 0, the absolute value of the slope is large, but when $|x|$ is large, the slope is near 0.

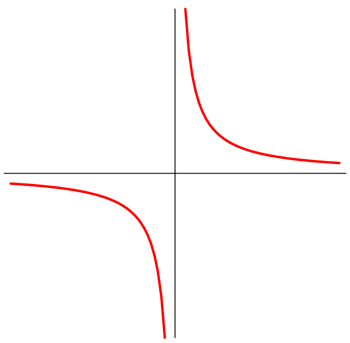


Figure 3.2.3

Now, let us find the derivative of $1/x$:

$$\begin{aligned} D(1/x) &= \lim_{t \to x} \frac{1/t - 1/x}{t - x} \\ &= \lim_{t \to x} \frac{1}{t-x}\left(\frac{x-t}{xt}\right) \qquad \left(\text{ since } \frac{1}{t} - \frac{1}{x} = \frac{x-t}{xt}\ \right) \\ &= \lim_{t \to x} \frac{-1}{xt} \\ &= -\frac{1}{x^2}. \end{aligned}$$

As a check, we see that $-1/x^2$ is always negative, has large absolute value when x is near 0, and is near 0 when $|x|$ is large. $\diamond$

It is worth memorizing that

Derivative of x^{-1}

$$D\left(\frac{1}{x}\right) = -\frac{1}{x^2}.$$

Or, in exponential notation,

$$D(x^{-1}) = -x^{-2}.$$

The second form fits the pattern established for positive integers n, $D(x^n) = nx^{n-1}$.

EXAMPLE 5 Find the derivative of $x^{2/3}$.
SOLUTION Once again we use the definition of the derivative:

$$D(x^{2/3}) = \lim_{t \to x} \frac{t^{2/3} - x^{2/3}}{t - x}.$$

A bit of algebra will help us find the limit. We write the four terms $t^{2/3}$, $x^{2/3}$, t, and x as powers of $t^{1/3}$ and $x^{1/3}$. Thus

$$D(x^{2/3}) \quad = \quad \lim_{t\to x} \frac{\left(t^{1/3}\right)^2 - \left(x^{1/3}\right)^2}{\left(t^{1/3}\right)^3 - \left(x^{1/3}\right)^3}.$$

Because $a^2 - b^2 = (a-b)(a+b)$ and $a^3 - b^3 = (a-b)(a^2+ab+b^2)$, we find

$$\begin{aligned}
D(x^{2/3}) &= \lim_{t\to x} \frac{\left(\left(t^{1/3}\right) - \left(x^{1/3}\right)\right)\left(\left(t^{1/3}\right) + \left(x^{1/3}\right)\right)}{\left(\left(t^{1/3}\right) - \left(x^{1/3}\right)\right)\left(\left(t^{1/3}\right)^2 + \left(t^{1/3}\right)\left(x^{1/3}\right) + \left(x^{1/3}\right)^2\right)} \\
&= \lim_{t\to x} \frac{\left(t^{1/3}\right) + \left(x^{1/3}\right)}{\left(t^{1/3}\right)^2 + \left(t^{1/3}\right)\left(x^{1/3}\right) + \left(x^{1/3}\right)^2} \\
&= \frac{\left(x^{1/3}\right) + \left(x^{1/3}\right)}{\left(x^{1/3}\right)^2 + \left(x^{1/3}\right)\left(x^{1/3}\right) + \left(x^{1/3}\right)^2} \\
&= \tfrac{2x^{1/3}}{3x^{2/3}} = \tfrac{2}{3}x^{-1/3}.
\end{aligned}$$

In short,

$$D(x^{2/3}) = \frac{2}{3}x^{-1/3}.$$

This formula follows the pattern we found for $D(x^n)$ for $n = 1, 2, 3, \ldots$ and $n = -1$. The exponent of x becomes the coefficient and the exponent of x is lowered by 1. ◇

The method used in Example 5 applies to any positive rational exponent. In the next two sections we will show how the result extends first to negative rational exponents and then to irrational exponents. In all cases the formula will be the same. We state the general result here, but remember that — so far — we have justified it only for positive rational exponents and for -1.

Derivative of Power Functions x^a

For any fixed number a, $D(x^a) = ax^{a-1}$.

The formula holds for values of x where both x^a and x^{a-1} are defined. For instance, $x^{1/2} = \sqrt{x}$ is defined for $x \geq 0$, but its derivative $\frac{1}{2}x^{-1/2}$ is defined only for $x > 0$.

Because the derivative of the square root function occurs so often we emphasize its formula

Derivative of Square Root Function (as Power Function)

$$D(x^{1/2}) = \frac{1}{2}x^{-1/2}.$$

or, in terms of the usual square root sign,

Derivative of Square Root Function (Square Root Sign)

$$D(\sqrt{x}) = \frac{1}{2\sqrt{x}}.$$

Another Notation for the Derivative

We have used the notations f' and $D(f)$ for the derivative of a function f. There is another notation that is also convenient.

If $y = f(x)$, the derivative is denoted by the symbols

$$\frac{dy}{dx} \text{ or } \frac{df}{dx}.$$

The symbol $\frac{dy}{dx}$ is read as "the derivative of y with respect to x" or "dee y, dee x."

In this notation the derivative of x^3 is written

$$\frac{d\left(x^3\right)}{dx}.$$

If the function is expressed in terms of another letter, such as t, we would write

$$\frac{d\left(t^3\right)}{dt}.$$

Keep in mind that in the notations df/dx and dy/dx, the symbols df, dy, and dx have no meaning by themselves. The symbol dy/dx should be thought of as a single entity, like the numeral 8, which we do not think of as formed of two 0's.

In the study of motion, Newton's **dot notation** is often used. If x is a function of time t, then $\dot{x}$ denotes the derivative dx/dt.

Summary

In this section we saw why limits are important in calculus. We need them to define the derivative of a function. The definition can be stated in several ways, but each one says, informally, "look at how a small change in input changes the output." Here is the formal definition, in various notations:

$$\begin{aligned} f'(a) &= \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \qquad & f'(x) &= \lim_{x \to 0} \frac{f(x+h) - f(x)}{h} \\ f'(x) &= \lim_{t \to x} \frac{f(t) - f(x)}{t - x} \qquad & f'(x) &= \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}. \end{aligned}$$

The following derivatives should be memorized. However, if one is forgotten it can be recovered by using the definition as a limit.

Function	Derivative
$f(x)$	$f'(x)$
x^a	ax^{a-1}
e^x	e^x
$\sin(x)$	$\cos(x)$
$\cos(x)$	$-\sin(x)$

EXERCISES for Section 3.2

1. Show that $D(\cos(x)) = -\sin(x)$. ($\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$.)

Using the definition of the derivative, compute the appropriate limit to find the derivatives of the functions in Exercises 2 to 12.

2. $1/(x+2)$

3. $2x - x^2$

4. 3^x. (Use your calculator to estimate the coefficient that appears.)

5. $6x^3$

6. $x^{4/3}$

7. $5x^2$

8. $4\sin(x)$

9. $2e^x + \sin(x)$

10. $x^2 + x^3$

11. $1/(2x+1)$

12. $1/x^2$

13. Use the formulas obtained for the derivatives of e^x, x^a, $\sin(x)$, and $\cos(x)$ to evaluate the derivatives of the given function at the given input.

(a) e^x at -1

(b) $x^{1/3}$ at -8

(c) $\sqrt[3]{x}$ at 27

(d) $\cos(x)$ at $\pi/4$

(e) $\sin(x)$ at $2\pi/3$

14. Use the formulas obtained for the derivatives of e^x, x^a, $\sin(x)$, and $\cos(x)$ to evaluate the derivatives of the given function at the given input.

(a) e^x at 0

(b) $x^{2/3}$ at -1

(c) $\sqrt{x}$ at 25

(d) $\cos(x)$ at $-\pi$

(e) $\sin(x)$ at $\pi/3$

15. State the definition of the derivative of a function in words, using no mathematical symbols.

16. State the definition of the derivative of $g(t)$ at b as a mathematical formula, with no words.

In Exercises 17 to 22 use the definition of the derivative to show that the given equation is correct.

17. $D\left(e^{-x}\right) = -e^{-x}$

18. $D\left(e^{3x}\right) = 3e^{3x}$

19. $D(1/\cos(x)) = \sin(x)/\cos^2(x)$

20. $D(\tan(x)) = 1 + \tan^2(x) = \sec^2(x)$
(Use the identity $\tan(A-B) = \frac{\tan(A)-\tan(B)}{1+\tan(A)\tan(B)}$)

21. $D(\sin(2x)) = 2\cos(2x)$

22. $D(\cos(x/2)) = -1/2\sin(x/2)$

23. This Exercise shows why in calculus angles are measured in radians. Let $\mathrm{Sin}(x)$ denote the sine of an angle of x degrees and let $\mathrm{Cos}(x)$ denote the cosine of an angle of x degrees.

(a) Graph $y = \mathrm{Sin}(x)$ on the interval $[-180, 360]$, using the same scale on the x- and y-axes.

(b) Find $\lim_{x\to 0} \frac{\mathrm{Sin}(x)}{x}$.

(c) Find $\lim_{x\to 0} \frac{1-\mathrm{Cos}(x)}{x}$.

(d) Using the definition of the derivative, differentiate $\mathrm{Sin}(x)$.

24. Use a limit to show that $D\left((x^{-5}\right) = -5x^{-6}$.

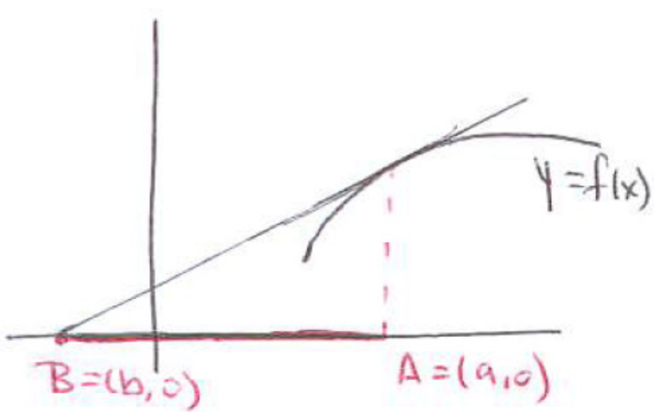


Figure 3.2.4

Let f be a differentiable function and a a number such that $f'(a)$ is not zero. The tangent to the graph of f at $A = (a, f(a))$ meets the x-axis at a point $B = (b, 0)$, see Figure 3.2.4. The **subtangent** of f is the line AB. Its length is $|a - b|$.

Exercises 25 and 26 involve the subtangent of a function.

25. Show that for e^x the length of the subtangent is the same for all values of a.

26. Find the length of the subtangent at $(a, f(a))$ for any differentiable function f. (Assume $f'(a)$ is not zero.)

3.3 Shortcuts for Computing Derivatives

This section develops methods for finding the derivative of a function, or what is called **differentiating** a function. Using them will make it easy to find, for instance, the derivative of

$$\frac{(3+4x+5x^2)e^x}{\sin(x)}$$

without going back to the definition of the derivative and finding the limit of a complicated quotient.

We first find the derivative of a constant function.

The Derivative of a Constant Function

Constant Rule

The derivative of a constant function $f(x) = C$ is 0.

$$\frac{dC}{dx} = (C)' = 0$$

Proof

Let C be a number and let f be the constant function, $f(x) = C$ for all inputs x. By the definition of a derivative,

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}.$$

Since f has the same output C for all inputs,

Δx is another name for h

$$f(x+\Delta x) = C \text{ and } f(x) = C.$$

Thus

$$\begin{aligned} f'(x) &= \lim_{\Delta x \to 0} \frac{C - C}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{0}{\Delta x} \\ &= \lim_{\Delta x \to 0} 0 \qquad (\text{ since } \Delta x \neq 0 \text{ }) \\ &= 0. \end{aligned}$$

This shows the derivative of a constant function is 0 for all x. •

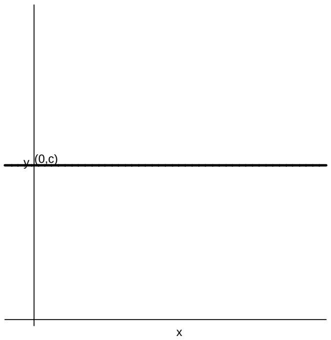


Figure 3.3.1

From two points of view, the constant rule is no surprise. Since the graph of $f(x) = C$ is a horizontal line, it coincides with each of its tangent lines, which have slope 0, as can be seen in Figure 3.3.1. Also, if we think of x as time and $f(x)$ as the position of a particle at time x, the constant rule implies that a stationary particle has zero velocity.

Derivatives of $f + g$ and $f - g$

The next theorem asserts that if the functions, f and g have derivatives, so does their sum $f + g$ and

$$\frac{d}{dx}(f + g) = \frac{df}{dx} + \frac{dg}{dx}.$$

In other words, "the derivative of the sum is the sum of the derivatives." Equivalently, $(f + g)' = f' + g'$ and $D(f + g) = D(f) + D(g)$. Similar formulas hold for the derivative of $f - g$.

Sum Rule and Difference Rule

If f and g are differentiable functions, then so are $f + g$ and $f - g$. The **sum rule** and **difference rule** for computing their derivatives are

$$\begin{array}{rcll} (f + g)' & = & f' + g' & (\text{ sum rule }) \\ (f - g)' & = & f' - g'. & (\text{ difference rule }) \end{array}$$

Proof

To justify this we go back to the definition of the derivative. To begin, we give the function $f + g$ the name u, that is, $u(x) = f(x) + g(x)$. We will examine

$$\lim_{\Delta x \to 0} \frac{u(x + \Delta x) - u(x)}{\Delta x}$$

or, equivalently,

$$\lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x}. \tag{3.3.1}$$

To evaluate (3.3.1), we will express Δu in terms of Δf and Δg. Here are the details:

$$\begin{array}{rcll} \Delta u & = & u(x + \Delta x) - u(x) & \\ & = & (f(x + \Delta x) + g(x + \Delta x)) - (f(x) + g(x)) & (\text{ definition of } u\text{ }) \\ & = & (f(x) + \Delta f) + (g(x) + \Delta g) - (f(x) + g(x)) & (\text{ definition of } \Delta f \text{ and } \Delta g\text{ }) \\ & = & \Delta f + \Delta g & \end{array}$$

All told, $\Delta u = \Delta f + \Delta g$. The change in u is the change in f plus the change in g.

We can now evaluate (3.3.1):

$$\lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta f + \Delta g}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} + \lim_{\Delta x \to 0} \frac{\Delta g}{\Delta x} = f'(x) + g'(x).$$

Thus, $u = f + g$ is differentiable and

$$u'(x) = f'(x) + g'(x).$$

A similar argument applies to $f - g$. •

The sum and difference rules extend to any finite number of differentiable functions. For example,

$$\begin{aligned}(f+g+h)' &= f' + g' + h' \\ (f-g+h)' &= f' - g' + h'.\end{aligned}$$

EXAMPLE 1 Using the sum rule, differentiate $x^2 + x^3 + \cos(x) + 3$.
SOLUTION

$$\begin{aligned} D\left(x^2 + x^3 + \cos(x) + 3\right) &= D(x^2) + D(x^3) + D(\cos(x)) + D(3) \\ &= 2x^{2-1} + 3x^{3-1} + (-\sin(x)) + 0 \\ &= 2x + 3x^2 - \sin(x). \end{aligned}$$

◇

EXAMPLE 2 Differentiate $x^4 - \sqrt{x} - e^x$.
SOLUTION

$$\begin{aligned} \tfrac{d}{dx}\left(x^4 - \sqrt{x} - e^x\right) &= \tfrac{d}{dx}\left(x^4\right) - \tfrac{d}{dx}\left(\sqrt{x}\right) - \tfrac{d}{dx}\left(e^x\right) \\ &= 4x^3 - \tfrac{1}{2\sqrt{x}} - e^x. \end{aligned}$$

◇

The Derivative of fg

The following theorem, concerning the derivative of the product of two functions, may be surprising, for it turns out that the derivative of the product is not the product of the derivatives. The formula is more complicated than the one for the derivative of the sum. It asserts that "the derivative of the product is the derivative of the first function times the second plus the first function times the derivative of the second."

Product Rule

If f and g are differentiable functions, then so is their product fg. Its derivative is given by

$$(fg)' \quad = \quad f'g + fg'.$$

Proof

The proof is similar to that for the sum and difference rules. This time we give the product fg the name u. Then we express Δu in terms of Δf and Δg. Finally, we determine $u'(x)$ by examining $\lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x}$. These steps are practically forced upon us.

We have

$$u(x) = f(x)g(x) \quad \text{and} \quad u(x + \Delta x) = f(x + \Delta x)g(x + \Delta x).$$

Rather than subtract $u(x)$ from $u(x + \Delta x)$ directly, we write

$$f(x + \Delta x) = f(x) + \Delta f \quad \text{and} \quad g(x + \Delta x) = g(x) + \Delta g.$$

Then

$$\begin{aligned} u(x + \Delta x) &= (f(x + \Delta x))\,(g(x + \Delta x)) \\ &= (f(x) + \Delta f)\,(g(x) + \Delta g) \\ &= f(x)g(x) + (\Delta f)g(x) + f(x)\Delta g + (\Delta f)(\Delta g). \end{aligned}$$

Hence

$$\begin{aligned} \Delta u &= u(x + \Delta x) - u(x) \\ &= f(x)g(x) + (\Delta f)g(x) + f(x)(\Delta g) + (\Delta f)(\Delta g) - f(x)g(x) \\ &= (\Delta f)g(x) + f(x)(\Delta g) + (\Delta f)(\Delta g) \end{aligned}$$

and

$$\begin{aligned} \frac{\Delta u}{\Delta x} &= \frac{(\Delta f)g(x) + f(x)(\Delta g) + (\Delta f)(\Delta g)}{\Delta x} \\ &= \frac{\Delta f}{\Delta x}g(x) + f(x)\frac{\Delta g}{\Delta x} + \Delta f \frac{\Delta g}{\Delta x}. \end{aligned}$$

As $\Delta x \to 0$, $\Delta g/\Delta x \to g'(x)$ and $\Delta f/\Delta x \to f'(x)$. Furthermore, because f is differentiable and hence continuous, $\Delta f \to 0$ as $x \to 0$. It follows that

$$\lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} = f'(x)g(x) + f(x)g'(x) + 0 \cdot g'(x).$$

The formula for $(fg)'$ was discovered by Leibniz in 1676. His first guess was wrong.

Therefore, u is differentiable and

$$u' = f'g + fg'.$$

•

Remark: Figure 3.3.2 illustrates the product rule and its proof. With f, Δf, g, and Δg taken to be positive, the inner rectangle has area $u = fg$ and the whole rectangle has area $u + \Delta u = (f+\Delta f)(g+\Delta g)$. The shaded region whose area is Δu is made up of rectangles of areas $f \cdot (\Delta g)$, $(\Delta f) \cdot g$, and $(\Delta f) \cdot (\Delta g)$. The little corner rectangle, of area $(\Delta f) \cdot (\Delta g)$, is negligible in comparison with the other two rectangles. Thus, $\Delta u \approx (\Delta f)g + f(\Delta g)$, which suggests the formula for the derivative of a product.

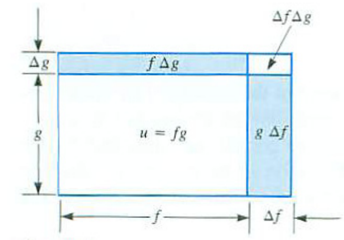


Figure 3.3.2

EXAMPLE 3 Find $D\left((x^2 + x^3 + \cos(x) + 3)\left(x^4 - \sqrt{x} - e^x\right)\right)$.
SOLUTION The function to be differentiated is the product of the functions differentiated in Examples 1 and 2. By the product rule,

$$\begin{aligned}
&D\left(\left(x^2 + x^3 + \cos(x) + 3\right)\left(x^4 - \sqrt{x} - e^x\right)\right)\\
&= \left(D\left(x^2 + x^3 + \cos(x) + 3\right)\right)\left(x^4 - \sqrt{x} - e^x\right)\\
&\quad + \left(x^2 + x^3 + \cos(x) + 3\right)\left(D\left(x^4 - \sqrt{x} - e^x\right)\right)\\
&= \left(2x + 3x^2 - \sin(x)\right)\left(x^4 - \sqrt{x} - e^x\right)\\
&\quad + \left(x^2 + x^3 + \cos(x) + 3\right)\left(4x^3 - \frac{1}{2\sqrt{x}} - e^x\right).
\end{aligned}$$

◇

Derivative of Constant Times f

A special case of the formula for the product rule occurs so frequently that it is singled out in the constant multiple rule.

Constant Multiple Rule

If C is a constant function and f is a differentiable function, then Cf is differentiable and its derivative is given by

$$(Cf)' = C(f').$$

In other notations, $\frac{d(Cf)}{dx} = C\frac{df}{dx}$ and $D(Cf) = CD(f)$.

In words, the derivative of a constant times a function is the constant times the derivative of the function.

Proof

Because we have a product of two differentiable functions, C and f, we may use the product rule. We have

$$\begin{aligned}
(Cf)' &= (C')f + C(f') && \text{(derivative of a product)}\\
&= 0 \cdot f + Cf' && \text{(derivative of constant is 0)}\\
&= C(f').
\end{aligned}$$

•

The constant multiple rule asserts that "it is legal to move a constant factor outside the derivative symbol."

EXAMPLE 4 Find $D(6x^3)$.
SOLUTION

$$\begin{aligned} D(6x^3) &= 6D(x^3) && (\text{ 6 is a constant }) \\ &= 6 \cdot 3x^2 && (\ D(x^n) = nx^{n-1}\ ) \\ &= 18x^2. \end{aligned}$$

With a little practice, one could immediately write $D(6x^3) = 18x^2$. ◇

EXAMPLE 5 Find $D\left(\sqrt{x}/11\right)$.
SOLUTION

$$D\left(\frac{\sqrt{x}}{11}\right) = D\left(\tfrac{1}{11}\sqrt{x}\right) = \tfrac{1}{11}D(\sqrt{x}) = \tfrac{1}{11}\tfrac{1}{2\sqrt{x}} = \tfrac{1}{22}x^{-1/2}.$$

◇

Example 5 generalizes to the fact that for a nonzero C,

Constant Division Rule

$$\left(\frac{f}{C}\right)' = \frac{f'}{C} \qquad C \neq 0.$$

The formula for the derivative of the product extends to the product of several differentiable functions. For instance,

Generalized Product Rule

$$(fgh)' = (f')gh + f(g')h + fg(h').$$

In each summand only one derivative appears. The next example illustrates the use of this formula.

EXAMPLE 6 Differentiate $\sqrt{x}e^x \sin(x)$.
SOLUTION

$$\begin{aligned} &\left(\sqrt{x}e^x \sin(x)\right)' \\ &\quad= (\sqrt{x})'e^x \sin(x) + \sqrt{x}(e^x)' \sin(x) + \sqrt{x}e^x(\sin(x))' \\ &\quad= \left(\frac{1}{2\sqrt{x}}\right) e^x \sin(x) + \sqrt{x}e^x \sin(x) + \sqrt{x}e^x \cos(x) \end{aligned}$$

◇

Any polynomial can be differentiated by the methods already developed.

EXAMPLE 7 Differentiate $6t^8 - t^3 + 5t^2 + \pi^3$.

SOLUTION The independent variable in this polynomial is t, and the polynomial is to be differentiated with respect to t.

Differentiate a polynomial "term-by-term". Note that π^3 is a constant.

$$\begin{aligned}\frac{d}{dt}\left(6t^8 - t^3 + 5t^2 + \pi^3\right) &= \frac{d}{dt}\left(6t^8\right) - \frac{d}{dt}\left(t^3\right) + \frac{d}{dt}\left(5t^2\right) + \frac{d}{dt}\left(\pi^3\right)\\ &= 48t^7 - 3t^2 + 10t + 0\\ &= 48t^7 - 3t^2 + 10t.\end{aligned}$$

◇

Derivative of $1/g$

Often one needs the derivative of the reciprocal of a function g, that is, $(1/g)'$.

Reciprocal Rule

If g is a differentiable function, then

$$\left(\frac{1}{g}\right)' = -\frac{g'}{g^2}, \qquad \text{where } g(x) \neq 0.$$

Proof

Again we go back to the definition of the derivative.

Assume $g(x) \neq 0$ and let $u(x) = 1/g(x)$. Then $u(x+\Delta x) = 1/g(x+\Delta x) = 1/(g(x) + \Delta g)$. Thus

$$\begin{aligned}\Delta u &= u(x+\Delta x) - u(x)\\ &= \frac{1}{g(x)+\Delta g} - \frac{1}{g(x)}\\ &= \frac{g(x) - (g(x)+\Delta g)}{g(x)(g(x)+\Delta g)} && \text{(common denominator)}\\ &= \frac{-\Delta g}{g(x)(g(x)+\Delta g)} && \text{(cancellation in numerator).}\end{aligned}$$

Then

$$
\begin{aligned}
u'(x) &= \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{-\Delta g / \left(g(x)(g(x)+\Delta g)\right)}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{-\Delta g/\Delta x}{g(x)(g(x)+\Delta g)} && \left(\text{ algebra: } \tfrac{(a/b)}{c} = \tfrac{(a/c)}{b} \right) \\
&= \frac{\lim\limits_{\Delta x \to 0} \left(\dfrac{-\Delta g}{\Delta x}\right)}{\lim\limits_{\Delta x \to 0} \left(g(x)(g(x)+\Delta g)\right)} && (\text{ quotient rule for limits }) \\
&= \frac{-g'(x)}{g(x)^2}. && \left(\ g(x) \text{ is continuous, } \lim_{\Delta x \to 0} \Delta g = 0 \ \right)
\end{aligned}
$$

•

EXAMPLE 8 Find $D\left(\frac{1}{\cos(x)}\right)$.
SOLUTION In this case, $g(x) = \cos(x)$ and $g'(x) = -\sin(x)$. Therefore,

$$
\begin{aligned}
D\left(\frac{1}{\cos(x)}\right) &= \frac{-(-\sin(x))}{(\cos(x))^2} \\
&= \frac{\sin(x)}{\cos^2(x)} \qquad \text{for all } x \text{ with } \cos(x) \neq 0.
\end{aligned}
$$

◇

Example 8 gives a formula for the derivative of $\sec(x)$, which is defined as $1/\cos(x)$.

$$
D(\sec(x)) = D\left(\tfrac{1}{\cos(x)}\right) = \frac{\sin(x)}{\cos^2(x)} = \frac{\sin(x)}{\cos(x)}\frac{1}{\cos(x)} = \tan(x)\sec(x).
$$

Therefore,

Derivative of $\sec(x)$
$D(\sec(x)) = \sec(x)\tan(x)$

The reciprocal rule allows us to complete the justification of the power rule for exponents that are negative rational numbers.

EXAMPLE 9 Show that the power rule, in Section 3.2, is valid when a is a negative rational number. That is, show that $D(x^{-p/q}) = (-p/q)x^{(-p/q)-1}$

for any positive integers p and q, with $q \neq 0$.
SOLUTION The reciprocal rule to find the derivative of $x^{-p/q}$ Write $x^{-p/q}$ as $1/x^{p/q}$, then

$$D\left(x^{-p/q}\right) = D\left(\frac{1}{x^{p/q}}\right) = \frac{-D(x^{p/q})}{(x^{p/q})^2} = \frac{-\frac{p}{q}x^{\frac{p}{q}-1}}{x^{\frac{2p}{q}}} = -\frac{p}{q}x^{\left(\frac{p}{q}\right)-1-2\left(\frac{p}{q}\right)} = -\frac{p}{q}x^{-(p/q)-1}.$$

◇

The Derivative of f/g

EXAMPLE 10 Derive a formula for the derivative of the quotient f/g.
SOLUTION Write the quotient f/g as a product, $f \cdot \frac{1}{g}$. Assuming f and g are differentiable functions, we may use the product and reciprocal rules to find

$$\begin{aligned}
\left(\frac{f(x)}{g(x)}\right)' &= \left(f(x)\frac{1}{g(x)}\right)' && \text{(rewrite quotient as product)}\\
&= f'(x)\left(\frac{1}{g(x)}\right) + f(x)\left(\frac{1}{g(x)}\right)' && \text{(product rule)}\\
&= f'(x)\left(\frac{1}{g(x)}\right) + f(x)\left(\frac{-g'(x)}{g(x)^2}\right) && \text{(reciprocal rule, assuming } g(x) \neq 0 \text{)}\\
&= \frac{f'(x)}{g(x)} - \frac{f(x)g'(x)}{g(x)^2} && \text{(algebra)}\\
&= \frac{g(x)f'(x)-f(x)g'(x)}{g(x)^2}. && \text{(algebra: common denominator)}
\end{aligned}$$

◇

Example 10 is the proof of the **quotient rule**. The quotient rule should be committed to memory. A simple case of the quotient rule has already been used to find the derivative of $\sec(x) = \frac{1}{\cos(x)}$ (Example 8). The quotient rule will be used to find the derivative of $\tan(x) = \frac{\sin(x)}{\cos(x)}$ (Example 11). Because the quotient rule is used so often, it should be memorized.

Quotient Rule

Let f and g be differentiable functions at x, and assume $g(x) \neq 0$. Then the quotient f/g is differentiable at x, and

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2} \qquad \text{where } g(x) \neq 0.$$

Because the numerator in the quotient rule is a difference, it is important to get the terms in the correct order. Here is an easy way to remember the quotient rule.

Step 1. Write down the parts where g^2 and g appear:

$$\frac{g\qquad\quad}{g^2}.$$

This ensures that the denominator is correct and one has a good start on the numerator.

Step 2. To complete the numerator, remember that it has a minus sign:

$$\frac{gf' - fg'}{g^2}.$$

EXAMPLE 11 Find the derivative of the tangent function.
SOLUTION

$$\begin{aligned}
(\tan(x))' &= \left(\frac{\sin(x)}{\cos(x)}\right)' \\
&= \frac{\cos(x)(\sin(x))' - \sin(x)(\cos(x))'}{(\cos(x))^2} && (\text{ quotient rule }) \\
&= \frac{(\cos(x))\cos(x) - \sin(x)(-\sin(x))}{(\cos(x))^2} \\
&= \frac{\cos^2(x) + \sin^2(x)}{\cos^2(x)} \\
&= \frac{1}{\cos^2(x)} && (\ \sin^2(x) + \cos^2(x) = 1\ ) \\
&= \sec^2(x). && (\ \sec(x) = 1/\cos(x)\ )
\end{aligned}$$

This result is valid whenever $\cos(x) \neq 0$, and should be memorized. ◇

Derivative of $\tan(x)$

$$D(\tan(x)) = \sec^2(x) \qquad \text{for all } x \text{ in the domain of } \tan(x).$$

EXAMPLE 12 Compute $(x^2/(x^3+1))'$, showing each step.
SOLUTION

$$\begin{aligned}
\left(\frac{x^2}{x^3+1}\right)' &= \frac{(x^3+1)\cdots}{(x^3+1)^2} && \text{(write denominator and start numerator)}\\
&= \frac{(x^3+1)(x^2)'-(x^2)(x^3+1)'}{(x^3+1)^2} && \text{(complete numerator, remembering the minus sign)}\\
&= \frac{(x^3+1)(2x)-(x^2)(3x^2)}{(x^3+1)^2} && \text{(compute derivatives)}\\
&= \frac{2x^4+2x-3x^4}{(x^3+1)^2} && \text{(algebra)}\\
&= \frac{2x-x^4}{(x^3+1)^2}. && \text{(algebra: collecting)}
\end{aligned}$$

◇

As Example 12 illustrates, the techniques for differentiating polynomials and quotients can be combined to differentiate any rational function, that is, any quotient of polynomials.

Summary

Let f and g be two differentiable functions and let C be a constant function. We obtained formulas for differentiating $f+g$, $f-g$, fg, Cf, $1/g$, and f/g.

Rule	Formula	Comment
Constant Rule	$C'=0$	C a constant
Sum Rule	$(f+g)'=f'+g'$	
Difference Rule	$(f-g)'=f'-g'$	
Product Rule	$(fg)'=f'g+fg'$	
Constant Multiple Rule	$(Cf)'=Cf'$	
Reciprocal Rule	$\left(\frac{1}{g}\right)'=\frac{-g'}{g^2}$	$g(x)\neq 0$
Quotient Rule	$\left(\frac{f}{g}\right)'=\frac{gf'-fg'}{g^2}$	$g(x)\neq 0$

Table 3.3.1

With the aid of the formulas in Table 3.3.1, we can differentiate $\sec(x)$, $\csc(x)$, $\tan(x)$, and $\cot(x)$ using $(\sin(x))'=\cos(x)$ and $(\cos(x))'=-\sin(x)$. Exercises 17(a) and 17(c) concern the differentiation of $\cot(x)$ and $\csc(x)$. We also have shown that $D(x^a)=ax^{a-1}$ for any fixed rational number a.

Function	Derivative	Comment
x^a	ax^{a-1}	a is a fixed number
$\tan(x)$	$\sec^2(x)$	for all x except odd multiples of $\pi/2$
$\sec(x)$	$\sec(x)\tan(x)$	for all x except odd multiples of $\pi/2$

Table 3.3.2

EXERCISES for Section 3.3

In Exercises 1 to 15 differentiate the given function. Use only the formulas presented in this and earlier sections.

1. $5x^3$

2. $5x^3 - 7x + 2^3$

3. $3\sqrt{x} - \sqrt[3]{x}$

4. $1/\sqrt{x}$

5. $(5+x)(x^2 - x + 7)$

6. $\sin(x)\cos(x)$

7. $3\tan(x)$

8. $3(\tan(x))^2$ (Write $(\tan(x))^2$ as $\tan(x)\tan(x)$.)

9. $\dfrac{x^3-1}{2x+1}$

10. $\dfrac{\sin(x)}{e^x}$

11. $\dfrac{3x^2 + x + \sqrt{2}}{\cos(x)}$

12. $\dfrac{2}{x^3} + \dfrac{3}{x^4}$

13. $x^2\sin(x)e^x$

14. $\sqrt{x}\sin(x)$

15. $\sqrt{x}/e^x$

16. Differentiate the following functions:

(a) $\dfrac{(1+\sqrt{x})(x^3 + \sin(x))}{x^2 + 5x + 3e^x}$

(b) $\dfrac{(3 + 4x + 5x^2)e^x}{\sin(x)}$

17. Use the quotient rule to obtain the following derivatives.

(a) $D(\cot(x)) = -(\csc(x))^2$

(b) $D(\sec(x)) = \sec(x)\tan(x)$

(c) $D(\csc(x)) = -\csc(x)\cot(x)$

There is a pattern here: the minus sign goes with each "co" function (cos, cot, csc).

18. Find $\left(e^{2x}\right)'$ by writing e^{2x} as $e^x e^x$.

19. Find $\left(e^{3x}\right)'$ by writing e^{3x} as $e^x e^x e^x$.

20. Find $(e^{-x})'$ by writing e^{-x} as $\frac{1}{e^x}$.

21. Find $\left(e^{-2x}\right)'$ by writing $e^{-2x} = e^{-x} \cdot e^{-x}$. (See Exercise 20.)

22. Find $\left(e^{-2x}\right)'$ by writing $e^{-2x} = \frac{1}{e^{2x}}$. (See Exercise 18.)

In Exercises 23 to 41 find the derivative of the function using formulas from this section.

23. $2^3 - \sqrt{\pi}$

24. $(x - x^{-1})^2$

25. $3\sin(x) - 5\cos(x)$

26. $5\tan(x)$

27. $u^5 - 6u^3 + u - 7$

28. $t^8/8$

29. $s^{-7}/(-7)$

30. $\sqrt{t}(t+4)$

31. $5/u^5$

32. $(x^3)^{1/2}$

33. $6\tan(x)$

34. $3\sec(x) - 4\cos(x)$

35. $\sec^2(\theta) - \tan^2(\theta)$ Simplify your answer.

36. $(3x)^4$

37. $u^2 e^u$

38. $e^t \sin(t)/\sqrt{t}$

39. $(3 + x^5)e^{-x}\tan(x)$

40. $(x - x^2)^3$ (multiply it out first.)

41. $\sqrt[3]{x}/\sqrt[5]{x}$

42. In Section 3.1 we showed that $D(1/x) = -1/x^2$. Obtain this formula by using the Quotient Rule.

43. If you had lots of time, how would you differentiate $(1 + 2x)^{100}$ using the formulas developed so far? In Section 3.5 we will obtain a shortcut for differentiating this function.

44. At what point on the graph of $y = xe^{-x}$ is the tangent horizontal?

45. Using the formula for the derivative of a product, obtain the formula for $(fgh)'$. (First write fgh as $(f)(gh)$. Then use the product rule twice.)

46. Obtain the formula for $(f-g)'$ by first writing $f-g$ as $f+(-1)g$.

47. Using the definition of the derivative as a limit, show that $(f-g)' = f' - g'$.

48. Using the version of the definition of the derivative that makes use of both x and $x+h$, obtain the formula for differentiating the sum of two functions.

49. Using the version of the definition of the derivative in the form $\lim\limits_{x \to a} \dfrac{f(x) - f(a)}{x-a}$, obtain the formula for differentiating the product of two functions.

Exercises 50 to 52 are examples of **proof by mathematical induction**. In this technique the truth of the statement for n is used to prove the truth of the statement for $n+1$.

50. In Section 3.2 we showed that $D(x^n) = nx^{n-1}$, when n is a positive integer. Now that we have the formula for the derivative of a product of two functions we can obtain this result more easily.

(a) Show, using the definition of the derivative, that the formula $D(x^n) = nx^{n-1}$ holds when $n = 1$.

(b) Using (a) and the formula for the derivative of a product, show that the formula holds when $n = 2$. ($x^2 = x \cdot x$.)

(c) Using (b) and the formula for the derivative of a product, show that it holds when $n = 3$.

(d) Show that if it holds for some positive integer n, it also holds for $n+1$.

(e) Combine (c) and (d) to show that the formula holds for $n = 4$.

(f) Why must it hold for $n = 5$?

(g) Why must it hold for all positive integers?

51. Using induction, as in Exercise 50, show that for each positive integer n, $D\left(x^{-n}\right) = -nx^{-n-1}$.

52. Using induction, as in Exercise 50, show that for each positive integer n, $D(\sin^n(x)) = n \sin^{n-1}(x) \cos(x)$.

53. We obtained the formula for $(f/g)'$ by writing f/g as the product of f and $1/g$. Obtain $(f/g)'$ directly from the definition of the derivative. (Review how we obtained the formula for the derivative of a product.)

3.4 The Chain Rule

We come now to the most frequently used formula for computing derivatives. For example, it will help us to find the derivative of $(1+x^2)^{100}$ without having to multiply out one hundred copies of $(1+x^2)$. You might expect the derivative of $(1+x^2)^{100}$ to be $100(1+x^2)^{99}$. This cannot be right.

When you expand $(1+x^2)^{100}$ you get a polynomial of degree 200, so its derivative is a polynomial of degree 199. But when you expand $(1+x^2)^{99}$ you get a polynomial of degree 198. Something is wrong.

At this point we know the derivative of $\sin(x)$, but what is the derivative of $\sin(x^2)$? It is not the cosine of x^2. In this section we obtain a way to differentiate these functions easily.

The key is that both $(1+x^2)^{100}$ and $\sin(x^2)$ are composite functions. This section shows how to differentiate composite functions.

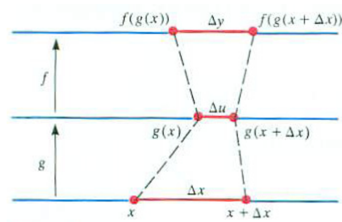


Figure 3.4.1

How to Differentiate a Composite Function

The composite function $y=(f\circ g)(x)=f(g(x))$ is built up by setting $u=g(x)$ and $y=f(u)$. The derivative of y with respect to x is the limit of $\Delta y/\Delta x$ as Δx approaches 0. The change in Δx causes a change Δu in u, which in turn causes the change Δy in y. (See Figure 3.4.1.) If Δu is not zero, then we may write

$$\frac{\Delta y}{\Delta x}=\frac{\Delta y}{\Delta u}\frac{\Delta u}{\Delta x}. \tag{3.4.1}$$

It could happen that $\Delta u = 0$, as it would, for instance, if g were a constant function. This special case is treated in Exercise 75.

Then

$$(f\circ g)'(x) = \lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta u}\frac{\Delta u}{\Delta x}=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta u}\lim_{\Delta x\to 0}\frac{\Delta u}{\Delta x}.$$

Since g is continuous, $\Delta u\to 0$ as $\Delta x\to 0$. So we have

$$(f\circ g)'(x) = \lim_{\Delta u\to 0}\frac{\Delta y}{\Delta u}\lim_{\Delta x\to 0}\frac{\Delta u}{\Delta x}=f'(u)g'(x).$$

Which gives us

Chain Rule
Let g be differentiable at x and f be differentiable at $g(x)$. Then

$$(f\circ g)'(x)=f'(g(x))g'(x).$$

The formula tells us how to differentiate a composite function, $f\circ g$:

Step 1. Compute the derivative of the outer function f, evaluated at the inner function. This is $f'(g(x))$.

Step 2. Compute the derivative of the inner function, $g'(x)$.

Step 3. Multiply the derivatives found in Steps 1 and 2, obtaining $f'(g(x))g'(x)$.

In short, to differentiate $f(g(x))$, think of g as the inner function and f as the outer function. Then the derivative of $f \circ g$ is

$$\underbrace{f'(g(x))}_{\text{derivative of outer function evaluated at inner function}} \qquad \text{times} \qquad \underbrace{g'(x)}_{\text{derivative of inner function}}$$

Examples

EXAMPLE 1 Find $D\left((1+x^2)^{100}\right)$.
SOLUTION Here $g(x) = 1 + x^2$ (the inner function) and $f(u) = u^{100}$ (the outer function). The first step is to compute $f'(u) = 100u^{99}$, which gives us $f'(g(x)) = 100(1 + x^2)^{99}$. The second step is to find $g'(x) = 2x$. Then,

$$(f\circ g)'(x) = f'(\underbrace{u}_{u=g(x)})g'(x) = \underbrace{100u^{99}}_{f'(u)} \cdot \underbrace{2x}_{g'(x)} = \underbrace{100(1+x^2)^{99}}_{f'(g(x))} \cdot 2x = 200x(1+x^2)^{99}.$$

The answer is not just $100(1 + x^2)^{99}$. There is a factor of $2x$ that comes from the derivative of the inner function, so the derivative of $(1 + x^2)^{100}$ has degree 199, as expected. ◇

The same example, done with Leibniz notation, looks like this:

$$y = (1+x^2)^{100} = u^{100}, \qquad u = 1 + x^2.$$

Then the Chain Rule reads

$$\frac{dy}{dx} = \underbrace{\frac{dy}{du}\frac{du}{dx}}_{\text{Chain Rule}} = 100u^{99} \cdot 2x = \underbrace{100(1+x^2)^{99}(2x)}_{\text{Using } u=1+x^2} = 200x(1+x^2)^{99}.$$

WARNING We avoided using Leibniz notation earlier, in particular, during the derivation of the Chain Rule, because it tempts the reader to cancel the du's in (3.4.1). However, the expressions dy, du, and dx are meaningless in themselves. In Leibniz's time in the late seventeenth century their meaning was fuzzy, standing for a quantity that was zero and also vanishingly small at the same time. Bishop Berkeley poked fun at this, asking "may we not call them the ghosts of departed quantities?"

George Berkeley, 1734, *The Analyst: A Discourse Addressed to an Infidel Mathematician.*

With practice, you will be able to do the calculation without introducing extra symbols, such as u, which do not apear in the final answer. You will be writing just

$$D\left((1+x^2)^{100}\right) = 100(1+x^2)^{99}\cdot 2x = 200x(1+x^2)^{99}.$$

Developing this skill, like playing the guitar, takes practice, which the exercises at the end of this section (and chapter) provide.

When we write $\frac{dy}{du}$ and $\frac{du}{dx}$, the u serves two roles. In $\frac{dy}{du}$ it denotes an independent variable while in $\frac{du}{dx}$, u is a dependent variable. The double role usually causes no problem in computing derivatives.

EXAMPLE 2 If $y = \sin(x^2)$, find $\frac{dy}{dx}$.
SOLUTION Starting from the outer function, let $y = \sin(u)$ and $u = x^2$. Then, by the chain rule,

$$\left(\sin(x^2)\right)' = \frac{dy}{dx} = \underbrace{\frac{dy}{du}\frac{du}{dx}}_{\text{chain rule}} = \cos(u)\cdot 2x = \cos(x^2)\cdot 2x = 2x\cos(x^2).$$

The outer function is the sine and the inner function is x^2. So we have, in short,

$$(\underbrace{\sin}_{\text{outer}}\ \underbrace{(x^2)}_{\text{inner}})' = \underbrace{\cos(x^2)}_{\text{derivative of outer function evaluated at inner function}} \quad \text{times} \quad \underbrace{2x}_{\text{derivative of inner function}} = 2x\cos(x^2).$$

◇

The chain rule holds for compositions of more than two functions. We illustrate this in the next example.

EXAMPLE 3 Differentiate $y = \sqrt{\sin(x^2)}$.
SOLUTION Tthe function is the composition of three functions:

$$u = x^2 \qquad v = \sin(u) \qquad y = \sqrt{v}.$$

Do this example yourself *without introducing any auxiliary symbols (u, v, and y).*

Then

$$\frac{dy}{dx} = \underbrace{\frac{dy}{dv}\frac{dv}{dx}}_{\text{chain rule}} = \underbrace{\frac{dy}{dv}\frac{dv}{du}\frac{du}{dx}}_{\text{chain rule, again}} = \frac{1}{2\sqrt{v}} \cdot \cos(u) \cdot 2x$$

$$= \frac{1}{2\sqrt{\sin(x^2)}} \cdot \cos(x^2) \cdot 2x = \frac{x\cos(x^2)}{\sqrt{\sin(x^2)}}.$$

◇

$b = e^{\ln(b)}$ for any $b > 0$

EXAMPLE 4 Let $y = 2^x$. Find y'.

SOLUTION As it stands, 2^x is not a composite function. However, we can write $2 = e^{\ln(2)}$ and then 2^x equals $(e^{\ln(2)})^x = e^{\ln(2)x}$, so 2^x can be expressed as the composite function

$$y = e^u, \text{ where } u = (\ln(2))x.$$

Then

$$y' = \frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = e^u \cdot \ln(2) = e^{\ln(2)x}\ln(2) = 2^x\ln(2).$$

In Example 2 (Section 3.2), using a calculator, we found $D(2^x) \approx (0.693)2^x$. We have just seen that the exact formula is $D(2^x) = 2^x\ln(2)$. This means that 0.693 is an approximation of $\ln(2)$. ◇

The next Example shows how the chain rule is used in combination with other differentiation rules such as the product and quotient rules.

EXAMPLE 5 Find $D\left(x^3\tan(x^2)\right)$.

SOLUTION The function $x^3\tan(x^2)$ is the product of two functions. We first apply the product rule to obtain

product rule: $(fg)' = f' \cdot g + f \cdot g'$

$$\begin{aligned} D\left(x^3\tan(x^2)\right) &= (x^3)'\tan(x^2) + x^3\left(\tan(x^2)\right)' \\ &= 3x^2\tan(x^2) + x^3\left(\tan(x^2)\right)'. \end{aligned}$$

$(\tan(x))' = \sec^2(x)$

Since the derivative of the tangent is the square of the secant, the chain rule tells us that

$$\left(\tan(x^2)\right)' = \sec^2(x^2)(x^2)' = 2x\sec^2(x^2).$$

Thus,

$$\begin{aligned} D\left(x^3\tan(x^2)\right) &= 3x^2\tan(x^2) + x^3\left(\tan(x^2)\right)' \\ &= 3x^2\tan(x^2) + x^3\left(2x\sec^2(x^2)\right) \\ &= 3x^2\tan(x^2) + 2x^4\sec^2(x^2). \end{aligned}$$

◇

In the computation of $D(\tan(x^2))$ we did not introduce any new symbols. That is how your computations will look, once you become proficient using the chain rule.

Famous Composite Functions

Certain types of composite functions occur so often that it is worthwhile memorizing their derivatives. This table lists the most common ones.

Function	Derivative	Example
$(g(x))^n$	$ng(x)^{n-1}g'(x)$	$\left((1+x^2)^{100}\right)' = 100(1+x^2)^{99}(2x)$
$\frac{1}{g(x)}$	$\frac{-g'(x)}{(g(x))^2}$	$D\left(\frac{1}{\cos(x)}\right) = \frac{-(-\sin(x))}{(\cos(x))^2}$
$\sqrt{g(x)}$	$\frac{g'(x)}{2\sqrt{g(x)}}$	$\left(\sqrt{\tan(x)}\right)' = \frac{(\sec(x))^2}{2\sqrt{\tan(x)}}$
$e^{g(x)}$	$e^{g(x)}g'(x)$	$\left(e^{x^2}\right)' = e^{x^2}(2x)$

Table 3.4.1

Summary

This section presented the single most important tool for computing derivatives: the chain rule, which says that the derivative of $f \circ g$ at x is

$$\underbrace{f'(g(x))}_{\text{derivative of outer function evaluated at the inner function}} \quad \text{times} \quad \underbrace{g'(x)}_{\text{derivative of inner function}}$$

Introducing the symbol u, we described the chain rule for $y = f(u)$ and $u = g(x)$ with the brief notation

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}.$$

When the function is built up from more than two functions, such as $y = f(u)$, $u = g(v)$, and $v = h(x)$, we have

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dv}\frac{dv}{dx},$$

a chain of more derivatives.

With practice, applying the chain rule becomes second nature.

EXERCISES for Section 3.4

In Exercises 1 to 4, repeat the specified example from this section without introducing an extra variable (such as u).

1. Example 1

2. Example 2

3. Example 3

4. Example 4

In Exercises 5 to 40 find the derivative of the function. Simplify your answers.

5. $(x^3 + 2)^5$

6. $(x^2 + 3x + 1)^4$

7. $\sqrt{\cos(x^3)}$

8. $\sqrt{\tan(x^2)}$

9. $\left(\frac{1}{x}\right)^{10}$

10. $\cos(3x)\sin(2x)$

11. $x^2 \tan(x^3)$

12. $(1 + 2x)\sin(x^4)$

13. $5(\tan(x^3))^2$

14. $\frac{\cos^3(2x)}{x^5}$

15. $\sin(2\exp(x))$

16. $e^{\cos(x)}$

17. $\frac{(1+2x)^2}{x^3}$

18. $(\sec(5x))(\cos(5x))$

19. $(5x^2 + 3)^{10}$

20. $(\sin(3x))^3$

21. $\frac{1}{5t^2+t+2}$

22. $\frac{1}{e^{5s}+s}$

23. $\sqrt{4 + u^2}$

24. $\left(\sqrt{\cos(2\theta)}\right)^3$

25. e^{5x^3}

26. $\sin^2(3x)$

27. $e^{\tan(3t)}$

28. $\sqrt{\tan(2u)}$

29. $\sqrt[3]{\tan(s^2)}$

30. $v^3 \tan(2v)$

31. $e^{2r}\sin(3r)$

32. $\frac{\sec(2x)}{x^2}$

33. $\exp(\sin(2x))$

34. $\frac{(3t+2)^4}{\sin(2t)}$

35. $e^{-5s}\tan(3s)$

36. e^{x^2}

37. $(\sin(2u))^5(\cos(3u))^6$

38. $\left(x+3^{3x}\right)^2(\sin(\sqrt{x}))^3$

39. $\frac{t^3}{\left(t+\sin^2(3t)\right)}$

40. $\frac{(3x+2)^4}{(x^3+x+1)^2}$

Learning to use the chain rule takes practice. Exercises 41 to 68 offer opportunities. They also show that sometimes the derivative of a function can be simpler than the function. In each case show that the derivative of the first function is the second function, the functions being separated by a semi-colon. The letters a, b, and c denote constants.

41. $\frac{b}{2a^2(ax+b)^2}-\frac{1}{a^2(ax+b)}$; $\frac{x}{(ax+b)^2}$

42. $\frac{-1}{2a(ax+b)^2}$; $\frac{1}{(ax+b)^3}$

43. $\frac{2}{3a}\sqrt{(ax+b)^3}$; $\sqrt{ax+b}$

44. $\frac{2(3ax-2b)}{15a^2}\sqrt{(ax+b)^3}$; $x\sqrt{ax+b}$

45. $\frac{-\sqrt{ax^2+c}}{cx}$; $\frac{1}{x^2\sqrt{ax^2+c}}$

46. $\frac{x}{c\sqrt{ax^2+c}}$; $(ax^2+c)^{-3/2}$

47. $\frac{1}{a}\sin(ax)-\frac{1}{3a}\sin^3(ax)$; $\cos^3(ax)$

48. $\frac{1}{a(n+1)}\sin^{n+1}(ax)$; $\sin^n(ax)\ \cos(ax)$

49. $\frac{2(ax-2b)}{3a^2}\sqrt{ax+b}$; $\frac{x}{\sqrt{ax+b}}$

50. $\frac{2(3a^2x^2-4abx+8b^2)}{15a^3}\sqrt{ax+b}$; $\frac{x^2}{\sqrt{ax+b}}$

51. $\frac{-\sqrt{ax^2+c}}{cx}$; $\frac{1}{x^2\sqrt{ax^2+c}}$

52. $\frac{-x^2}{a\sqrt{ax^2+c}}+\frac{2}{a^2}\sqrt{ax^2+c}$; $\frac{x^3}{(ax^2+c)^{3/2}}$

53. $\frac{-1}{a}\cos(ax)+\frac{1}{3a}\cos^3(ax)$; $\sin^3(ax)$

54. $\frac{3x}{8}-\frac{3\sin(2ax)}{16a}-\frac{\sin^3(ax)\cos(ax)}{4a}$; $\sin^4(ax)$

55. $\frac{\sin((a-b)x)}{2(a-b)}-\frac{\sin((a+b)x)}{2(a+b)}$; $\sin(ax)\sin(bx)$ (Assume $a^2\neq b^2$.)

56. $\frac{x}{2}+\frac{\sin(2ax)}{3a}$; $\cos^3(ax)$

57. $\frac{1}{a}\tan(ax)$; $\frac{1}{\cos^2(ax)}$

58. $\frac{1}{a}\tan\left(\frac{ax}{2}\right)$; $\frac{1}{1+\cos(ax)}$

59. $2\sqrt{2}\sin\left(\frac{x}{2}\right)$; $\sqrt{1+\cos(x)}$ You will need to use a trigonometric identity.

60. $\frac{\sin((a-b)x}{2(a-b)} + \frac{\sin((a+b)x)}{2(a+b)}$; $\cos(ax)\cos(bx)$ (Assume $a^2 \neq b^2$.)

61. $\frac{1}{a}\left(\tan(ax) - \cot(ax)\right)$; $\frac{1}{\sin^2(ax)\cos^2(ax)}$

62. $\frac{1}{a}\tan(ax) - 1$; $\tan^2(ax)$

63. $\frac{\sec^n(ax)}{an}$; $\tan(ax)\sec^n(ax)$ (Assume $n \neq 0$.)

64. $\frac{\sin(ax)}{a^2} - \frac{x\cos(ax)}{a}$; $x\sin(ax)$

65. $\frac{\cos(ax)}{a^2} + \frac{x\sin(ax)}{a}$; $x\cos(ax)$

66. $\frac{1}{a^2}e^{ax}(ax-1)$; xe^{ax}

67. $\frac{1}{a^3}e^{ax}(a^2x^2 - 2ax + 2)$; x^2e^{ax}

68. $\frac{e^{ax}(a\sin(bx) - b\cos(bx))}{a^2+b^2}$; $e^{ax}\sin(bx)$

Exercises 69 and 70 illustrate how differentiation can be used to obtain one trigonometry identity from another.

69.

(a) Differentiate both sides of the identity $\sin^2(x) = \frac{1}{2}(1-\cos(2x))$. What trigonometric identity do you get?

(b) Differentiate the identity found in (a) to obtain another trigonometric identity. What identity is obtained?

(c) Does this process continued forever produce new identities?

70. Let k be a constant. Differentiate both sides of the identity $\sin(x+k) = \sin(x)\cos(k) + \cos(x)\sin(k)$ to obtain the corresponding identity for $\cos(x+k)$.

71. Differentiate $(e^x)^3$

(a) Directly, by the chain rule

(b) After writing it as $e^x \cdot e^x \cdot e^x$ and using the product rule

(c) After writing it as e^{3x} and using the chain rule

(d) Which of these approaches do you prefer? Why?

72. In Section 3.3 we obtained the derivative of $1/g(x)$ by using the definition of the derivative. Obtain that formula for the reciprocal rule by using the chain rule.

73. In our proof of the chain rule we had to assume that Δu is not 0 when Δx is sufficiently small. Show that if the derivative of g is not 0 at the argument x, then the proof is valid.

74. Here is an example of a differentiable g not covered by the proof of the chain rule given in the text. Define $g(x)$ to be $x^2 \sin\left(\frac{1}{x}\right)$ for x different from 0 and $g(0)$ to be 0.

(a) Sketch the part of the graph of g near the origin.

(b) Show that there are arbitrarily small values of Δx such that

$$\Delta u = g(\Delta x) - g(0) = 0.$$

(c) Show that g is differentiable at 0.

75. Here is a proof of the chain rule that manages to avoid division by $\Delta u = 0$. Let $f(u)$ be differentiable at $g(a)$, where g is differentiable at a. Let $\Delta f = f(g(a) + \Delta u) - f(g(a))$. Then $\frac{\Delta f}{\Delta u} - f'(g(a))$ is a function of Δu, which we call $p(\Delta u)$. It is defined for $\Delta u \neq 0$. By the definition of f', $p(\Delta u)$ tends to 0 as Δu approaches 0. Define $p(0)$ to be 0. Note that p is continuous at 0.

(a) Show that $\Delta f = f'(g(a))\Delta u + p(\Delta u)\Delta u$ when Δu is different than 0, and also when $\Delta u = 0$.

(b) Define $q(\Delta x) = \frac{\Delta u}{\Delta x} - g'(a)$. Observe that $q(\Delta x)$ approaches 0 as Δx approaches 0. Show that $\Delta u = g'(a)\Delta x + q(\Delta x)\Delta x$ when Δx is not 0.

(c) Combine (a) and (b) to show that

$$\Delta f = f'(g(a))\left(g'(a)\Delta x + q(\Delta x)\Delta x\right) + p(\Delta u)\Delta u.$$

(d) Using (c), show that

$$\lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = f'(g(a))g'(a).$$

(e) Why did we have to define $p(0)$ but not $q(0)$?

3.5 Derivative of an Inverse Function

In this section we obtain the derivatives of the inverse functions of e^x and of the six trigonometric functions. This will complete the inventory of basic derivatives. The chain rule will be our main tool.

Differentiability of Inverse Functions

As mentioned in Section 1.1, the graph of an inverse function is a copy of the graph of the original function. One graph is obtained from the other by reflection across the line $y = x$. If the original function, f, is differentiable at a point (a, b), $b = f(a)$, then the graph of $y = f(x)$ has a tangent line at (a, b). The reflection across $y = x$ of the tangent line to the graph of f is the tangent line to the inverse function at (b, a). Thus, we expect that the inverse function, f^{-1}, is differentiable at (b, a), and we will assume it is.

$a = f^{-1}(b)$ means $b = f(a)$

First, the chain rule will be used to find the derivative of $\log_e(x)$.

The Derivative of $\log_e(x)$

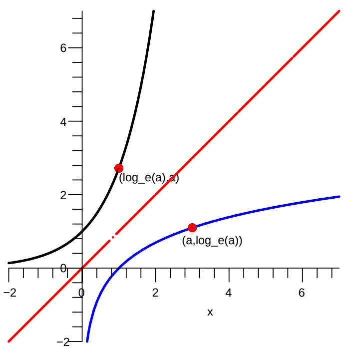


Figure 3.5.1

Let $y = \log_e(x)$. Figure 3.5.1 shows the graphs of $y = e^x$ and the inverse function $y = \log_e(x)$. We want to find $y' = \frac{dy}{dx}$. By the definition of logarithm as the inverse of the exponential function,

$$x = e^y. \tag{3.5.1}$$

We differentiate both sides of (3.5.1) with respect to x:

$$\begin{aligned}
\frac{d(x)}{dx} &= \frac{d(e^y)}{dx} && \text{(e^y is a function of x, since y is a function of x}\\
1 &= \frac{d(e^y))}{dx} && \left(\tfrac{dx}{dx} = 1 \right)\\
1 &= e^y \frac{dy}{dx} && \text{(chain rule.)}
\end{aligned}$$

Solving for $\frac{dy}{dx}$, we obtain

$$\frac{dy}{dx} = \frac{1}{e^y} = \frac{1}{x}.$$

This is another important differentiation rule.

Derivative of $\log_e(x)$
$$(\log_e(x))' = \frac{1}{x}, \qquad x > 0.$$

It may come as a surprise that such a complicated function has such a simple derivative. It may also be a surprise that $\log_e(x)$ is one of the most important functions in calculus, mainly because it has the derivative $1/x$.

EXAMPLE 1 Find $(\log_b(x))'$ for any $b > 0$.
SOLUTION The function $\log_b x$ is a constant times $\log_e(x)$:

$$\log_b(x) = (\log_b(e)) \log_e(x).$$

Therefore

$$(\log_b(x))' = (\log_b(e)) \frac{1}{x}.$$

If b is not e, then $\log_b(e)$ is not 1. If e is chosen as the base for logarithms, then the coefficient of $\frac{1}{x}$ becomes $\log_e(e) = 1$. That is another reason why we prefer e as the base for logarithms in calculus ◇

Recall from Section 2.2 that $\log_e(x)$ is the natural logarithm, which is denoted $\ln(x)$.

> **WARNING** *(Logarithm Notation)* $\ln(x)$ is often written simply as $\log(x)$, with the base understood to be e. All references in this book to the base-10 logarithm will use the notation $\log_{10}$.

The Derivative of $\arcsin(x)$

For x in $[-\pi/2, \pi/2]$ $\sin(x)$ is one-to-one and therefore has an inverse function, $\arcsin(x)$, which gives the angle, in radians, if you know the sine of the angle. For instance, $\arcsin(1) = \pi/2$, $\arcsin(\sqrt{2}/2) = \pi/4$, $\arcsin(-1/2) = -\pi/6$, and $\arcsin(-1) = -\pi/2$. The domain of $\arcsin(x)$ is $[-1, 1]$ and its range is $[-\pi/2, \pi/2]$. For convenience we include the graphs of $y = \sin(x)$ and $y = \arcsin(x)$ in Figure 3.5.2, but will not need them to find $(\arcsin(x))'$.

Inverse trigonometric functions were introduced in Section 1.2.

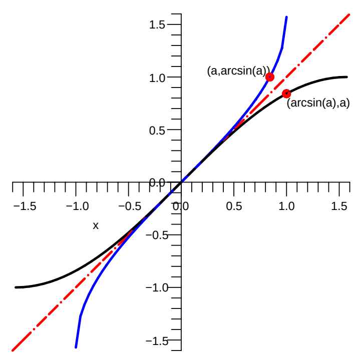


Figure 3.5.2

To find $(\arcsin(x))'$, we proceed as we did when finding $(\log_e(x))'$. Let $y = \arcsin(x)$, so

$$x = \sin(y).$$

We then have

$$\begin{aligned} x &= \sin(y). \\ \frac{d(x)}{dx} &= \frac{d(\sin(y))}{dx} && (\text{ differentiate with respect to } x\text{ }) \\ 1 &= (\cos(y))\, y' && (\text{ chain rule }) \\ y' &= \frac{1}{\cos(y)} && (\text{ algebra. }) \end{aligned}$$

Figure 3.5.3 displays the diagram that defines the sine of an angle. The line segment AB represents $\cos(y)$ and the line segment BC represents $\sin(y)$.

The cosine is positive for angles y in $\left(\frac{-\pi}{2}, \frac{\pi}{2}\right)$, the first and fourth quadrants. When $x = \sin(y)$, $x^2 + \cos^2(y) = 1$ gives $\cos(y) = \pm\sqrt{1-x^2}$. We use the positive value: $\cos(y) = \sqrt{1-x^2}$ because arcsin is an increasing function. Consequently, we find

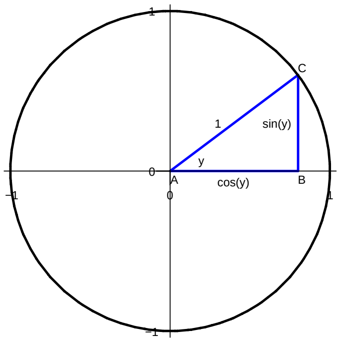


Figure 3.5.3

Derivative of $\arcsin(x)$

$$\frac{d}{dx}\left(\arcsin(x)\right) = \frac{1}{\sqrt{1-x^2}}, \qquad |x| < 1.$$

The formula for the derivative of the inverse sine should be memorized.

At $x = 1$ or at $x = -1$, the derivative is not defined. However, for x near 1 or -1 the derivative is very large (in absolute value), telling us that the graph of the arcsine function is very steep near its two ends. That is a reflection of the fact that the graph of $\sin(x)$ is horizontal at $x = -\pi/2$ and $x = \pi/2$.

Functions such as $x^3 - x$, $x^{2/7}$, and $\frac{1}{\sqrt{1-x^2}}$ that can be written in terms of the algebraic operations of addition, subtraction, multiplication, division, raising to a power, and extracting a root are called **algebraic functions**. Functions that cannot be written in this way, including e^x, $\cos(x)$, and $\arcsin(x)$, are known as **transcendental functions**. The derivative of $\ln(x)$ and $\arcsin(x)$ shows that the derivative of a transcendental function can be an algebraic function. The derivative of an algebraic function will always be algebraic.

An algebraic function always has an algebraic derivative.

EXAMPLE 2 Differentiate $\arcsin\left(x^2\right)$.
SOLUTION By the chain rule,

$$\frac{d}{dx}\left(\arcsin\left(x^2\right)\right) = \frac{1}{\sqrt{1-\left(x^2\right)^2}} \cdot \frac{d}{dx}\left(x^2\right) = \frac{2x}{\sqrt{1-x^4}}.$$

◇

EXAMPLE 3 Differentiate $\frac{1}{2}\left(x\sqrt{a^2-x^2} + a^2 \arcsin\left(\frac{x}{a}\right)\right)$ where a is a constant.

SOLUTION

$$\begin{aligned}
&D\left(\frac{1}{2}\left(x\sqrt{a^2-x^2}+a^2\arcsin\left(\frac{x}{a}\right)\right)\right) && \\
&= \tfrac{1}{2}D\left(\left(x\sqrt{a^2-x^2}+a^2\arcsin\left(\tfrac{x}{a}\right)\right)\right) && \\
&= \tfrac{1}{2}\left(D\left(x\sqrt{a^2-x^2}\right)+a^2D\left(\arcsin\left(\tfrac{x}{a}\right)\right)\right) && \\
&= \tfrac{1}{2}\left(\left((1)\sqrt{a^2-x^2}\right)+\left(x\left(\frac{\left(\frac{1}{2}\right)(-2x)}{\sqrt{a^2-x^2}}\right)\right)\right. && (\text{ product and chain rules }) \\
&\qquad \left.+a^2\left(\frac{\frac{1}{a}}{\sqrt{1-\left(\frac{x}{a}\right)^2}}\right)\right) && \left(\ D(\arcsin(x))=\frac{1}{\sqrt{1-x^2}}\ \right) \\
&= \tfrac{1}{2}\left(\sqrt{a^2-x^2}+\frac{-x^2}{\sqrt{a^2-x^2}}+\frac{a^2}{\sqrt{a^2-x^2}}\right) && (\text{ algebra }) \\
&= \tfrac{1}{2}\left(\frac{a^2-x^2-x^2+a^2}{\sqrt{x^2-a^2}}\right) && (\text{ common denominator }) \\
&= \sqrt{a^2-x^2}.
\end{aligned}$$

A complicated function can have a simple derivative. ◇

The Derivative of $\arctan(x)$

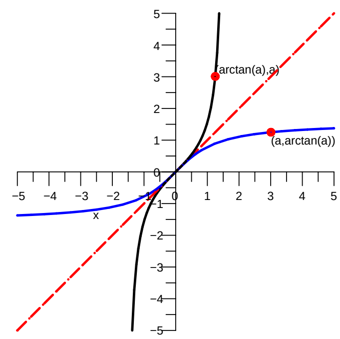


Figure 3.5.4

See Exercise 82.

For x in $(-\pi/2, \pi/2)$ the function $\tan(x)$ is one-to-one and has an inverse function, $\arctan(x)$, which tells us the angle, in radians, if we know the tangent of the angle. For instance, $\arctan(1) = \pi/4$, $\arctan(0) = 0$, and $\arctan(-1) = -\pi/4$. When x is a large positive number, $\arctan(x)$ is near, and smaller than, $\pi/2$. When x is a large negative number, $\arctan(x)$ is near, and larger than, $-\pi/2$. Figure 3.5.4 shows the graph of $y = \arctan(x)$ and $y = \tan(x)$. We will not need this graph when differentiating $\arctan(x)$, but it serves as a check on the formula.

To find $(\arctan(x))'$, we again use the chain rule. Starting with

$$y = \arctan(x),$$

we proceed as before:

$$\begin{aligned}
x &= \tan(y). && \\
\frac{d(x)}{dx} &= \frac{d(\tan(y))}{dx} && (\text{ differentiate with respect to } x\ ) \\
1 &= \left(\sec^2(y)\right)y' && (\text{ chain rule }) \\
y' &= \frac{1}{\sec^2(y)} && (\text{ algebra }) \\
y' &= \frac{1}{1+\tan^2(y)} && (\text{ trigonometric identity }) \\
y' &= \frac{1}{1+x^2} && (\ x=\tan(y)\ ).
\end{aligned}$$

This derivation is summarized by a simple formula, which should be memorized.

Derivative of $\arctan(x)$

$$D(\arctan(x)) = \frac{1}{1+x^2} \qquad \text{for all inputs } x.$$

EXAMPLE 4 Find $D(\arctan(3x))$.
SOLUTION By the chain rule

$$D(\arctan(3x)) = \frac{1}{1+(3x)^2}\frac{d(3x)}{dx} = \frac{3}{1+9x^2}.$$

◇

EXAMPLE 5 Find $D\left(x\tan^{-1}(x) - \frac{1}{2}\ln\left(1+x^2\right)\right)$.
SOLUTION

$$\begin{aligned} D\left(x\tan^{-1}(x) - \tfrac{1}{2}\ln\left(1+x^2\right)\right) &= D\left(x\tan^{-1}(x)\right) - \tfrac{1}{2}D\left(\ln\left(1+x^2\right)\right) \\ &= \left(\tan^{-1}(x) + \frac{x}{1+x^2}\right) - \frac{1}{2}\frac{2x}{1+x^2} \\ &= \tan^{-1}(x). \end{aligned}$$

◇

More on $\ln(x)$

Recall that $\ln(x)$ is not defined for $x < 0$.

An **antiderivative** of a function, $f(x)$, is another function, $F(x)$, whose derivative is equal to $f(x)$. That is, $F'(x) = f(x)$, and so $\ln(x)$ is an antiderivative of $1/x$. We showed that for $x > 0$, $\ln(x)$ is an antiderivative of $1/x$. But what if we needed an antiderivative of $1/x$ for negative x? The next example answers this question.

EXAMPLE 6 Show that for negative x, $\ln(-x)$ is an antiderivative of $1/x$.
SOLUTION Let $y = \ln(-x)$. By the chain rule,

$$\frac{dy}{dx} = \left(\frac{1}{-x}\right)\frac{d(-x)}{dx} = \frac{1}{-x}(-1) = \frac{1}{x}.$$

So $\ln(-x)$ is an antiderivative of $1/x$ when x is negative. ◇

In view of Example 6, $\ln|x|$ is an antiderivative of $1/x$, whether x is positive or negative.

Derivative of $\ln|x|$

$$D(\ln|x|) = \frac{1}{x} \qquad \text{for } x \neq 0.$$

We know the derivative of x^a for any rational number a. To extend this result to x^k for any number k, and positive x, we write x as $e^{\ln(x)}$.

EXAMPLE 7 Find $D(x^k)$ for $x > 0$ and any constant $k \neq 0$, rational or irrational.

SOLUTION For $x > 0$ we can write $x = e^{\ln(x)}$. Then

$$x^k = \left(e^{\ln(x)}\right)^k = e^{k\ln(x)}.$$

Looking at $y = e^{k\ln(x)}$ as a composite function, $y = e^u$ where $u = k\ln(x)$, we have

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = e^u\frac{k}{x} = x^k\frac{k}{x} = kx^{k-1}.$$

◇

The example shows that for positive x and any fixed exponent k, $\left(x^k\right)' = kx^{k-1}$. It probably does not come as a surprise. You may wonder why we worked so hard to get the derivative of x^a when a is an integer or rational number when this example covers all exponents. We had two reasons for treating the special cases. First, they include cases when x is negative. Second, they were simpler and helped introduce the derivative.

The Derivatives of the Six Inverse Trigonometric Functions

Of the six inverse trigonometric functions, the most important are arcsin and arctan. The other four are treated in Exercises 71 to 74. Table 3.5.1 summarizes all six derivatives. There is no reason to memorize the formulas. If we need, say, an antiderivative of $\frac{-1}{1+x^2}$, we do not have to use $\operatorname{arccot}(x)$. Instead, $-\arctan(x)$ would do. For finding antiderivatives, we don't need arccot, or any of the inverse co-functions. The formulas for the derivatives of arcsin, arctan, and arcsec suffice.

$$
\begin{array}{rclrcll}
D(\arcsin(x)) & = & \dfrac{1}{\sqrt{1-x^2}} & D(\arccos(x)) & = & \dfrac{-1}{\sqrt{1-x^2}} & (-1 < x < 1) \\
D(\arctan(x)) & = & \dfrac{1}{1+x^2} & D(\operatorname{arccot}(x)) & = & \dfrac{-1}{1+x^2} & (-\infty < x < \infty) \\
D(\operatorname{arcsec}(x)) & = & \dfrac{1}{x\sqrt{x^2-1}} & D(\operatorname{arccsc}(x)) & = & \dfrac{-1}{x\sqrt{x^2-1}} & (x > 1 \text{ or } x < -1)
\end{array}
$$

Table 3.5.1 Derivatives of the six inverse trigonometric functions.

Another View of e

For each choice of the base b ($b > 0$), we obtain a value for $\lim_{x\to 0} \frac{b^x - 1}{x}$. We defined e to be the base for which the limit is as simple as possible, namely 1: $\lim_{x\to 0} \frac{e^x - 1}{x} = 1$.

Now that we know that the derivative of $\ln x = \log_e x$ is $1/x$, we can obtain a new view of e.

The derivative of $\ln(x)$ at 1 is $1/1 = 1$. By the definition of the derivative,

$$\lim_{h\to 0} \frac{\ln(1+h) - \ln(1)}{h} = 1.$$

Since $\ln(1) = 0$, we have

$$\lim_{h\to 0} \frac{\ln(1+h)}{h} = 1.$$

By a property of logarithms, we may rewrite the limit as

$$\lim_{h\to 0} \ln\left((1+h)^{1/h}\right) = 1.$$

Writing e^x as $\exp(x)$ for convenience, we conclude that

$$\exp\left(\lim_{h\to 0} \ln\left((1+h)^{1/h}\right)\right) = \exp(1) = e.$$

Since exp is a continuous function, we may switch exp and lim, getting

$$\lim_{h\to 0} \left(\exp\left(\ln\left((1+h)^{1/h}\right)\right)\right) = e.$$

But $\exp(\ln(p)) = p$ for any positive number, by the definition of a logarithm. That tells us that

$$\lim_{h\to 0} (1+h)^{1/h} = e.$$

This is a more direct view of e than the one in Section 2.2. As a check, let $h = 1/1000 = 0.001$. Then $(1 + 1/1000)^{1000} \approx 2.717$, and values of h closer to 0 give better estimates for e, whose decimal expansion begins 2.718.

Summary

A geometric argument suggested that the inverse of every differentiable function is differentiable. The chain rule then helped find the derivatives of $\ln(x)$, $\arcsin(x)$, and $\arctan(x)$ and of the other four inverse trigonometric functions.

EXERCISES for Section 3.5

In Exercises 1 to 6 evaluate the function and its derivative at the given argument.

1. $\arcsin(x)$; $1/2$

2. $\arcsin(x)$; $-1/2$

3. $\arctan(x)$; -1

4. $\arctan(x)$; $\sqrt{3}$

5. $\ln(x)$; e

6. $\ln(x)$; 1

In Exercises 7 to 28 differentiate the function.

7. $\arcsin(3x)\sin(3x)$

8. $\arctan(5x)\tan(5x)$

9. $e^{2x}\ln(3x)$

10. $e^{\left(\ln(3x)x^{\sqrt{2}}\right)}$

11. $x^2\arcsin(x^2)$

12. $(\arcsin(3x))^2$

13. $\frac{\arctan(2x)}{1+x^2}$

14. $\frac{x^3}{\arctan(6x)}$

15. $\log_{10}(x)$ (Express $\log_{10}$ in terms of the natural logarithm.)

16. $\log_x(10)$ (Express $\log_x$ in terms of the natural logarithm.)

17. $\arcsin(x^3)$

18. $\arctan(x^2)$

19. $(\arctan(3x))^2$

20. $(\arccos(5x))^3$

21. $\frac{\arcsin(1+x^2)}{1+3x}$

22. $\operatorname{arcsec}(x^3)$

23. $x^2\arcsin(3x)$

24. $\frac{\arctan(3x)}{\tan(2x)}$

25. $\frac{\arctan(x^3)}{\arctan(x)}$

26. $\ln(\sin(3x))$

27. $\ln\left((\sin(x)^3)\right)$

28. $\ln(\exp(4x))$

In Exercises 29 to 65 check that the derivative of the first function is the second, separated by a semi-colon. The letters a, b, and c denote constants.

29. $\frac{1}{cn}\ln\left(\frac{x^n}{ax^n+c}\right)$; $\frac{1}{x(ax^n+c)}$ (To simplify the calculation, first use the property $\ln(p/q)=\ln(p)-\ln(q)$.)

30. $\frac{1}{nc}\ln\left(\frac{\sqrt{ax^n+c}-\sqrt{c}}{\sqrt{ax^n+c}+\sqrt{c}}\right)$; $\frac{1}{x\sqrt{ax^n+c}}$ (Assume $c>0$.)

31. $\frac{2}{n\sqrt{-c}}\,\mathrm{arcsec}\left(\sqrt{\frac{ax^n}{-c}}\right)$; $\frac{1}{x\sqrt{ax^n+c}}$ (Assume $c<0$.)

32. $\sqrt{ax^2+c}+\sqrt{c}\ln\left(\frac{\sqrt{ax^2+c}-\sqrt{c}}{x}\right)$; $\frac{\sqrt{ax^2+c}}{x}$ (Assume $c>0$.)

33. $\sqrt{ax^2+c}-\sqrt{-c}\arctan\left(\frac{\sqrt{ax^2+c}}{\sqrt{-c}}\right)$; $\frac{\sqrt{ax^2+c}}{x}$ (Assume $c<0$.)

34. $\frac{2}{\sqrt{4ac-b^2}}\arctan\left(\frac{2ax+b}{\sqrt{4ac-b^2}}\right)$; $\frac{1}{ax^2+bx+c}$ (Assume $b^2<4ac$.)

35. $\frac{-2}{2ax+b}$; $\frac{1}{ax^2+bx+c}$ (Assume $b^2=4ac$.)

36. $\frac{1}{\sqrt{b^2-4ac}}\ln\left(\frac{2ax+b-\sqrt{b^2-4ac}}{2ax+b+\sqrt{b^2-4ac}}\right)$; $\frac{1}{ax^2+bx+c}$ (Assume $b^2>4ac$. Use properties of ln before differentiating.)

37. $\frac{1}{2}\left((x-a)\sqrt{2ax-x^2}+a^2\arcsin\left(\frac{x-a}{a}\right)\right)$; $\sqrt{2ax-x^2}$

38. $\arccos\left(\frac{a-x}{a}\right)$; $\frac{1}{\sqrt{2ax-x^2}}$

39. $\arcsin(x)-\sqrt{1-x^2}$; $\sqrt{\frac{1+x}{1-x}}$

40. $2\arcsin\left(\sqrt{\frac{x-b}{a-b}}\right)$; $\frac{1}{\sqrt{x-b}\sqrt{x-a}}$

41. $\frac{1}{a}\ln\left(\tan\left(\frac{ax}{2}\right)\right)$; $\frac{1}{\sin(ax)}$

42. $\ln(\ln(ax))$; $\frac{1}{x\ln(ax)}$

43. $\frac{-1}{(n-1)(\ln(ax))^{n-1}}$; $\frac{1}{x(\ln(ax))^n}$

44. $x\arcsin(ax)+\frac{1}{a}\sqrt{1-a^2x^2}$; $\arcsin(ax)$

45. $x\left(\arcsin(ax)\right)^2-2x+\frac{2}{a}\sqrt{1-a^2x^2}\arcsin(ax)$; $\left(\arcsin(ax)\right)^2$

46. $\frac{1}{ab}\left(ax-\ln\left(b+ce^{ax}\right)\right)$; $\frac{1}{b+ce^{ax}}$

47. $\frac{1}{a\sqrt{bc}}\arctan\left(e^{ax}\sqrt{\frac{b}{c}}\right)$; $\frac{1}{be^ax+ce^{-ax}}$ (Assume b, $c>0$.)

48. $x\left(\ln(ax)\right)^2-2x\ln(ax)+2x$; $\ln^2(ax)=\left(\ln(ax)\right)^2$

49. $-\frac{1}{2}\ln\left(\frac{1+\cos(x)}{1-\cos(x)}\right)$; $\frac{1}{\sin(x)}=\csc(x)$

50. $\frac{1}{b^2}\left(a+bx-a\ln(a+bx)\right)$; $\frac{x}{ax+b}$ (Assume $a+bx>0$.)

51. $\frac{1}{b^3}\left(a+bx-2a\ln(a+bx)-\frac{a^2}{a+bx}\right)$; $\frac{x^2}{(a+bx)^2}$, (Assume $a+bx>0$.)

52. $\frac{1}{ab}\arctan\left(\frac{bx}{a}\right)$; $\frac{1}{a^2+b^2x^2}$

53. $\frac{x}{2a^2(a^2+x^2)}+\frac{1}{2a^2}\arctan\left(\frac{x}{a}\right)$; $\frac{1}{(a^2+x^2)^2}$

54. $\frac{1}{2a^2}\arctan\left(\frac{x^2}{a^2}\right)$; $\frac{x}{a^4+x^4}$

55. $\frac{2\sqrt{x}}{b} - 2\frac{a}{b^3}\arctan\left(\frac{b\sqrt{x}}{a}\right)$; $\frac{\sqrt{x}}{a^2+b^2x}$

56. $x\,\mathrm{arcsec}(ax) - \frac{1}{a}\ln\left(ax + \sqrt{a^2x^2-1}\right)$; $\mathrm{arcsec}(ax)$

57. $x\arctan(ax) - \frac{1}{2a}\ln\left(1+a^2x^2\right)$; $\arctan(ax)$

58. $x\arccos(ax) - \frac{1}{a}\sqrt{1-a^2x^2}$; $\arccos(ax)$

59. $\frac{x^2}{2}\arcsin(ax) - \frac{1}{4a^2}\arcsin(ax) + \frac{x}{2a}\sqrt{1-a^2x^2}$; $x\arcsin(ax)$

60. $x\left(\arcsin(ax)\right)^2 - 2x + \frac{2}{a}\sqrt{1-a^2x^2}\arcsin(ax)$; $\left(\arcsin(ax)\right)^2$

61. $\frac{1}{a^2}\cos(ax) + \frac{x}{a}\sin(ax)$; $x\cos(ax)$

62. $\frac{1}{a^3}e^{ax}\left(a^2x^2 - 2ax + 2\right)$; x^2e^{ax}

63. $\frac{1}{ab}\left(ax - \ln\left(b + ce^{ax}\right)\right)$; $\frac{1}{b+ce^{ax}}$

64. $\frac{1}{a^2+b^2}e^{ax}\left(a\sin(bx) - b\cos(bx)\right)$; $e^{ax}\sin(bx)$

65. $\ln\left(\sec(x)+\tan(x)\right)$; $\sec(x)$

66. Find $D(\ln^3(x))$

(a) By the chain rule.

(b) By writing $\ln^3(x)$ as $\ln(x)\cdot\ln(x)\cdot\ln(x)$.

Which method do you prefer? Why?

67. We have used the equation $\sec^2(x) = 1 + \tan^2(x)$.

(a) Derive it from the equation $\cos^2(x) + \sin^2(x) = 1$.

(b) Derive $\cos^2(x) + \sin^2(x) = 1$ from the Pythagorean Theorem.

68. Find two antiderivatives of

(a) $2x$

(b) x^2

(c) $1/x$

(d) $\sqrt{x}$

69. Find two antiderivatives of

(a) e^{3x}

(b) $\cos(x)$

(c) $\sin(x)$

(d) $1/(1+x^2)$

70. This problem provides some additional experience with the development of the formula $\log_b(x) = \log_b(e)\log_e(x)$. Let $b > 0$. Recall that $\log_b(a) = \frac{\log_e(a)}{\log_e(b)}$.

(a) Show that $\log_b(e) = 1/\log_e(b)$.

(b) Conclude that $\log_b(x) = \log_b(e)\log_e(x)$.

This was used in Example 1.

In Exercises 71 to 74 use the chain rule to obtain the derivative.

71. $(\arccos(x))' = \frac{-1}{\sqrt{1-x^2}}$

72. $(\operatorname{arcsec}(x))' = \frac{1}{x\sqrt{x^2-1}}$

73. $(\operatorname{arccot}(x))' = \frac{-1}{1+x^2}$

74. $(\operatorname{arccsc}(x))' = \frac{-1}{x\sqrt{x^2-1}}$

75. Verify that $D\left(2(\sqrt{x}-1)e^{\sqrt{x}}\right) = e^{\sqrt{x}}$.

76.

Sam: I say that $D(\log_b(x)) = \frac{1}{x\ln(b)}$. It's simple. Let $y = \log_b(x)$. That tells me $x = b^y$. I differentiate both sides of that, getting $1 = b^y(\ln(b))y'$. So $y' = \frac{1}{b^y\ln(b)} = \frac{1}{x\ln(b)}$.

Jane: Well, not so fast. I start with the equation $\log_b(x) = (\log_b(e))\ln(x)$. So $D(\log_b(x)) = \frac{\log_b(e)}{x}$.

Sam: Something is wrong. Where did you get that equation you started with?

Jane: Just take $\log_b$ of both sides of $x = e^{\ln(x)}$.

Sam: I hope this won't be on the next midterm.

Settle this argument.

We did not need the chain rule to find the derivatives of inverse functions. Instead, we could have taken a geometric approach, using the interpretation of the derivative of the slope of the tangent line. When we reflect the graph of f around the line $y = x$ to obtain the graph of f^{-1}, the reflection of the tangent line to the graph of f with slope m is the tangent line to the graph of f^{-1} with slope $1/m$. (See Section 1.1.) Exercises 77 to 81 use this approach to develop formulas obtained in this section.

77. Let $f(x) = \ln(x)$. The slope of the graph of $y = \ln(x)$ at $(a, \ln(a))$, $a > 0$, is the reciprocal of the slope of the graph of $y = e^x$ at $(\ln(a), a)$. Use this to show that the slope of the graph of $y = \ln(x)$ when $x = a$ is $1/a$.

In Exercises 78 to 81 use the technique illustrated in Exercise 77 to differentiate the function.

78. $f(x) = \arctan(x)$

79. $f(x) = \arcsin(x)$

80. $f(x) = \operatorname{arcsec}(x)$

81. $f(x) = \arccos(x)$

82.

(a) Evaluate $\lim_{x\to\infty} \frac{1}{1+x^2}$ and $\lim_{x\to-\infty} \frac{1}{1+x^2}$.

(b) What do these results tell you about the graph of the arctangent function?

83.

Sam: I can get the formula for $(fg)'$ real easy if I assume fg is differentiable when f and g are.

Jane: How?

Sam: Start with $\ln(fg) = \ln(f) + \ln(g)$, which is OK if $f(x)$ and $g(x)$ are positive. Then differentiate like mad, using the chain rule:

$$\frac{1}{fg}(fg)' = \frac{f'}{f} + \frac{g'}{g}.$$

Jane: So?

Sam: Then solve for $(fg)'$ and out pops $(fg)' = fg' + gf'$.

Jane: I wonder why the book used all those Δs instead.

Why didn't the book use Sam's approach?

84. Use the assumptions and methods in Exercise 83 to find $D(f/g)$.

85.

Sam: In Exercise 83 I assumed that fg is differentiable if f and g are. I can get around that by using the differentiability of exp and ln.

Jane: How so?

Sam: Again I assume $f(x)$ and $g(x)$ are positive and I write fg as $\exp(\ln(fg))$.

Jane: So?

Sam: But $\ln(fg) = \ln(f) + \ln(g)$, and that does it.

Jane: I'm lost.

Sam: Well, $fg = \exp(\ln(f) + \ln(g))$ and use the chain rule. It's good for more than grinding out derivatives. In fact, when you differentiate both sides of my equation, you get that fg is differentiable and $(fg)'$ is $f'g + fg'$.

Jane: Why wouldn't the authors use this approach?

Sam: It would make things too easy and reveal that calculus is all about e, exponentials, and logarithms. I peeked at Chapter 12 and saw that you can even get sine and cosine out of e^x.

Is Sam's argument correct?

3.6 Antiderivatives and Slope Fields

So far in this chapter we have started with a function and found its derivative. In this section we will go in the opposite direction: given a function f, we will be interested in finding a function F whose derivative is f. Why? Because going from the derivative back to the function plays a central role in **integral calculus**, as we will see in Chapter 5. Chapter 6 describes several ways to find antiderivatives.

Some Antiderivatives

EXAMPLE 1 Find an antiderivative of x^6.
SOLUTION When we differentiate x^a we get ax^{a-1}. The exponent in the derivative, $a-1$, is one less than the original exponent, a. So we expect an antiderivative of x^6 to involve x^7.

Because $(x^7)' = 7x^6$, x^7 is an antiderivative of $7x^6$, not of x^6. We want to get rid of that coefficient 7 in front of x^6. If we divide x^7 by 7 we have

$$\begin{aligned} \left(\frac{x^7}{7}\right)' &= \frac{7x^6}{7} && \left(\left(\frac{f}{C}\right)' = \frac{f'}{C} \right) \\ &= x^6 && (\text{ canceling the 7s }). \end{aligned}$$

We conclude that $\frac{1}{7}x^7$ is an antiderivative of x^6.

However, $\frac{1}{7}x^7$ is not the only antiderivative of x^6. For instance,

$$\left(\frac{1}{7}x^7 + 2012\right)' = \frac{1}{7}7x^6 + 0 = x^6.$$

A constant added to any antiderivative of a function f gives another antiderivative of f.

We can add any constant to $\frac{1}{7}x^7$ and the result is always an antiderivative of x^6. ◇

As Example 1 suggests, if $F(x)$ is an antiderivative of $f(x)$ so is $F(x) + C$ for any constant C.

The reasoning in this example suggests that $\frac{1}{a+1}x^{a+1}$ is an antiderivative of x^a. This formula is meaningless when $a + 1 = 0$. We have to expect a different formula for antiderivatives of $x^{-1} = \frac{1}{x}$. In Section 3.5 we saw that $(\ln(x))' = 1/x$. That's one reason the function $\ln(x)$ is so important: it provides an antiderivative for $1/x$.

Antiderivatives of x^a

For any number a, except -1, the antiderivatives of x^a are

$$\frac{1}{a+1}x^{a+1} + C \quad \text{for any constant } C.$$

The antiderivatives of $x^{-1} = \frac{1}{x}$ are, when $x > 0$,

$$\ln(x) + C \quad \text{for any constant } C.$$

Every time you compute a derivative, you are also finding an antiderivative. For instance, since $D(\sin(x)) = \cos(x)$, $\sin(x)$ is an antiderivative of $\cos(x)$. So is $\sin(x) + C$ for any constant C. There are tables of antiderivatives that go on for hundreds of pages. Here is a small table with entries corresponding to the derivatives that we have found so far.

Search Google for "antiderivative table".

Function (f)	Antiderivative (F)	Comment
x^a	$\frac{1}{a+1}x^{a+1}$	for $a \neq -1$
$x^{-1} = \frac{1}{x}$	$\ln(x)$	
e^x	e^x	
$\cos(x)$	$\sin(x)$	
$\sin(x)$	$-\cos(x)$	
$\sec^2(x)$	$\tan(x)$	Example 8 in Section 3.3
$\sec(x)\tan(x)$	$\sec(x)$	Example 11 in Section 3.3
$\frac{1}{\sqrt{1-x^2}}$	$\arcsin(x)$	Section 3.4
$\frac{1}{1+x^2}$	$\arctan(x)$	Section 3.4

Table 3.6.1 Table of antiderivatives ($F' = f$).

An **elementary function** is a function that can be expressed in terms of polynomials, powers, trigonometric functions, exponentials, logarithms, and the functions obtained from them by algebra and by composition of functions. The derivative of an elementary function is elementary. We might expect that every elementary function would have an elementary antiderivative.

In 1833 Joseph Liouville proved that there are elementary functions that do not have elementary antiderivatives. Here are five examples of such functions:

Joseph Liouville (1809–1882)

e^{-x^2} is important in statisticians' **bell curve**

$$e^{x^2} \qquad \frac{\sin(x)}{x} \qquad x\tan(x) \qquad \sqrt{x}\sqrt[3]{1+x} \qquad \sqrt[4]{1+x^2}$$

There are two types of elementary functions: **algebraic** and **transcendental**. Algebraic functions, defined in Section 3.5, consist of polynomials, quotients of polynomials (the rational functions), and functions that can be built up by

The four operations of algebra are $+$, $-$, $\times$ and $/$.

the four operations of algebra and taking roots. For instance, $\dfrac{\sqrt{x+\sqrt[3]{x}}+x^2}{(1+2x)^5}$ is algebraic while functions such as $\sin(x)$ and 2^x are not algebraic. These functions are called **transcendental**.

It is difficult to tell whether a given elementary function has an elementary antiderivative. For instance, $x\sin(x)$ does, namely $-x\cos(x)+\sin(x)$, as may be checked, but $x\tan(x)$ does not. The function e^{x^2} does not, as mentioned earlier. However, $e^{\sqrt{x}}$, which looks more complicated, has an elementary antiderivative. (See Exercise 75.)

The importance of antiderivatives will be revealed in Chapter 5. Techniques for finding them are developed in Chapter 8.

Picturing Antiderivatives

If it is not possible to find an explicit formula for the antiderivative of many elementary functions, why do we believe that they have antiderivatives? The following shows why.

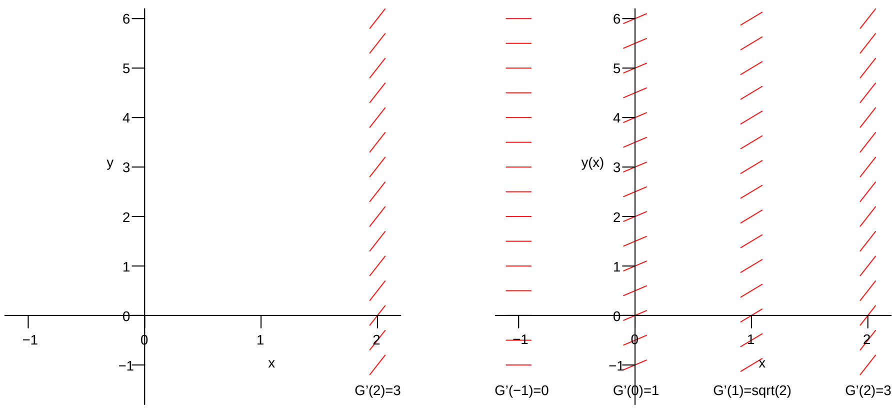
The **slope field** for a function $f(x)$ is made of short line segments with slope $f(x)$ at a few points whose x-coordinate is x. By drawing a slope field you can convince yourself that an antiderivative exists, and see the shape of its graph.

EXAMPLE 2 Imagine that we are looking for an antiderivative $F(x)$ of $\sqrt{1+x^3}$. We want $F'(x)$ to be $\sqrt{1+x^3}$. Or, to put it geometrically, we want the slope of the curve $y = F(x)$ to be $\sqrt{1+x^3}$. For instance, when $x = 2$, we want the slope to be $\sqrt{1+x^3} = 3$. We do not know what $F(2)$ is, but at least we can draw a short piece of the tangent line at all points for which $x = 2$: they all have slope 3. (See Figure 3.6.1(a).) When $x = 1$, $\sqrt{1+x^3} = \sqrt{2} \approx 1.4$. So we draw short lines with slope $\sqrt{2}$ on the vertical line $x = 1$. When $x = 0$, $\sqrt{1+x^3} = 1$: the tangent lines for $x = 0$ all have slope 1. When $x = -1$, the slopes are $\sqrt{1+x^3} = 0$ so the tangent lines are all horizontal. (See Figure 3.6.1(b).)

For a sample of available resources, search Google for "calculus slope field plot".

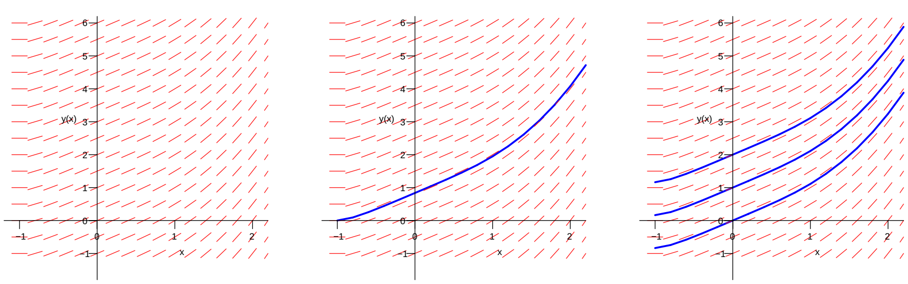
The plot of a slope field is most commonly made with the aid of specialized software on a graphing calculator or computer. A typical slope field, showing more segments of tangent lines than we have the patience to draw by hand, is in Figure 3.6.2(a) which shows a computer-generated direction field for $f(x) = \sqrt{1+x^3}$ that has many more segments of tangent lines than Figure 3.6.1(a).

We can imagine the curves that follow the slope field for $f(x) = \sqrt{1+x^3}$. Start at a point, say $(-1,0)$. There the slope is $F'(-1) = f(-1) = 0$, and the curve starts moving horizontally to the right. As soon as the curve leaves this initial point the slope, as given by $F'(x) = f(x)$, becomes slightly positive. This pushes the curve upward. The slope continues to increase as x increases.



(a) (b)

Figure 3.6.1 Constructing the slope field for $f(x) = \sqrt{1 + x^3}$. (a) For $x = 2$ all slopes are $f(2) = 3$. (b) For $x = -1$ all slopes are $f(1) = 0$, for $x = 0$, the slopes are $f(0) = 1$, and for $x = 1$ all slopes are $f(1) = \sqrt{2}$.



(a) (b) (c)

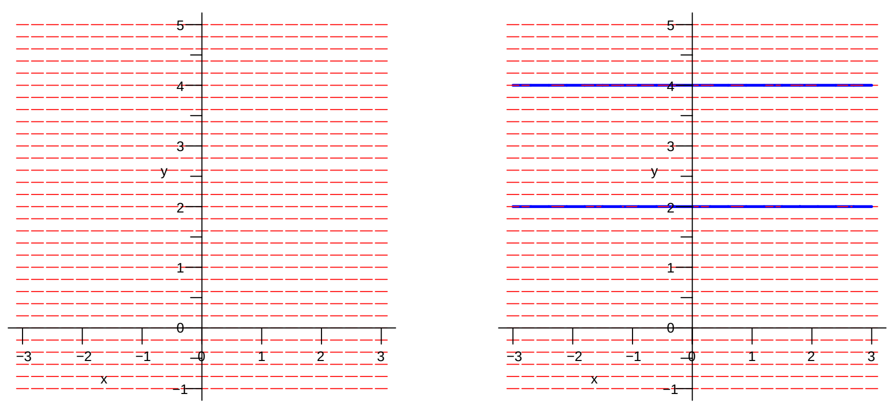
Figure 3.6.2 (a) Slope field for $f(x) = \sqrt{1 + x^3}$. (b) The antiderivative of $f(x)$ that passes through $(-1, 0)$. (c) Three more antiderivatives of $f(x)$, passing through $(0, 0)$, $(0, 1)$, and $(0, 2)$.

The curve in Figure 3.6.2(b) is the graph of the antiderivative of $f(x) = \sqrt{1+x^3}$ that equals 0 when x is -1.

Starting at a different initial point will produce a different antiderivative. Three antiderivatives are shown in Figure 3.6.2(c). Many other antiderivatives for $f(x) = \sqrt{1+x^3}$ are visible in the slope field. None is elementary. ◇

Example 2 suggests that different antiderivatives of a function differ by a constant: the graph of one is the graph of the other raised or lowered by their constant difference. The next example suggests that the constant functions are the only antiderivatives of the zero function. Both suggestions are correct, as will be shown in Section 3.7.

EXAMPLE 3 Draw the slope field for $\frac{dy}{dx} = 0$.
SOLUTION Since the slope is 0 everywhere, each of the tangent lines is represented by a horizontal line segment, as in Figure 3.6.3(a). In Figure 3.6.3(b)



(a) (b)

Figure 3.6.3

two possible antiderivatives of 0 are shown, namely the constant functions $f(x) = 2$ and $g(x) = 4$. ◇

We will assume from now on that

Every antiderivative of the zero function on an interval is constant. That is, if $f'(x) = 0$ for all x in an interval, then $f(x) = C$ for some constant C. Two antiderivatives of a function on an interval differ by a constant. That is, if $F'(x) = G'(x)$ for all x in an interval, then $F(x) = G(x) + C$ for some constant C.

These results will be established using the definitions and theorems of calculus in Section 3.7.

How computers find antiderivatives

There are algorithms implemented in software on computers, hand-held devices, and calculators that determine if a given elementary function has an elementary antiderivative. The most well-known is the **Risch algorithm**, developed in 1968, based on differential equations and abstract algebra. A web search for "risch antiderivative elementary symbolic" produces links related to the Risch algorithm.

Summary

If $F' = f$, then F is an antiderivative of f; so is $F + C$ for any constant C.

We introduced the notion of an *elementary function*. Such a function is built up from polynomials, logarithms, exponentials, and the trigonometric functions by algebraic operations and the most important operation, composition. While the derivative of an elementary function is elementary, its antiderivative does not need to be. Each elementary function is either algebraic or transcendental.

We showed how a slope field can help analyze an antiderivative even though we may not know a formula for it. Slope fields will be used later for other purposes.

EXERCISES for Section 3.6

1.

(a) Verify that $-x\cos(x)+\sin(x)$ is an antiderivative of $x\sin(x)$.

(b) Spend at least one minute and at most ten minutes trying to find an antiderivative of $x\tan(x)$.

In Exercises 2 to 11 give two antiderivatives for each function.

2. x^3

3. x^4

4. x^{-2}

5. $\frac{1}{x^3}$

6. $\sqrt[3]{x}$

7. $\frac{2}{x}$

8. $\sec(x)\tan(x)$

9. $\sin(x)$

10. e^{-x}

11. $\sin(2x)$

In Exercises 12 to 20

(a) Draw the slope field for the given derivative,

(b) Use it to draw the graphs of two possible antiderivatives $F(x)$.

12. $F'(x) = 2$

13. $F'(x) = x$

14. $F'(x) = \frac{-x}{2}$

15. $F'(x) = \frac{1}{x}$, $x > 0$

16. $F'(x) = \cos(x)$

17. $F'(x) = \sqrt{x}$

18. $F'(x) = e^{-x}$, $x > 0$

19. $F'(x) = 1/x^2$, $x \neq 0$

20. $F'(x) = 1/(x-1)$, $x \neq 1$

In Exercises 21 to 30 use differentiation to check that the first function is an antiderivative of the second function.

21. $2x\sin(x) - (x^2-2)\cos(x)$; $x^2\sin(x)$

22. $(4x^3 - 24x)\sin(x) - (x^4 - 12x^2 + 24)\cos(x)$; $x^4 \sin(x)$

23. $\frac{-1}{2x^2}$; $\frac{1}{x^3}$

24. $\frac{-2}{\sqrt{x}}$; $\frac{1}{x^{3/2}}$

25. $(x-1)e^x$; xe^x

26. $(x^2 - 2x + 2)e^x$; x^2e^x

27. $\frac{1}{2}e^u(\sin(u) - \cos(u))$; $e^u \sin(u)$

28. $\frac{1}{2}e^u(\sin(u) + \cos(u))$; $e^u \cos(u)$

29. $\frac{x}{2} - \frac{\sin(x)\cos(x)}{2}$; $\sin^2(x)$

30. $2x\cos(x) - (x^2 - 2)\sin(x)$; $x^2\cos(x)$

31.

(a) Draw the slope field for $\frac{dy}{dx} = e^{-x^2}$.

(b) Draw the graph of the antiderivative of e^{-x^2} that passes through the point $(0, 1)$.

32.

(a) Draw the slope field for $\frac{dy}{dx} = \frac{\sin(x)}{x}$ when $x \neq 0$. For the slopes when $x = 0$, use 1, which is $\lim_{x\to 0} \frac{\sin(x)}{x}$.

(b) What is the slope for any point on the y-axis?

(c) Draw the graph of the antiderivative of $f(x)$ that passes through the point $(0, 1)$.

33. A table of antiderivatives lists two antiderivatives of $\frac{1}{x^2(a+bx)}$, where a and b are constants, namely

$$\frac{-1}{a^2}\left(\frac{a+bx}{x} - b\ln\left(\frac{a+bx}{x}\right)\right) \quad \text{and} \quad -\frac{1}{ax} + \frac{b}{a^2}\ln\left(\frac{a+bx}{x}\right).$$

Assume $\frac{a+bx}{x} > 0$.

(a) By differentiating the expressions, show that both are correct.

(b) Show that they differ by a constant by finding their difference.

34. If $F(x)$ is an antiderivative of $f(x)$, find a function that is an antiderivative of

(a) $g(x) = 2f(x)$

(b) $h(x) = f(2x)$

35.

(a) Draw the slope field for $dy/dx = -y$.

(b) Draw the graph of the function $y = F(x)$ such that $F(0) = 1$ and $F'(X) = -y$.

(c) What do you think $\lim_{x\to\infty} F(x)$ is?

3.7 Motion and the Second Derivative

In a drag race Melanie Troxel reached a speed of 324 miles per hour, which is about 475 feet per second, in a mere 4.539 seconds. By comparison, a 1968 Fiat 850 Idromatic could reach a speed of 60 miles per hour in 25 seconds and a 1997 Porsche 911 Turbo S in 3.6 seconds.

Numerical acceleration data can be found with a web search for "automobile acceleration."

Since Troxel increased her speed from 0 feet per second to 475 feet per second in 4.539 seconds her speed was increasing at the rate of $\frac{475}{4.539} \approx 105$ feet per second per second, assuming she kept the motor at constant power throughout the time interval. That acceleration is more than three times the acceleration due to gravity at sea level (32 feet per second per second). Ms. Troxel must have felt quite a force as her seat pressed against her back.

This brings us to the definition of **acceleration** and an introduction to higher derivatives.

Acceleration

Velocity is the rate at which the position of an object moving on a line changes. The rate at which velocity changes is called **acceleration**, denoted a. Thus if $y = f(t)$ denotes position on the y-axis at time t, then the derivative $\frac{dy}{dt}$ equals the velocity, and the derivative of the derivative equals the acceleration. That is,

The sign of the velocity indicates direction. **Speed**, the absolute value of velocity, does not indicate direction.

$$v = \frac{dy}{dt} \quad \text{and} \quad a = \frac{dv}{dt} = \frac{d}{dt}\left(\frac{dy}{dt}\right).$$

The derivative of the derivative of a function $y = f(x)$ is called the **second derivative**. It is denoted in many different ways, including

$$\frac{d^2y}{dx^2}, \quad D^2y, \quad y'', \quad f'', \quad D^2f, \quad f^{(2)}, \quad \text{and} \quad \frac{d^2f}{dx^2}.$$

If $y = f(t)$, where t denotes time, the first and second derivatives dy/dt and d^2y/dt^2 are sometimes denoted $\dot{y}$ and $\ddot{y}$, respectively.

For instance, if $y = x^3$,

$$\frac{dy}{dx} = 3x^2 \quad \text{and} \quad \frac{d^2y}{dx^2} = 6x.$$

Other ways of denoting the second derivative are

$$D^2(x^3) = 6x, \; \frac{d^2(x^3)}{dx^2} = 6x, \quad \text{and} \quad (x^3)'' = 6x.$$

The table in the margin lists the function and its first and second derivatives.

y	$\frac{dy}{dx}$	$\frac{d^2y}{dx^2}$
x^3	$3x^2$	$6x$
$\frac{1}{x}$	$\frac{-1}{x^2}$	$\frac{2}{x^3}$
$\sin(5x)$	$5\cos(5x)$	$-25\sin(5x)$

Most functions met in applications of calculus can be differentiated repeatedly in the sense that Df exists, the derivative of Df, namely, D^2f, exists, the derivative of D^2f exists, and so on.

The derivative of the second derivative is called the **third derivative** and is denoted many ways, such as

$$\frac{d^3y}{dx^3}, \quad D^3y, \quad y''', \quad f''', \quad f^{(3)}, \quad \text{and} \quad \frac{d^3f}{dx^3}.$$

The fourth derivative is defined similarly, as the derivative of the third derivative. In the same way we can define the n^{th} derivative for any positive integer n and denote this by such symbols as

$$\frac{d^ny}{dx^n}, \quad D^ny, \quad f^{(n)}, \quad \text{and} \quad \frac{d^nf}{dx^n}.$$

It is read as "the n^{th} derivative of y with respect to x." For instance, if $f(x) = 2x^3 + x^2 - x + 5$, we have

$$\begin{aligned} f^{(1)}(x) &= 6x^2 + 2x - 1 \\ f^{(2)}(x) &= 12x + 2 \\ f^{(3)}(x) &= 12 \\ f^{(4)}(x) &= 0 \\ f^{(n)}(x) &= 0 \qquad\qquad \text{for } n \geq 5. \end{aligned}$$

EXAMPLE 1 Find $D^n(e^{-2x})$ for each positive integer n.
SOLUTION

$$\begin{aligned} D^1\left(e^{-2x}\right) &= D\left(e^{-2x}\right) = -2e^{-2x} \\ D^2\left(e^{-2x}\right) &= D\left(-2e^{-2x}\right) = (-2)^2e^{-2x} \\ D^3\left(e^{-2x}\right) &= D\left((-2)^2e^{-2x}\right) = (-2)^3e^{-2x} \end{aligned}$$

At each differentiation another (-2) becomes part of the coefficient. Thus

$$D^n\left(e^{-2x}\right) = (-2)^ne^{-2x}.$$

This can also be written

$$D^n\left(e^{-2x}\right) = (-1)^n2^ne^{-2x},$$

because the power $(-1)^n$ records a plus if n is even and a minus if n is odd. $\diamond$

Finding Velocity and Acceleration from Position

EXAMPLE 2 A falling rock drops $16t^2$ feet in the first t seconds. Find its velocity and acceleration.

SOLUTION Place the y-axis in the usual position, with 0 at the beginning of the fall and the part with positive values above 0, as in Figure 3.7.1. At time t the object has the y coordinate

$$y = -16t^2.$$

The velocity is $v = (-16t^2)' = -32t$ feet per second, and the acceleration is $a = (-32t)' = -32$ feet per second per second. The velocity changes at a constant rate. That is, the acceleration is constant. ◇

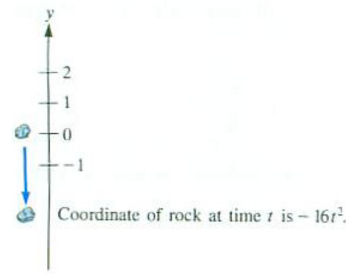


Figure 3.7.1

Finding Position from Velocity and Acceleration

To calculate the position of a moving object it is enough to know the object's acceleration, its initial position, and its initial velocity. This will be demonstrated in the next two examples in the special case that the acceleration is constant. In the first example, the acceleration is 0.

EXAMPLE 3 In the simplest motion, no forces act on a moving particle, hence its acceleration is 0. Assume that a particle is moving on the x-axis and no forces act on it. Let its location at time t seconds be $x = f(t)$ feet. See Figure 3.7.2. If at time $t = 0$, $x = 3$ feet and the velocity is 5 feet per second, determine $f(t)$.

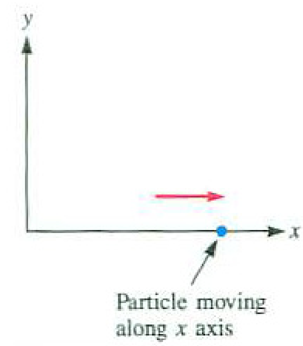


Figure 3.7.2

SOLUTION The assumption that no force operates on the particle tells us that there is no acceleration: $d^2x/dt^2 = 0$. Call the velocity v. Then

$$\frac{dv}{dt} = \frac{d}{dt}\left(\frac{dx}{dt}\right) = \frac{d^2x}{dt^2} = 0$$

That is, v is a function of time whose derivative is 0. At the end of Section 3.6 we saw that constant functions are the antiderivatives of 0. Thus, v must be constant:

$$v(t) = C \qquad \text{for some constant } C.$$

Since $v(0) = 5$, the constant C must be 5.

We know that the derivative of the position x is the velocity v. Hence

$$\frac{dx}{dt} = 5.$$

Similar reasoning tells us that $x = f(t)$ has the form

$$x = 5t + K \qquad \text{for some constant } K.$$

When $t = 0$, $x = 3$, so $K = 3$. Thus at time t seconds the particle is at $x = 5t + 3$ feet. ◇

In the next example the acceleration is constant, but not zero.

EXAMPLE 4 A ball is thrown straight up, with an initial speed of 64 feet per second, from a cliff 96 feet above a beach. Where is the ball t seconds later? When does it reach its maximum height? How high above the beach does the ball rise? When does the ball hit the beach? Assume that there is no air resistance and that the acceleration due to gravity is constant.

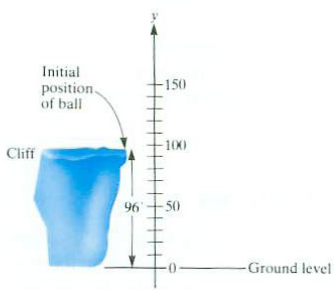


Figure 3.7.3

SOLUTION Introduce a vertical coordinate axis to describe the position of the ball. It is more natural to call it the y-axis, and so the velocity is dy/dt and acceleration is d^2y/dt^2. Place the origin at ground level and let the positive part of the y-axis be above the ground, as in Figure 3.7.3. At time $t = 0$, the velocity dy/dt is 64, since the ball is thrown up at a speed of 64 feet per second. As time increases, dy/dt decreases from 64 to 0 (when the ball reaches the top of its path and begins its descent) and continues to decrease through larger and larger negative values as the ball falls to the ground. Since v is decreasing, the acceleration dv/dt is negative. The constant value of dv/dt, gravitational acceleration, is approximately -32 feet per second per second.

If it had been thrown down dy/dt would be -64.

From the equation

$$a = \frac{dv}{dt} = -32$$

it follows that

Velocity is an antiderivative of acceleration.

$$v = -32t + C,$$

where C is some constant. To find C, we use the fact that $v = 64$ when $t = 0$ so

$$64 = -32 \cdot 0 + C$$

and $C = 64$. Hence $v = -32t + 64$ for any time t until the ball hits the beach.

So we have

$$\frac{dy}{dt} = v = -32t + 64.$$

Since the position function y is an antiderivative of the velocity, $-32t + 64$, we have

$$y(t) = -16t^2 + 64t + K,$$

where K is a constant. To find K, recall that $y = 96$ when $t = 0$. Thus

$$96 = -16 \cdot 0^2 + 64 \cdot 0 + K,$$

and $K = 96$.

We have obtained a complete description of the position of the ball at any time t while it is in the air:

$$y = -16t^2 + 64t + 96.$$

This, together with $v = -32t + 64$, provides answers to many questions about the ball's flight. (As a check, note that when $t = 0$, $y = 96$, the initial height.)

When does it reach its maximum height? When it is neither rising nor falling. That is, the velocity is neither positive nor negative, and so must be 0. The velocity is zero when $-32t + 64 = 0$, which is when $t = 2$ seconds.

How high above the ground does the ball rise? Compute y when $t = 2$. This gives $-16 \cdot 2^2 + 64 \cdot 2 + 96 = 160$ feet. (See Figure 3.7.4.)

When does the ball hit the beach? When $y = 0$. Find t such that

$$y = -16t^2 + 64t + 96 = 0.$$

Division by -16 yields $t^2 - 4t - 6 = 0$, which has the solutions

$$t = \frac{4 \pm \sqrt{16 + 24}}{2} = 2 \pm \sqrt{10}.$$

Since $2 - \sqrt{10}$ is negative and the ball cannot hit the beach before it is thrown, the only physically meaningful solution is $2 + \sqrt{10}$. The ball lands $2 + \sqrt{10}$ seconds after it is thrown, so it is in the air for about 5.2 seconds.

The graphs of position, velocity, and acceleration as functions of time provide another perspective on the motion of the ball, as shown in Figure 3.7.4.

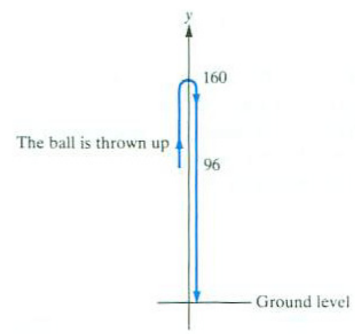
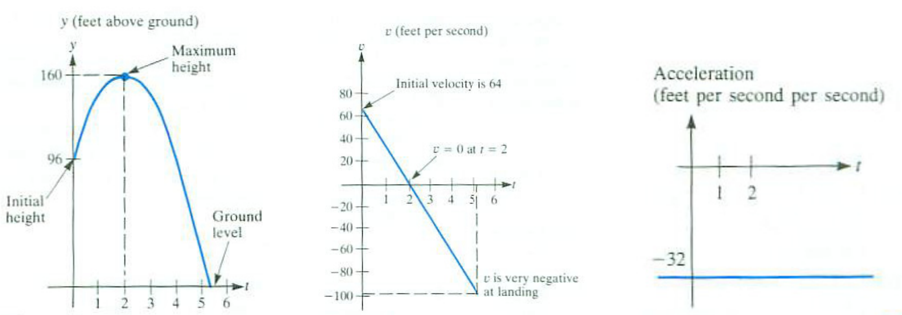


Figure 3.7.4



(a) (b) (c)

Figure 3.7.5 (a) position, (b) velocity, and (c) acceleration for the object in Example 4.

◇

Reasoning like that in Examples 3 and 4 establishes the following description of motion in all cases where the acceleration is constant.

Motion Under Constant Acceleration

Assume that a particle moving on the y-axis has a constant acceleration a. Assume that at time $t = 0$ it has the initial velocity v_0 and has the initial y-coordinate y_0. Then at any time $t \geq 0$ its y-coordinate is

$$y = \frac{a}{2}t^2 + v_0 t + y_0.$$

In Example 3, $a = 0$, $v_0 = 5$, and $y_0 = 3$ and in Example 4, $a = -32$ $v_0 = 64$, and $y_0 = 96$. The data must be given in consistent units, for instance, all in meters or all in feet.

Summary

We defined the higher derivatives of a function. They are obtained by repeatedly differentiating. The second derivative is the derivative of the derivative, the third derivative is the derivative of the second derivative, and so on. The first and second derivatives, $D(f)$ and $D^2(f)$, are used in many applications. We used them to analyze motion under constant acceleration.

EXERCISES for Section 3.7

In Exercises 1 to 16 find the first and second derivatives of the functions.

1. $y = 2x + 3$

2. $y = e^{-x^3}$

3. $y = x^5$

4. $y = \ln(6x + 1)$

5. $y = \sin(\pi x)$

6. $y = 4x^3 - x^2 + x$

7. $y = \frac{x}{x+1}$

8. $y = \frac{x^2}{x-1}$

9. $y = x\cos(x^2)$

10. $y = \frac{x}{\tan(3x)}$

11. $y = (x - 2)^4$

12. $y = (x + 1)^3$

13. $y = e^{3x}$

14. $y = \tan(x^2)$

15. $y = x^2 \arctan(3x)$

16. $y = -\frac{\arcsin(2x)}{x^2}$

17. Use calculus, specifically derivatives, to restate the following report about the Leaning Tower of Pisa.

> Until 2001, the tower's angle from the vertical was increasing more rapidly.

(Let $\theta = f(t)$ be the angle of deviation from the vertical at time t. Incidently, the tower, begun in 1174 and completed in 1350, is 179 feet tall and leans about 14 feet from the vertical. Each day it leaned on the average, another $\frac{1}{5000}$ inch until the tower was propped up in 2001.)

Exercises 18 to 20 concern Example 4.

18.

(a) How long after the ball in Example 4 is thrown does it pass by the top of the hill?

(b) What are its speed and velocity at this instant?

19. Suppose the ball in Example 4 had simply been dropped from the cliff. Find the position y as a function of time. How long would it take the ball to reach the beach?

20. In view of the result of Exercise 19, provide a physical interpretation of the three terms on the right-hand side of the formula $y = -16t^2 + 64t + 96$.

21. At time $t = 0$ a particle is at $y = 3$ feet and has a velocity of -3 feet per second; it has a constant acceleration of 6 feet per second per second. Find its position at any time t.

22. At time $t = 0$ a particle is at $y = 10$ feet and has a velocity of 8 feet per second and has a constant acceleration of -8 feet per second per second.

(a) Find its position at any time t.

(b) What is its maximum y coordinate?

23. At time $t = 0$ a particle is at $y = 0$ feet and has a velocity of 0 feet per second. Find its position at any time t if its acceleration is always -32 feet per second per second.

24. At time $t = 0$ a particle is at $y = -4$ feet and has a velocity of 6 feet per second and it has a constant acceleration of -32 feet per second per second.

(a) Find its position at any time t.

(b) What is its largest y coordinate.

In Exercises 25 to 34 find the given derivatives.

25. $D^3\left(5x^2 - 2x + 7\right)$

26. $D^4\left(\sin(2x)\right)$

27. $D^n\left(e^x\right)$

28. $D(\sin(x))$, $D^2(\sin(x))$, $D^3(\sin(x))$, and $D^4(\sin(x))$

29. $D(\cos(x))$, $D^2(\cos(x))$, $D^3(\cos(x))$, and $D^4(\cos(x))$

30. $D(\ln(x))$, $D^2(\ln(x))$, $D^3(\ln(x))$, and $D^4(\ln(x))$

31. $D^4(x^4)$ and $D^5(x^4)$

32. $D^{200}(\sin(x))$

33. $D^{200}(e^x)$

34. $D^2(5^x)$

35. Find all functions f such that $D^2(f) = 0$ for all x.

36. Find all functions f such that $D^3(f) = 0$ for all x.

37. A jetliner begins its descent 120 miles from an airport. Its velocity when the descent begins is 500 miles per hour and its landing velocity is 180 miles per hour. Assuming a constant deceleration, how long does the descent take?

38. Let $y = f(t)$ describe the motion on the y-axis of an object whose acceleration has the constant value a. Show that

$$y = \frac{a}{2}t^2 + v_0 t + y_0,$$

where v_0 is the velocity when $t = 0$ and y_0 is the position when $t = 0$.

39. Which has the highest acceleration? Melanie Troxel's dragster, a 1997 Porsche 911 Turbo S, or an airplane being launched from an aircraft carrier? The plane reaches a velocity of 180 miles per hour in 2.5 seconds, within a distance of 300 feet. (Assume each acceleration is constant.)

40. Why do engineers call the third derivative of position with respect to time the **jerk**?

41. Give two functions f such that $D^2(f) = 9f$, neither a constant multiple of the other.

42. Give two functions f such that $D^2(f) = -4f$, neither a constant multiple of the other.

43. A car accelerates with constant acceleration from 0 (rest) to 60 miles per hour in 15 seconds. How far does it travel in this period? Be sure to do your computations either all in seconds, or all in hours. 60 miles per hour is 88 feet per second.

44. Show that a ball thrown straight up from the ground takes as long to rise as to fall back to its initial position. Disregard air resistance. How does the velocity with which it strikes the ground compare with its initial velocity? How do the initial and landing speeds compare?

3.8 Precise Definition of Limits at Infinity: $\lim_{x\to\infty} f(x)$ L

One day a teacher drew on the board the graph of $y = x/2 + \sin(x)$, shown in Figure 3.8.1. Then the class was asked whether they thought that

$$\lim_{x\to\infty} f(x) = \infty.$$

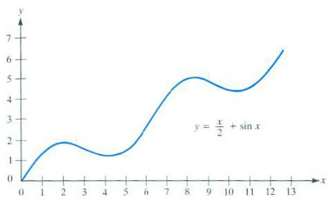


Figure 3.8.1

A third of the class voted “No” because “it keeps going up and down.” A third voted ”Yes” because ”the function tends to get very large as x increases.” A third didn’t vote. Such a variety of views on such a fundamental concept suggests that we need a more precise definition of a limit than the ones developed in Sections 2.2 and 2.3. (How would you vote?)

The definitions of the limits considered in Chapter 2 used such phrases as “x approaches a,” “$f(x)$ approaches a specific number,” “as x gets larger,” and “$f(x)$ becomes and remains arbitrarily large.” Such phrases, although appealing to the intuition and conveying the sense of a limit, are not precise. The definitions seem to suggest moving objects and call to mind the motion of a pencil point as it traces out the graph of a function.

The informal approach was adequate during the early development of calculus, from Leibniz and Newton in the seventeenth century through the Bernoullis, Euler, and Gauss in the eighteenth and early nineteenth centuries. By the mid-nineteenth century, mathematicians, facing more complicated functions and more difficult theorems, no longer could depend solely on intuition. They realized that glancing at a graph was no longer adequate to understand the behavior of functions, especially if theorems covering a broad class of functions were needed.

It was Weierstrass who developed, over the period 1841–1856, a way to define limits without any reference to motion or pencils tracing out graphs. His approach, on which he lectured after joining the faculty at the University of Berlin in 1859, has since been followed by pure and applied mathematicians throughout the world. Even an undergraduate advanced calculus course depends on Weierstrass’s approach.

In this section we examine how Weierstrass would define the limits at infinity:

$$\lim_{x\to\infty} f(x) = \infty \quad \text{and} \quad \lim_{x\to\infty} f(x) = L.$$

In the next section we consider limits at finite points:

$$\lim_{x\to a} f(x) = L.$$

The Precise Definition of $\lim_{x\to\infty} f(x) = \infty$

First we treat the case in which the limit is infinite, which includes the example that introduces this section. We had a definition of $\lim_{x\to\infty} f(x) = \infty$ in Section 2.2.

Informal definition of $\lim_{x\to\infty} f(x) = \infty$

1. $f(x)$ is defined for all x beyond some number.
2. As x gets large through positive values, $f(x)$ becomes and remains arbitrarily large and positive.

To take us part way to the precise definition, let us reword the informal definition, paraphrasing it in the following definition, which is still informal.

Reworded informal definition of $\lim_{x\to\infty} f(x) = \infty$

1. Assume that $f(x)$ is defined for all x greater than some number c.
2. If x is sufficiently large and positive, then $f(x)$ is necessarily large and positive.

The precise definition parallels the reworded definition.

DEFINITION *(Precise definition of* $\lim_{x\to\infty} f(x) = \infty$*)*

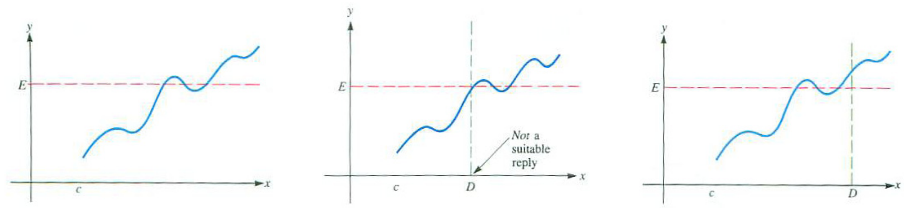
1. Assume the $f(x)$ is defined for all x greater than some number c.
2. For each number E there is a number D such that for all $x > D$ it is true that $f(x) > E$.

If those two conditions are met, then $\lim_{x\to\infty} f(x) = \infty$.

The challenge and reply approach to limits. Think of E as the "enemy" and D as the "defense."

Think of the number E as a challenge and D as the reply. The larger E is, the larger D must usually be. Only if a number D (which depends on E) can he found for every number E can we make the claim that $\lim_{x\to\infty} f(x) = \infty$. In other words, D could be expressed as a function of E. To picture the idea behind the precise definition, consider the graph in Figure 3.8.2(a) of a function f for which $\lim_{x\to\infty} f(x) = \infty$. For each possible choice of a horizontal line, say at height E, if you are far enough to the right on the graph of f, you stay above that horizontal line. That is, there is a number D such that if $x > D$, then $f(x) > E$.

The number D in Figure 3.8.2(b) is not a suitable reply. It is too small since there are some values of $x > D$ such that $f(x) \leq E$. The number D in



(a) (b) (c)

Figure 3.8.2

Figure 3.8.2(c) does fulfill the second part of the definition. For every value of $x > D$, $f(x) > E$.

Examples 1 and 2 illustrate how the precise definition is used.

EXAMPLE 1 Using the precise definition, show that $\lim_{x\to\infty} 2x = \infty$.
SOLUTION Let E be any positive number. We must show that there is a number D such that whenever $x > D$ it follows that $2x > E$. (For example, if $E = 100$, then $D = 50$ would do because if $x > 50$, then $2x > 100$.) The number D will depend on E. Our goal is to find a formula for D for any value of E.

The inequality $2x > E$ is equivalent to

$$x > \frac{E}{2}.$$

D depends on E

So if $x > E/2$, then $2x > E$. Choosing $D = E/2$ will suffice. To verify this, when $x > D$ $(= E/2)$, $2x > 2D = 2\frac{E}{2} = E$. This allows us to conclude that

$$\lim_{x\to\infty} 2x = \infty.$$

◇

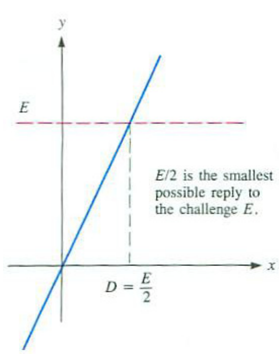


Figure 3.8.3

In Example 1 a formula was provided for a suitable D in terms of E, namely, $D = E/2$ (see Figure 3.8.3). When challenged with $E = 1000$, the response $D = 500$ suffices. Any larger value of D also is suitable. If $x > 600$, it is still the case that $2x > 1000$ (since $2x > 1200$). If one value of D is a satisfactory response to a given challenge E, then any larger value of D also is a satisfactory response.

Now that we have a precise definition of $\lim_{x\to\infty} f(x) = \infty$ we can settle the question, is $\lim_{x\to\infty}(x/2 + \sin(x)) = \infty$?

EXAMPLE 2 Using the precise definition, show that $\lim_{x\to\infty} \frac{x}{2} + \sin(x) = \infty$.
SOLUTION Let E be any number. We need to exhibit a number D, depending on E, such that $x > D$ implies

$$\frac{x}{2} + \sin(x) > E. \tag{3.8.1}$$

Now, $\sin(x) \geq -1$ for all x. So, if we can force

$$\frac{x}{2} + (-1) > E$$

then it will follow that

$$\frac{x}{2} + \sin(x) > E.$$

The smallest value of x that satisfies (3.8.1) can be found as follows:

$$\begin{array}{rcll} \frac{x}{2} & > & E+1 & \text{(add 1 to both sides of (3.8.1))} \\ x & > & 2(E+1). & \text{(multiply by a positive constant)} \end{array}$$

Thus $D = 2(E+1)$ will suffice. That is,

D depends on E

$$\text{If } x > 2(E+1), \text{ then } \frac{x}{2} + \sin(x) > E.$$

To verify this we check that $D = 2(E+1)$ is a satisfactory reply to E. Assume that $x > D = 2(E+1)$. Then

$$\frac{x}{2} > E + 1$$

and

$$\sin(x) \geq -1.$$

Adding the inequalities gives

If $a > b$ and $c \geq d$, then $a + c > b + d$.

$$\frac{x}{2} + \sin(x) > (E+1) + (-1)$$

or

$$\frac{x}{2} + \sin(x) > E,$$

which is inequality (3.8.1). Therefore we can conclude that

$$\lim_{x\to\infty} \left(\frac{x}{2} + \sin(x)\right) = \infty.$$

As x increases, the function does become and remain large, despite the small dips downward. ◇

The Precise Definition of $\lim_{x\to\infty} f(x) = L$

We defined $\lim_{x\to\infty} f(x) = L$ informally in Section 2.2.

L is a finite number.

Informal definition of $\lim_{x\to\infty} f(x) = L$

1. $f(x)$ is defined for all x beyond some number.
2. As x gets large through positive values, $f(x)$ approaches L.

Again we reword this definition before offering the precise definition.

Reworded informal definition of $\lim_{x\to\infty} f(x) = L$

1. Assume that $f(x)$ is defined for all x greater than some number c.
2. If x is sufficiently large, then $f(x)$ is necessarily near L.

Once again, the precise definition parallels the reworded definition. In order to make precise the phrase "$f(x)$ is necessarily near L," we shall use the absolute value of $f(x) - L$ to measure the distance from $f(x)$ to L. The following definition says that if x is large enough, then $|f(x) - L|$ is as small as we please.

DEFINITION *(Precise definition of* $\lim_{x\to\infty} f(x) = L$*)*

1. Assume that $f(x)$ is defined for all x beyond some number c.
2. For each positive number ϵ there is a number D such that for all $x > D$ it is true that

$$|f(x) - L| < \epsilon.$$

If both conditions are met, then $\lim_{x\to\infty} f(x) = L$.

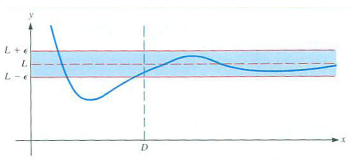


Figure 3.8.4

This definition can also be interpreted graphically. Draw two lines parallel to the x-axis, one of height $L+\epsilon$ and one of height $L-\epsilon$. They are the two edges of a band of width 2ϵ centered at $y = L$. Assume that for each positive ϵ, a number D can be found such that the part of the graph to the right of $x = D$ lies within the band. Then we say that as x approaches ∞, $f(x)$ approaches L and write

$$\lim_{x\to\infty} f(x) = L.$$

The positive number ϵ is the challenge, and D is a reply. The smaller ϵ is, the narrower the band is, and the larger D usually must be chosen. The geometric meaning of the precise definition of $\lim_{x\to\infty} f(x) = L$ is shown in Figure 3.8.4.

"ϵ" (epsilon) is the Greek letter corresponding to the English letter "e".

EXAMPLE 3 Use the precise definition of $\lim_{x\to\infty} f(x) = L$ to show that

$$\lim_{x\to\infty}\left(1 + \frac{1}{x}\right) = 1.$$

SOLUTION Here $f(x) = 1 + \frac{1}{x}$, which is defined for all $x \neq 0$. The number L is 1. We must show that for each positive number ϵ, however small, there is a number D such that, for all $x > D$,

$$\left|\left(1 + \frac{1}{x}\right) - 1\right| < \epsilon. \tag{3.8.2}$$

Inequality (3.8.2) reduces to

$$\left|\frac{1}{x}\right| < \epsilon.$$

Since we may consider only $x > 0$, it is equivalent to

$$\frac{1}{x} < \epsilon. \tag{3.8.3}$$

Multiplying inequality (3.8.3) by the positive number x yields the equivalent inequality

$$1 < x\epsilon. \tag{3.8.4}$$

Dividing inequality (3.8.4) by the positive number ϵ yields

$$\frac{1}{\epsilon} < x \quad \text{or} \quad x > \frac{1}{\epsilon}.$$

The steps are reversible, which shows that $D = 1/\epsilon$ is a suitable reply to the challenge ϵ. If $x > 1/\epsilon$, then

D depends on ϵ.

$$\left|\left(1 + \frac{1}{x}\right) - 1\right| < \epsilon.$$

That is, (3.8.2) is satisfied.

According to the precise definition of $\lim_{x\to\infty} f(x) = L$, we conclude that

$$\lim_{x\to\infty}\left(1 + \frac{1}{x}\right) = 1.$$

◇

The graph of $f(x) = 1 + \frac{1}{x}$, shown in Figure 3.8.5, reinforces the argument. It seems plausible that no matter how narrow a band someone may place around the line $y = 1$, it will always be possible to find a number D such that the part of the graph to the right of $x = D$ stays within it. In Figure 3.8.5 a typical band is shown shaded.

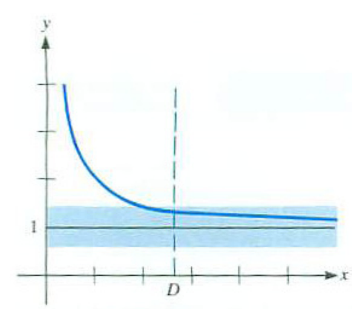


Figure 3.8.5

The precise definitions can also be used to show that some claim about an alleged limit is false. The next example illustrates this.

EXAMPLE 4 Show that the claim that $\lim_{x\to\infty}\sin(x) = 0$ is false.
SOLUTION To show that the claim is false, we must exhibit a challenge $\epsilon > 0$ for which no response D can be found. That is, we exhibit a positive number ϵ such that no D exists for which $|\sin(x) - 0| < \epsilon$ for all $x > D$.

Because $\sin(x) = 1$ whenever $x = \frac{\pi}{2} + 2n\pi$ for any integer n, there are arbitrarily large values of x for which $\sin(x) = 1$. This suggests how to exhibit an $\epsilon > 0$ for which no response D can be found. Pick ϵ to be a positive number less than or equal to 1. For instance, $\epsilon = 0.7$ will do.

For any number D there is always $x^* > D$ such that we have $\sin(x^*) = 1$. This means that $|\sin(x^*) - 0| = 1 > 0.7$. Hence no response can he found for $\epsilon = 0.7$. Thus the claim that $\lim_{x\to\infty}\sin(x) = 0$ is false. ◇

To conclude this section, we show how the precise definition of the limit can be used to obtain information about new limits.

EXAMPLE 5 Use the precise definition of $\lim_{x\to\infty} f(x) = L$ to show that if f and g are defined everywhere, $\lim_{x\to\infty} f(x) = 2$, and $\lim_{x\to\infty} g(x) = 3$, then $\lim_{x\to\infty}(f(x) + g(x)) = 5$.
SOLUTION The objective is to show that for each positive number ϵ, however small, there is a number D such that, for all $x > D$,

$$|(f(x) + g(x)) - 5| < \epsilon.$$

Because $|(f(x) + g(x)) - 5|$ can be written as $|(f(x) - 2) + (g(x) - 3))|$ it is no larger than $|f(x) - 2| + |g(x) - 3|$. If we can show that for all x sufficiently large both $|f(x) - 2| < \epsilon/2$ and $|g(x) - 3| < \epsilon/2$, then their sum will be no larger than $\epsilon/2 + \epsilon/2 = \epsilon$.

Here is how we can do this.

Because $\lim_{x\to\infty} f(x) = 2$ we know that for any $\epsilon > 0$ there exists a number D_1 with the property that $|f(x) - 2| < \epsilon/2$ for all $x > D_1$. (In this case $\epsilon/2$ is the challenge and D_1 is the response.) Likewise, because $\lim_{c\to\infty} g(x) = 3$ we know that for any $\epsilon > 0$ there exists a number D_2 with the property that $|g(x) - 2| < \epsilon/2$ for all $x > D_2$.

Let D be the larger of D_1 and D_2. For any x greater than D we know that

$$|f(x) + g(x) - 5| < |f(x) - 2| + |g(x) - 3| < \epsilon/2 + \epsilon/2 = \epsilon.$$

According to the precise definition of a limit at infinity, we conclude that

$$\lim_{x\to\infty}(f(x) + g(x)) = 2 + 3 = 5.$$

◇

Summary

We developed a precise definition of the limit of a function as the argument becomes arbitrarily large: $\lim_{x \to \infty} f(x)$. The definition involves being able to respond to a challenge. In the case of an infinite limit, the challenge is a large number. In the case of a finite limit, the challenge is a small number used to describe a narrow horizontal band.

EXERCISES for Section 3.8

1. Let $f(x) = 3x$.

(a) Find a number D such that $x > D$ implies $f(x) > 600$.

(b) Find another number D such that $x > D$ implies $f(x) > 600$.

(c) What is the smallest number D such that $x > D$ implies $f(x) > 600$?

2. Let $f(x) = 4x$.

(a) Find a number D such that $x > D$ implies $f(x) > 1000$.

(b) Find another number D such that $x > D$ implies $f(x) > 1000$.

(c) What is the smallest number D such that $x > D$ implies $f(x) > 1000$?

3. Let $f(x) = 5x$. Find a number D such that, for all $x > D$,

(a) $f(x) > 2000$

(b) $f(x) > 10,000$

4. Let $f(x) = 6x$. Find a number D such that, for all $x > D$,

(a) $f(x) > 1200$

(b) $f(x) > 1800$

In Exercises 5 to 12 use the precise definition of $\lim_{x\to\infty} f(x) = \infty$ to establish

5. $\lim_{x\to\infty} 3x = \infty$

6. $\lim_{x\to\infty} 4x = \infty$

7. $\lim_{x\to\infty} (x + 5) = \infty$

8. $\lim_{x\to\infty} (x - 600) = \infty$

9. $\lim_{x\to\infty} (2x + 4) = \infty$

10. $\lim_{x\to\infty} (3x - 1200) = \infty$

11. $\lim_{x\to\infty} (4x + 100\cos(x)) = \infty$

12. $\lim_{x\to\infty} (2x - 300\cos(x)) = \infty$

13. Let $f(x) = x^2$.

(a) Find a number D such that, for all $x > D$, $f(x) > 100$.

(b) Let E be a nonnegative number. Find a number D such that, for all $x > D$, it follows that $f(x) > E$.

(c) Let E be a negative number. Find a number D such that, for all $x > D$, it follows that $f(x) > E$.

(d) Using the precise definition of $\lim_{x\to\infty} f(x) = \infty$, show that $\lim_{x\to\infty} x^2 = \infty$.

14. Using the precise definition of $\lim_{x\to\infty} f(x) = \infty$, show that $\lim_{x\to\infty} x^3 = \infty$. (See Exercise 13.)

Exercises 15 to 26 concern the precise definition of $\lim_{x\to\infty} f(x) = L$.

15. Let $f(x) = 3 + 1/x$ if $x \neq 0$.

(a) Find a number D such that $x > D$ implies $|f(x) - 3| < \frac{1}{10}$.

(b) Find another number D such that $x > D$ implies $|f(x) - 3| < \frac{1}{10}$.

(c) What is the smallest number D such that $x > D$ implies $|f(x) - 3| < \frac{1}{10}$?

(d) Using the precise definition of $\lim_{x\to\infty} f(x) = L$, show that $\lim_{x\to\infty} (3 + 1/x) = 3$.

16. Let $f(x) = 2/x$ if $x \neq 0$.

(a) Find a number D such that $x > D$ implies $|f(x) - 0| < \frac{1}{100}$.

(b) Find another number D such that $x > D$ implies $|f(x) - 0| < \frac{1}{100}$.

(c) What is the smallest number D such that $x > D$ implies $|f(x) - 0| < \frac{1}{100}$?

(d) Using the precise definition of $\lim_{x\to\infty} f(x) = L$, show that $\lim_{x\to\infty} (2/x) = 0$.

In Exercises 17 to 22 use the precise definition of $\lim_{x\to\infty} f(x) = L$ to establish

17. $\lim_{x\to\infty} \dfrac{\sin(x)}{x} = 0$ ($|\sin(x)| \leq 1$ for all x.)

18. $\lim_{x\to\infty} \dfrac{x + \cos(x)}{x} = 1$

19. $\lim_{x\to\infty} \dfrac{4}{x^2} = 0$

20. $\lim_{x\to\infty} \dfrac{2x + 3}{x} = 2$

21. $\lim_{x\to\infty} \frac{1}{x-100} = 0$

22. $\lim_{x\to\infty} \frac{2x+10}{3x-5} = \frac{2}{3}$

In Exercises 23 to 26 use the precise definition to prove each statement.

23. $\lim_{x\to\infty} x/(x+1) = \infty$ is false.

24. $\lim_{x\to\infty} \sin(x) = \frac{1}{2}$ is false.

25. $\lim_{x\to\infty} 3x = 6$ is false.

26. $\lim_{x\to\infty} 2x = L$ is false.

Exercises 27 to 30 develop precise definitions of the limit. Phrase them in terms of a challenge number E or ϵ and a reply D. Show the geometric meaning of your definition on a graph.

27. $\lim_{x\to\infty} f(x) = -\infty$

28. $\lim_{x\to-\infty} f(x) = \infty$

29. $\lim_{x\to-\infty} f(x) = -\infty$

30. $\lim_{x\to-\infty} f(x) = L$

31. Let $f(x) = 5$ for all x. (See Exercise 30.) Using a precise definition, show that

(a) $\lim_{x\to\infty} f(x) = 5$.

(b) $\lim_{x\to-\infty} f(x) = 5$.

32. Is this argument correct?
I will prove that $\lim_{x\to\infty} (2x+\cos(x)) = \infty$. Let E be given. I want

$$\begin{array}{rrcl} & 2x+\cos(x) & > & E \\ \text{or} & 2x & > & E-\cos(x) \\ \text{so} & x & > & \frac{E-\cos(x)}{2}. \end{array}$$

Thus, if $D = \frac{E-\cos(x)}{2}$, then $2x+\cos(x) > E$.

33. Use the precise definition of $\lim_{x\to\infty} f(x) = L$, to prove this version of the sum law for limits: if $\lim_{x\to\infty} f(x) = A$ and $\lim_{x\to\infty} g(x) = B$, then $\lim_{x\to\infty}(f(x)+g(x)) = A+B$. (See Example 5.)

34. Use the precise definition of $\lim\limits_{x\to\infty} f(x) = L$, to prove this version of the product law for limits: if $\lim_{x\to\infty} f(x) = A$, then $\lim_{x\to\infty}(f(x)^2) = A^2$. ($f(x)^2 - A^2 = (f(x) - A)(f(x) + A)$, and control the size of each factor.)

35. Use the precise definition of $\lim\limits_{x\to\infty} f(x) = L$, to prove this version of the product law for limits: if $\lim_{x\to\infty} f(x) = A$ and $\lim_{x\to\infty} g(x) = B$, then $\lim_{x\to\infty}(f(x)g(x)) = AB$. (To use the two limits, write $f(x)$ as $A+(f(x)-A)$ and $g(x)$ as $B+(g(x)-B)$.)

36. Assume that $\lim_{x\to\infty} f(x) = 5$. Is there necessarily a number c such that for $x > c$, $f(x)$ stays in the closed interval $[4.5, 5]$? Explain in detail.

37. Assume that $\lim_{x\to\infty} f(x) = 5$. Is there necessarily a number c such that for $x > c$, $f(x)$ stays in the open interval $(4.9, 5.3)$? Explain in detail.

38.

Sam: I got lost in Example 5 when $\epsilon/2$ came out of nowhere.

Jane: It's just another small number. They were looking ahead to what they needed.

Sam: Why must the two numbers add up to ϵ?

Jane: They don't have to. They could add up to ϵ divided by 12 for instance.

Sam: What if they added up to 12ϵ? Would that work too?

Jane: No.

Sam: I'm getting a headache.

Explain Jane's explanation for Sam's benefit.

3.9 Precise Definition of Limits at a Finite Point: $\lim\limits_{x \to a} f(x) = L$

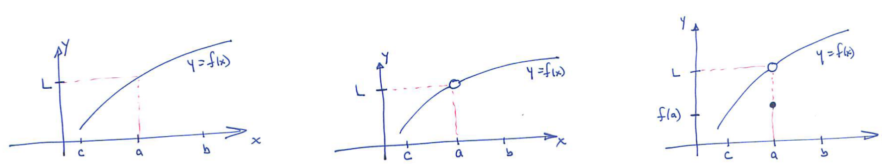
To conclude the discussion of limits, we extend the ideas developed in Section 3.8 to the limit of a function at a number a.

Informal definition of $\lim_{x \to a} f(x) = L$

Let f be a function and a some fixed number.

1. Assume that the domain of f contains open intervals (c, a) and (a, b) for some $c < a$ and some $b > a$.
2. If, as x approaches a, either from the left or from the right, $f(x)$ approaches a number L, then L is called the **limit** of $f(x)$ as x approaches a. This is written

$$\lim_{x \to a} f(x) = L.$$



(a) (b) (c)

Figure 3.9.1 Three possibilities for $\lim_{x \to a} f(x) = L$.

Figure 3.9.1 illustrates three possibilities for $\lim_{x \to a} f(x) = L$: (a) a is in the domain of f and $f(a) = L$, (b) a is in the domain of f and $f(a) \neq L$, and (c) a is not in the domain of f. These remind us that a need not be in the domain of f. And, even if it is, the value of $f(a)$ plays no role in determining whether $\lim_{x \to a} f(x)$ exists.

Reworded informal definition of $\lim_{x \to a} f(x) = L$

Let f be a function and a some number.

1. Assume that the domain of f contains open intervals (c, a) and (a, b) for some $c < a$ and some $b > a$.
2. If x is is sufficiently close to a but not equal to a, then $f(x)$ is necessarily near L.

The following precise definition parallels the reworded informal definition.

δ (delta) is the lower case version of the Greek letter Δ; they correspond to the English letters d and D.

DEFINITION *(Precise definition of* $\lim_{x\to a} f(x) = L$*)* Let f be a function and a some fixed number.

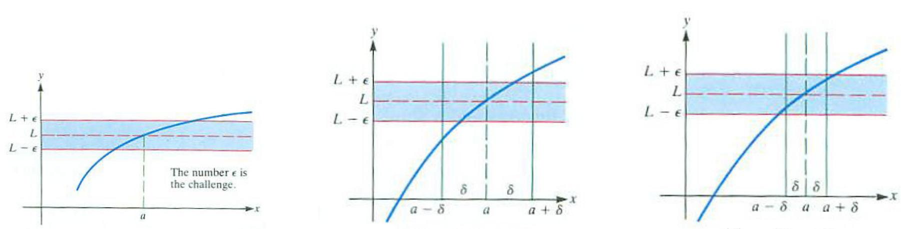
1. Assume that the domain of f contains open intervals (c, a) and (a, b) for some $c < a$ and some $b > a$.
2. For each positive number ϵ there is a positive number δ such that

$$\begin{array}{ll} \text{for all } x \text{ that satisfy the inequality} & 0 < |x - a| < \delta \\ \text{it is true that} & |f(x) - L| < \epsilon. \end{array}$$

The meaning of $0 < |x - a| < \delta$

The inequality $0 < |x - a|$ that appears in the definition is a way of saying "x is not a." The inequality $|x - a| < \delta$ asserts that x is within a distance δ of a. The two inequalities are combined as the single statement $0 < |x - a| < \delta$, which describes the open interval $(a - \delta, a + \delta)$ from which a is deleted. This deletion is made since $f(a)$ plays no role in the definition of $\lim_{x\to a} f(x)$.

Once again ϵ is the challenge. The reply is δ. Usually, the smaller ϵ is, the smaller δ will have to be.



(a) (b) (c)

Figure 3.9.2 (a) The number ϵ is the challenge. (b) δ is not small enough. (c) δ is small enough.

The geometric significance of the precise definition of $\lim_{x\to a} f(x) = L$ is shown in Figure 3.9.2. The narrow horizontal band of width 2ϵ is the challenge (see Figure 3.9.2(a)). The desired response is a sufficiently narrow vertical band, of width 2δ, such that the part of the graph within the vertical band (except perhaps at $x = a$) also lies in the horizontal band of width 2ϵ. In Figure 3.9.2(b) the vertical band is not narrow enough to meet the challenge of the horizontal band, but the vertical band in Figure 3.9.2(c) is narrow enough.

Assume that for each positive number ϵ it is possible to find a positive number δ such that the parts of the graph between $x = a - \delta$ and $x = a$ and between $x = a$ and $x = a + \delta$ lie within the given horizontal band. Then we say that "as x approaches a, $f(x)$ approaches L". The narrower the horizontal band around the line $y = L$, the smaller δ usually must be.

EXAMPLE 1 Use the precise definition of $\lim_{x \to a} f(x) = L$ to show that $\lim_{x \to 2}(3x + 5) = 11$.

SOLUTION Here $f(x) = 3x + 5$, $a = 2$, and $L = 11$. Let ϵ be a positive number. We wish to find a number $\delta > 0$ such that for $0 < |x - 2| < \delta$ we have $|(3x + 5) - 11| < \epsilon$.

Let us find out for which x it is true that $|(3x + 5) - 11| < \epsilon$. This is equivalent to

$$|3x - 6| < \epsilon$$

or

$$3|x - 2| < \epsilon$$

or

$$|x - 2| < \frac{\epsilon}{3}.$$

Any positive number less than $\epsilon/3$ is also a suitable response.

Thus $\delta = \epsilon/3$ is a suitable response. If $0 < |x-2| < \epsilon/3$, then $|(3x+5)-11| < \epsilon$. $\diamond$

The algebra of finding a response δ can be more involved for other functions, such as $f(x) = x^2$. The precise definition of limit can actually be easier to apply in more general situations where f and a are not given explicitly. To illustrate, we present a proof of the permanence property.

When the permanence property was introduced in Section 2.5, the only justification we provided was a picture and an appeal to intuition that a continuous function cannot jump instantaneously from a positive value to zero or a negative value — the function has to remain positive on some open interval. Mathematicians call this a "proof by handwaving." We can prove without the use of intuition or handwaving that there must be an open interval around a given input, a, such that for any x in that interval $f(x)$ stays near $f(a)$.

EXAMPLE 2 Prove the permanence property: Assume that f is continuous in an open interval that contains a and that $f(a) = p > 0$. Then for any $q < p$ there is an open interval I containing a such that $f(x) > q$ for all x in I.

SOLUTION Let $p = f(a) > 0$ and let q be a number less than p. Pick $\epsilon = p - q$. (The reason for this choice for ϵ will become clear in a moment.) Because f is continuous at a, $\lim_{x \to a} f(x) = f(a)$. By the precise definition of $\lim_{x \to a} f(x) = L$, when $\epsilon = p - q$ there is a positive number δ such that

$$|f(a) - f(x)| < p - q \qquad \text{for } a - \delta < x < a + \delta.$$

Thus
$$-(p-q) < f(a) - f(x) < p - q.$$
In particular,
$$f(a) - f(x) < p - q \tag{3.9.1}$$
Because $f(a) = p$, (3.9.1) can be rewritten as
$$p - f(x) < p - q$$
or
$$f(x) > q.$$
Thus $f(x)$ is greater than q if x is in the interval $I = (a - \delta, a + \delta)$.

◇

One of the common uses of the permanence property is to say that if a continuous function is positive at a then there is an interval containing a on which the function remains positive. (This corresponds to $p = f(a) > 0$ and $q = 0$.)

Summary

This section developed a precise definition of the limit of a function as the argument approaches a fixed number: $\lim_{x\to a} f(x)$. It involves being able to respond to an arbitrary challenge number. For a finite limit, the challenge is a small positive number. The smaller that number, the harder it is to meet the challenge.

In addition, it also gave a rigorous proof of the permanence principle. That we could deduce it from the precise definition of a limit reassures us that the precise definition expresses what we feel the word "limit" means.

EXERCISES for Section 3.9

In Exercises 1 to 4 use the precise definition of $\lim_{x\to a} f(x) = L$ to justify

1. $\lim_{x\to 2} 3x = 6$

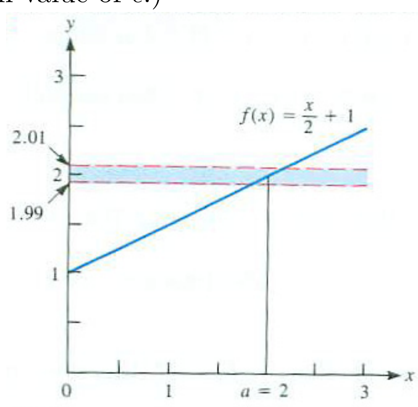
2. $\lim_{x\to 3}(4x-1) = 11$

3. $\lim_{x\to 1}(x+2) = 3$

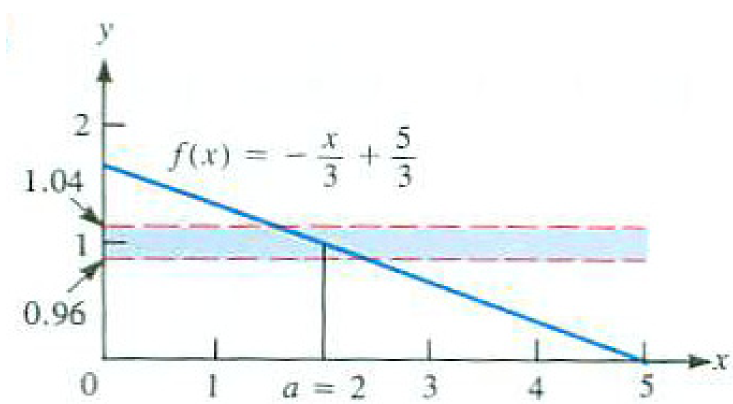
4. $\lim_{x\to 5}(2x-3) = 7$

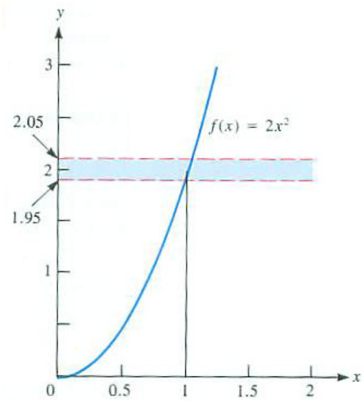
In Exercises 5 to 8 find a number δ such that the point $(x, f(x))$ lies in the shaded band for all x in the interval $(a-\delta, a+\delta)$. (Draw a suitable vertical band for the given value of ϵ.)

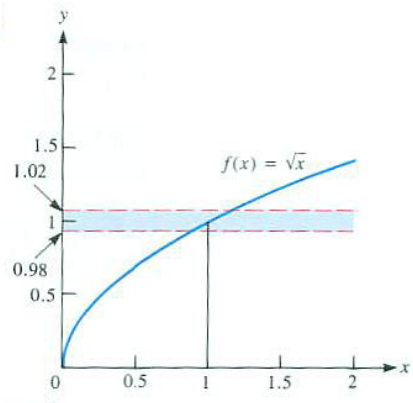
5.



6.



7. 

8. 

In Exercises 9 to 12 use the precise definition of $\lim_{x\to a} f(x) = L$ to justify

9. $\lim_{x\to 1}(3x+5) = 8$

10. $\lim_{x\to 1} \frac{5x+3}{4} = 2$

11. $\lim_{x\to 0} \frac{x^2}{4} = 0$

12. $\lim_{x\to 0} 4x^2 = 0$

13. Give an example of a number $\delta > 0$ such that $|x^2 - 4| < 1$ if $0 < |x-2| < \delta$.

14. Give an example of a number $\delta > 0$ such that $|x^2+x-2| < 0.5$ if $0 < |x-1| < \delta$.

Develop precise definitions of the limits in Exercises 15 to 20. Phrase your definitions in terms of a challenge, E or ϵ, and a response, δ.

15. $\lim_{x\to a^+} f(x) = L$

16. $\lim_{x\to a^-} f(x) = L$

17. $\lim_{x\to a} f(x) = \infty$

18. $\lim_{x \to a} f(x) = -\infty$

19. $\lim_{x \to a^+} f(x) = \infty$

20. $\lim_{x \to a^-} f(x) = \infty$

21. Let $f(x) = 9x^2$.

(a) Find $\delta > 0$ such that, for $0 < |x - 0| < \delta$, it follows that $|9x^2 - 0| < \frac{1}{100}$.

(b) Let ϵ be any positive number. Find a positive number δ such that, for $0 < |x - 0| < \delta$ we have $|9x^2 - 0| < \epsilon$.

(c) Show that $\lim_{x \to 0} 9x^2 = 0$.

22. Let $f(x) = x^3$.

(a) Find $\delta > 0$ such that, for $0 < |x - 0| < \delta$, it follows that $|x^3 - 0| < \frac{1}{1000}$.

(b) Show that $\lim_{x \to 0} x^3 = 0$.

23. Show that the assertion that $\lim_{x \to 2} 3x = 5$ is false. To do this, it is necessary to exhibit a positive number ϵ such that there is no response number $\delta > 0$. (Draw a picture.)

24. Show that the assertion"$\lim_{x \to 2} x^2 = 3$" is false.

25. In the proof of the permanence property given in Example 2, $p = f(a) > 0$ and $q < p$.

(a) Would the argument have worked if we had used $\epsilon = 2(p - q)$?

(b) Would the argument have worked if we had used $\epsilon = \frac{1}{2}(p - q)$?

(c) Would the argument have worked if we had used $\epsilon = q$?

(d) What is the largest value of ϵ for which the proof of the permanence property works?

26. The permanence property discussed in Example 2 and Exercise 25 pertains to limits at a finite point a. State, and prove, a version of the permanence property that is valid when a is replaced by ∞.

27.

(a) Show that if $0 < \delta < 1$ and $|x-3| < \delta$ then $|x^2 - 9| < 7\delta$. (Factor $x^2 - 9$.)

(b) Use (a) to deduce that $\lim_{x\to 3} x^2 = 9$.

28.

(a) Show that if $0 < \delta < 1$ and $|x-4| < \delta$ then

$$|\sqrt{x} - 2| < \frac{\delta}{\sqrt{3}+2}.$$

(b) Use (a) to deduce that $\lim_{x\to 4} \sqrt{x} = 2$.

29.

(a) Show that if $0 < \delta < 1$ and $|x-3| < \delta$ then $|x^2 + 5x - 24| < 12\delta$. (Factor $x^2 + 5x - 24$.)

(b) Use (a) to deduce that $\lim_{x\to 3}(x^2 + 5x) = 24$.

30.

(a) Show that if $0 < \delta < 1$ and $|x-2| < \delta$ then

$$\left|\frac{1}{x} - \frac{1}{2}\right| < \frac{\delta}{2}.$$

(b) Use (a) to deduce that $\lim_{x\to 2} \frac{1}{x} = \frac{1}{2}$.

31. Use a precise definition of limit to prove: if f is defined in an open interval including a and f is continuous at a, so is $3f$.

32. Use a precise definition of limit to prove: if f and g are both defined in an open interval including a and both are continuous at a, so is $f + g$.

33. Use a precise definition of limit to prove: if f and g are both continuous at a, then their product, fg, is also continuous at a. Assume that both functions are defined at least in an open interval around a.

34. Assume that $f(x)$ is continuous at a and is defined on an open interval containing a. Assume that $f(x) = p > 0$ and that q is a number greater than p. Using the precise definition of a limit, show that there is an open interval, I, containing a such that $f(x) < q$ for all x in I.

3.S Chapter Summary and Look Ahead

In this chapter we defined the derivative of a function, developed ways to compute derivatives, and applied them to graphs and motion.

The derivative of a function f at a number $x = a$ is defined as the limit of the slopes of secants through the points $(a, f(a))$ and $(b, f(b))$ as the input b is taken closer and closer to a.

Algebraically, the derivative is the limit of a quotient, the change in the output divided by the change in the input. It is usually written as

$$\lim_{x\to a} \frac{f(x) - f(a)}{x - a}, \quad \text{or} \quad \lim_{h\to 0} \frac{f(a+h) - f(a)}{h}, \quad \text{or} \quad \lim_{\Delta x\to 0} \frac{\Delta y}{\Delta x}.$$

The derivative is denoted in several ways, such as

$$f', \quad \text{or} \quad f'(x), \quad \text{or} \quad \frac{df}{dx}, \quad \text{or} \quad \frac{dy}{dx}, \quad \text{or} \quad D(f).$$

For functions most frequently encountered in applications, derivatives exist. Geometrically, the derivative exists whenever the graph of the function on a small interval looks almost like a straight line.

The derivative records how fast something changes. The velocity of a moving object is defined as the derivative of the object's position. Also, the derivative gives the slope of the tangent line to the graph of a function.

We then developed ways to compute the derivative of functions expressible in terms of the functions met in algebra and trigonometry, as well as exponentials with a fixed base and logarithms, the elementary functions. They were based on three limits:

$$\begin{aligned} \lim_{x\to a} \frac{x^n - a^n}{x - a} &= na^{n-1}, \qquad n \text{ a positive integer} \\ \lim_{x\to 0} \frac{e^x - 1}{x} &= 1 \\ \lim_{x\to 0} \frac{\sin(x)}{x} &= 1. \end{aligned}$$

Using them, we obtained the derivatives of x^n, e^x, and $\sin(x)$. We showed, if we knew the derivatives of two functions, how to compute the derivatives of their sum, difference, product, and quotient.

The next step was the development of a most important computational tool: the chain rule. It enables us to differentiate a composite function, such as $\cos^3(x^2)$, telling us that its derivative is $3\cos^2(x^2)(-\sin(x^2))(2x)$.

Differentiating inverse functions enabled us to show that the derivative of $\ln(x)$ is $\frac{1}{x}$ for $x > 0$ and the derivative of $\arcsin(x)$ is $\frac{1}{\sqrt{1-x^2}}$ for $-1 < x < 1$.

The following list of derivatives of key functions should be memorized.

Function	Derivative
x^a (a constant)	ax^{a-1}
$\sin(x)$	$\cos(x)$
$\cos(x)$	$-\sin(x)$
e^x	e^x
a^x (a constant)	write $a^x = e^{x(\ln(a))}$
$\ln(x)$ $(x > 0)$	$1/x$
$\ln\|x\|$ $(x \neq 0)$	$1/x$
$\tan(x)$	$\sec^2(x)$
$\sec(x)$	$\sec(x)\tan(x)$
$\arcsin(x)$	$\frac{1}{\sqrt{1-x^2}}$
$\arctan(x)$	$\frac{1}{1+x^2}$
$\sqrt{x}$	$\frac{1}{2\sqrt{x}} = \frac{1}{2}x^{-1/2}$
$\frac{1}{x}$	$-1/x^2$

Table 3.S.1 Table of Common Functions and Derivatives.

As you work with derivatives you may begin to think of them as slope or velocity or rate of change, and forget their underlying definition as limits. However, we will return to the definition in terms of limits as we develop more applications of the derivative.

We also introduced the antiderivative and, related to it, the slope field. While the derivative of an elementary function is again elementary, an antiderivative often is not. For instance, $\sqrt{1+x^3}$ does not have an elementary antiderivative. However, as we will see in Chapter 6, it does have an antiderivative. Chapter 8 will present different ways to find antiderivatives. Slope fields will be met again in Chapter 13.

The derivative of the derivative is the second derivative. For motion, the second derivative describes acceleration. It can be denoted several ways, such as $D^2 f$, $\frac{d^2 f}{dx^2}$, f'', and $f^{(2)}$. While the first and second derivatives suffice for most applications, higher derivatives of all orders are used in Chapter 5 to estimate the error when approximating a function by a polynomial.

The final two sections provided precise definitions of limits and a proof of the permanence property.

EXERCISES for 3.S

In Exercises 1 to 19 differentiate the expression.

1. $\exp(x^2)$

2. 2^{x^2}

3. $x^3 \sin(4x)$

4. $\frac{1+x^2}{1+x^3}$

5. $\ln(x^3)$

6. $\ln(x^3+1)$

7. $\cos^4(x^2)\tan(2x)$

8. $\sqrt{5x^2+x}$

9. $\arcsin(\sqrt{3+2x})$

10. $x^2 \arctan(2x) e^{3x}$

11. $\sec^2(3x)$

12. $\sec^2(3x) - \tan^2(3x)$

13. $\left(\frac{3+2x}{4+5x}\right)^3$

14. $\frac{1}{1+2e^{-x}}$

15. $\frac{x}{\sqrt{x^2+1}}$

16. $(\arcsin(3x))^2$

17. $x^2 \arctan(3x)$

18. $\sin^5(3x^2)$

19. $\frac{1}{(2^x+3^x)^{20}}$

In Exercises 20 to 29 give an antiderivative of the expression. Use differentiation to check each answer.

20. $4x^3$

21. x^3

22. $3/x^2$

23. $\cos(x)$

24. $\cos(2x)$

25. $\sin^{100}(x)\cos(x)$

26. $1/(x+1)$

27. $5e^{4x}$

28. $1/e^x$

29. 2^x

In Exercises 30 to 51 evaluate the derivative to verify each equation. The letters a, b, c, and d denote constants. These exercises, based on tables of antiderivatives, provide practice in differentiation and algebra.

30. $\frac{d}{dx}\left(\frac{1}{a}\tan^{-1}\left(\frac{x}{a}\right)\right) = \frac{1}{a^2+x^2}$

31. $D\left(\frac{1}{2a}\ln\left(\frac{a+x}{a-x}\right)\right) = \frac{1}{a^2-x^2}$

32. $\left(\ln\left(x+\sqrt{a^2+x^2}\right)\right)' = \frac{1}{\sqrt{a^2+x^2}}$

33. $\frac{d}{dx}\left(\frac{1}{a}\ln\left(\frac{x+\sqrt{a^2-x^2}}{x}\right)\right) = \frac{1}{x\sqrt{a^2+x^2}}$

34. $D\left(\frac{-1}{b(a+bx)}\right) = \frac{1}{(a+bx)^2}$

35. $\left(\frac{1}{b^2}\left(a+bx-a\ln(a+bx)\right)\right)' = \frac{x}{a+bx}$

36. $\frac{d}{dx}\left(\frac{1}{b^2}\left(\frac{a}{2(a+bx)^2}-\frac{1}{a+bx}\right)\right) = \frac{x}{(a+bx)^3}$

37. $D\left(\frac{1}{ad-bc}\ln\left(\frac{c+dx}{a+bx}\right)\right) = \frac{1}{(a+bx)(c+dx)}$

38. $\left(\frac{2}{\sqrt{4ac-b^2}}\arctan\left(\frac{2cx+b}{\sqrt{4ac-b^2}}\right)\right)' = \frac{1}{a+bx+cx^2}$ $\quad (4ac > b^2)$

39. $\frac{d}{dx}\left(\frac{-2}{\sqrt{b^2-4ac}}\ln\left(\frac{2cx+b-\sqrt{b^2-4ac}}{2cx+b+\sqrt{b^2-4ac}}\right)\right) = \frac{1}{a+bx+cx^2}$ $\quad (4ac < b^2)$

40. $D\left(\frac{1}{a}\cos^{-1}\left(\frac{a}{x}\right)\right) = \frac{1}{x\sqrt{x^2-a^2}}$

41. $\left(\frac{1}{2}\left(x\sqrt{a^2-x^2}+a^2\arcsin\left(\frac{x}{a}\right)\right)\right)' = \sqrt{a^2-x^2}$ $\quad (|x| < |a|)$

42. $\frac{d}{dx}\left(\frac{-x}{2}\sqrt{a^2-x^2}+\frac{a^2}{2}\arcsin\left(\frac{x}{a}\right)\right) = \frac{x^2}{\sqrt{a^2-x^2}}$ $\quad (|x| < |a|)$

43. $D\left(-\frac{\sqrt{a^2-x^2}}{x}-\arcsin\left(\frac{x}{a}\right)\right) = \frac{\sqrt{a^2-x^2}}{x^2}$ $\quad (|x| < |a|)$

44. $\left(\arcsin(x)-\sqrt{1-x^2}\right)' = \sqrt{\frac{1+x}{1-x}}$ $\quad (|x| < 1)$

45. $\frac{d}{dx}\left(\frac{x}{2}-\frac{1}{2}\cos(x)\sin(x)\right) = \sin^2(x)$

46. $D\left(x\arcsin x+\sqrt{1-x^2}\right) = \arcsin(x)$ $\quad (|x| < 1)$

47. $\left(x\tan^{-1}(x)-\frac{1}{2}\ln(1+x^2)\right)' = \arctan(x)$

48. $\frac{d}{dx}\left(\frac{e^{ax}}{a^2}\left(a^2-1\right)\right) = xe^{ax}$

49. $D\left(x-\ln(1+e^x)\right) = \frac{1}{1+e^x}$

50. $\left(\frac{x}{2}\left(\sin(\ln(ax))-\cos(\ln(ax))\right)\right)' = \sin(\ln(ax))$

51. $\left(\frac{e^{ax}(a\sin(bx)-b\cos(bx))}{a^2+b^2}\right)' = e^{ax}\sin(bx)$

In Exercises 52 to 55 give two antiderivatives for each expression.

52. xe^{x^2}

53. $(x^2+x)e^{x^3+3x}$

54. $\cos^3(x)\sin(x)$

55. $\sin(2x)$

56. Verify that $2(\sqrt{x}-1)e^{\sqrt{x}}$ is an antiderivative of $e^{\sqrt{x}}$.

In Exercises 57 to 60 (a) sketch the slope field and (b) draw the solution curve through the point $(0, 1)$.

57. $dy/dx = 1/(x + 1)$

58. $dy/dx = e^{-x^2}$

59. $dy/dx = -y$

60. $dy/dx = y - x$

61. Sam threw a baseball straight up and caught it 6 seconds later.

(a) How high above his head did it rise?

(b) How fast was it going as it left his hand?

(c) How fast was it going when he caught it?

(d) Translate the answers in (b) and (c) to miles per hour. (60 mph = 88 fps.)

62. Assuming that $D(x^4) = 4x^3$ and $D(x^7) = 7x^6$, you could find $D(x^3)$ from them by viewing x^3 as x^7/x^4 and using the formula for differentiating a quotient. Show how you could use them to find

(a) $D(x^{11})$

(b) $D(x^{-4})$

(c) $D(x^{28})$

(d) $D(x^8)$

63. Let $y = x^{m/n}$, where $x > 0$ and m and $n \neq 0$ are integers. Assuming that y is differentiable, show that $\frac{dy}{dx} = \frac{m}{n} x^{\frac{m}{n}-1}$ by starting with $y^n = x^m$ and differentiating both sides with respect to x. (Think of y as $y(x)$ and remember to use the chain rule when differentiating y^n with respect to x.)

64. A spherical balloon is being filled with helium at the rate of 3 cubic feet per minute. At what rate is the radius increasing when the radius is (a) 2 feet? (b) 3 feet? (The volume of a ball of radius r is $\frac{4}{3}\pi r^3$.)

65. An object at the end of a vertical spring is at rest. When pulled down and released, it goes up and down for a while. With the origin of the y-axis at the rest position, the position of the object t seconds later is $3e^{-2t}\cos(2\pi t)$ inches.

(a) What is the physical significance of 3 in the formula?

(b) What does e^{-2t} tell us?

(c) What does $\cos(2\pi t)$ tell us?

(d) How long does it take the object to complete a full cycle (go from its rest position, down, up, then down to its rest position)?

(e) What happens to the object after a long time?

66. The motor on a moving motor boat is turned off. It then coasts along the x-axis. Its position, in meters, at time t (seconds) is $500 - 50e^{-3t}$.

(a) Where is it at time $t = 0$?

(b) What is its velocity at time t?

(c) What is its acceleration at time t?

(d) How far does it coast?

(e) Show that its acceleration is proportional to its velocity. This means the force of the water slowing the boat is proportional to the velocity of the boat. (See also Exercise 78.)

67. It is safe to switch the "sin" and "lim" in $\sin\left(\lim_{x\to 0}\frac{e^x-1}{x}\right) = \lim_{x\to 0}\left(\sin\left(\frac{e^x-1}{x}\right)\right)$. However, such a switch sometimes is not correct. Consider f defined by $f(x) = 2$ for $x \neq 1$ and $f(1) = 0$.

(a) Show that $f\left(\lim\limits_{x\to 0}\dfrac{e^x-1}{x}\right)$ is not equal to $\lim_{x\to 0} f\left(\frac{e^x-1}{x}\right)$.

(b) What property of $\sin(x)$ permits us to switch it with lim?

It is important to keep in mind the definition of a derivative as a limit. Exercises 68 to 72 are intended to reinforce the definition.

68. Define the derivative of the function $g(x)$ at $x = a$ in (a) the x and $x + h$ notation, (b) the x and a notation, and (c) the Δy and Δx notation.

69. We obtained the derivative of $\sin(x)$ using the x and $x + h$ notation and the addition identity for $\sin(x + h)$. Instead, obtain the derivative of $\sin(x)$ using the x and a notation. That is, find

$$\lim_{x\to a}\frac{\sin(x) - \sin(a)}{x - a}.$$

(a) Show that $\sin(x) - \sin(y) = 2\sin\left(\frac{1}{2}(x-y)\right)\cos\left(\frac{1}{2}(x+y)\right)$.

(b) Use the identity in (a) to find the limit.

70. We obtained the derivative for $\tan(x)$ by writing it as $\sin(x)/\cos(x)$. Instead, obtain it directly by finding

$$\lim_{h\to 0} \frac{\tan(x+h) - \tan(x)}{h}.$$

(The identity $\tan(a+b) = \frac{\tan(a)+\tan(b)}{1-\tan(a)\tan(b)}$ will help.)

71. Show that $\frac{\tan(a)}{\tan(b)} > \frac{a}{b} > \frac{\sin(a)}{\sin(b)}$ for all angles a and b in the first quadrant with $a > b$. (Use the two inequalities that squeezed $\sin(x)/x$ toward 1.)

72. We obtained the derivative of $\ln(x)$, $x > 0$, by viewing it as the inverse of $\exp(x)$. Instead, find it directly from the definition. (Use the x and h notation.)

Exercises 73 and 74 show how we could have predicted that $\ln(x)$ would provide an antiderivative for $1/x$.

73. The antiderivative of $1/x$ that passes through $(1,0)$ is $\ln(x)$. One would expect that for t near 1, the antiderivative of $1/x^t$ that passes through $(1,0)$ would look much like $\ln(x)$ when x is near 1. To verify that this is true

(a) graph the slope field for $1/x^t$ with $t = 1.1$

(b) graph the antiderivative of $1/x^t$ that passes through $(1,0)$ for $t = 1.1$

(c) repeat (a) and (b) for $t = 0.9$

(d) repeat (a) and (b) for $t = 1.01$

(e) repeat (a) and (b) for $t = 0.99$

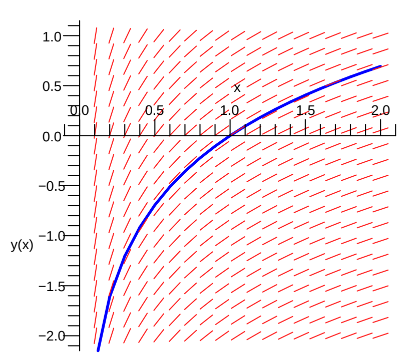


Figure 3.S.1

The slope field for $1/x$ and the antiderivative of $1/x$ passing through $(1,0)$ are shown in Figure 3.S.1.

74. (See Exercise 73.)

(a) Verify that for $t \neq 1$ the antiderivative of $1/x^t$ that passes through $(1,0)$ is $\frac{x^{1-t}-1}{1-t}$.

(b) Holding x fixed and letting t approach 1, show that

$$\lim_{t\to 1} \frac{x^{1-t}-1}{1-t} = \ln(x).$$

(Recognize the limit as the derivative of a function at a certain input. Keep in mind that x is constant in this limit.)

75. Define f as

$$f(x) = \begin{cases} x & \text{if } x \text{ is rational,} \\ -x & \text{if } x \text{ is irrational.} \end{cases}$$

(a) What does the graph of f look like? A dotted curve may be used to indicate that points are missing.

(b) Does $\lim\limits_{x\to 0} f(x)$ exist?

(c) Does $\lim\limits_{x\to 1} f(x)$ exist?

(d) Does $\lim\limits_{x\to \sqrt{2}} f(x)$ exist?

(e) For which numbers a does $\lim\limits_{x\to a} f(x)$ exist?

76. Define f as

$$f(x) = \begin{cases} x^2 & \text{if } x \text{ is rational,} \\ x^3 & \text{if } x \text{ is irrational.} \end{cases}$$

(a) What does the graph of f look like? A dotted curve may be used to indicate that points are missing.

(b) Does $\lim\limits_{x\to 0} f(x)$ exist?

(c) Does $\lim\limits_{x\to 1} f(x)$ exist?

(d) Does $\lim_{x \to \sqrt{2}} f(x)$ exist?

(e) For which numbers a does $\lim_{x \to a} f(x)$ exist?

77. A heavy block rests on a horizontal table covered with thick oil. The block, which is at the origin of the x-axis, is given an initial velocity v_0 at time $t = 0$. It then coasts along the positive x-axis.
Assume that its acceleration is $-k\sqrt{v(t)}$, where $v(t)$ is the velocity at time t and k is a constant. (That means it meets a resistance force proportional to the square root of its velocity.)

(a) Show that $\frac{dv}{dt} = -kv^{1/2}$.

(b) Is k positive or negative? Explain.

(c) Show that $2v^{1/2}$ and $-kt$ have the same derivative with respect to t.

(d) Show that $2v^{1/2} = -kt + 2v_0^{1/2}$.

(e) When, in terms of v_0 and k, does the block come to a rest?

(f) How far, in terms of v_0 and k, does the block slide?

78. A motorboat traveling along the x-axis at the speed v_0 stops its motor at time $t = 0$ when it is at the origin. It then coasts along the positive x-aixis.
Assume the resistance force of the water is proportional to the velocity. That implies the acceleration of the boat is proportional to its velocity, $v(t)$. (See also Exercise 66.)

(a) Show that there is a constant k such that $\frac{dv}{dt} = -kv(t)$.

(b) Is k positive or negative? Explain.

(c) Deduce that $\ln(v)$ and $-kt$ have the same derivative with respect to t.

(d) Deduce that $\ln(v(t)) = -kt + \ln(v_0)$.

(e) Deduce that $v(t) = v_0 e^{-kt}$.

(f) According to (e), how long, in terms of v_0 and k, does the boat continue to move?

(g) How far, in terms of v_0 and k, does it move during that time?

79. Archimedes used the following property of a parabola in his study of the equilibrium of floating bodies. Let P be a point on the parabola $y = x^2$ other than the origin. The line perpendicular to the parabola at P meets the y-axis in a point Q. The line through P and parallel to the x-axis meets the y-axis in a point R. Show that the length of QR is constant, independent of the choice of P. This problem introduces the **subnormal** of the graph; compare this with Exercises 25 and 26 in Section 3.2.

Calculus is Everywhere # 4
Solar Cookers

A satellite dish is parabolic in shape. It is formed by rotating a parabola about its axis. The reason is that all radio waves parallel to the axis of the parabola, after bouncing off the parabola, pass through a common point, called the **focus** of the parabola. (See Figure C.4.1.) The reflector behind a flashlight bulb is also parabolic.

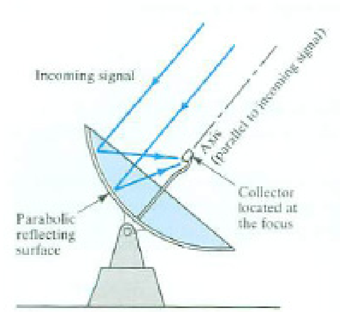


Figure C.4.1

An ellipse also has a reflection property. Light, or sound, or heat radiating from a focus, after bouncing off the ellipse, goes through the other focus. This fact is applied in the construction of computer chips where it is necessary to bake a photomask onto the surface of a silicon wafer. The heat is focused at the mask by placing a heat source at one focus of an ellipse and positioning the wafer at the other focus, as in Figure C.4.2.

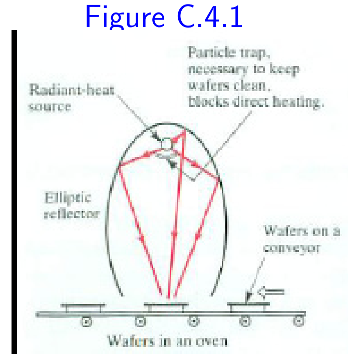


Figure C.4.2

The reflection property is used in wind tunnel tests of aircraft noise. The test is run in an elliptical chamber, with the aircraft model at one focus and a microphone at the other.

Whispering rooms, such as the rotunda in the Capitol in Washington, D.C., are based on the same principle. A person talking quietly at one focus can be heard easily at the other focus but not at points between the foci. (The whisper would be unintelligible except for the additional property that all the paths of the sound from one focus to the other have the same length.)

An ellipsoidal reflector cup is used for crushing kidney stones. (An ellipsoid is formed by rotating an ellipse about the line through its foci.) An electrode is placed at one focus and an ellipsoid positioned so that the stone is at the other focus. Shock waves generated at the electrode bounce off the ellipsoid, concentrate on the other focus, and pulverize the stones without damaging other parts of the body. The patient recovers in three to four days instead of the two to three weeks required after surgery. This advance also reduced the mortality rate from kidney stones by a factor of 200, from 1 in 50 to 1 in 10,000.

The reflecting property of the ellipse also is used in the study of air pollution. One way to detect air pollution is by light scattering. A laser is aimed through one focus of a shiny ellipsoid. When a particle passes through this focus, the light is reflected to the other focus where a light detector is located. The number of particles detected is used to determine the amount of pollution in the air.

The Angle Between Two Lines

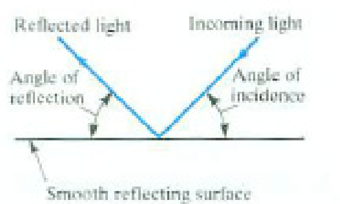
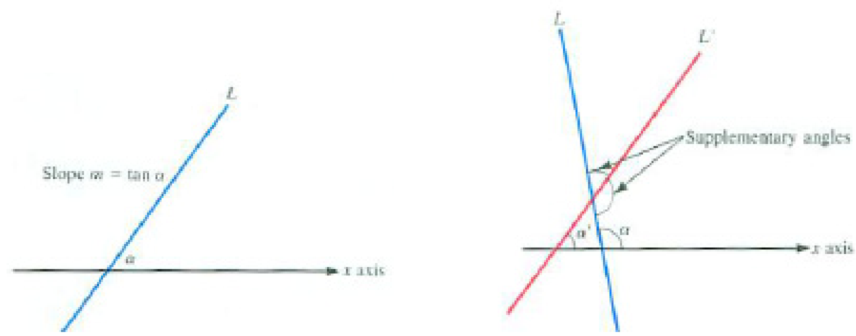


Figure C.4.3

To establish the reflection properties we will use the principle that the angle of reflection equals the angle of incidence, as in Figure C.4.3, and work with the angle between two lines, given their slopes.

Consider a line L in the xy-plane. It forms an **angle of inclination** α, $0 \le \alpha < \pi$, with the positive x-axis. The slope of L is $\tan(\alpha)$. (See Figure C.4.4(a).) If $\alpha = \pi/2$, the slope is not defined.



(a) (b)

Figure C.4.4

Two lines L and L' with angles of inclination α and α' and slopes m and m', respectively, as in Figure C.4.4(b) intersect so that there are two supplementary angles between them. The following definition distinguishes one as the angle between L and L'.

> **DEFINITION** *(Angle between two lines.)* Let L and L' be two lines in the xy-plane, named so that L has the larger angle of inclination, $\alpha > \alpha'$. The angle θ between L and L' is defined to be
>
> $$\theta = \alpha - \alpha'.$$
>
> If L and L' are parallel, θ is defined to be 0.

So, θ is the counterclockwise angle from L' to L and $0 \le \theta < \pi$. The tangent of θ can be expressed in terms of the slopes m of L and m' of L':

$$\begin{aligned}
\tan(\theta) &= \tan(\alpha - \alpha') && (\text{ definition of } \theta \text{ }) \\
&= \frac{\tan(\alpha) - \tan(\alpha')}{1 + \tan(\alpha)\tan(\alpha')} && (\text{ by the identity for } \tan(A-B) \text{ }) \\
&= \frac{m - m'}{1 + mm'}. && (\text{ definition of } m \text{ and } m' \text{ })
\end{aligned}$$

Thus

$$\tan(\theta) = \frac{m - m'}{1 + mm'}. \tag{C.4.1}$$

The Reflection Property of a Parabola

Given the parabola $y = x^2$. We wish to show that angles A and B at the typical point (a, a^2) on the parabola are equal. (See Figure C.4.5.) We will do this by showing that $\tan(A) = \tan(B)$.

First of all, $\tan(C) = 2a$, the slope of the parabola at (a, a^2). Since A is the complement of C,

$$\tan(A) = \tan(\frac{\pi}{2} - C) = \frac{1}{\tan(C)} = 1/(2a).$$

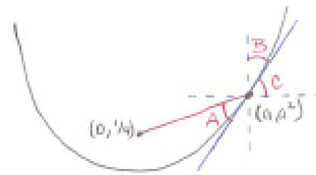


Figure C.4.5

The slope of the line through the focus $(0, \frac{1}{4})$ and a point on the parabola (a, a^2) is

$$\frac{a^2 - \frac{1}{4}}{a - 0} = \frac{4a^2 - 1}{4a}.$$

Therefore,

$$\tan(B) = \frac{2a - \frac{4a^2-1}{4a}}{1 + 2a\left(\frac{4a^2-1}{4a}\right)}.$$

Exercise 1 asks you to supply the algebraic steps to complete the proof that $\tan(B) = \tan(A)$.

The Reflection Property of an Ellipse

An ellipse consists of points such that the sum of the distances from a point to two fixed points is constant. Let the two fixed points, called the **foci** of the ellipse, be a distance $2c$ apart, and the sum of the distances be $2a$, where $a > c$. If the foci are at $(c, 0)$ and $(-c, 0)$ and $b^2 = a^2 - c^2$, the equation of the ellipse is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

(See Figure C.4.6.)

As for the parabola, one shows $\tan(A) = \tan(B)$.

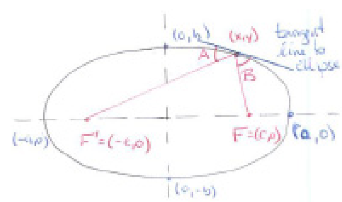


Figure C.4.6

Exercise 3 asks you to carry out the calculation, which uses the same approach as was used for the parabola. One reason to do Exercise 3 is to appreciate more fully the power of vector calculus, developed in Chapter 15, for with that tool you can establish the reflection property of either the parabola or the ellipse in one line.

Diocles, in his book *On Burning Mirrors*, written around 190 B.C., studied spherical and parabolic reflectors, both of which had been considered by earlier writers. Some had thought that a spherical reflector focuses incoming light at a single point. This is false, and Diocles showed that a spherical reflector subtending an angle of 60° reflects light that is parallel to its axis of symmetry to points on the axis that occupy about one-thirteenth of the radius. He proposed an experiment, "Perhaps you would like to make two examples of a burning-mirror, one spherical, one parabolic, so that you can measure the burning power of each." Though the reflection property of a parabola was already known, *On Burning Mirrors* contains the first known proof.

Exercise 4 shows that a spherical oven is fairly effective. After all, a potato or hamburger is not a point.

EXERCISES

1. Do the algebra to complete the proof that $\tan(A) = \tan(B)$ in the case of the parabola.

2. Let C be an angle with $\frac{-\pi}{2} < C < \frac{\pi}{2}$. Show that $\tan(\frac{\pi}{2} - C) = \tan(C)$.

3. This exercise establishes the reflection property of an ellipse. Refer to Figure C.4.6 for the meaning of the notation.

(a) Find the slope of the tangent line at (x, y).

(b) Find the slope of the line through $F = (c, 0)$ and (x, y).

(c) Find $\tan(B)$.

(d) Find the slope of the line through $F' = (c', 0)$ and (x, y).

(e) Find $\tan(A)$.

(f) Check that $\tan(A) = \tan(B)$.

4. Use trigonometry to show that a spherical mirror of radius r and subtending an angle of 60° causes light parallel to its axis of symmetry to reflect and meet the axis in an interval of length $\left(\frac{1}{\sqrt{3}} - \frac{1}{2}\right) r \approx r/12.9$. (See Figure C.4.7.)

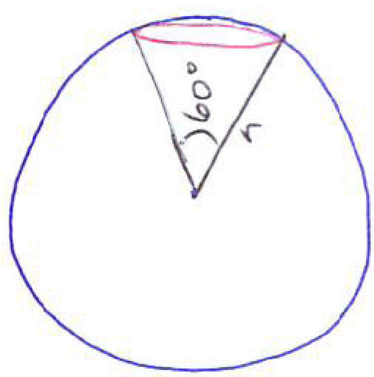


Figure C.4.7

Chapter 4

Derivatives and Curve Sketching

When you graph a function you typically plot a few points and connect them with (generally) straight line segments. Most electronic graphing devices use the same approach, and obtain better results by plotting more points and using shorter segments. The more points used, the smoother the graph will appear. This chapter will show you how to choose the key points.

Three properties of the derivative developed in Section 4.1, and proved in Section 4.4, are used in Section 4.2 to help graph a function. In Section 4.3 we see what the second derivative tells about a graph.

4.1 Three Theorems about the Derivative

This section is based on plausible observations about the graphs of differentiable functions, which we restate as theorems. These ideas will then be combined, in Section 4.2, to sketch graphs of functions.

An effective approach to sketching graphs of functions is to find the extreme values of the function, that is, where the function takes on its largest and smallest values.

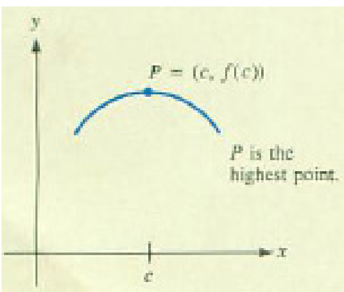


Figure 4.1.1

OBSERVATION *(Tangent Line at an Extreme Value)* Suppose that a function $f(x)$ attains its largest value when $x = c$, that is, $f(c)$ is the largest value of $f(x)$ over a given open interval that contains c. Figure 4.1.1 illustrates this. The maximum occurs at a point $(c, f(c))$, which we call P. If $f(x)$ is differentiable at c, then the tangent line at P will exist. What can we say about it?

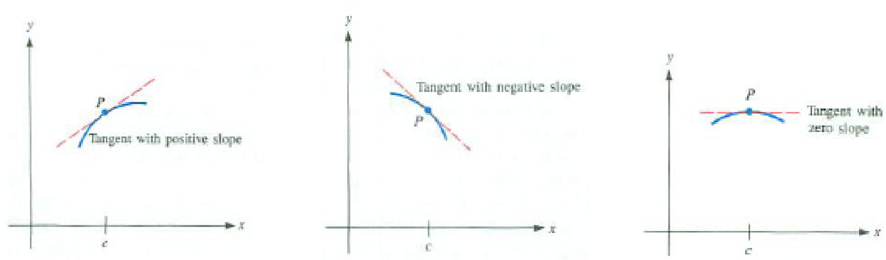


Figure 4.1.2

If the tangent at P were *not* horizontal (that is, not parallel to the x-axis), then it would be tilted. So a small piece of the graph around P which appears to be almost straight — would look as shown in Figure 4.1.2(a) or (b).

In the first case P could not be the highest point on the curve because there would be higher points *to the right* of P. In the second case P could not be the highest point because there would be higher points *to the left* of P. Therefore the *tangent at P must be horizontal*, as shown in Figure 4.1.2(c). That is, $f'(c) = 0$.

This observation suggests a simple criterion for identifying local extrema.

Theorem of the Interior Extremum

Theorem 4.1.1 (Theorem of the Interior Extremum)**.** *Let f be a function defined at least on the open interval (a, b). If f takes on an extreme value at a number c in this interval, then either*

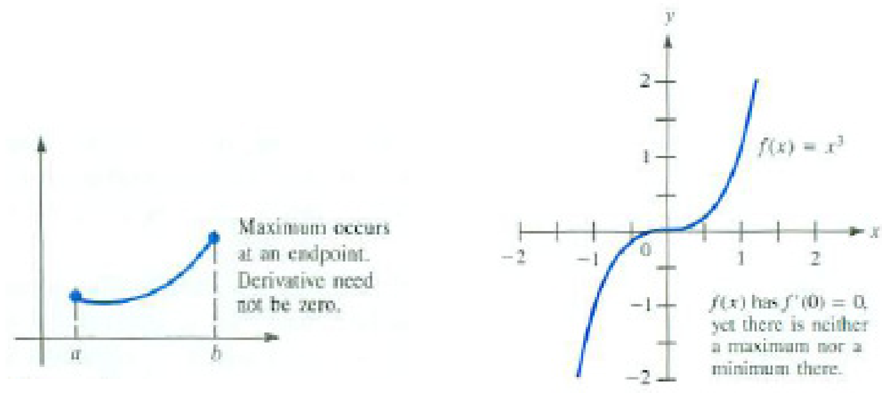
1. $f'(c) = 0$ *or*

2. $f'(c)$ *does not exist.*

If an extreme value occurs within an open interval and the derivative exists there, the derivative must be 0 there. This idea will be used in Section 4.2 to find the maximum and minimum values of a function.

WARNING *(Two Cautions about Theorem 4.1.1)*

1. If in Theorem 4.1.1 the open interval (a, b) is replaced by a closed interval $[a, b]$ the conclusion may not hold. A glance at Figure 4.1.3(a) shows why — the extreme value could occur at an endpoint ($x = a$ or $x = b$).



(a) (b)

Figure 4.1.3

2. The converse of Theorem 4.1.1 is not true. Having the derivative equal to 0 at a point does *not* guarantee that there is an extremum at this point. The graph of $y = x^3$, Figure 4.1.3(b), shows why. Since $f'(x) = 3x^2$, $f'(0) = 0$. While the tangent line is indeed horizontal at $(0, 0)$, it crosses the curve at this point. The graph has neither a maximum nor a minimum at the origin.

Though the next observation is phrased in terms of slopes, we will see that it has implications for velocity and any changing quantity.

OBSERVATION *(Chord and Tangent Line with Same Slope)* Let $A = (a, f(a))$ and $B = (b, f(b))$ be two points on the graph of a differentiable function f defined at least on the interval $[a, b]$, as

A line segment that joins two points on the graph of a function f is called a **chord** of f.

shown in Figure 4.1.4(a). Draw the line segment AB joining A and B. Assume part of the graph lies above that line. Imagine holding a ruler parallel to AB and lowering it until it just touches the graph of $y = f(x)$, as in Figure 4.1.4(b). The ruler touches the

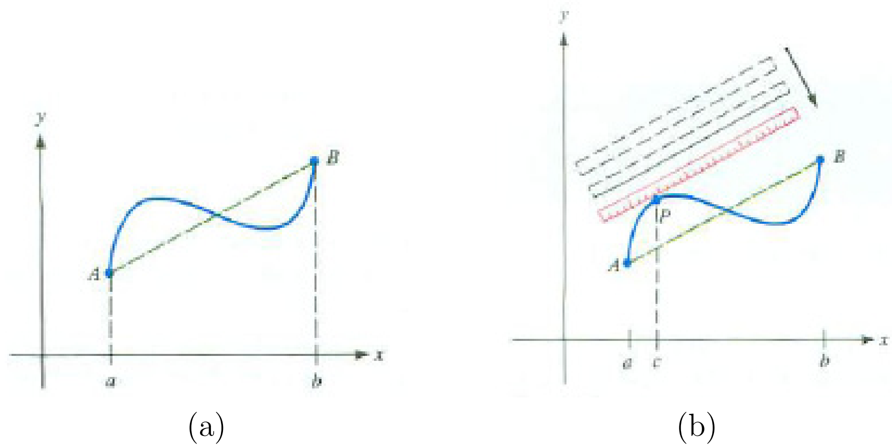


Figure 4.1.4

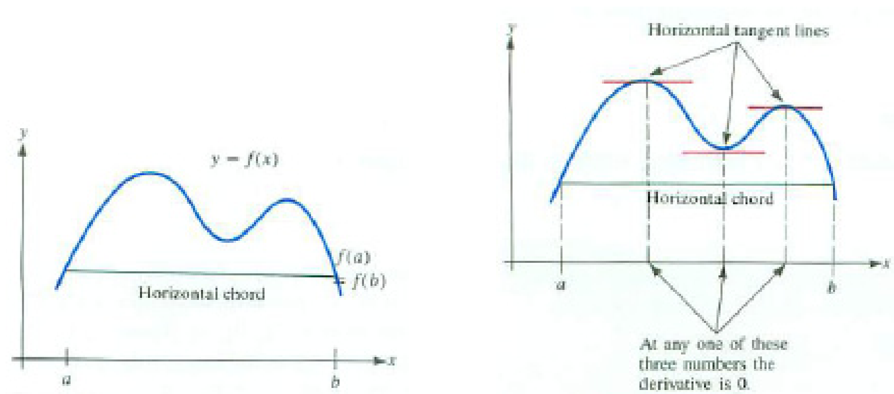
curve at a point P and lies along the tangent at P. At that point $f'(c)$ is equal to the slope of AB. (In Figure 4.1.4(b) there is a second number between a and b where the tangent line to $y = f(x)$ is parallel to the chord between A and B.)

It is customary to state two separate theorems based on the observation about chords and tangent lines. The first, Rolle's Theorem, is a special case of the second, the Mean-Value Theorem.

Rolle's Theorem

The next theorem is suggested by a special case of the second observation. When the points A and B in Figure 4.1.4(a) have the same y coordinate, the chord AB has slope 0. (See Figure 4.1.5.) In this case, the observation tells us there must be a horizontal tangent to the graph. Expressed in terms of derivatives, this suggests Rolle's Theorem[0]

[0]Michel Rolle (1652–1719) was a French mathematician and an early critic of calculus before later changing his opinion. In addition to his discovery of Rolle's Theorem in 1691, he is the first person known to have placed the index in the opening of a radical to denote the n^{th} root of a number: $\sqrt[n]{x}$. Source: Cajori, A History of Mathematical Notation, Dover Publ., 1993 and `http://en.wikipedia.org/wiki/Michel_Rolle`.



(a) (b)

Figure 4.1.5

Theorem 4.1.2 (Rolle's Theorem)**.** *Let f be a continuous function on the closed interval $[a,b]$ and have a derivative at all x in the open interval (a,b). If $f(a)=f(b)$, then there is at least one number c in (a,b) such that $f'(c)=0$.*

EXAMPLE 1 Verify Rolle's Theorem for the case with $f(t)=(t^2-1)\ln\left(\frac{t}{\pi}\right)$ on $[1,\pi]$.

SOLUTION The function $f(t)$ is defined for $t>0$ and is differentiable. In particular, $f(t)$ is differentiable on the closed interval $[1,\pi]$. Notice that $f(1)=0$ and, because $\ln(1)=0$, $f(\pi)=0$. Therefore, by Rolle's Theorem, there must be a value of c between 1 and π where $f'(c)=0$.

The derivative $f'(t)=2t\ln\left(\frac{t}{\pi}\right)+\frac{t^2-1}{t}$ is a pretty complicated function. Even though it is not possible to find the exact value of c with $f'(c)=0$, Rolle's Theorem guarantees that there is at least one such number c. Figure 4.1.6 indicates that there is only one solution to $f'(c)=0$ on $[1,\pi]$. In Exercise 3 (at the end of Chapter 10 on page 907) you will find that this critical number is approximately 2.128. ◇

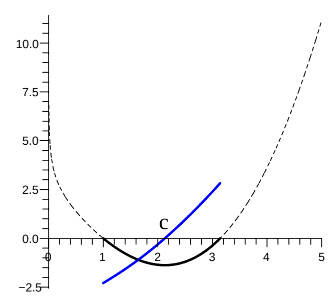


Figure 4.1.6 Graph of $y=f(t)$ (black) and $y=f'(t)$ (blue).

> **Remark:** Assume that $f(x)$ is a differentiable function such that $f'(x)$ is never 0 for x in an interval. Then the equation $f(x)=0$ *can have at most one solution in that interval.* See Figure 4.1.7. (If it had two solutions, a and b, then $f(a)=0$ and $f(b)=0$, and we could apply Rolle's Theorem on $[a,b]$.)

This justifies the observation:

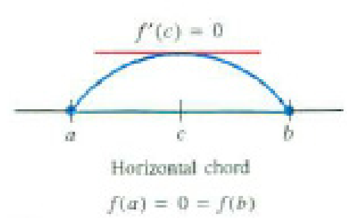


Figure 4.1.7

In an interval in which the derivative $f'(x)$ is never 0, the graph of $y=f(x)$ can have no more than one x-intercept.

Example 2 applies this.

EXAMPLE 2 Use Rolle's Theorem to determine how many real roots there are for the equation

$$x^3 - 6x^2 + 15x + 3 = 0. \tag{4.1.1}$$

SOLUTION Recall that the Intermediate Value Theorem guarantees that an odd degree polynomial, $f(x)$, such as $f(x) = x^3 - 6x^2 + 15x + 3$, has at least one real solution to the equation $f(x) = 0$. Call it r. Could there be another root, s? If so, by Rolle's Theorem, there would be a number c (between r and s) at which $f'(c) = 0$.

To check, we compute the derivative of $f(x)$ and see if it is ever equal to 0. We have $f'(x) = 3x^2 - 12x + 15$. To find when $f'(x)$ is 0, we solve the equation $3x^2 - 12x + 15 = 0$ by the quadratic formula, obtaining

$$x = \frac{-(-12) \pm \sqrt{(-12)^2 - 4(3)(15)}}{6} = \frac{12 \pm \sqrt{-36}}{6} = 2 \pm \sqrt{-1}.$$

Thus the equation $x^3 - 6x^2 + 15x + 3$ has only one real root. In Exercise 4 (at the end of Chapter 10) you will find that the only real solution to (4.1.1) in approximately -0.186. ◇

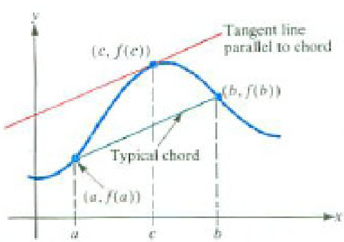


Figure 4.1.8

Mean-Value Theorem

The "mean-value" theorem is a generalization of Rolle's Theorem in that it applies to any chord, not just horizontal chords.

In geometric terms, the theorem asserts that if you draw a chord for the graph of a well-behaved function (as in Figure 4.1.8), then somewhere above or below that chord the graph has at least one tangent line parallel to the chord. (See Figure 4.1.4(a).) Let us translate this geometric statement into the language of functions. Call the ends of the chord $(a, f(a))$ and $(b, f(b))$. The slope of the chord is

$$\frac{f(b) - f(a)}{b - a}.$$

Since the tangent line and the chord are parallel, they have the same slopes. If the tangent line is at the point $(c, f(c))$, then

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

This observation suggests

Theorem 4.1.3 (Mean-Value Theorem). *Let f be a continuous function on the closed interval $[a,b]$ and have a derivative at every x in the open interval (a,b). Then there is at least one number c in the open interval (a,b) such that*

$$f'(c) = \frac{f(b)-f(a)}{b-a}.$$

EXAMPLE 3 Verify the Mean-Value Theorem for $f(t) = \sqrt{4-t^2}$ on the interval $[0,2]$.

SOLUTION Because $4-t^2 \geq 0$ for t between -2 and 2 (including these two endpoints), f is continuous on $[0,2]$ and is differentiable on $(0,2)$. The slope of the chord through $(a,f(a)) = (0,2)$ and $(b,f(b)) = (2,0)$ is

$$\frac{f(b)-f(a)}{b-a} = \frac{0-2}{2-0} = -1.$$

According to the Mean-Value Theorem, there is at least one number c between 0 and 2 where $f'(c)$ is -1.

Let us try to find c. Since $f'(t) = \dfrac{-2t}{2\sqrt{4-t^2}}$, we need to solve the equation

$$\begin{array}{rcll}
\dfrac{-c}{\sqrt{4-c^2}} & = & -1 & \\
-c & = & -\sqrt{4-c^2} & \text{multiply both sides by } \sqrt{4-c^2} \\
c^2 & = & 4-c^2 & \text{square both sides} \\
2c^2 & = & 4 & \text{add } c^2 \text{ to both sides} \\
c^2 & = & 2 & \text{divide both sides by 2.}
\end{array}$$

While both $c = \sqrt{2}$ and $c = -\sqrt{2}$ satisfy $c^2 - 2$, $f'(\sqrt{2}) = -1$, only $c = \sqrt{2}$ lies in $[0,2]$. The other root, $c = -\sqrt{2}$, is of no interest for two reasons: it is not in $[0,2]$ and $f(-\sqrt{2})$ is not -1. Thus $c = \sqrt{2}$ is the number in $[0,2]$ whose existence is guaranteed by the Mean-Value Theorem. ◇

The interpretation of the derivative as slope suggested the Mean-Value Theorem. What does the Mean-Value Theorem say when the function describes the position of a moving object, and the derivative, its velocity? This is answered in Example 4.

EXAMPLE 4 A car moving on the x-axis has the x-coordinate $x = f(t)$ at time t. At time a its position is $f(a)$. At some later time b its position is $f(b)$. What does the Mean-Value Theorem assert for this car?

SOLUTION In this case the quotient

$$\frac{f(b)-f(a)}{b-a} \quad \text{equals} \quad \frac{\text{Change in position}}{\text{Change in time}}.$$

The Mean-Value Theorem asserts that at some time c, $f'(c)$ is equal to the quotient $\dfrac{f(b)-f(a)}{b-a}$. This says that the velocity at time c is the same as the average velocity during the time interval $[a, b]$. To be specific, if a car travels 210 miles in 5 hours, then at some time its speedometer must read 42 miles per hour. $\diamond$

Consequences of the Mean-Value Theorem

There are several ways of writing the Mean-Value Theorem. For example, the equation

$$f'(c) = \frac{f(b)-f(a)}{b-a}$$

is equivalent to

$$f(b) - f(a) = (b-a)f'(c)$$

and hence to

$$f(b) = f(a) + (b-a)f'(c).$$

In this last form, the Mean-Value Theorem asserts that $f(b)$ is equal to $f(a)$ plus a quantity that involves the derivative f' at some number c between a and b. The following important corollaries are based on this alternative view of the Mean-Value Theorem.

Corollary 4.1.4. *If the derivative of a function is 0 throughout an interval I, then the function is constant on the interval.*

Proof

Let a and b be any two numbers in the interval I and let the function be denoted by f. To prove this corollary, it suffices to prove that $f(a) = f(b)$, for that is the defining property of a constant function.

By the Mean-Value Theorem in the form (2), there is a number c between a and b such that

$$f(b) = f(a) + (b-a)f'(c).$$

But $f'(c) = 0$, since $f'(x) = 0$ for all x in I. Hence

$$f(b) = f(a) + (b-a)(0)$$

which proves that

$$f(b) = f(a).$$

•

When Corollary 4.1.4 is interpreted in terms of motion, it is quite plausible. It asserts that if an object has zero velocity for a period of time, then it does not move during that time.

EXAMPLE 5 Use calculus to show that $f(x) = (e^x + e^{-x})^2 - e^{2x} - e^{-2x}$ is a constant. Find the constant.

SOLUTION The function f is differentiable for all numbers x. Its derivative is

$$\begin{aligned} f'(x) &= 2(e^x + e^{-x})(e^x - e^{-x}) - 2e^{2x} + 2e^{-2x} \\ &= 2(e^{2x} - e^{-2x}) - 2e^{2x} + 2e^{-2x} \\ &= 0 \end{aligned}$$

Because $f'(x)$ is always zero, f must be a constant.

To find the constant, just evaluate $f(x)$ for any convenient value of x. For simplicity we choose $x = 0$: $f(0) = (e^0 + e^0)^2 - e^0 - e^0 = 2^2 - 2 = 2$. Thus,

$$(e^x + e^{-x})^2 - e^{2x} - e^{-2x} = 2 \qquad \text{for all numbers } x.$$

This result can also be obtained by squaring $e^x + e^{-x}$. ⋄

Corollary 4.1.5. *If two functions have the same derivatives throughout an interval, then they differ by a constant. That is, if $F'(x) = G'(x)$ for all x in an interval, then there is a constant C such that $F(x) = G(x) + C$.*

Proof

Define a third function h by the equation $h(x) = F(x) - G(x)$. Then, because $F'(x) = G'(x)$,

$$h'(x) = F'(x) - G'(x) = 0.$$

Since the derivative of h is 0, Corollary 4.1.4 implies that h is constant, that is, $h(x) = C$ for some fixed number C. Thus

$$F(x) - G(x) = C \quad \text{or} \quad F(x) = G(x) + C,$$

and Corollary 4.1.5 is proved. •

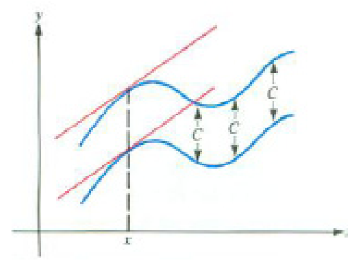


Figure 4.1.9

Is Corollary 4.1.5 plausible when the derivative is interpreted as slope? In this case, the corollary asserts that if the graphs of two functions have the property that their tangent lines at points with the same x coordinate are parallel, then one graph can be obtained from the other by raising (or lowering) it by a constant amount C. If you sketch two such graphs (as in Figure 4.1.9), you will see that the corollary is reasonable.

EXAMPLE 6 What functions have a derivative equal to $2x$ everywhere?

SOLUTION One such solution is x^2; another is $x^2 + 25$. For any constant

In the language of Section 3.5, any antiderivative of $2x$ must be of the form $x^2 + C$.

C, $D(x^2+C) = 2x$. Are there any other possibilities? Corollary 4.1.5 tells us there are not, for if F is a function such that $F'(x) = 2x$, then $F'(x) = (x^2)'$ for all x. Thus the functions F and x^2 differ by a constant, say C, that is,

$$F(x) = x^2 + C.$$

The only antiderivatives of $2x$ are of the form $x^2 + C$. ◇

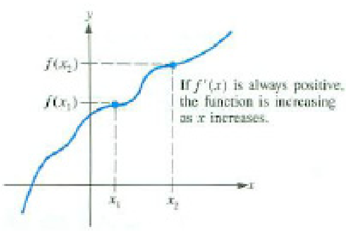


Figure 4.1.10

Corollary 4.1.4 asserts that if $f'(x) = 0$ for all x, then f is a constant. What can be said about f if $f'(x)$ is *positive* for all x in an interval? In terms of the graph of f, this assumption implies that all the tangent lines slope upward. It is reasonable to expect that as we move from left to right on the graph in Figure 4.1.10, the y-coordinate increases, that is, the function is increasing. (See Section 1.1.)

Corollary 4.1.6. *If f is continuous on the closed interval $[a, b]$ and has a positive derivative on the open interval (a, b), then f is increasing on the interval $[a, b]$.*

If f is continuous on the closed interval $[a, b]$ and has a negative derivative on the open interval (a, b), then f is decreasing on the interval $[a, b]$.

Proof

We prove the "increasing" case; the other case is handled in Exercise 44. Take two numbers x_1 and x_2 such that

$$a \le x_1 < x_2 \le b.$$

The goal is to show that $f(x_2) > f(x_1)$.

By the Mean-Value Theorem, there is some number c between x_1 and x_2 such that

$$f(x_2) = f(x_1) + (x_2 - x_1)f'(c).$$

Now, since $x_2 > x_1$, we know $x_2 - x_1$ is positive. Since $f'(c)$ is assumed to be positive, and the product of two positive numbers is positive, it follows that

$$(x_2 - x_1)f'(c) > 0.$$

Thus, $f(x_2) > f(x_1)$, and so $f(x)$ is an increasing function. •

EXAMPLE 7 Determine whether $2x + \sin(x)$ is an increasing function, a decreasing function, or neither.

SOLUTION The funcion $2x + \sin(x)$ is the sum of two simpler functions: $2x$ and $\sin(x)$. The "$2x$" part is an increasing function. The second term, "$\sin(x)$", increases for x between 0 and $\pi/2$ and decreases for x between $\pi/2$

and π. It is not clear what type of function you will get when you add $2x$ and $\sin(x)$. Let's see what Corollary 4.1.6 tells us.

The derivative of $2x + \sin(x)$ is $2 + \cos(x)$. Since $\cos(x) \geq -1$ for all x,

$$(2x + \sin(x))' = 2 + \cos(x) \geq 2 + (-1) = 1.$$

Because $(2x+\sin(x))'$ is positive for all numbers x, $2x+\sin(x)$ is an increasing function. Figure 4.1.11 shows the graph of $2x+\sin(x)$ together with the graphs of $2x$ and $\sin(x)$. ⋄

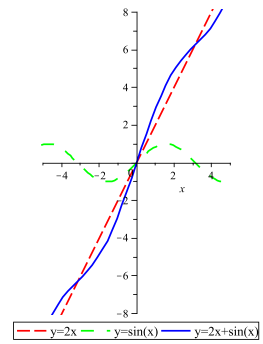


Figure 4.1.11

Remark: Increasing/Decreasing at a Point

1. Corollary 4.1.6, and the definitions of increasing and decreasing, are stated in terms of intervals. When we talk about a function f increasing (or decreasing) "at a point c," we mean *there is an interval* (a, b) *with* $a < c < b$ *where* f *is increasing.* In addition, "a function is increasing at c" is shorthand for "a function is increasing in an interval that contains c."
2. When $f'(c) > 0$ and f' is continuous, the Permanence Property (in Section 2.5) tells us there is an interval (a, b) containing c where $f'(x)$ remains positive for all numbers x in (a, b). Thus, f is increasing on (a, b), and hence increasing at c.

More generally, if $f'(x)$ is never negative, that is $f'(x) \geq 0$ for all inputs x, then f is non-decreasing. In the same manner, if $f'(x) \leq 0$ for all inputs x, then f is a non-increasing function.

Summary

This section focused on three theorems, which we state informally. For the assumptions on the functions, see the formal statements in this section.

The Theorem of the Interior Extremum says that at a local extreme the derivative must be zero. (The converse is not true.)

Rolle's Theorem asserts that if a function has equal values at two inputs, its derivative must equal zero at least at one number between these inputs. The Mean-Value Theorem, a generalization of Rolle's Theorem, asserts that for any chord on the graph of a function, there is a tangent line parallel to it. This means that for $a < b$ there is c in (a, b) such that $f'(c) = \frac{f(b)-f(a)}{b-a}$, or in a more useful form $f(b) = f(a) + f'(c)(b - a)$.

From the Mean-Value Theorem it follows that where a derivative is positive, a function is increasing; where it is negative it is decreasing; and where it stays at the value zero, it is constant. The last assertion implies that two antiderivatives of the same function differ by a constant (which may be zero).

EXERCISES for Section 4.1

1. State Rolle's Theorem in words, using as few mathematical symbols as you can.

2. Draw a graph illustrating Rolle's Theorem. Be sure to identify the critical parts of the graph.

3. Draw a graph illustrating the Mean-Value Theorem. Be sure to identify the critical parts of the graph.

4. Express the Mean-Value Theorem in words, using no symbols to denote the function or the interval.

5. Express the Mean-Value Theorem in symbols, where the function is denoted g and the interval is $[e, f]$.

6. Which of the corollaries to the Mean-Value Theorem implies that

(a) if two cars on a straight road have the same velocity at every instant, they remain a fixed distance apart?

(b) If all tangents to a curve are horizontal, the curve is a horizontal line.

Explain each answer.

Exercises 7 to 12 concern the Theorem of the Interior Extremum.

7. Consider the function $f(x) = x^2$ only for x in $[-1, 2]$.

(a) Graph $f(x)$ for x in $[-1, 2]$.

(b) What is the maximum value of $f(x)$ for x in the interval $[-1, 2]$?

(c) Does $f'(x)$ exist at the maximum?

(d) Does $f'(x)$ equal zero at the maximum?

(e) Does $f'(x)$ equal zero at the minimum?

8. Consider $f(x) = \sin(x)$ only for x in $[0, \pi]$.

(a) Graph the function $f(x)$ for x in $[0, \pi]$.

(b) What is the maximum value of $f(x)$ for x in the interval $[0, \pi]$?

(c) Does $f'(x)$ exist at the maximum?

(d) Does $f'(x)$ equal zero at the maximum?

(e) Does $f'(x)$ equal zero at the minimum?

9.

(a) Repeat Exercise 7 on the interval $[1, 2]$.

(b) Repeat Exercise 7 on the interval $(-1, 2)$.

(c) Repeat Exercise 7 on the interval $(1, 2)$.

(d) Repeat Exercise 8 on the interval $[0, 2\pi]$.

(e) Repeat Exercise 8 on the interval $(0, \pi)$.

(f) Repeat Exercise 8 on the interval $(0, 2\pi)$.

10.

(a) Graph $y = -x^2 + 3x + 2$ for x in $[0, 2]$.

(b) Looking at the graph, estimate the x coordinate where the maximum value of y occurs for x in $[0, 2]$.

(c) Find where $dy/dx = 0$.

(d) Using (c), determine exactly where the maximum occurs.

11.

(a) Graph $y = 2x^2 - 3x + 1$ for x in $[0, 1]$.

(b) Looking at the graph, estimate the x coordinate where the maximum value of y occurs for x in $[0, 1]$. At which value of x does it occur?

(c) Looking at the graph, estimate the x coordinate where the minimum value of y occurs for x in $[0, 1]$.

(d) Find where $dy/dx = 0$.

(e) Using (d), determine exactly where the minimum occurs.

12. For each of the following functions, (a) show that the derivative of the given function is 0 when $x = 0$ and (b) decide whether the function has an extremum at $x = 0$.

(a) $x^2 \sin(x)$

(b) $1 - \cos(x)$

(c) $e^x - x$

(d) $x^2 - x^3$

Exercises 13 to 21 concern Rolle's Theorem.

13.

(a) Graph $f(x) = x^{2/3}$ for x in $[-1, 1]$.

(b) Show that $f(-1) = f(1)$.

(c) Is there a number c in $(-1, 1)$ such that $f'(c) = 0$?

(d) Why does this not contradict Rolle's Theorem?

14.

(a) Graph $f(x) = 1/x^2$ for x in $[-1, 1]$.

(b) Show that $f(-1) = f(1)$.

(c) Is there a number c in $(-1, 1)$ such that $f'(c) = 0$?

(d) Why does this not contradict Rolle's Theorem?

In Exercises 15 to 20, verify that the given function satisfies Rolle's Theorem for the given interval. Find all numbers c that satisfy the conclusion of the theorem.

15. $f(x) = x^2 - 2x - 3$ and $[0, 2]$

16. $f(x) = x^3 - x$ and $[-1, 1]$

17. $f(x) = x^4 - 2x^2 + 1$ and $[-2, 2]$

18. $f(x) = \sin(x) + \cos(x)$ and $[0, 4\pi]$

19. $f(x) = e^x + e^{-x}$ and $[-2, 2]$

20. $f(x) = x^2 e^{-x^2}$ and $[-2, 2]$

21. Let $f(x) = \ln(x^2)$. Note that $f(-1) = f(1)$. Is there a number c in $(-1, 1)$ such that $f'(c) = 0$? If so, find at least one such number. If not, why is this not a contradiction of Rolle's Theorem?

Exercises 22 to 27 concern the Mean-Value Theorem. In Exercises 22 to 25, find explicitly all values of c which satisfy the Mean-Value Theorem for the given functions and intervals.

22. $f(x) = x^2 - 3x$ and $[1, 4]$

23. $f(x) = 2x^2 + x + 1$ and $[-2, 3]$

24. $f(x) = 3x + 5$ and $[1, 3]$

25. $f(x) = 5x - 7$ and $[0, 4]$

26.

(a) Graph $y = \sin(x)$ for x in $[\pi/2, 7\pi/2]$.

(b) Draw the chord joining $(\pi/2, f(\pi/2))$ and $(7\pi/2, f(7\pi/2))$.

(c) Draw all tangents to the graph parallel to the chord drawn in (b).

(d) Using (c), determine how many numbers c there are in $(\pi, 7\pi/2)$ such that

$$f'(c) = \frac{f(7\pi/2) - f(\pi/2)}{7\pi/2 - \pi/2}.$$

(e) Use the graph to estimate the values of the c's.

27.

(a) Graph $y = \cos(x)$ for x in $[0, 9\pi/2]$.

(b) Draw the chord joining $(0, f(0))$ and $(9\pi/2, f(9\pi/2))$.

(c) Draw all tangents to the graph that are parallel to the chord drawn in (b).

(d) Using (c), determine how many numbers c there are in $(0, 9\pi/2)$ such that

$$f'(c) = \frac{f(9\pi/2) - f(0)}{9\pi/2 - 0}.$$

(e) Use the graph to estimate the values of the c's.

28. At time t seconds a thrown ball has the height $f(t) = -16t^2 + 32t + 40$ feet.

(a) What is the initial height? That is, the height when t is zero.

(b) Show that after 2 seconds it returns to its initial height.

(c) What does Rolle's Theorem imply about the velocity of the ball?

(d) Verify Rolle's Theorem in this case by computing the numbers c which it asserts exist.

29. Find all points where $f(x) = 2x^3(x-1)$ can have an extreme value on the following intervals

(a) $(-1/2, 1)$

(b) $[-1/2, 1]$

(c) $[-1/2, 1/2]$

(d) $(-1/2, 1/2)$

30. Let $f(x) = |2x-1|$.

(a) Explain why $f'(1/2)$ does not exist.

(b) Find $f'(x)$. (Write the absolute value in two parts, one for $x < 1/2$ and the other for $x > 1/2$.)

(c) Does the Mean-Value Theorem apply on the interval $[-1, 2]$?

31. The year is 2015. Because a gallon of gas costs six dollars and Highway 80 is full of tire-wrecking potholes, the California Highway Patrol no longer patrols the 77 miles between Sacramento and Berkeley. Instead it uses two cameras. One, in Sacramento, records the license number and time of a car on the freeway, and another does the same in Berkeley. A computer processes the data instantly. Assume that the two cameras show that a car that was in Sacramento at 10:45 reached Berkeley at 11:40. Show that the Mean-Value Theorem justifies giving the driver a ticket for exceeding the 70 mile-per-hour speed limit. (Of course, intuition justifies the ticket, but mentioning the Mean-Value Theorem is likely to impress a judge who studied calculus.)
While it makes a nice story to suggest that mentioning the Mean-Value Theorem will impress a judge who studied calculus, reality is that the California Vehicle Code forbids this way to catch speeders. It reads, "No speed trap shall be used in securing

evidence as to the speed of any vehicle. A 'speed trap' is a particlar section of highway measured as to distance in order that the speed of a vehicle may be calculated by securing the time it takes the vehicle to travel the known distance." It sounds as though the lawmakers who wrote this law studied calculus.

32. What is the shortest time for the trip from Berkeley to Sacramento for which the Mean-Value Theorem does not convict the driver of speeding? See Exercise 31.

33. Verify the Mean-Value Theorem for $f(t) = x^2 e^{-x/3}$ on $[1, 10]$. See Example 1.

34. Find all antiderivatives of each of the following functions. Check your answer by differentiation.

(a) $3x^2$

(b) $\sin(x)$

(c) $\frac{1}{1+x^2}$

(d) e^x

35. Find all antiderivatives of each of the following functions. Check your answer by differentiation.

(a) $\cos(x)$

(b) $\sec(x)\tan(x)$

(c) $1/x$ $(x > 0)$

(d) $\sqrt{x}$ $(x > 0)$

36.

(a) Differentiate $\sec^2(x)$ and $\tan^2(x)$.

(b) The derivatives in (a) are equal. Corollary 4.1.5 then asserts that there exists a constant C such that $\sec^2(x) = \tan^2(x) + C$. Find the constant.

37. Show by differentiation that $f(x) = \ln(x/5) - \ln(5x)$ is a constant for all positive x. Find the constant.

38. Find all functions whose second derivative is 0 for all x in $(-\infty, \infty)$.

39. Use Rolle's Theorem to determine how many real roots there are for the equation $x^3 - 6x^2 + 15x + 3 = 0$.

40. Use Rolle's Theorem to determine how many real roots there are for the equation $3x^4+4x^3-12x^2+4=0$. Give intervals on which there is exactly one root.

41. Use Rolle's Theorem to determine how many real roots there are for the polynomial $f(x) = 3x^4 + 4x^3 - 12x^2 + A$. That number may depend on A. For which A is there exactly one root? Are there any values of A for which there is an odd number of real roots? Exercise 40 uses this equation with $A = 4$.

42. Consider the equation $x^3 - ax^2 + 15x + 3 = 0$. The number of real roots to this equation depends on the value of a.

(a) Find all values of a when the equation has 3 real roots.

(b) Find all values of a when the equation has 1 real root.

(c) Are there any values of a with exactly two real roots?

Exercise 39 uses this equation with $a = 6$.

43. If f is differentiable for all real numbers and $f'(x) = 0$ has three solutions, what can be said about the number of solutions of $f(x) = 0$? of $f(x) = 5$?

44. Prove the "decreasing" case of Corollary 4.1.6.

45. For which values of the constant k is the function $7x + k\sin(2x)$ always increasing?

46. If two functions have the same second derivative for all x in $(-\infty, \infty)$, what can be said about the relation between them?

47. If a function f is differentiable for all x and c is a number, is there necessarily a chord of the graph of f that is parallel to the tangent line at $(c, f(c))$? Explain.

48. Sketch a graph of a continuous function $f(x)$ defined for all numbers such that $f'(1)$ is 2, yet there is no open interval around 1 on which f is increasing.

Exercises 49 to 52 involve the **hyperbolic functions**. The **hyperbolic sine** function is $\sinh(x) = \frac{e^x - e^{-x}}{2}$ and the **hyperbolic cosine** function is $\cosh(x) = \frac{e^x + e^{-x}}{2}$. Hyperbolic functions are discussed in greater detail in Section 5.8.

49.

(a) Show that $\frac{d}{dx}\sinh(x) = \cosh(x)$.

(b) Show that $\frac{d}{dx}\cosh(x) = \sinh(x)$.

50. Define $\operatorname{sech}(x) = \dfrac{1}{\cosh(x)} = \dfrac{2}{e^x + e^{-x}}$ and $\tanh(x) = \dfrac{\sinh(x)}{\cosh(x)} = \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$.

(a) Show that $\frac{d}{dx}\tanh(x) = (\operatorname{sech}(x))^2$.

(b) Show that $\frac{d}{dx}\operatorname{sech}(x) = -\operatorname{sech}(x)\tanh(x)$.

51. Use calculus to show that $(\cosh(x))^2 - (\sinh(x))^2$ is a constant. Find the constant.

52. Use calculus to show that $(\operatorname{sech}(x))^2 + (\tanh(x))^2$ is a constant. Find the constant.

4.2 The First-Derivative and Graphing

Section 4.1 showed the connection between extrema and the places where the derivative is zero. In this section we use this connection to find high and low points on a graph.

The graph of a differentiable function f is shown in Figure 4.2.1. The points P, Q, R, and S are of special interest. S is the highest point on the graph *for all x in the domain*. We call it a *global maximum* or *absolute maximum*. The point P is higher than all points *near it* on the graph; it is called a *local maximum* or *relative maximum*. Similarly, Q is called a *local minimum* or *relative minimum*. The point R is neither a relative maximum nor a relative minimum.

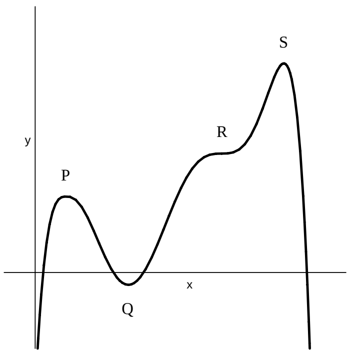


Figure 4.2.1

A point that is either a maximum or minimum is called an **extremum**.

If you were to walk left to right along the graph in Figure 4.2.1, you would call P the top of a hill, Q the bottom of a valley, and S the highest point on your walk (it is also a top of a hill). You might notice R, for you get a momentary break from climbing from Q to S. For just this one instant it would be like walking along a horizontal path.

These important aspects of a function and its graph are made precise in the following definitions which are phrased in terms of a general domain. In most cases the domain of the function will be an interval — open, closed, or half-open.

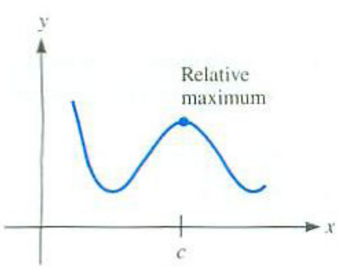


Figure 4.2.2

DEFINITION *(Relative Maximum (Local Maximum))* The function f has a **relative maximum** (or **local maximum**) at a number c if there is an open interval around c such that $f(c) \geq f(x)$ for all x in that interval that lie in the domain of f.

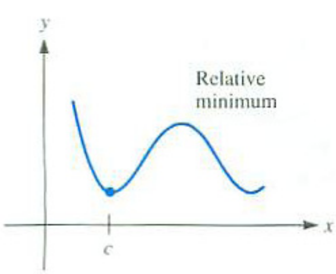


Figure 4.2.3

DEFINITION *(Relative Minimum (Local Minimum))* The function f has a **relative minimum** (or **local minimum**) at a number c if there is an open interval around c such that $f(c) \leq f(x)$ for all x in that interval that lie in the domain of f.

DEFINITION *(Absolute Maximum (Global Maximum))* The function f has an **absolute maximum** (or **global maximum**) at a number c if $f(c) \geq f(x)$ for all x in the domain of f.

Each global extremum is also a local extremum.

DEFINITION *(Absolute Minimum (Global Minimum))* The function f has an **absolute minimum** (or **global minimum**) at a number c if $f(c) \leq f(x)$ for all x in the domain of f.

A local extremum is like the summit of a single mountain or the lowest point in a valley. A global maximum corresponds to Mt. Everest at more than 29,000 feet above sea level; a global minimum corresponds to the Mariana Trench in the Pacific Ocean 36,000 feet below sea level, the lowest point on the Earth's crust.

In this section it is assumed that the functions are differentiable. If a function is not differentiable at an isolated point, this point will need to be considered separately.

DEFINITION *(Critical Number and Critical Point)* A number c at which $f'(c) = 0$ is called a **critical number** for the function f. The corresponding point $(c, f(c))$ on the graph of f is a **critical point**.

Remark: Some texts define a critical number as a number where the derivative is 0 or else is not defined. Since we emphasize differentiable functions, a critical number is defined to be a number where the derivative is 0.

The Theorem of the Interior Extremum, in Section 4.1, says that every local maximum and minimum of a function f occurs where the tangent line to the curve either is horizontal or does not exist.

Some functions have extreme values, and others do not. The following theorem gives simple conditions under which both a global maximum and a global minimum are guaranteed to exist. To convince yourself that this is plausible, imagine drawing the graph of the function. Somewhere your pencil will reach a highest point and elsewhere a lowest point.

Theorem 4.2.1 (Extreme Value Theorem)**.** *Let f be a continuous function on a closed interval $[a,b]$. Then f attains an absolute maximum value $M = f(c)$ and an absolute minimum value $m = f(d)$ at some numbers c and d in $[a,b]$.*

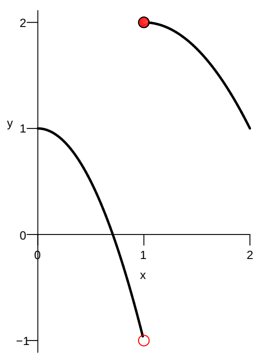


Figure 4.2.4

EXAMPLE 1 Find the absolute extrema on the interval $[0,2]$ of the function whose graph is shown in Figure 4.2.4.
SOLUTION The function has an absolute maximum value of 2 but no absolute minimum value. The range is $(-1,2]$. This function takes on values that are arbitrarily close to -1, but -1 is not in the range of this function. This can occur because the function is not continuous at $x = 1$. ◇

Recall that Corollary 4.1.6 provides a convenient test to determine if a function is increasing or decreasing at a point: if $f'(c) > 0$ then f is increasing at $x = c$ and if $f'(c) < 0$ then f is decreasing at $x = c$.

WARNING Differentiable implies continuous, so "not continuous" implies "not differentiable."

EXAMPLE 2 Let $f(x) = x\ln(x)$ for all $x > 0$. Determine the intervals on which f is increasing, decreasing, or neither.
SOLUTION The function is increasing at numbers x where $f'(x) > 0$ and decreasing where $f'(x) < 0$. More effort is needed to determine the behavior at points where $f'(x) = 0$ (or does not exist). (Observe that the domain of f is $x > 0$.) The Product Rule allows us to find

$$f'(x) = \ln(x) + x\left(\frac{1}{x}\right) = \ln(x) + 1.$$

In order to find where $f'(x)$ is positive or is negative, we first find where it is zero. At such numbers the derivative may switch sign, and the function switch between increasing and decreasing. So we solve the equation:

$$\begin{aligned} f'(x) &= 0 \\ \ln(x) + 1 &= 0 \\ \ln(x) &= -1 \\ e^{\ln(x)} &= e^{-1} \\ x &= e^{-1}. \end{aligned}$$

$e^{-1} \approx 0.367879$

When x is larger than e^{-1}, $\ln(x)$ is larger than -1 so that $f'(x) = \ln(x) + 1$ is positive and f is increasing. Finally, f is decreasing when x is between 0 and

e^{-1} because $\ln(x) < -1$, which makes $f'(x) = \ln(x)+1$ negative. The graph of $y = x\ln(x)$ in Figure 4.2.5 confirms these findings. In addition, observe that $x = e^{-1}$ provides a minimum of $e^{-1}\ln\left(e^{-1}\right) = -1/e$. ◇

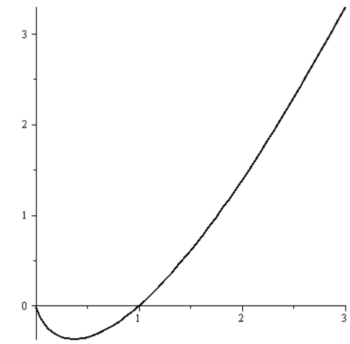


Figure 4.2.5

Using Critical Numbers to Identify Local Extrema

The previous examples show there is a close connection between critical points and local extrema. Notice that, generally, just to the left of a local maximum the function is increasing, while just to the right it is decreasing. The opposite holds for a local minimum. The First-Derivative Test for a Local Extreme Value at $x = c$ gives a precise statement of this result.

First-Derivative Test for a Local Extreme Value at $x = c$

Theorem 4.2.2. *Let f be a function and let c be a number in its domain. Suppose f is continuous on an open interval that contains c and is differentiable on that interval, except possibly at c. Then:*

1. *If f' changes from positive to negative as x moves from left to right through the value c,*
   then f has a local maximum at c.
2. *If f' changes from negative to positive as x moves from left to right through the value c,*
   then f has a local minimum at c.
3. *If f' does not change sign at c,*
   then f does not have a local extremum at $x = c$.

EXAMPLE 3 Classify each critical number of $f(x) = 3x^5 - 20x^3 + 10$ as a local maximum, local minimum, or neither.
SOLUTION To identify the critical numbers of f, we find and factor the derivative:

$$f'(x) = 15x^4 - 60x^2 = 15x^2(x^2 - 4) = 15x^2(x - 2)(x + 2).$$

The critical numbers of f are $x = 0$, $x = 2$, and $x = -2$. To determine if any of these numbers provide local extrema it is necessary to know where f is increasing and where it is decreasing.

Because f' is continuous the three critical numbers are the only places the sign of f' can possibly change. As a result, on each of the intervals $(-\infty, -2)$,

$(-2, 0)$, $(0, 2)$, and $(2, \infty)$, f is either increasing or decreasing; all that remains is to determine which. This is easily answered from the table of function values shown in the first two rows of Table 4.2.1. Observe that $f(-2) = 74 > 10 = f(0)$; this means f is decreasing on $(-2, 0)$. Likewise, f must be decreasing on $(0, 2)$ because $f(0) = 10 > -54 = f(2)$. For the two unbounded intervals, limits at $\pm\infty$ must be used but the overall idea is the same. Since $\lim_{x\to-\infty} f(x) = -\infty$, the function must be increasing on $(-\infty, -2)$. Likewise, in order to have $\lim_{x\to\infty} f(x) = +\infty$, f must be increasing on $(2, \infty)$. (See Figure 4.2.6.)

x	$\to -\infty$	-2	0	2	$\to \infty$
$f(x)$	$\to -\infty$	74	10	-54	$\to \infty$
$f'(x)$		0	0	0	

Table 4.2.1

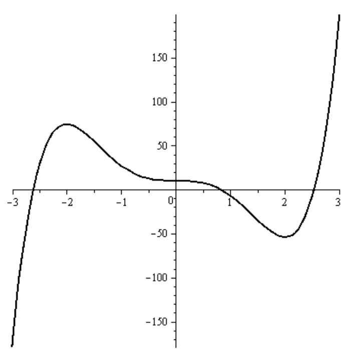


Figure 4.2.6

To conclude, because the graph of f changes from increasing to decreasing at $x = -2$, there is a local maximum at $(-2, 74)$. At $x = 2$ the graph changes from decreasing to increasing, so a local minimum occurs at $(2, -54)$. Because the derivative does not change sign at $x = 0$, this critical number is not a local extreme. ◇

EXAMPLE 4 Find all local extrema of $f(x) = (x + 1)^{2/7}e^{-x}$.

SOLUTION (Observe that the domain of f is $(-\infty, \infty)$.) The Product and Chain Rules for derivatives can be used to obtain

$$\begin{aligned} f'(x) &= \tfrac{2}{7}(x+1)^{-5/7}e^{-x} + (x+1)^{2/7}e^{-x}(-1) \\ &= \tfrac{2}{7}(x+1)^{-5/7}e^{-x} - (x+1)^{2/7}e^{-x} \\ &= (x+1)^{-5/7}e^{-x}\left(\tfrac{2}{7} - (x+1)\right) \\ &= (x+1)^{-5/7}e^{-x}\left(-x - \tfrac{5}{7}\right) \\ &= \tfrac{-x-\frac{5}{7}}{(x+1)^{5/7}e^{x}}. \end{aligned}$$

The only solution to $f'(x) = 0$ is $x = -5/7$, so $c = -5/7$ is the only critical number. In addition, because the denominator of $f'(x)$ is zero when $x = -1$, f is not differentiable for $x = -1$. Using the information in Table 4.2.2, we conclude f is decreasing on $(-\infty, -1)$, increasing on $(-1, -5/7)$, and decreasing on $(-5/7, \infty)$. By the First-Derivative Test, f has a local minimum at $(-1, 0)$ and a local maximum at $(-5/7, (2/7)^{(2/7)}e^{5/7}) \approx (-0.71429, 1.42811)$.

x	$\to -\infty$	-1	$-5/7$	$\to \infty$
$f(x)$	$\to \infty$	0	$(2/7)^{(2/7)}e^{5/7} \approx 1.42811$	$\to 0$
$f'(x)$		dne	0	

Table 4.2.2 Note that "dne" means the limit does not exist.

Notice that the First-Derivative Test applies at $x = -1$ even though f is not differentiable for $x = -1$. A graph of $y = f(x)$ is shown in Figure 4.2.7. (See also Exercise 40 in Section 4.3.) ◇

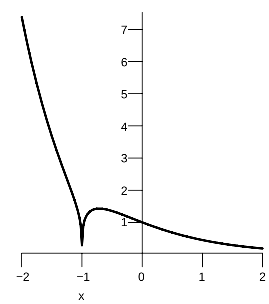


Figure 4.2.7

Extreme Values on a Closed Interval

Many applied problems involve a continuous function whose domain is a closed interval $[a, b]$. (See Section 4.1.)

The Extreme Value Theorem guarantees the function attains both a maximum and a minimum at some point in the interval. The extreme values occur either at

1. an endpoint ($x = a$ or $x = b$),
2. a critical number ($x = c$ where $f'(c) = 0$), or
3. where f is not differentiable ($x = c$ where $f'(c)$ is not defined).

EXAMPLE 5 Find the absolute maximum and minimum values of $f(x) = x^4 - 8x^2 + 1$ on the interval $[-1, 3]$.

SOLUTION The function is continuous on a closed and bounded interval. The absolute maximum and minimum values occur either at a critical point or at an endpoint of the interval. The endpoints are $x = -1$ and $x = 3$. To find the critical points we solve $f'(x) = 0$:

$$f'(x) = 4x^3 - 16x = 4x(x^2 - 4) = 4x(x - 2)(x + 2) = 0.$$

There are three critical numbers, $x = 0$, 2, and -2, but only $x = 0$ and $x = 2$ are in the interval. The intervals where the graph of $y = f(x)$ is increasing and decreasing can be determined from the information in Table 4.2.3.

x	-1	0	2	3
$f(x)$	-6	1	-15	10
$f'(x)$		0	0	0

Table 4.2.3

Since we are looking only for global extrema on a closed interval, it is unnecessary to determine these intervals or to classify critical points as local extrema. Instead, we simply scan the list of function values at the endpoints and at the critical numbers – row 2 of Table 4.2.3 – for the largest and smallest values of $f(x)$. The largest value is 10, so the global maximum occurs at $x = 3$. The smallest value is -15, so the global minimum occurs at $x = 2$. (See Figure 4.2.8.) ◇

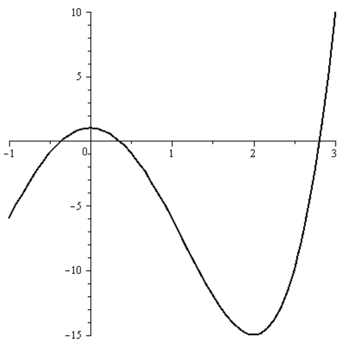


Figure 4.2.8

In Example 5 it was not necessary to determine the intervals on which the function is increasing and decreasing, nor did we need to identify the local extreme values. (See also Exercise 5.)

Summary

This section shows how to use the first derivative to find extreme values of a function. Namely, identify when the derivative is zero, positive, and negative, and where it changes sign.

A continuous function on a closed and bounded interval always has a maximum and a minimum. All extrema occur either at an endpoint, a critical number (where $f'(c) = 0$), or where f is not differentiable.

EXERCISES for Section 4.2

In Exercises 1 to 28, sketch the graph of the given function. Find all intercepts and critical points, determine the intervals where the function is increasing and where it is decreasing, and identify all local extreme values.

1. $f(x) = x^5$

2. $f(x) = (x-1)^4$

3. $f(x) = 3x^4 + x^3$

4. $f(x) = 2x^3 + 3x^2$

5. $f(x) = x^4 - 8x^2 + 1$

6. $f(x) = x^3 - 3x^2 + 3x$

7. $f(x) = x^4 - 4x + 3$

8. $f(x) = 2x^2 + 3x + 5$

9. $f(x) = x^4 + 2x^3 - 3x^2$

10. $f(x) = 2x^3 + 3x^2 - 6x$

11. $f(x) = xe^{-x/2}$

12. $f(x) = xe^{x/3}$

13. $f(x) = e^{-x^2}$

14. $f(x) = xe^{-x^2/2}$

15. $f(x) = x\sin(x) + \cos(x)$

16. $f(x) = x\cos(x) - \sin(x)$

17. $f(x) = \frac{\cos(x)-1}{x^2}$

18. $f(x) = x\ln(x)$

19. $f(x) = \frac{\ln(x)}{x}$

20. $f(x) = \frac{e^x-1}{x}$

21. $f(x) = \frac{e^{-x}}{x}$

22. $f(x) = \frac{x-\arctan(x)}{x^3}$

23. $f(x) = \dfrac{3x+1}{3x-1}$

24. $f(x) = \dfrac{x}{x^2+1}$

25. $f(x) = \dfrac{x}{x^2-1}$

26. $f(x) = \dfrac{1}{2x^2-x}$

27. $f(x) = \dfrac{1}{x^2-3x+2}$

28. $f(x) = \dfrac{\sqrt{x^2+1}}{x}$

In Exercises 29 to 36 sketch the general shape of the graph, using the given information. Assume the function and its derivative are defined for all x and are continuous. Explain your reasoning.

29. Critical point $(1, 2)$, $f'(x) < 0$ for $x < 1$ and $f'(x) > 0$ for $x > 1$.

30. Critical point $(1, 2)$ and $f'(x) < 0$ for all x except $x = 1$.

31. x intercept -1, critical points $(1, 3)$ and $(2, 1)$, $\lim_{x\to\infty} f(x) = 4$, $\lim_{x\to-\infty} f(x) = -1$.

32. y intercept 3, critical point $(1, 2)$, $\lim_{x\to\infty} f(x) = \infty$, $\lim_{x\to-\infty} f(x) = 4$.

33. x intercept -1, critical points $(1, 5)$ and $(2, 4)$, $\lim_{x\to\infty} f(x) = 5$, $\lim_{x\to-\infty} f(x) = -\infty$.

34. x intercept 1, y intercept 2, critical points $(1, 0)$ and $(4, 4)$, $\lim_{x\to\infty} f(x) = 3$, $\lim_{x\to-\infty} f(x) = \infty$.

35. x intercepts 2 and 4, y intercept 2, critical points $(1, 3)$ and $(3, -1)$, $\lim_{x\to\infty} f(x) = \infty$, $\lim_{x\to-\infty} f(x) = 1$.

36. No x intercepts, y intercept 1, no critical points, $\lim_{x\to\infty} f(x) = 2$, $\lim_{x\to-\infty} f(x) = 0$.

Exercises 37 to 52 concern functions whose domains are restricted to closed intervals. In each, find the maximum and minimum values for the given function on the given interval.

37. $f(x) = x^2 - x^4$ on $[0, 1]$

38. $f(x) = 4x - x^2$ on $[0, 5]$

39. $f(x) = 2x^2 - 5x$ on $[-1, 1]$

40. $f(x) = x^3 - 2x^2 + 5x$ on $[-1, 3]$

41. $f(x) = \dfrac{x}{x^2 + 1}$ on $[0, 3]$

42. $f(x) = x^2 + x^4$ on $[0, 1]$

43. $f(x) = \dfrac{x + 1}{\sqrt{x^2 + 1}}$ on $[0, 3]$

44. $f(x) = \sin(x) + \cos(x)$ on $[0, \pi]$

45. $f(x) = \sin(x) - \cos(x)$ on $[0, \pi]$

46. $f(x) = x + \sin(x)$ on $[-\pi/2, \pi/2]$

47. $f(x) = x + \sin(x)$ on $[-\pi, 2\pi]$

48. $f(x) = x/2 + \sin(x)$ on $[-\pi, 2\pi]$

49. $f(x) = 2\sin(x) - \sin(2x)$ on $[-\pi, \pi]$

50. $f(x) = \sin(x^2) + \cos(x^2)$ on $[0, \sqrt{2\pi}]$

51. $f(x) = \sin(x) - \cos(x)$ on $[-2\pi, 2\pi]$

52. $f(x) = \sin^2(x) - \cos^2(x)$ on $[-2\pi, 2\pi]$

In Exercises 53 to 59 graph the function.

53. $f(x) = \dfrac{\sin(x)}{1 + 2\cos(x)}$

54. $f(x) = \dfrac{\sqrt{x^2-1}}{x}$

55. $f(x) = \dfrac{1}{(x-1)^2(x-2)}$

56. $f(x) = \dfrac{3x^2+5}{x^2-1}$

57. $f(x) = 2x^{1/3} + x^{4/3}$

58. $f(x) = \dfrac{3x^2+5}{x^2+1}$

59. $f(x) = \sqrt{3}\sin(x) + \cos(x)$

60. Graph $f(x) = (x^2-9)^{1/3}e^{-x}$. (This function is difficult to graph in one picture. Instead, create separate sketches for $x > 0$ and for $x < 0$. Watch out for the points where f is not differentiable.)

61. A certain differentiable function has $f'(x) < 0$ for $x < 1$ and $f'(x) > 0$ for $x > 1$. Moreover, $f(0) = 3$, $f(1) = 1$, and $f(2) = 2$.

(a) What is the minimum value of $f(x)$ for x in $[0, 2]$? Why?

(b) What is the maximum value of $f(x)$ for x in $[0, 2]$? Why?

In Exercises 62 to 64 decide if there is a function that meets all of the stated conditions. If you think there is such a function, sketch its possible graph. Otherwise, explain why a function cannot meet all of the conditions.

62. $f(x) > 0$ for all x, $f'(x) < 0$ for all x

63. $f(3) = 1$, $f(5) = 1$, $f'(x) > 0$ for x in $[3, 5]$

64. $f'(x) \neq 0$ for all x except $x = 3$ and 5, when $f'(x) = 0$ and $f(x) = 0$ for $x = -2$, 4, and 5

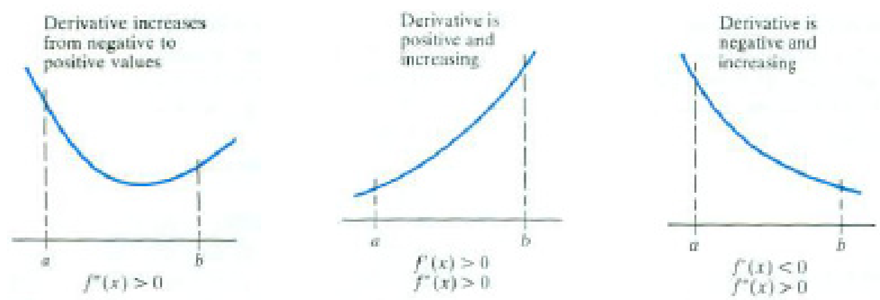
65. What is the minimum value of $y = (x^3 - x)/(x^2 - 4)$ for $x > 2$?

4.3 The Second Derivative and Graphing

The sign of the first derivative tells whether a function is increasing or decreasing. In this section we examine what the sign of the second derivative tells us about a function and its graph. This information will be used to help graph functions and also to provide an additional way to test whether a critical point is a maximum or minimum.

Concavity and Points of Inflection

The second derivative is the derivative of the first derivative. Thus, the sign of the second derivative determines if the first derivative is increasing or decreasing. For example, if $f''(x)$ is positive for all x in an interval (a, b), then f' is an increasing function throughout the interval (a, b). In other words, the slope of the graph of $y = f(x)$ increases as x increases from left to right. The slope may



(a) (b) (c)

Figure 4.3.1

increase from negative values to zero to positive values, as in Figure 4.3.1(a). Or the slope may be positive throughout (a, b), as in Figure 4.3.1(b). Or the slope may be negative throughout (a, b), as in Figure 4.3.1(c).

In the same way, if $f''(x)$ is negative on the interval (a, b) then f' is decreasing on (a, b). The slope of the graph of $y = f(x)$ decreases as x increases from left to right on that part of the graph corresponding to (a, b).

As you drive along the graph of a concave up function, going from left to right, the car veers to the left.

DEFINITION *(Concave Up and Concave Down)*

A function f whose first derivative is increasing throughout the open interval (a, b) is called **concave up** in that interval.

A function f whose first derivative is decreasing throughout the open interval (a, b) is called **concave down** in that interval.

When a curve is concave up, it lies above its tangent lines and below its chords. The graph of a <u>c</u>oncave <u>up</u> function is shaped like a <u>*cup*</u>. See Figure 4.3.2.

When a curve is concave down, it lies below its tangent lines and above its chords. The graph of a concave <u>*down*</u> function is shaped like a <u>*frown*</u>. See Figure 4.3.3.

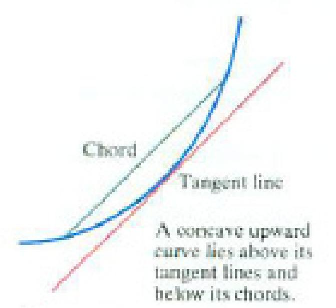


Figure 4.3.2

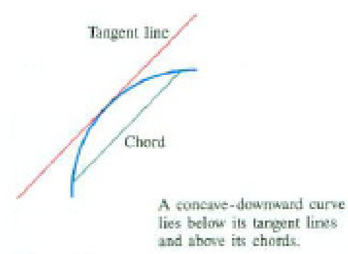


Figure 4.3.3

Convex and Concave Sets

In more advanced courses "concave up" is called "convex." This is because the set in the xy-plane above this part of a graph is a convex set. (A convex set is a set with the property that for any two points P and Q in the set the line segment joining them also lies in the set. See also Exercises 26 to 32 in Section 2.5.) In the same way "concave down" is called "concave." For instance, the part of the graph of $y = x^3$ to the right of the x-axis is convex and the part to the left is concave.

EXAMPLE 1 Where is the graph of $f(x) = x^3$ concave up? concave down?

SOLUTION First, compute the second derivative: $f''(x) = 6x$, which is positive when x is positive and negative when x is negative. Thus, the graph is concave up for $x > 0$ and is concave down for $x < 0$. Note that the sense of concavity changes at $x = 0$, where $f''(x) = 0$. (See Figure 4.3.4.) ◇

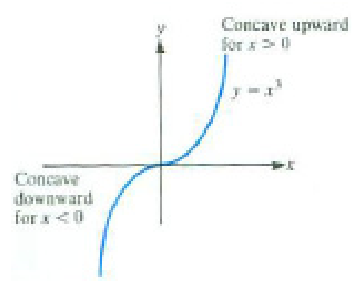


Figure 4.3.4

In an interval where $f''(x)$ is positive, the function $f'(x)$ is increasing, and so the function f is concave up. However, if a function is concave up, $f''(x)$ need not be positive for all x in the interval. For instance, consider $y = x^4$. Even though the second derivative $12x^2$ is zero for $x = 0$, the first derivative $4x^3$ is increasing on any interval, so the graph is concave up over any interval.

Any point where the graph of a function changes concavity is important.

DEFINITION *(Inflection Number and Inflection Point)* Let f be a function and let a be a number. Assume there are numbers b and c such that $b < a < c$ and

1. f is continuous on the open interval (b, c)
2. f is concave up on (b, a) and concave down on (a, c)
   or
   f is concave down on (b, a) and concave up on (a, c).

Then, the point $(a, f(a))$ is called an **inflection point** or **point of inflection** of f. The number a is called an **inflection number** of f.

Notice that $f''(a) = 0$ does not automatically make a an inflection number of f. To be an inflection number, the concavity has to change at a.

Observe that if the second derivative changes sign at the number a, then a is an inflection number. If the second derivative exists at an inflection number, it must be 0. But there can be an inflection point if $f''(a)$ is not defined. This is illustrated in the next example.

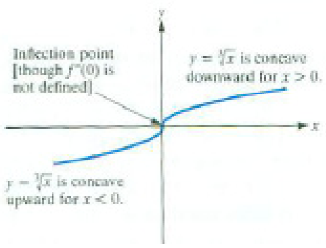


Figure 4.3.5

EXAMPLE 2 Examine the concavity of the graph of $y = x^{1/3}$.
SOLUTION Here $y' = \frac{1}{3}x^{-2/3}$ and $y'' = \frac{1}{3}\left(\frac{-2}{3}\right)x^{-5/3}$. Although $x = 0$ is in the domain of this function, neither y' nor y'' is defined for $x = 0$. When x is negative, y'' is positive; when x is positive, y'' is negative. Thus, the concavity changes from concave up to concave down at $x = 0$. This means $x = 0$ is an inflection number and $(0, 0)$ is an inflection point. See Figure 4.3.5. ◇

The simplest way to look for inflection points is to use both the first and second derivatives:

To find inflection points of $y = f(x)$:

1. Compute $f'(x)$ and $f''(x)$.
2. Look for numbers a such that f'' is not defined at a.
3. Look for numbers a such that $f''(a) = 0$
4. For each interval defined by the numbers found in Steps 2 and 3, determine the sign of $f''(x)$.

This process can be implemented using the same ideas used in Section 4.2 to identify critical points, as Example 3 shows.

EXAMPLE 3 Find the inflection point(s) of $f(x) = x^4 - 8x^3 + 18x^2$.
SOLUTION First, $f'(x) = 4x^3 - 24x^2 + 36x$ and

$$f''(x) \quad = \quad 12x^2 - 48x + 36 = 12(x^2 - 4x + 3) = 12(x-1)(x-3).$$

Because f'' is defined for all real numbers, the only candidates for inflection numbers are the solutions to $f''(x) = 0$. Solving $f''(x) = 0$ yields:

$$0 \quad = \quad 12(x-1)(x-3).$$

Hence $x - 1 = 0$ or $x - 3 = 0$, and $x = 1$ or $x = 3$.

To decide whether 1 and 3 are inflection numbers of f, look at the sign of $f''(x) = 12(x-1)(x-3)$. For $x > 3$ both $x - 1$ and $x - 3$ are positive, so $f''(x)$

is positive. For x in $(1,3)$, $x-1$ is positive and $x-3$ is negative, so $f''(x)$ is negative. For $x<1$, both $x-1$ and $x-3$ are negative, so $f''(x)$ is positive. This is recorded in Table 4.3.1. Since sign changes in $f''(x)$ correspond to

x	$(-\infty,1)$	1	$(1,3)$	3	$(3,\infty)$
$f''(x)$	$+$	0	$-$	0	$+$

Table 4.3.1

changes in concavity of the graph of f, this function has two inflection points: $(1,11)$ and $(3,27)$. (See Figure 4.3.6.)

◇

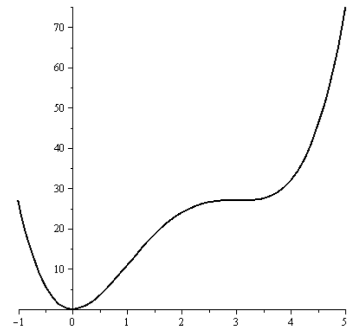


Figure 4.3.6

Using Concavity in Graphing

The second derivative, together with the first derivative and the other tools of graphing, can help us sketch the graph of a function. Example 4 continues Example 3.

EXAMPLE 4 Graph $f(x)=x^4-8x^3+18x^2$.

SOLUTION Because the function is a polynomial of degree at least two it has no asymptotes. Since $f(0)=0^4-8(0^3)+18(0^2)$, its y intercept is 0. To find its x intercepts we look for solutions to the equation

$$\begin{aligned} x^4-8x^3+18x^2 &= 0 \\ x^2(x^2-8x+18) &= 0. \end{aligned}$$

Thus $x=0$ or $x^2-8x+18=0$, which can be solved by the quadratic formula. The discriminant is $(-8)^2-4(1)(18)=-8$ which is negative, so there are no real solutions of $x^2-8x+18=0$. The only x intercept is $x=0$.

The **discriminant** of ax^2+bx+c is b^2-4ac.

In Example 3 the first derivative was found:

Analysis based on $f'(x)$

$$f'(x) = 4x^3-24x^2+36x = 4x(x^2-6x+9) = 4x(x-3)^2.$$

Thus, $f'(x)=0$ only when $x=0$ and $x=3$. The two critical points are $(0,f(0))=(0,0)$ and $(3,f(3))=(3,27)$. The information in Table 4.3.2 allows us to conclude that the function f is decreasing on $(-\infty,0)$ and increasing on $(0,\infty)$ with a local minimum at $(0,0)$.

x	$(-\infty,0)$	0	$(0,3)$	3	$(3,\infty)$
$f'(x)$	$-$	0	$+$	0	$+$

Table 4.3.2

By Example 3, the graph is concave up on $(-\infty, 1)$ and $(3, \infty)$ and concave down on $(1, 3)$.

To begin to sketch the graph of $y = f(x)$, plot the three points $(0, f(0)) = (0, 0)$, $(1, f(1)) = (1, 11)$, and $(3, f(3)) = (3, 27)$. These three points divide the domain into four intervals. On $(-\infty, 0)$ the function is decreasing and concave up; on $(0, 1)$ it is increasing and concave up; on $(1, 3)$ it is increasing and concave down; and on $(3, \infty)$ it is once again increasing and concave up. The final graph is shown in Figure 4.3.7. ◇

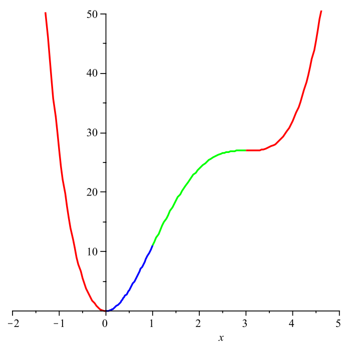


Figure 4.3.7

The procedure demonstrated in Example 4 has several advantages. Note that it was necessary to evaluate $f(x)$ only at a few "important" inputs x. These inputs cut the domain into intervals where neither the first derivative nor the second derivative changes sign. On each of these intervals the graph of the function will have one of the four shapes shown in Figure 4.3.8. A graph usually is made up of these four shapes.

(a) (b) (c) (d)

Figure 4.3.8 The general shape of a function that is (a) increasing and concave up, (b) increasing and concave down, (c) decreasing and concave up, and (d) decreasing and concave down

Local Extrema and the Second-Derivative Test

The second derivative is also useful in testing whether a critical number corresponds to a relative minimum or relative maximum. For this, we will use the relationships between concavity and tangent lines shown in Figures 4.3.2 and 4.3.3.

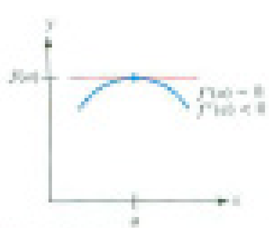


Figure 4.3.9

Let a be a critical number for the function f. Assume, for instance, that $f''(a)$ is negative. If f'' is continuous in some open interval that contains a, then (by the Permanence Property) $f''(x)$ remains negative for a suitably small open interval that contains a. This means the graph of f is concave down near $(a, f(a))$; it lies below its tangent lines. In particular, it lies below the horizontal tangent line at the critical point $(a, f(a))$, as illustrated in Figure 4.3.9. Thus the function f has a *relative maximum* at the critical number

a. Similarly, if $f'(a) = 0$ and $f''(a) > 0$, the critical point $(a, f(a))$ is a relative minimum because the graph of f is concave up and lies above the horizontal tangent line at $(a, f(a))$. These observations justify the following test for a relative extremum.

Theorem 4.3.1. *[Second-Derivative Test for Relative Extrema] Let f be a function such that $f'(x)$ is defined at least on some open interval containing the number a. Assume that $f''(x)$ is continuous and $f''(a)$ is defined.*

If $f'(a) = 0$ and $f''(a) > 0$, then f has a relative minimum at $(a, f(a))$.

If $f'(a) = 0$ and $f''(a) < 0$, then f has a relative maximum at $(a, f(a))$.

EXAMPLE 5 Use the Second-Derivative Test to classify all local extrema of the function $f(x) = x^4 - 8x^3 + 18x^2$.

Compare with Examples 3 and 4.

SOLUTION This is the same function analyzed in Examples 3 and 4. The two critical points are $(0, 0)$ and $(3, 27)$. The second derivative is $f''(x) = 12x^2 - 48x + 36$. At $x = 0$ we have

$$f''(0) = 12(0^2) - 48(0) + 36 = 36,$$

which is positive. Since $f'(0) = 0$ and $f''(0) > 0$, f has a local minimum at $(0, 0)$. At $x = 3$ we have

$$f''(3) = 12(3^2) - 48(3) + 36 = 0.$$

Since $f''(3) = 0$, the Second-Derivative Test tells us nothing about the critical number 3.

This is consistent with our previous findings. The point at $(3, 27)$ is an inflection point and not a local extreme point. ⋄

Summary

Table 4.3.3 shows the meaning of the signs of $f(x)$, $f'(x)$, and $f''(x)$ in terms of the graph of $y = f(x)$.

	is positive (> 0).	is negative (< 0).	changes sign.
Where the ordinate $f(x)$	the graph is above the x-axis.	the graph is below the x-axis.	the graph crosses x-axis.
Where the slope $f'(x)$	the graph slopes upward.	the graph slopes downward.	the graph has a zontal tangent a relative extremu
Where $f''(x)$	the graph is concave up (like a cup).	the graph is concave down (like a frown).	the graph has a flection point.

Table 4.3.3 EDITOR: This table should appear after the first, short, paragraph

The graph has a critical point at $(a, f(a))$ whenever $f'(a) = 0$. This critical point is an extremum of f if the first derivative changes sign at $x = a$; a maximum if the first derivative changes from positive to negative and a minimum if the first derivative changes from negative to positive.

Keep in mind that the graph has an inflection point at $(a, f(a))$ when the sign of $f''(x)$ changes at $x = a$. This can occur when either $f''(a) = 0$ or when $f''(a)$ is not defined. Similarly, a graph can have a maximum or minimum at $(a, f(a))$ when either $f'(a) = 0$ or $f'(a)$ is not defined.

EXERCISES for Section 4.3

In Exercises 1 to 16 describe the intervals where the function is concave up and concave down and give any inflection points.

1. $f(x) = x^3 - 3x^2 + 2$

2. $f(x) = x^3 - 6x^2 + 1$

3. $f(x) = x^2 + x + 1$

4. $f(x) = 2x^2 - 5x$

5. $f(x) = x^4 - 4x^3$

6. $f(x) = 3x^5 - 5x^4$

7. $f(x) = \frac{1}{1+x^2}$

8. $f(x) = \frac{1}{1+x^4}$

9. $f(x) = x^3 + 6x^2 - 15x$

10. $f(x) = \frac{x^2}{2} + \frac{1}{x}$

11. $f(x) = e^{-x^2}$

12. $f(x) = xe^x$

13. $f(x) = \tan(x)$

14. $f(x) = \sin(x) + \sqrt{3}\cos(x)$

15. $f(x) = \cos(x)$

16. $f(x) = \cos(x) + \sin(x)$

In Exercises 17 to 29 graph the functions, showing critical points, inflection points, and intercepts.

17. $f(x) = x^3 + 3x^2$

18. $f(x) = 2x^3 + 9x^2$

19. $f(x) = x^4 - 4x^3 + 6x^2$

20. $f(x) = x^4 + 4x^3 + 6x^2 - 2$

21. $f(x) = x^4 - 6x^3 + 12x^2$

22. $f(x) = 2x^6 - 10x^4 + 10$

23. $f(x) = 2x^6 + 3x^5 - 10x^4$

24. $f(x) = 3x^4 + 4x^3 - 12x^2 + 4$

25. $f(x) = xe^{-x}$

26. $f(x) = e^{x^3}$

27. $f(x) = 3x^5 - 20x^3 + 10$ This function was first met in Example 3 in Section 4.2.

28. $f(x) = 3x^4 + 4x^3 - 12x^2 + 4$

29. $f(x) = 2x^6 - 15x^4 + 20x^3 - 20x + 10$

In each of Exercises 30 to 37 sketch the general appearance of the graph of the given function near $(1, 1)$ on the basis of the information given. Assume that f, f', and f'' are continuous.

30. $f(1) = 1,\ f'(1) = 0,\ f''(1) = 1$

31. $f(1) = 1,\ f'(1) = 0,\ f''(1) = -1$

32. $f(1) = 1,\ f'(1) = 0,\ f''(1) = 0$ Sketch four quite different possibilities.

33. $f(1) = 1,\ f'(1) = 0,\ f''(1) = 0,\ f''(x) < 0$ for $x < 1$ and $f''(x) > 0$ for $x > 1$

34. $f(1) = 1,\ f'(1) = 0,\ f''(1) = 0$ and $f''(x) < 0$ for x near 1

35. $f(1) = 1,\ f'(1) = 1,\ f''(1) = -1$

36. $f(1) = 1,\ f'(1) = 1,\ f''(1) = 0,\ f''(x) < 0$ for $x < 1$ and $f''(x) > 0$ for $x > 1$

37. $f(1) = 1,\ f'(1) = 1,\ f''(1) = 0$ and $f''(x) > 0$ for x near 1

38. Find all inflection points of $f(x) = x\ln(x)$. On what intervals is the graph of $y = f(x)$ concave up? concave down? Graph $y = f(x)$ on an interval large enough to clearly show all interesting features of the graph. On what intervals is the graph increasing? decreasing? This graph appeared in Example 2 of Section 4.2.

39. Find all inflection points of $f(x) = x + \ln(x)$. On what intervals is the graph of $y = f(x)$ concave up? concave down? Graph $y = f(x)$ on an interval large enough to show all interesting features of the graph. On what intervals is the function increasing? decreasing?

40. Find all inflection points of $f(x) = (x+1)^{2/7}e^{-x}$. On what intervals is the graph of $y = f(x)$ concave up? concave down? On what intervals is the function increasing? decreasing? This function was first met in Example 4 of Section 4.2.

41. Find the critical points and inflection points of $f(x) = x^2e^{-x/3}$. See Example 1 of Section 4.1.

In Exercises 42 to 43 sketch a graph of a hypothetical function that meets the given conditions. Assume f' and f'' are continuous. Explain your reasoning.

42. Critical point $(2, 4)$; inflection points $(3, 1)$ and $(1, 1)$; $\lim_{x\to\infty} f(x) = 0$ and $\lim_{x\to-\infty} f(x) = 0$

43. Critical points $(-1, 1)$ and $(3, 2)$; inflection point $(4, 1)$; $\lim_{x\to0^+} f(x) = -\infty$ and $\lim_{x\to0^-} f(x) = \infty$; $\lim_{x\to\infty} f(x) = 0$ and $\lim_{x\to-\infty} f(x) = \infty$

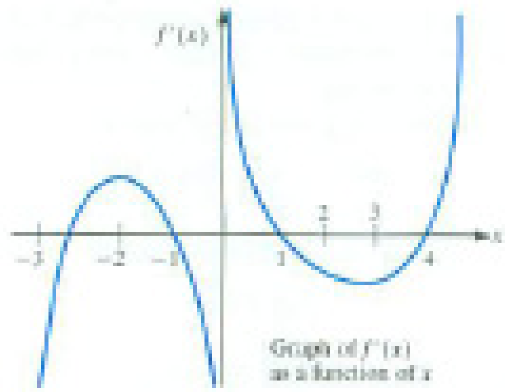


Figure 4.3.10

44. *(Contributed by David Hayes)* Let f be a function that is continuous for all x and differentiable for all x other than 0. Figure 4.3.10 is the graph of its derivative $f'(x)$ as a function of x.

(a) Answer the following questions about f (*not* about f'). Where is f increasing? decreasing? concave up? concave down? What are the critical numbers? Where do any relative extrema occur? Explain.

(b) Assuming that $f(0) = 1$, graph a hypothetical function f that satisfies the conditions given.

(c) Graph $f''(x)$.

45. Graph $y = 2(x-1)^{5/3} + 5(x-1)^{2/3}$, paying particular attention to points where y' does not exist.

46. Graph $y = x + (x+1)^{1/3}$.

47. Find the critical points and inflection points in $[0, 2\pi]$ of $f(x) = \sin^2(x)\cos(x)$.

48. Can a polynomial of degree 6 have (a) no inflection points? (b) exactly one inflection point? Explain.

49. Can a polynomial of degree 5 have (a) no inflection points? (b) exactly one inflection point? Explain.

50. Let f be a function such that $f''(x) = (x-1)(x-2)$.

(a) For which x is f concave up?

(b) For which x is f concave down?

(c) List its inflection number(s).

(d) Find a specific function f whose second derivative is $(x-1)(x-2)$.

51. In the theory of **inhibited growth** it is assumed that the growing quantity y approaches some limiting size M. Specifically, one assumes that the rate of growth is proportional both to the amount present and to the amount left to grow:

$$\frac{dy}{dt} = ky(M-y),$$

where k is a positive number. Prove that the graph of y as a function of time has an inflection point when the amount y is exactly half the limiting amount M.

52. A certain function $y = f(x)$ has the property that

$$y' = \sin(y) + 2y + x.$$

Show that at a critical number the function has a local minimum.

53. Assume that the domain of $f(x)$ is the entire x-axis, and $f'(x)$ and $f''(x)$ are continuous. Assume that $(1,1)$ is the only critical point and that $\lim_{x\to\infty} f(x) = 0$.

(a) Can $f(x)$ be negative for some $x > 1$?

(b) Must $f(x)$ be decreasing for $x > 1$?

(c) Must $f(x)$ have an inflection point?

4.4 Proofs of the Three Theorems

In Section 4.1 two observations about tangent lines led to the Theorem of the Interior Extremum, Rolle's Theorem, and the Mean-Value Theorem. Now, using the definition of the derivative, and no pictures, we prove them. That the proofs go through based only on the definition of the derivative as a limit reassures us that this definition is suitable to serve as part of the foundation of calculus.

Proof of Theorem 4.1.1:

Proof of the Theorem of the Interior Extremum

Suppose the maximum of f on the open interval (a, b) occurs at the number c. This means that $f(c) \geq f(x)$ for each number x between a and b.

Our challenge is to use only this information and the definition of the derivative as a limit to show that either $f'(c) = 0$ or f is not differentiable at c.

Assume that f is differentiable at c. We will show that $f'(c) \geq 0$ and $f'(c) \leq 0$, forcing $f'(c)$ to be zero.

Recall that

$$f'(c) = \lim_{\Delta x \to 0} \frac{f(c + \Delta x) - f(c)}{\Delta x}.$$

The assumption that f is differentiable on (a, b) means that $f'(c)$ exists. Consider the difference quotient

$$\frac{f(c + \Delta x) - f(c)}{\Delta x}. \tag{4.4.1}$$

when Δx is so small that $c+\Delta x$ is in the interval (a, b). Then $f(c+\Delta x) \leq f(c)$. Hence $f(c + \Delta x) - f(c) \leq 0$. Therefore, when Δx is positive, the difference quotient in (4.4.1) will be negative or zero. Consequently, as $\Delta x \to 0$ through *positive* values,

$\frac{\text{negative}}{\text{positive}} = \text{negative}$

$$f'(c) = \lim_{\Delta x \to 0^+} \frac{f(c + \Delta x) - f(c)}{\Delta x} \leq 0. \tag{4.4.2}$$

If, on the other hand, Δx is negative, then the difference quotient in (4.4.1) will be positive or zero. Hence, as $\Delta x \to 0$ through *negative* values,

$\frac{\text{negative}}{\text{negative}} = \text{positive}$

$$f'(c) = \lim_{\Delta x \to 0^-} \frac{f(c + \Delta x) - f(c)}{\Delta x} \geq 0. \tag{4.4.3}$$

The only way $f'(c) \leq 0$ and $f'(c) \geq 0$ can both hold is when $f'(c) = 0$. This proves that if f has a maximum on (a, b), then $f'(c) = 0$.

The proof for the case when f has a minimum on (a, b) is essentially the same. •

See Exercise 12.

The proofs of Rolle's Theorem and the Mean-Value Theorem are related. Suppose f is continuous on $[a, b]$ and differentiable on (a, b).

Proof of Rolle's Theorem

Proof of Theorem 4.1.2:

If $f(a) = f(b)$, then $f'(c) = 0$ for at least one number between a and b.

The goal here is to use the facts that f is continuous on $[a, b]$, differentiable on (a, b) and $f(a) = f(b)$ to conclude that there must be a number c in (a, b) with $f'(c) = 0$.

Since f is continuous on the closed interval $[a, b]$, it has a maximum value M and a minimum value m on that interval. There are two cases to consider: $m < M$ and $m = M$.

Case 1: If $m = M$, f is constant and $f'(x) = 0$ for all x in $[a, b]$. Then any number in (a, b) will serve as the desired number c.

Case 2: Suppose $m < M$. Because $f(a) = f(b)$ the minimum and maximum cannot both occur at the ends of the interval. At least one of the extrema occurs at a number c strictly between a and b. By assumption, f is differentiable at c, so $f'(c)$ exists. Thus, by the Theorem of the Interior Extremum, $f'(c) = 0$. This completes the proof of Rolle's Theorem. •

The idea behind the proof of the Mean-Value Theorem is to define a function to which Rolle's Theorem can be applied.

Proof of the Mean-Value Theorem

Proof of Theorem 4.1.3:

$f'(c) = \frac{f(b)-f(a)}{b-a}$ for at least one number between a and b.

Let $y = L(x)$ be the equation of the chord through the two points $(a, f(a))$ and $(b, f(b))$. The slope of this line is $L'(x) = \dfrac{f(b) - f(a)}{b - a}$. Define $h(x) = f(x) - L(x)$. Note that $h(a) = h(b) = 0$ because $f(a) = L(a)$ and $f(b) = L(b)$.

By assumption, f is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b). So h, being the difference of f and L, is also continuous on $[a, b]$ and differentiable on (a, b).

Rolle's Theorem applies to h on the interval $[a, b]$. Therefore, there is at least one number c in (a, b) where $h'(c) = 0$. Now, $h'(c) = f'(c) - L'(c)$, so that

$$f'(c) = L'(c) = \frac{f(b) - f(a)}{b - a}.$$

•

Summary

Using only the definition of the derivative and the assumption that a continuous function defined on a closed interval assumes maximum and minimum values, we proved the Theorem of the Interior Extremum, Rolle's Theorem, and the Mean-Value Theorem. Note that we did not appeal to any pictures or to our geometric intuition.

EXERCISES for Section 4.4

In each of Exercises 1 to 3 sketch a graph of a differentiable function that meets the given conditions. (Just draw the graph; there is no need to give a formula for the function.)

1. $f'(x) < 0$ for all x

2. $f'(3) = 0$ and $f'(x) < 0$ for x not equal to 3

3. $f'(x) = 0$ only when $x = 1$ or 4; $f(1) = 3$, $f(4) = 1$; $f'(x) > 0$ for $x < 1$ and for $x > 4$

In Exercises 4 to 5 explain why no differentiable function satisfies all the conditions.

4. $f(1) = 3$, $f(2) = 4$, $f'(x) < 0$ for all x

5. $f(x) = 2$ only when $x = 0$, 1, and 3; $f'(x) = 0$ only when $x = \frac{1}{4}$, $\frac{3}{4}$, and 4.

6. In "Surely You're Joking, Mr. Feynman!," Norton, New York, 1985, Nobel laureate Richard P. Feynman writes:

> I often liked to play tricks on people when I was at MIT. One time, in mechanical drawing class, some joker picked up a French curve (a piece of plastic for drawing smooth curves — a curly funny-looking thing) and said, "I wonder if the curves on that thing have some special formula?"
>
> I thought for a moment and said, "Sure they do. The curves are very special curves. Lemme show ya," and I picked up my French curve and began to turn it slowly. "The French curve is made so that at the lowest point on each curve, no matter how you turn it, the tangent is horizontal."
>
> All the guys in the class were holding their French curve up at different angles, holding their pencil up to it at the lowest point and laying it down, and discovering that, sure enough, the tangent is horizontal.

How was Feynman playing a trick on his classmates?

7. What can be said about the number of solutions of the equation $f(x) = 3$ for a differentiable function if

(a) $f'(x) > 0$ for all x?

(b) $f'(x) > 0$ for $x < 7$ and $f'(x) < 0$ for $x > 7$?

8. Consider the function $f(x) = x^3 + ax^2 + c$. Show that if $a < 0$ and $c > 0$, then f has exactly one negative solution.

9. With the book closed, obtain the Mean-Value Theorem from Rolle's Theorem.

10.

(a) Recall the definition of $L(x)$ in the proof of the Mean-Value Theorem, and show that

$$L(x) = f(a) + \frac{x-a}{b-a}\left(f(b) - f(a)\right).$$

(b) Using (a), show that

$$L'(x) = \frac{f(b) - f(a)}{b-a}.$$

11. Show that Rolle's Theorem is a special case of the Mean-Value Theorem.

12. Prove the Theorem of the Interior Extremum when the minimum of f on (a, b) occurs at c.

13. This exercise shows that a polynomial $f(x)$ of degree n, $n \geq 1$, can have at most n distinct real roots, that is, solutions to the equation $f(x) = 0$.

(a) Use algebra to show that the statement holds for $n = 1$ and $n = 2$.

(b) Use calculus to show that the statement then holds for $n = 3$.

(c) Use calculus to show that the statement continues to hold for $n = 4$ and $n = 5$.

(d) Why does it hold for all positive integers n?

14. Is this proposed proof of the Mean-Value Theorem correct?
Proof

Tilt the x and y axes and the graph of the function until the x-axis is parallel to the given chord. The chord is now "horizontal," and we may apply Rolle's Theorem. •

15. Is there a differentiable function f whose domain is the x-axis such that f is increasing and yet the derivative is *not* positive for all x?

16. Prove: If f has a negative derivative on (a, b) then f is decreasing on the interval $[a, b]$.

17. This Exercise provides an analytic justification for the first part of the statement, in Section 4.3, that "[W]hen a curve is concave up, it lies above its tangent lines and below its chords." The second part is proven in Exercise 49 of the Chapter 4 Summary.

Show that in an open interval in which f'' is positive, tangents to the graph of f lie below the curve. (Why do you want to show that if a and x are in the interval, then $f(x) > f(a)+f'(a)(x-a)$? Treat the cases $a < x$ and $x > a$ separately.)

18. We stated, in Section 4.3, that if $f(x)$ is defined in an open interval around the critical number a and $f''(a)$ is negative, then $f(x)$ has a relative maximum at a. Explain why this is so, following these steps.

(a) Why is $\lim_{\Delta x \to 0} \dfrac{f'(a+\Delta x) - f'(a)}{\Delta x}$ negative?

(b) Deduce that if Δx is small and positive, then $f'(a+\Delta x)$ is negative.

(c) Show that if Δx is small and negative, then $f'(a+\Delta x)$ is positive.

(d) Show that $f'(x)$ changes sign from positive to negative at a. By the First-Derivative Test for a Relative Maximum, $f(x)$ has a relative maximum at a.

Skill Drill

19. To keep your differentiation skills sharp, differentiate each of the following expressions:

(a) $\sqrt{1-x^2}\sin(3x)$

(b) $\dfrac{\sqrt[3]{x}}{x^2+1}$

(c) $\tan\left(\dfrac{1}{(2x+1)^2}\right)$

(d) $\ln\left(\dfrac{(x^2+1)^3\sqrt{1-x^2}}{\sec^2(x)}\right)$

(e) e^{x^4}

4.S Chapter Summary

In this chapter we saw that the sign of the function and of its first and second derivatives influenced the shape of its graph. In particular the derivatives show where the function is increasing or decreasing and is concave up or down. That enabled us to find extreme points and inflection points. (See Table 4.3.3 on page 318.)

We state here the main ideas informally for a function with continuous first and second derivatives.

If a function has an extremum at a number, then the derivative there is zero, or is not defined, or the number may be an end point of the domain. This narrows the search for extrema. If the derivative is zero and the second derivative is not zero, the function has an extremum there.

The utility of these tests becomes evident in Rolle's theorem, which says that if a differentiable function has the same value at two inputs on an interval in its domain, its derivative must be zero somewhere between them.

The Mean Value Theorem generalizes this idea. It says that between any two points on its graph there is a point on the graph where the tangent is parallel to the chord through those two points. We used this to show that if a and b are two numbers, then $f(b) = f(a) + f'(c)(b - a)$ for some number c between a and b.

If $f'(a)$ is positive and f' is continuous in some open interval containing a, then, by the Permanence Principle, $f'(x)$ remains positive for some open interval containing a. Specifically, if the derivative is positive at some number, then the function is increasing for inputs near that number. (A similar statement holds when $f'(a)$ is negative.)

Sam: Why bother me with limits? The authors say we need them to define derivatives.

Jane: Aren't you curious about why the formula for the derivative of a product is what it is?

Sam: No. It's been true for over three centuries. Just tell me what it is. If someone says the speed of light is 186,000 miles per second am I supposed to find a meter stick and clock and check it out?

Jane: But what if you forget the formula during a test?

Sam: That's not much of a reason.

Jane: But my physics class uses derivatives and limits to define basic concepts.

Sam: Oh?

Jane: Density of mass at a point and density of electric charge are defined as limits. And it uses derivatives all over the place. You will be lost if you don't know their definitions. Just look at the applications in Chapter 5.

Sam: O.K., O.K. enough. I'll look.

EXERCISES for 4.S

In each of Exercises 1 to 13 decide if it is possible for a single differentiable function to have all of the properties listed. If it is possible, sketch a graph of a differentiable function that meets the given conditions. (Just draw the graph; there is no need to come up with a formula for the function.) If it is not possible, explain why no differentiable function satisfies all of the conditions.

1. $f(0) = 1$, $f(x) > 0$ for all x, and $f'(x) < 0$ for all positive x

2. $f(0) = -1$, $f'(x) < 0$ for all x in $[0, 2]$, and $f(2) = 0$

3. x intercepts at 1 and 5; y intercept at 2; $f'(x) < 0$ for $x < 4$; $f'(x) > 0$ for $x > 4$

4. x intercepts at 2 and 5; y intercept at 3; $f'(x) > 0$ for $x < 1$ and for $x > 3$; $f'(x) < 0$ for x in $(1, 3)$

5. $f(0) = 1$, $f'(x) < 0$ for all positive x, and $\lim_{x\to\infty} f(x) = 1/2$

6. $f(2) = 5$, $f(3) = -1$, $f'(x) \geq 0$ for all x

7. x intercepts only at 1 and 2; $f(3) = -1$, $f(4) = 2$

8. $f'(x) = 0$ only when $x = 1$ or 4; $f(1) = 3$, $f(4) = 1$; $f'(x) < 0$ for $x < 1$; $f'(x) > 0$ for $x > 4$

9. $f(0) = f(1) = 1$ and $f'(0) = f'(1) = 1$

10. $f(0) = f(1) = 1$, $f'(0) = f'(1) = 1$, and $f(x) \neq 0$ for all x in $[0, 1]$

11. $f(0) = f(1) = 1$, $f'(0) = f'(1) = 1$, and $f(x) = 0$ for exactly one number x in $[0, 1]$

12. $f(0) = f(1) = 1$, $f'(0) = f'(1) = 1$, and $f(x)$ has exactly two inflection numbers in $[0, 1]$

13. $f(0) = f(1) = 1$, $f'(0) = f'(1) = 1$, and $f(x)$ has exactly two extrema in $[0, 1]$

14. State the assumptions and conclusions of the Theorem of the Interior Extremum for a function F defined on (a, b).

15. State the assumptions and conclusions of the Mean-Value Theorem for a function g defined on $[c, d]$.

16. The following discussion on higher derivatives in economics appears on page 124 of the College Mathematics Journal **37** (2006):

Charlie Marion of Shrub Oak, NY, submitted this excerpt from "Curses! The Second Derivative" by Jeremy J. Siegel in the October 2004 issue of *Kiplinger's* (p. 73):
"... I think what is bugging the market is something that I have seen happen many times before: the Curse of the Second Derivative. The second derivative, for all those readers who are a few years away from their college calculus class, is the rate of change of the rate of change — or, in this case, whether corporate earnings, which are still rising, are rising at a faster or slower pace."
In the October 1996 issue of the *Notices of the American Mathematical Society*, Hugo Rossi wrote, "In the fall of 1972 President Nixon announced that the rate of increase of inflation was decreasing. This was the first time a sitting president used the third derivative to advance his case for reelection."

Explain why the third derivative is involved in President Nixon's statement.

17. If you watch the tide come in and go out, you will notice at high tide and at low tide, the height of the tide seems to change very slowly. The same holds when you watch an outdoor thermometer: the temperature seems to change the slowest when it is at its highest or at its lowest. Why is that?

18.

(a) Graph $y = \sin^2(2\theta)\cos(2\theta)$ for θ in $[-\pi/2, \pi/2]$.

(b) What is the maximum value of y?

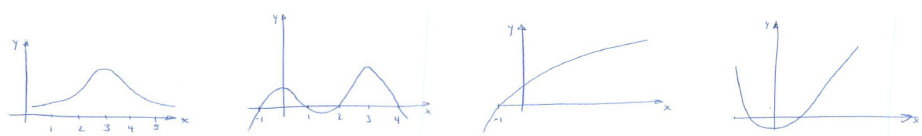
Exercises 19 to 22 display the graph of a function f with continuous f' and f''. Sketch a possible graph of f' and a possible graph of f''.

19. Figure 4.S.1(a)

20. Figure 4.S.1(b)

21. Figure 4.S.1(c)

22. Figure 4.S.1(d)



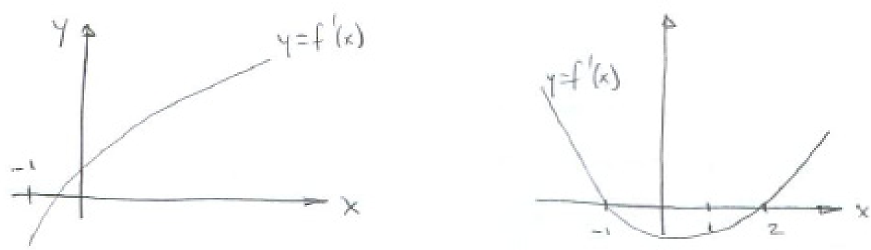
(a) (b) (c) (d)

Figure 4.S.1

In Exercises 23 and 24 sketch the graphs of two possible functions f whose derivative f' is graphed in the given figure.

23. Figure 4.S.2(a)

24. Figure 4.S.2(b)



(a) (b)

Figure 4.S.2

25. Sketch the graph of a function f whose second derivative is graphed in Figure 4.S.3.

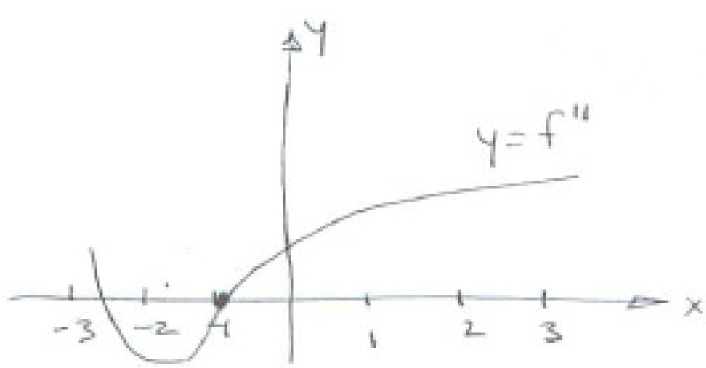
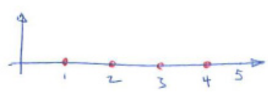


Figure 4.S.3

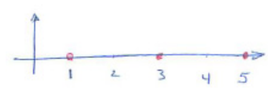
26. Figure 4.S.4(a) shows the only x-intercepts of a function f. Sketch the graph of possible f' and f''.

27. Figure 4.S.4(b) shows the only arguments at which $f'(x) = 0$. Sketch the graph of possible f and f''.

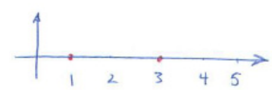
28. Figure 4.S.4(c) shows the only arguments at which $f''(x) = 0$. Sketch the graph of possible f and f'.



(a)



(b)



(c)

Figure 4.S.4

In Exercises 29 to 36 graph the given functions, showing extrema, inflection points, and asymptotes. **29.** $e^{-2x}\sin(x)$, x in $[0, 4\pi]$

30. $\frac{e^x}{1-e^x}$

31. $x^3 - 9x^2$

32. $x\sqrt{3-x}$

33. $\frac{x-1}{x-2}$

34. $\cos(x) - \sin(x)$, x in $[0, 2\pi]$

35. $x^{1/2} - x^{1/4}$

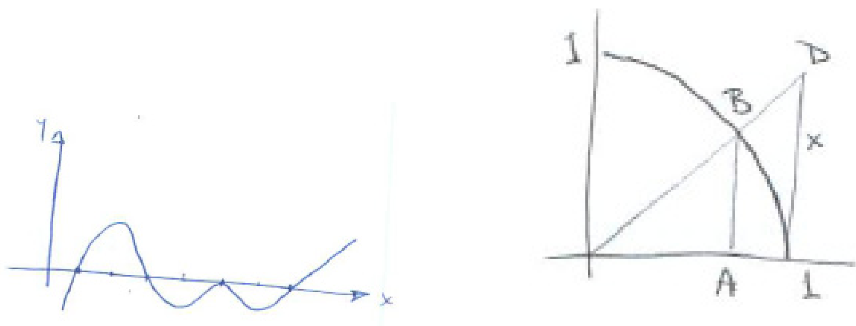
36. $\frac{x}{4-x^2}$

37. Figure 4.S.5(a) shows the graph of a function f. Estimate the arguments where

(a) f changes sign,

(b) f' changes sign,

(c) f'' changes sign.



(a) (b)

Figure 4.S.5

38. Assume the function f has continuous f' and f'' defined on an open interval.

(a) If $f'(a) = 0$ and $f''(a) = 0$, does f necessarily have an extrema at a? Explain.

(b) If $f''(a) = 0$, does f necessarily have an inflection point at $x = a$?

(c) If $f'(a) = 0$ and $f''(a) = 3$, does f necessarily have an extremum at a?

39. Find the maximum value of $e^{2\sqrt{3}x}\cos(2x)$ for x in $[0, \pi/4]$.

40.

(a) Show that the equation $5x - \cos(x) = 0$ has exactly one solution.

(b) Find a specific interval which contains the solution.

41. Consider the function f given by the formula $f(x) = x^3 - 3x$.

(a) At which numbers x is $f'(x) = 0$?

(b) Use the theorem of the Interior Extremum to show that the maximum value of $x^3 - 3x$ for x in $[1, 5]$ occurs either at 1 or at 5.

42. Let f and g be polynomials without a common root.

(a) Show that if the degree of g is odd, the graph of f/g has a vertical asymptote.

(b) Show that if the degree of f is less than or equal to the degree of g, then f/g has a horizontal asymptote.

43. If $\lim_{x\to\infty} f'(x) = 0$, does it follow that f has a horizontal asymptote? Explain.

44. Let f be a positive function on $(0,\infty)$ with f' and f'' both continuous. Let $g = f^2$.

(a) If f is increasing, is g?

(b) If f is concave up, is g?

45. Give an example of a positive function on $(0,\infty)$ that is concave down but f^2 is concave up.

46. Graph $\cos(2\theta) + 4\sin(\theta)$ for θ in $[0, 2\pi]$.

47. Graph $\cos(2\theta) + 2\sin(\theta)$ for θ in $[0, 2\pi]$.

48. Figure 4.S.5(b) shows part of a unit circle. The line segment CD is tangent to the circle and has length x. This exercise uses calculus to show that $AB < BC < CD$. (BC is the length of arc joining B and C.)

(a) Express AB and BC in terms of x.

(b) Using (a) and calculus, show that for $x > 0$, $AB < BC < CD$.

49. Assume that $f''(x)$ is positive for x in an open interval. Let $a < b$ be in the interval. In this exercise you will show that the chord joining $(a, f(a))$ to $(b, f(b))$ lies above the graph of f. ("A concave up curve has chords that lie above the curve.") Compare this Exercise with Exercise 17 in Section 4.4.

(a) Why does one want to prove that

$$f(a) + \frac{f(b) - f(a)}{b - a}(x - a) > f(x), \qquad \text{for } a < x < b?$$

(b) Why does one want to prove that

$$\frac{f(b) - f(a)}{b - a} > \frac{f(x) - f(a)}{x - a}? \tag{4.S.1}$$

(c) Show that the function on the right-hand side of the inequality in (b) is increasing for $a < x < b$. Why does this show that chords lie above the curve?

50.

Sam: I can do Exercise 49 more easily. I'll show that (4.S.1) is true. By the Mean-Value Theorem, I can write the left side as $f'(c)$ where c is in $[a, b]$ and the right side as $f'(d)$ where d is in $[a, x]$. Since $b > x$, I know $c > d$, hence $f'(c) > f'(d)$. Nothing to it.

Is Sam's reasoning correct?

51.

(a) Graph $y = \frac{\sin(x)}{x}$ showing intercepts and asymptotes.

(b) Graph $y = x$ and $y = \tan(x)$ relative to the same axes.

(c) Use (b) to find how many solutions there are to the equation $x = \tan(x)$.

(d) Write a short commentary on the critical points of $\sin(x)/x$. (Part (c) may come in handy.)

(e) Refine the graph produced in (a) to show several critical points.

52. Let $f(x) = ax^3 + bx^2 + cx + d$, where $a \neq 0$.

(a) Show that the graph of $y = f(x)$ always has exactly one inflection point.

(b) Show that the inflection point separates the graph of the cubic polynomial into two parts that are congruent. (Show the graph is symmetric with respect to the inflection point.) Why can one assume it is enough to show this for $a = 1$ and $d = 0$?

53. Find all functions $f(x)$ such that $f'(x) = 2$ for all x and $f(1) = 4$.

54. Find all differentiable functions such that $f(1) = 3$, $f'(1) = -1$, and $f''(1) = e^x$.

55.

(a) Graph $y = 1/(1 + 2^{-x})$.

(b) The point $(0, 1/2)$ is on the graph and divides it into two pieces. Are the two pieces congruent?

(Curves of this type model the depletion of a finite resource; x is time and y is the fraction used up to time x. See also Exercise 73 in the Chapter 5 Summary.)

56.

(a) If the graph of f has a horizontal asymptote (say, $\lim_{x\to\infty} f(x) = L$), does it follow that $\lim_{x\to\infty} f'(x)$ exists?

(b) If $\lim_{x\to\infty} f'(x)$ exists in (a), most it be 0?

57. Assume that f is continuous on $[1,3]$, $f(1) = 5$, $f(2) = 4$, and $f(3) = 5$. Show that the graph of f has a horizontal chord of length 1.

58. A function f defined on the whole x-axis has continuous first- and second-derivatives and exactly one inflection point. In at most how many points can a straight line intersect the graph of f? Explain. (x^n, n an odd integer greater than 1, are examples of such functions.)

59. Let f be an increasing function with continuous f' and f''. What, if anything, can be said about the concavity of $f \circ f$ if

(a) f is concave up?

(b) f is concave down?

60. Assume f has continuous f' and f''. Show that if f and $g = f^2$ have inflection points at the same argument a, then $f'(a) = 0$.

61. Graph $y = x^2 \ln(x)$, showing extrema and inflection points. Use the fact that $\lim_{x\to 0^+} x^2 \ln(x) = 0$; see Exercise 20 of Section 5.6.

62. Assume $\lim\limits_{x\to\infty} f'(x) = 3$. Show that for x sufficiently large, $f(x)$ is greater than $2x$. (Review the Mean-Value Theorem.)

63. Assume that f is differentiable for all numbers x.

(a) If f is an even function, what, if anything, can be said about $f'(0)$?

(b) If f is an odd function, what, if anything, can be said about $f'(0)$?

Explain your answers.

64. Graph $y = \sin(x^2)$ on the interval $[-\sqrt{\pi}, \sqrt{\pi}]$. Identify the extreme points and the inflection points.

65. Assume that $f(x)$ is a continuous function not identically 0 defined on $(-\infty, \infty)$ and that $f(x+y) = f(x) \cdot f(y)$ for all x and y.

(a) Show that f(0)=1.

(b) Show that $f(x)$ is never 0.

(c) Show that $f(x)$ is positive for all x.

(d) Letting $f(1) = a$, find $f(2)$, $f(1/2)$, and $f(-1)$.

(e) Show that $f(x) = a^x$ for all x.

66. Can a straight line meet the curve $y = x^5$ four times?

67. Assume $y = f(x)$ is a twice differentiable function with $f(0) = 1$ and $f''(x) < -1$ for all x. Is it possible that $f(x) > 0$ for all x in $(1, \infty)$?

68. If $\lim_{x\to\infty} f'(x) = 3$, does it follow that the graph of $y = f(x)$ is asymptotic to some line of the form $y = a + 3x$ for some constant a?

69. Assume that $f(x)$ is defined for all real numbers and has a continuous derivative. Assume that $f'(x)$ is positive for all x other than c and that $f'(c) = 0$.

(a) Give an two examples of functions with these properties.

(b) Must any function with these properties be increasing?

Calculus is Everywhere # 5
Calculus Reassures a Bicyclist

Both authors enjoy bicycling for pleasure and running errands in our flat towns. One of the authors (SS) often bicycles to campus through a parking lot. On each side of his route is a row of parked cars. At any moment a car can back into his path. Wanting to avoid a collision, he wondered where he should ride. The farther he rode from a row, the safer he was. However, the farther he rode from one row, the closer he would be to the other row. Where should he ride?

Instinct told him to ride midway between the two rows, an equal distance from both. But he had second thoughts. Maybe it's better to ride, say, one-third of the way from one row to the other, which is the same as two-thirds of the way from the other row. That would mean he had two safest routes, depending on which row he was nearer. Wanting a definite answer, he resorted to calculus.

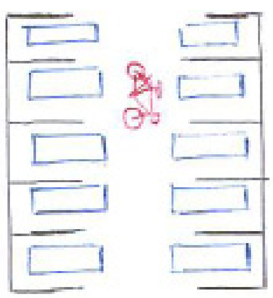


Figure C.5.1 ARTIST:picture of two rows of parked cars, with bicycle

He introduced a function, $f(x)$, which is the probability that he got through safely when his distance from one row was x, considering only cars in that row. Then he called the distance between the two rows d. When he was at a distance x from one row, he was at a distance $d-x$ from the other row. The probability that he did not collide with a car backing out from either row was then the product, $f(x)f(d-x)$. His intuition said that this was maximized when $x = d/2$, putting him midway between the two rows.

What did he know about f? First of all, the farther he rode from one line of cars, the safer he was. So f was an increasing function; thus f' is positive. Moreover, when he was very far from the cars, the probability of riding safely through the lot approached 1. So he assumed $\lim_{x\to\infty} f(x) = 1$ (which it turned out he did not need).

The derivative of f' measured the rate at which he gained safety as he increased his distance from the cars. When x was small, and he rode near the cars, $f'(x)$ was large: he gained a great deal of safety by increasing x. However, when he was far from the cars, he gained very little. That means that f' was a decreasing function. In other words f'' is negative.

Does that information about f imply that midway is the safest route?

In other words, does the maximum of $f(x)f(d-x)$ occur when $x = d/2$? Symbolically, is

$$f(d/2)f(d/2) \geq f(x)f(d-x)?$$

To begin, he took the logarithm of that expression, in order to replace a product by something easier, a sum. He wanted to see if

$$2\ln(f(d/2)) \geq \ln(f(x)) + \ln(f(d-x)).$$

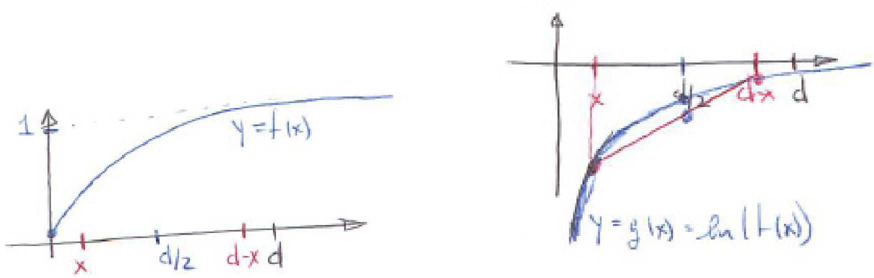
Letting $g(x)$ denote the composite function $\ln(f(x))$, he faced the inequality,

$$2g(d/2) \geq g(x) + g(d-x),$$

or

$$g(d/2) \geq \frac{1}{2}(g(x) + g(d-x)).$$

This inequality asserts that the point $(d/2, g(d/2))$ on the graph of g is at least



(a) (b)

Figure C.5.2

as high as the midpoint of the chord joining $(x, g(x))$ to $(d-x, g(d-x))$. This would be the case if the second derivative of g were negative, and the graph of g were concave down. He had to compute g'' and hope it was negative. First of all, $g'(x)$ was $f'(x)/f(x)$. Then $g''(x)$ was

$$\frac{f(x)f''(x) - (f'(x))^2}{f(x)^2}.$$

The denominator was positive. Because $f(x)$ is positive and concave down, the numerator was negative. So the quotient was negative. That means that the safest path was midway between the two rows. The bicyclist continued to follow that route, but, after these calculations, with more confidence that it was indeed the safest way.

Calculus is Everywhere # 6
Graphs in Economics

Elementary economics texts are full of graphs. They provide visual images of a variety of concepts, such as production, revenue, cost, supply, and demand. Here we show how economists use graphs to help analyze production as a function of the amount of labor, that is, the number of workers.

Let $P(L)$ be the amount of some product, such as cell phones, produced by a firm employing L workers. Since both workers and wireless network cards come in integer amounts, the graph of $P(L)$ is just a bunch of dots. In practice, these dots suggest a curve, and the economists use that curve in their analysis. So $P(L)$ is viewed as a differentiable function defined for some interval of the form $[0, b]$.

If there are no workers, there is no production, so $P(0) = 0$. When the first few workers are added, production may increase rapidly, but as more are hired, production may still increase, but not as rapidly. Figure C.6.1 is a typical **production curve**. It seems to have an inflection point when the gain from adding more workers begins to decline. The inflection point of $P(L)$ occurs at L_2 in Figure C.6.2.

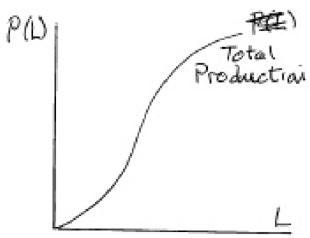


Figure C.6.1

When the firm employs L workers and adds one more, production increases by $P(L+1) - P(L)$, the marginal production. Economists manage to relate this to the derivative by a simple trick:

$$P(L+1) - P(L) = \frac{P(L+1) - P(L)}{(L+1) - L} \tag{C.6.1}$$

The right-hand side of (C.6.1) is "change in output" divided by "change in input," which is, by the definition of the derivative, an approximation to the derivative, $P'(L)$. For this reason economists define the **marginal production** as $P'(L)$, and think of it as the extra product produced by the "L plus first" worker. We denote the marginal product as $m(L)$, that is, $m(L) = P'(L)$.

The **average production** per worker when there are L workers is defined as the quotient $P(L)/L$, which we denote $a(L)$. We have three functions: $P(L)$, $m(L) = P'(L)$, and $a(L) = P(L)/L$.

Now the fun begins.

> *At what point on the graph of the production function is the average production a maximum?*

Since $a(L) = P(L)/L$, it is the slope of the line from the origin to the point $(L, P(L))$ on the graph. Therefore we are looking for the point on the graph where the slope is a maximum. One way to find that point is to rotate a straightedge around the origin, clockwise, starting at the vertical axis until it meets the graph, as in Figure C.6.2. Call the point of tangency $(L_1, P(L_1))$. For L less than L_1 or greater than L_1, average productivity is less than $a(L_1)$.

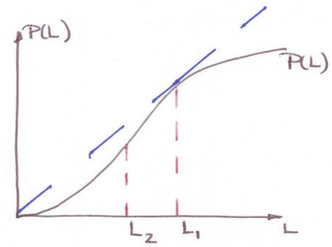


Figure C.6.2

Note that at L_1 the average product is the same as the marginal product, for the slope of the tangent at $(L_1, P(L_1))$ is both the quotient $P(L_1)/L_1$ and the derivative $P'(L_1)$. We can use calculus to obtain the same conclusion:

Since $a(L)$ has a maximum when the input is L_1, its derivative is 0 then. The derivative of $a(L)$ is

$$\frac{d}{dL}\left(\frac{P(L)}{L}\right) = \frac{LP'(L) - P(L)}{L^2}. \tag{C.6.2}$$

At L_1 the quotient in (C.6.2) is 0. Therefore, its numerator is 0, from which it follows that $P'(L_1) = P(L_1)/L_1$. (You might take a few minutes to see why this equation should hold, without using graphs or calculus.)

In any case, the graphs of $m(L)$ and $a(L)$ cross when L is L_1. For smaller values of L, the graph of $m(L)$ is above that of $a(L)$, and for larger values it is below, as shown in Figure C.6.3.

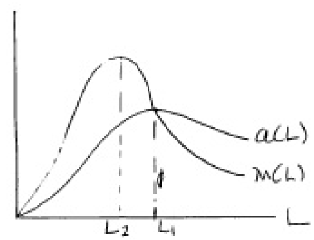


Figure C.6.3

What does the maximum point on the marginal product graph tell about the production graph?

Assume that $m(L)$ has a maximum at L_2. For smaller L than L_2 the derivative of $m(L)$ is positive. For L larger than L_2 the derivative of $m(L)$ is negative. Since $m(L)$ is defined as $P'(L)$, the second derivative of $P(L)$ switches from positive to negative at L_2, showing that the production curve has an inflection point at $(L_2, P(L_2))$.

Economists use similar techniques to deal with a variety of concepts, such as marginal and average cost or marginal and average revenue, viewed as functions of labor or of capital.

Chapter 5

More Applications of Derivatives

Chapter 2 constructed the foundation for derivatives, namely the concept of a limit. Chapters 3 and 4 developed the derivative and applied it to graphs of functions. The present chapter will apply the derivative in a variety of ways, such as: finding the most efficient method to accomplish a task (Section 5.1), connecting the rate one variable changes to the rate another changes (Sections 5.2 and 5.3), the approximation of functions such as e^x by polynomials (Sections 5.4 and 5.5), the evaluation of certain limits (Section 5.6), natural growth and decay (Section 5.7), and to presenting certain special functions (Section 5.8).

5.1 Applied Maximum and Minimum Problems

In Chapter 4, we saw how the derivative and second derivative are of use in finding the maxima and minima of a given function – the locally high and low points on its graph. Now we will use these same techniques to find extrema in applied problems. Though the examples will be drawn mainly from geometry they illustrate the general procedure. The main challenge in these situations is figuring out the formula for the function that describes the quantity to be maximized (or minimized).

The General Procedure

The general procedure runs something along these lines.

1. Get a feel for the problem (experiment with particular cases.)
2. Devise a formula for the function whose maximum or minimum you want to find.
3. Determine the domain of the function – that is, the inputs *that make sense in the application.*
4. Find the maximum or minimum of the function found in Step 2 for inputs that are in the domain identified in Step 3.

Additional worked examples can be found on the website for this book.

The most important step is finding a formula for the function. To become skillful at doing this takes practice. First, carefully read and study the three examples that comprise the remainder of this section.

A Large Garden

EXAMPLE 1 A couple have enough wire to construct 100 feet of fence. They wish to use it to form three sides of a rectangular garden, one side of which is along a building, as shown in Figure 5.1.1. What shape garden should they choose in order to enclose the largest possible area?

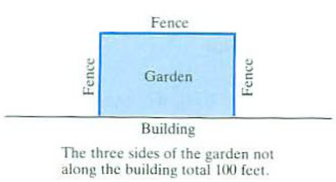
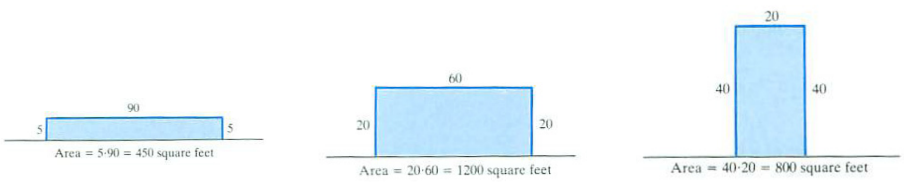


Figure 5.1.1

SOLUTION *Step* 1. First make a few experiments. Figure 5.1.2 shows some possible ways of laying out the 100 feet of fence. In the first case the side parallel to the building is very long, in an attempt to make a large area. However, doing this forces the other sides of the garden to be small. The area is $90 \times 5 = 450$ square feet. In the second case, the garden has a larger area, $60 \times 20 = 1200$ square feet. In the third case, the side parallel to the building



(a) (b) (c)

Figure 5.1.2

is only 20 feet long, but the other sides are longer. The area is $20 \times 40 = 800$ square feet.

In all three cases, once the length of the side parallel to the building is set, the other side lengths are known and the area can be computed.

Clearly, we may think of the *area of the garden as a function of the length of the side parallel to the building.*

Step 2. Let $A(x)$ be the area of the garden when the length of the side parallel to the building is x feet, as in Figure 5.1.3. The other sides of the garden have length y. But y is completely determined by x since the total length of the fence is 100 feet:

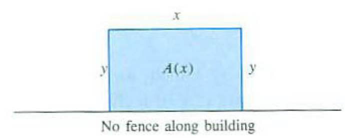


Figure 5.1.3

$$x + 2y = 100.$$

Thus $y = (100 - x)/2$.

Since the area of a rectangle is its length times its width,

$$A(x) \quad = \quad xy = x\left(\frac{100-x}{2}\right) = 50x - \frac{x^2}{2}.$$

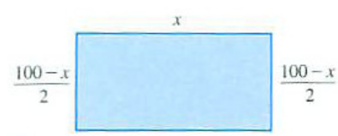


Figure 5.1.4

(See Figure 5.1.4.) We now have the function.

Step 3. Which values of x in (5.1.1) correspond to possible gardens?

Since there is only 100 feet of fence, $x \leq 100$. Furthermore, it makes no sense to have a negative amount of fence; hence $x \geq 0$. Therefore the domain on which we wish to consider the function (5.1.1) is the closed interval $[0, 100]$.

Step 4. To maximize $A(x) = 50x - x^2/2$ on $[0, 100]$ we examine $A(0)$, $A(100)$, and the value of $A(x)$ at any critical numbers.

To find critical numbers, differentiate $A(x)$:

$$A(x) = 50x - \frac{x^2}{2} \qquad \text{so} \qquad A'(x) = 50 - x$$

and solve $A'(x) = 0$ to find:

$$0 = 50 - x \qquad \text{or} \qquad x = 50.$$

There is one critical number, 50.

All that is left is to find the largest of $A(0)$, $A(100)$, and $A(50)$. We have

$$\begin{aligned} A(0) &= 50 \cdot 0 - \frac{0^2}{2} = 0, \\ A(100) &= 50 \cdot 100 - \frac{100^2}{2} = 0, \\ \text{and} \qquad A(50) &= 50 \cdot 50 - \frac{50^2}{2} = 1250. \end{aligned}$$

The maximum possible area is 1250 square feet, and the fence should be laid out as shown in Figure 5.1.5. ◇

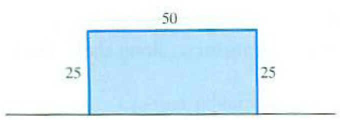


Figure 5.1.5

A Large Tray

EXAMPLE 2 Four congruent squares are cut out of the corners of a square piece of cardboard 12 inches on each side and the four remaining flaps can be folded up to obtain a tray without a top. (See Figure 5.1.6.) What size squares should be cut in order to maximize the volume of the tray?

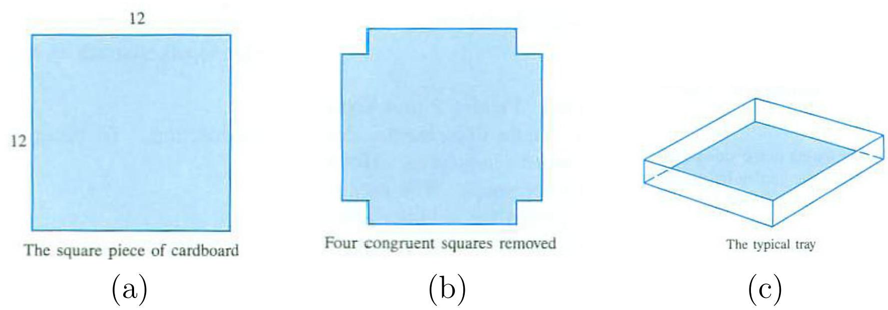
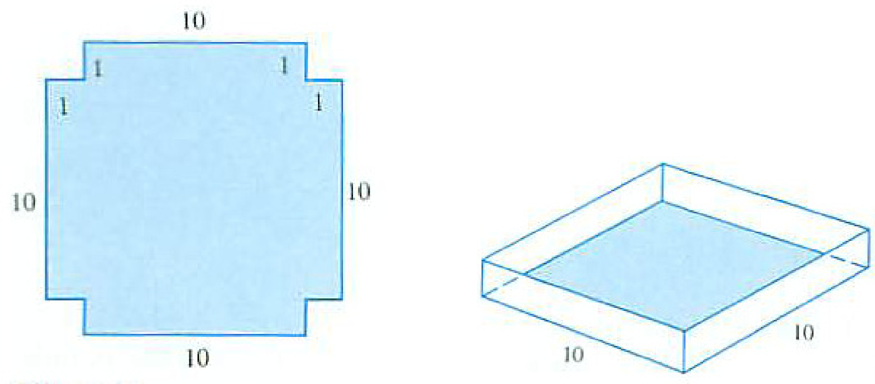


Figure 5.1.6

SOLUTION *Step* 1. First we get a feel for the problem. Let us make a couple of experiments.

Say that we remove small squares that are 1 inch by 1 inch, as in Figure 5.1.7(a). When we fold up the flaps we obtain a tray whose base is a 10-inch by 10-inch square and whose height is 1 inch, as in Figure 5.1.7(b). The volume of the tray is

$$\text{Area of base} \times \text{height} = \underbrace{10 \times 10}_{\text{base area}} \times \underbrace{1}_{\text{height}} = 100 \text{ cubic inches.}$$



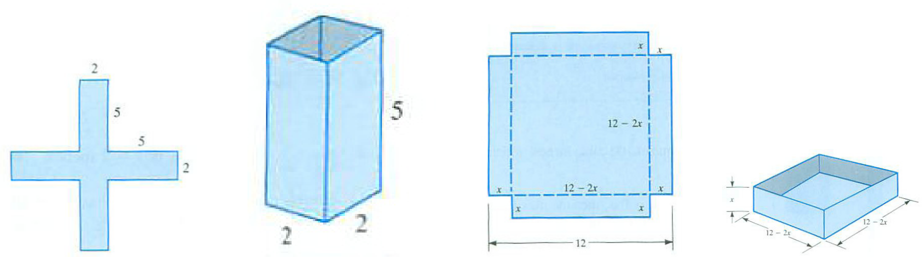
(a) (b)

Figure 5.1.7

For our second experiment, let's try cutting out a large square, say 5 inches by 5 inches, as in Figure 5.1.8(a). When we fold up the flaps, we get a very tall tray with a very small base, as in Figure 5.1.8(b). It volume is

$$\text{Area of base} \times \text{height} = 2 \times 2 \times 5 = 20 \text{ cubic inches.}$$

Clearly *volume depends on the size of the cut-out squares.* The function



(a) (b) (c) (d)

Figure 5.1.8

we will investigate is $V(x)$, the volume of the tray formed by removing four squares whose sides all have length x.

Step 2. To find the formula for $V(x)$ we make a *large*, clear diagram of the typical case, as in Figure 5.1.8(c) and Figure 5.1.8(d). Now

$$\text{Volume of tray} = \underbrace{(12-2x)}_{\text{length}}\underbrace{(12-2x)}_{\text{width}}\underbrace{x}\text{height} = (12-2x)^2 x,$$

hence

$$V(x) = (12 - 2x)^2 x = 4x^3 - 48x^2 + 144x. \tag{5.1.1}$$

We have obtained a formula for volume as a function of the length of the sides of the cut-out squares.

Step 3. Next determine the domain of the function $V(x)$ that is *meaningful* in the problem.

The smallest that x can be is 0. In this case the tray has height 0 and is just a flat piece of cardboard. (Its volume is 0.) The size of the cut is not more than 6 inches, since the cardboard has sides of length 12 inches. The cut can be as near 6 inches as we please, and the nearer it is to 6 inches, the smaller is the base of the tray. For convenience of our calculations, we allow cuts with $x = 6$, when the area of the base is 0 square inches and the height is 6 inches. (The volume is again 0 cubic inches.) Therefore the domain of the volume function $V(x)$ is the closed interval $[0, 6]$.

Step 4. To maximize $V(x) = 4x^3 - 48x^2 + 144x$ on $[0, 6]$, evaluate $V(x)$ at critical numbers in $[0, 6]$ and at the endpoints of $[0, 6]$.

We have

$$V'(x) \quad = \quad 12x^2 - 96x + 144 = 12(x^2 - 8x + 12) = 12(x - 2)(x - 6).$$

A critical number is a solution to the equation

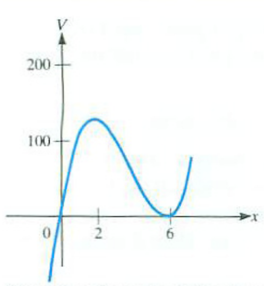
$$0 = 12(x - 2)(x - 6).$$

Hence $x - 2 = 0$ or $x - 6 = 0$. The critical numbers are 2 and 6.

The endpoints of the interval $[0, 6]$ are 0 and 6. Therefore the maximum value of $V(x)$ for x in $[0, 6]$ is the largest of $V(0)$, $V(2)$, and $V(6)$. Since $V(0) = 0$ and $V(6) = 0$, the largest value is

$$V(2) = 4(2^3) - 48(2^2) + 144 \cdot 2 = 128 \text{ cubic inches.}$$

The cut that produces the tray with the largest volume is $x = 2$ inches. ◇



Only values of x in the portion above $[0, 6]$ correspond to physically realizable trays.

Figure 5.1.9

As a matter of interest, we graph the function V, showing its behavior for all x, not just for values of x significant in the problem. Note in Figure 5.1.9 that at $x = 2$ and $x = 6$ the tangent is horizontal.

Remark: In Example 2 you might say $x = 0$ and $x = 6$ don't really correspond to what you would call a tray. If so, you would restrict the domain of $V(x)$ to the open interval $(0, 6)$. You would then have to examine the behavior of $V(x)$ for x *near* 0 and for x *near* 6. By making the domain $[0, 6]$ from the start, you avoid the extra work of examining $V(x)$ for x *near* the ends of the interval.

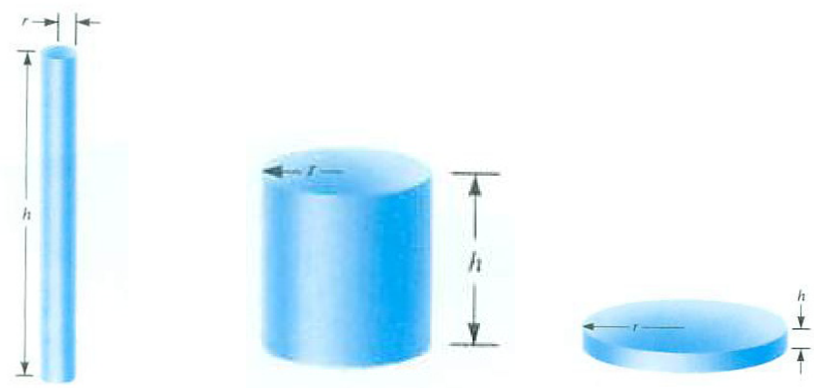
The key step in these two examples, and in any applied problem, is Step 2: findng a formula for the quantity whose extremum you are seeking. In case the problem is geometrical, the following chart may be of aid.

Setting Up the Function

1. Draw and label the appropriate diagrams.
   (Make them large enough so that there is room for labels.)
2. Label the various quantities by letters, such as x, y, A, V.
3. Identify the quantity to be maximized (or minimized).
4. Express the quantity to be maximized (or minimized) in terms of one or more of the other variables.
5. Finally, express that quantity in terms of only one variable.

An Economical Can

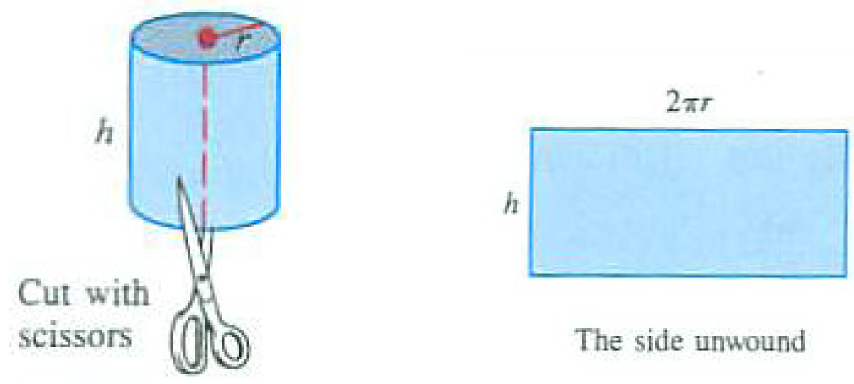
EXAMPLE 3 Of all the tin cans that enclose a volume of 100π cubic centimeters, which requires the least metal?



(a) (b) (c)

Figure 5.1.10

SOLUTION *Step* 1. The can may be flat or tall. If the can is flat, the side uses little metal, but then the top and bottom bases are large. If the can is shaped like a mailing tube, then the two bases require little metal, but the curved side requires a great deal of metal. (See Figure 5.1.10, where r denotes the radius and h the height of the can.) What is the ideal compromise between these two extremes?



(a) (b)

Figure 5.1.11

Step 2. The surface area S of the can is the sum of the area of the top, side, and bottom. The top and bottom are disks with radius r so their total area is $2\pi r^2$. Figure 5.1.11 shows why the area of the side is $2\pi rh$. The total surface area of the can is given by

$$S = 2\pi r^2 + 2\pi rh. \tag{5.1.2}$$

Since the amount of metal in the can is proportional to S, it suffices to minimize S.

Equation (5.1.2) gives S as a function of two variables, but we can express one of the variables in terms of the other. The radius and height are related by the equation

$$V = \pi r^2 h = 100\pi, \tag{5.1.3}$$

since the volume is 100π cubic centimeters. In order to express S as a function of one variable, use (5.1.3) to eliminate either r or h. Choosing to eliminate h, we solve (5.1.3) for h,

$$h = \frac{100}{r^2}.$$

Substitution into (5.1.2) yields

$$S = 2\pi r^2 + 2\pi r\frac{100}{r^2} \qquad \text{or} \qquad S = 2\pi r^2 + \frac{200}{r}\pi. \tag{5.1.4}$$

Equation (5.1.4) expresses S as a function of just one variable, r.

The cans have a positive radius as large as you please.

Step 3. The function $S(r)$ is continuous and differentiable on $(0, \infty)$.

Step 4. Compute dS/dr:

$$\frac{dS}{dr} = 4\pi r - \frac{200\pi}{r^2}\pi = \frac{4\pi r^3 - 200\pi}{r^2}. \tag{5.1.5}$$

Set the derivative equal to 0 to find any critical numbers. We have

$r = 0$ is *not* a critical number because it is not in the domain of V.

$$\begin{aligned} 0 &= \frac{4\pi r^3 - 200\pi}{r^2}, \\ \text{hence} \quad 0 &= 4\pi r^3 - 200\pi \\ \text{or} \quad 4\pi r^3 &= 200\pi \\ r^3 &= \frac{200}{4} \\ r &= \sqrt[3]{50} \approx 3.684. \end{aligned}$$

There is only one critical number. Does it provide a minimum? Let's check it two ways, first by the first-derivative test, then by the second-derivative test.

The first derivative is

$$\frac{dS}{dr} = \frac{4\pi r^3 - 200\pi}{r^2}. \tag{5.1.6}$$

When $r = \sqrt[3]{50}$, the numerator in (5.1.6) is 0. When $r < \sqrt[3]{50}$ the numerator is negative and when $r > \sqrt[3]{50}$ the numerator is positive. (The denominator is always positive.) Since $dS/dr < 0$ for $r < \sqrt[3]{50}$, and $dS/dr > 0$ for $r > \sqrt[3]{50}$, the function $S(r)$ decreases for $r < \sqrt[3]{50}$ and increases for $r > \sqrt[3]{50}$. That shows that a global minimum occurs at $\sqrt[3]{50}$. (See Figure 5.1.12(a).)

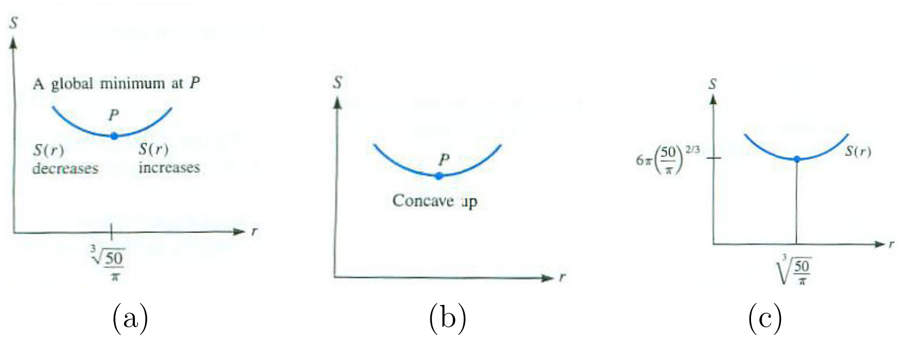


Figure 5.1.12

Let us instead use the second-derivative test. Differentiation of (5.1.5) gives

$$\frac{d^2S}{dr^2} = 4\pi + \frac{400}{r^3}\pi. \tag{5.1.7}$$

Inspection of (5.1.7) shows that for all meaningful values of r, that is r in $(0, \infty)$, d^2S/dr^2 is positive. (The function is concave up as shown in Figure 5.1.12(b).) Not only is P a relative minimum, it is a global minimum, since the graph lies above its tangents, in particular, the tangent at P.

The minimum of $S(r)$ is shown in Figure 5.1.12(c).

To find the height of the most economical can, solve (5.1.7) for h:

$$\begin{aligned} h = \tfrac{100}{r^2} &= \tfrac{100}{(\sqrt[3]{50})^2} \\ &= \tfrac{100}{(\sqrt[3]{50})^2}\tfrac{\sqrt[3]{50}}{\sqrt[3]{50}} && \text{rationalize the denominator} \\ &= \tfrac{100}{50}\sqrt[3]{50} = 2\sqrt[3]{50}. \end{aligned}$$

The height of the can is equal to twice its radius, that is, its diameter. The total surface area of the can is

$$S = 2\pi r^3 + \frac{200\pi}{r}\bigg|_{r=50^{1/3}} = (100 + 4\cdot 50^{2/3}) \approx 154.288 \text{ square centimeters.}$$

◇

Summary

We showed how to use calculus to solve applied problems: experiment, set up a function, find its domain, and its critical points. Then test the critical points and endpoints of the domain to determine the extrema.

1. Draw and label appropriate diagrams.
2. Express the quantity to be optimized as a function of one variable.
3. Determine the domain of the function.
4. Use the first or second derivative test to determine the maximum or minimum of the function in its domain.

If the interval is closed, the maximum or minimum will occur at a critical point or an endpoint. If the interval is not closed, a little more care is needed to confirm that a critical number provides an extremum.

With practice this process becomes second nature.

EXERCISES for Section 5.1

1. A gardener wants to make a rectangular garden with 100 feet of fence. What is the largest area the fence can enclose?

2. Of all rectangles with area 100 square feet, find the one with the shortest perimeter.

3. Solve Example 1, expressing A in terms of y instead of x.

4. A gardener is going to put a rectangular garden inside one arch of the cosine curve, as shown in Figure 5.1.13. What is the garden with the largest area.

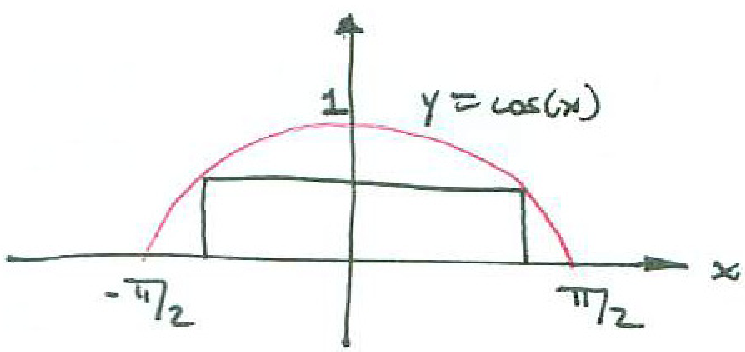


Figure 5.1.13

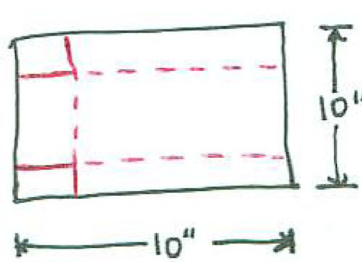
Exercises 5 to 8 are related to Example 2. In each case find the length of the cut that maximizes the volume of the tray. The dimensions of the cardboard are given.

5. 5 inches by 5 inches

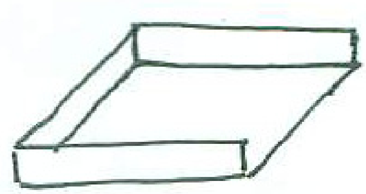
6. 5 inches by 7 inches

7. 4 inches by 8 inches,

8. 6 inches by 10 inches,



(a)



(b)

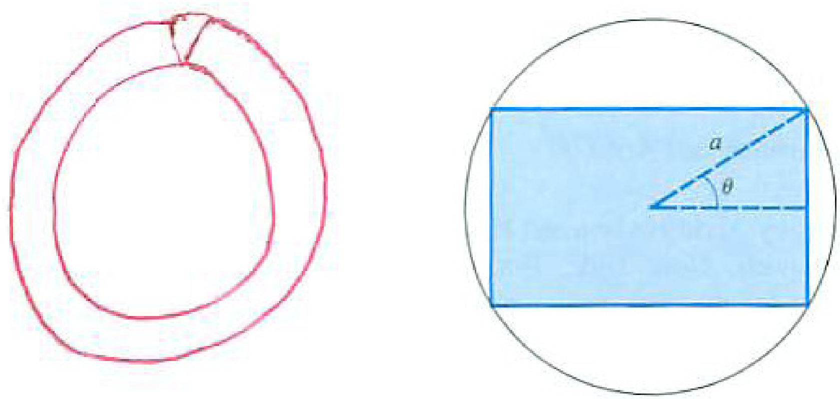
Figure 5.1.14

9. Starting with a square piece of paper $10''$ on a side, Sam wants to make a paper holder with three sides. The pattern he will use is shown in Figure 5.1.14 along with the tray. He will remove two squares and fold up three flaps.

(a) What size square maximizes the volume of the tray?

(b) What is that volume?

10. A chef wants to make a cake pan out of a circular piece of aluminum of radius 12 inches. To do this he plans to cut the circular base from the center of the piece and then cut the side from the remainder. What should the radius and height be to maximize the volume of the pan? (See Figure 5.1.15(a).)



(a) (b)

Figure 5.1.15

11. Solve Example 3, expressing S in terms of h instead of r.

12. Of all cylindrical tin cans *without a top* that contains 100 cubic inches, which requires the least material?

13. Of all enclosed rectangular boxes with square bases that have a volume of 1000 cubic inches, which uses the least material?

14. Of all topless rectangular boxes with square bases that have a volume of 1000 cubic inches, which uses the least material?

15. Find the dimensions of the rectangle of largest area that can be inscribed in a circle of radius a. The typical rectangle is shown in Figure 5.1.15(b). (Express the area in terms of the angle θ shown.)

16. Solve Exercise 15, expressing the area in terms of half the width of the rectangle, x. (Square the area to avoid square roots.)

17. Find the dimensions of the rectangle of largest perimeter that can be inscribed in a circle of radius a.

18. Show that of all rectangles of a given area, the square has the shortest perimeter. *Suggestion:* Call the fixed area A and keep in mind that it is a constant.

19. A rancher wants to construct a rectangular corral. He also wants to divide the corral by a fence parallel to one of the sides. He has 240 feet of fence. What are the dimensions of the corral of largest area he can enclose?

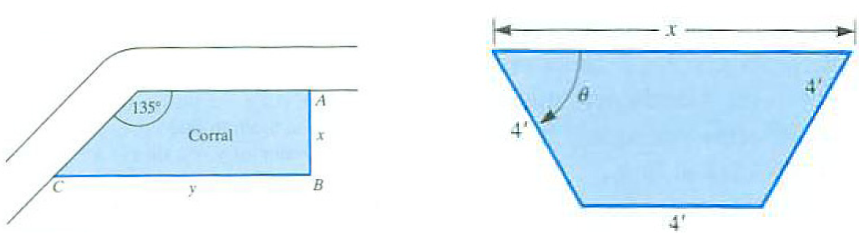
20. A river has a 45° turn, as indicated in Figure 5.1.16(a). A rancher wants to construct a corral bounded on two sides by the river and on two sides by 1 mile of fence ABC, as shown. Find the dimensions of the corral of largest area.

21.

(a) How should one choose two nonnegative numbers whose sum is 1 in order to maximize the sum of their squares?

(b) To minimize the sum of their squares?

22. How should one choose two nonnegative numbers whose sum is 1 in order to maximize the product of the square of one of them and the cube of the other?



(a) (b)

Figure 5.1.16

23. An irrigation channel made of concrete is to have a cross section in the form of an isosceles trapezoid, three of whose sides are 4 feet long. See Figure 5.1.16(b).

How should the trapezoid be shaped if it is to have the maximum possible area? (Consider the area as a function of x and solve.)

24.

(a) Solve Exercise 23 expressing the area as a function of θ instead of x.

(b) Do the answers in (a) and Exercise 23 agree? Explain.

In Exercises 25 to 28 use the fact that the combined length and girth (distance around) of a package to be sent through the mail by the United States Postal Service (USPS) cannot exceed 108 inches. The combined length and girth of a packages sent as "parcel post" is 130 inches. The United Parcel Service (UPS) limit is 165 inches for combined length and girth with the length not exceeding 108 inches.

25. Find the dimensions of the right circular cylinder of largest volume that can be sent through the mail.

26. Find the dimensions of the right circular cylinder of largest surface area that can be sent through the USPS.

27. Find the dimensions of the rectangular box with square base of largest volume that can be sent through the USPS.

28. Find the dimensions of the rectangular box with square base of largest surface area that can be sent through the USPS.

29.

(a) Repeat Exercise 25 with for a package sent by UPS.

(b) Generalize your solutions to Exercise 25 for a packages subject to a combined length and girth that does not exceed M inches.

30.

(a) Repeat Exercise 26 with for a package sent by UPS.

(b) Generalize your solutions to Exercise 26 for a packages subject to a combined length and girth that does not exceed M inches.

Exercises 31 to 38 concern "minimal cost" problems.

31. A cylindrical can is to be made to hold 100 cubic inches. The material for its top and bottom costs twice as much per square inch as the material for its side. Find the radius and height of the most economical can. *Warning:* This is not the same as Example 3.

(a) Would you expect the most economical can in this problem to be taller or shorter than the solution to Example 3? (Use common sense, not calculus.)

(b) For convenience, call the cost of 1 square inch of the material for the side k cents. Thus the cost of 1 square inch of the material for the top and bottom is $2k$ cents. (The precise value of k will not affect the answer.) Show that a can of radius r and height h costs

$$C = 4k\pi r^2 + 2k\pi rh \text{ cents.}$$

(c) Find r that minimizes the functions C in (b). Keep in mind during any differentiation that k is constant.

(d) Find the corresponding h.

32. A camper at A will walk to the river, put some water in a pail at P, and take it to the campsite at B.

(a) Express $AP + PB$ as a function of x.

(b) Use calculus to decide where P should be located to minimize the length of the walk, $AP + PB$? (See Figure 5.1.17.)

This exercise was first encountered as Exercise 34 in Section 1.1, where it was solved by geometry.

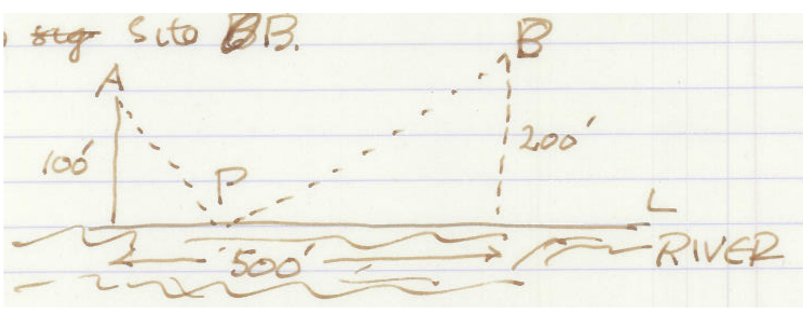


Figure 5.1.17 Sketch of situation in Exercise 32.

33. Sam is at the edge of a circular lake of radius one mile and Jane is at the edge, directly opposite. Sam wants to visit Jane. He can walk 3 miles per hour and he has a canoe. What mix of paddling and walking should Sam use to minimize the time needed to reach Jane if

(a) he paddles at least three miles an hour?

(b) he paddles at 1.5 miles per hour?

(c) he paddles at 2 miles per hour?

34. Consider a right triangle ABC, with C being at the right angle. There are two routes from A to B. One is direct, along the hypotenuse. The other is along the two legs, from A to C and then to B. Now, the shortest path between two points is the straight one. That raises this question: What is the largest percentage saving possible by walking along the hypotenuse instead of along the two legs? For which shape right triangle does this savings occur?

35. A rectangular box with a square base is to hold 100 cubic inches. Material for the top of the box costs 2 cents per square inch; material for the sides costs 3 cents per square inch; material for the bottom costs 5 cents per square inch. Find the dimensions of the most economical box.

36. The cost of operating a certain truck (for gasoline, oil, and depreciation) is $(20 + s/2)$ cents per mile when it travels at a speed of s miles per hour. A truck driver earns \$18 per hour. What is the most economical speed at which to operate the truck during a 600-mile trip?

(a) If you considered only the truck, would you want s to be small or large?

(b) If you, the employer, considered only the expense of the driver's wages, would you want s to be small or large?

(c) Express cost as a function of s and solve. (Be sure to put the costs all in terms of cents or all in terms of dollars.)

(d) Would the answer be different for a 1000-mile trip?

37. A government contractor who is removing earth from a large excavation can route trucks over either of two roads. There are $10,000$ cubic yards of earth to move. Each truck holds 10 cubic yards. On one road the cost per truckload is $1+2x^2$ cents, when x trucks use that raod; the function records the cost of congestion. On the other road the cost is $2 + x^2$ cents per truckload when x trucks use that road. How many trucks should be dispatched to each of the two roads?

38. On one side of a river 1 mile wide is an electric power station; on the other side, s miles upstream, is a factory. (See Figure 5.1.18.) It costs 3 dollars per foot to run cable over land and 5 dollars per foot under water. What is the most economical way to run cable from the station to the factory?

(a) Using no calculus, what do you think would be (approximately) the best route if s were very small? if s were very large?

(b) Solve with the aid of calculus, and draw the routes for $s = \frac{1}{2}$, $\frac{3}{4}$, 1, and 2.

(c) Solve for arbitrary s.

Warning: Minimizing the length of cable is *not* the same as minimizing its cost.

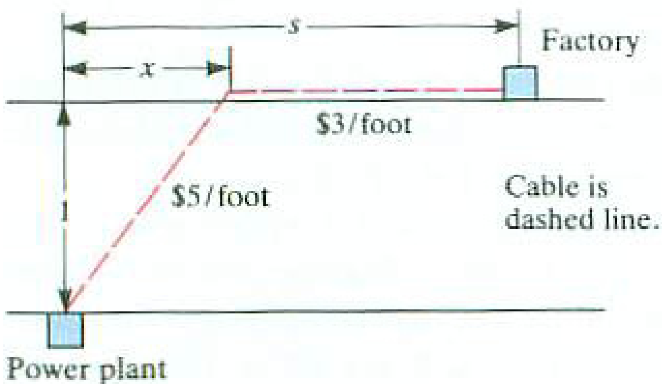


Figure 5.1.18

39. (From a text on the dynamics of airplanes.) "Recalling that

$$I = A\cos^2\theta + C\sin^2\theta - 2E\cos\theta\sin\theta,$$

we wish to find θ when I is a maximum or a minimum." Show that at an extremum of I,

$$\tan 2\theta = \frac{2\ E}{C - A}.\ (\text{assume that} A \neq C)$$

40. (From a physics text.) "By differentiating the equation for the horizontal range,

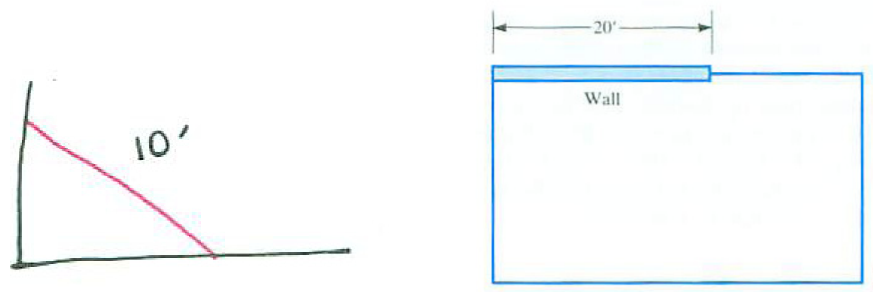
$$R = \frac{v_0^2 \sin(2\theta)}{g},$$

show that the initial elevation angle θ for maximum range is 45°." In the formula for R, v_0 and g are constants. (R is the horizontal distance a baseball covers if you throw it at an angle θ with speed v_0. Air resistance is disregarded.)

(a) Using calculus, show that the maximum range occurs when $\theta = 45°$.

(b) Solve the same problem without calculus.

41. A gardener has 10 feet of fence and wishes to make a triangular garden next to two buildings, as in Figure 5.1.19(a). How should he place the fence to enclose the maximum area?



(a) (b)

Figure 5.1.19

42. Fencing is to be added to an existing wall of length 20 feet, as shown in Figure 5.1.19(b). How should the extra fence be added to maximum the area of the enclosed rectangle if the additional fence is

(a) 40 feet long?

(b) 80 feet long?

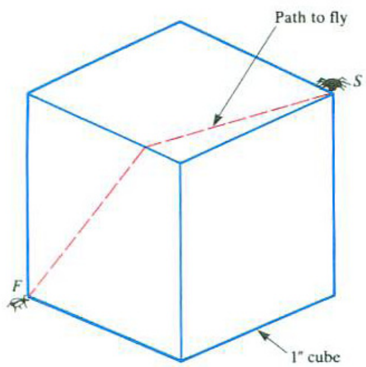
(c) 60 feet long?

43. Let A and B be constants. Find the maximum and mimimum values of $A\cos t + B\sin t$.

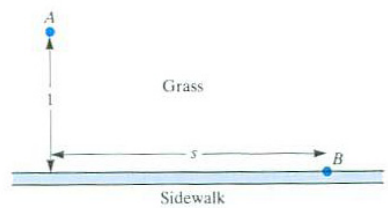
44. A spider at corner S of a cube of side 1 inch wishes to capture a fly at the opposite corner F. (See Figure 5.1.20(a).) The spider, who must walk on the surface of the solid cube, wishes to find the shortest path.

(a) Find a shortest path without the aid of calculus.

(b) Find a shortest path with calculus.



(a)



(b)

Figure 5.1.20

45. A ladder of length b leans against a wall of height a, $a < b$. What is the maximal horizontal distance that the ladder can extend beyond the wall if its base rests on the horizontal ground?

46. A woman can walk 3 miles per hour on grass and 5 miles per hour on sidewalk. She wishes to walk from point A to point B, shown in Figure 5.1.20(b), in the least time. What route should she follow if s is (a) $\frac{1}{2}$? (b) $\frac{3}{4}$? (c) 1?

47. The potential energy in a diatomic molecule is given by the formula

$$U(r) = u_0 \left(\left(\frac{r_0}{r} \right)^{12} - 2 \left(\frac{r_0}{r} \right)^{6} \right),$$

where U_0 and r_0 are constants and r is the distance between the atoms. For which value of r is $U(r)$ a minimum?

48. What are the dimensions of the right circular cylinder of largest volume that can be inscribed in a sphere of radius a?

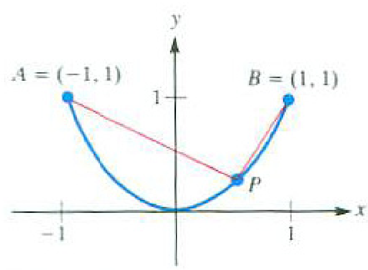
49. The stiffness of a rectangular beam is proportional to the product of the width and the cube of the height of its cross section. What shape beam should be cut from a log in the form of a right circular cylinder of radius r in order to maximize its stiffness.

50. A rectangular box-shaped house is to have a square floor. Three times as much heat per square foot enters through the roof as through the walls. What shape should the house be if it is to enclose a volume of 12,000 cubic feet and minimize heat entry. (Assume no heat enters through the floor.)

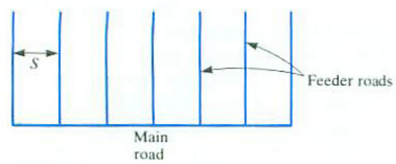
51. (See Figure 5.1.21(a).) Find the coordinates of the points $P = (x, y)$, with $y \leq 1$, on the parabola $y = x^2$, that

(a) minimize $\overline{PA}^2 + \overline{PB}^2$,

(b) maximize $\overline{PA}^2 + \overline{PB}^2$.



(a)



(b)

Figure 5.1.21

52. The speed of traffic through the Lincoln Tunnel in New York City depends on the amount of traffic. Let S be the speed in miles per hour and let D be the amount of traffic measured in vehicles per mile. The relation between S and D was seen to be approximated closely, for $D \leq 100$, by the formula

$$S = 42 - \frac{D}{3}.$$

(a) Express in terms of S and D the total number of vehicles that enter the tunnel in an hour.

(b) What value of D will maximize the flow in (a)?

53. When a tract of timber is to be logged, a main logging road is built from which small roads branch off as feeders. The question of how many feeders to build arises in practice. If too many are built, the cost of construction would be prohibitive. If too few are built, the time spent moving the logs to the roads would be prohibitive. The formula for total cost,

$$y = \frac{CS}{4} + \frac{R}{VS},$$

is used in a logger's manual to find how many feeder roads are to be built. R, C, and V are known constants: R is the cost of road at "unit spacing"; C is the cost of moving a log a unit distance; V is the value of timber per acre. S denotes the distance between the regularly spaced feeder roads. (See Figure 5.1.21(b).) Thus the cost y is a function of S, and the object is to find that value of S that minimizes y. The manual says, "To find the desired S set the two summands equal to each other and solve

$$\frac{CS}{4} = \frac{r}{VS}."$$

Show that the method if valid.

54. A delivery service is deciding how many warehouses to set up in a large city. The warehouses will serve similarly shaped regions of equal area A and, let us assume, an equal number of people.

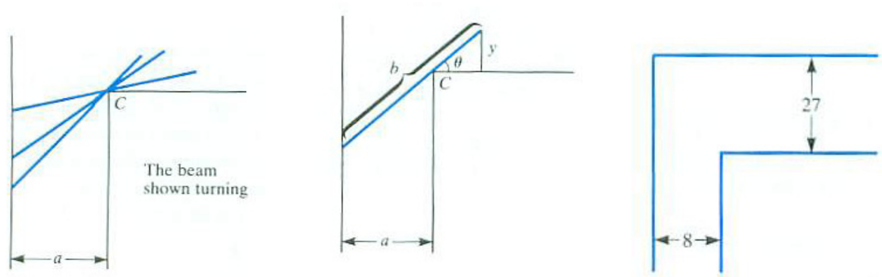
(a) Why would transportation costs per item presumably be proportional to $\sqrt{A}$?

(b) Assuming that the warehouse cost per item is inversely proportional to A, show that C, the cost of transportation and storage per item, is of the form $t\sqrt{A} + w/A$, where t and w are appropraite constants.

(c) Show that C is a minimum when $A = (2w/t)^{2/3}$.

Exercises 55 and 56 are related.

55. A pipe of length b is carried down a long corridor of width $a < b$ and then around corner C. (See Figure 5.1.22.) During the turn y starts out at 0, reaches a maximum, and then returns to 0. (Try this with a short stick.) Find that maximum in terms of a and b. *Suggestion:* Express y in terms of a, b, and θ; θ is a variable, while a and b are constants.

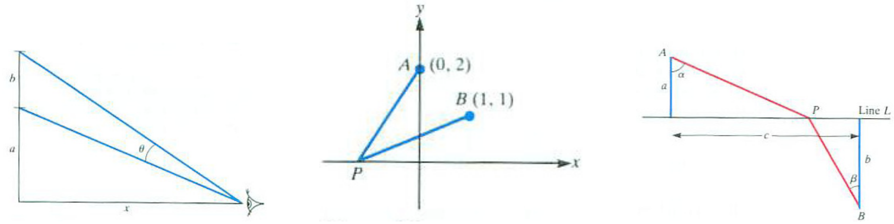


(a) (b) (c)

Figure 5.1.22

56. Figure 5.1.22(c) shows two corridors meeting at right angle. One has width 8; the other, width 27. Find the length of the longest pipe that can be carried horizontally from one hall, around the corner and into the other hall. *Suggestion:* Do Exercise 55 first.

57. The base of a painting on a wall is a feet above the eye of an observer, as shown in Figure 5.1.23(a). The vertical side of the painting is b feet long. How far from the wall should the ovserver stand to maximize the angle that the painting subtends? *Hint:* It is more convenient to maximize $\tan\theta$ than θ itself. (Recall that $\tan(A-B) = \frac{\tan A - \tan B}{1+\tan A \tan B}$.)



(a) (b) (c)

Figure 5.1.23

58. Find the point P on the x-axis such that the angle APB in Figure 5.1.23(b) is maximal. (See the hint in Exercise 57.)

59. *(Economics)* Let p denote the price of some commodity and y the number sold at that price. To be concrete, assume that $y = 250 - p$ for $0 \leq p \leq 250$. Assume that it costs the producer $100 + 10y$ dollars to manufacture y units. What price p should the producer choose in order to maximize total profit, that is, "revenue minus cost"?

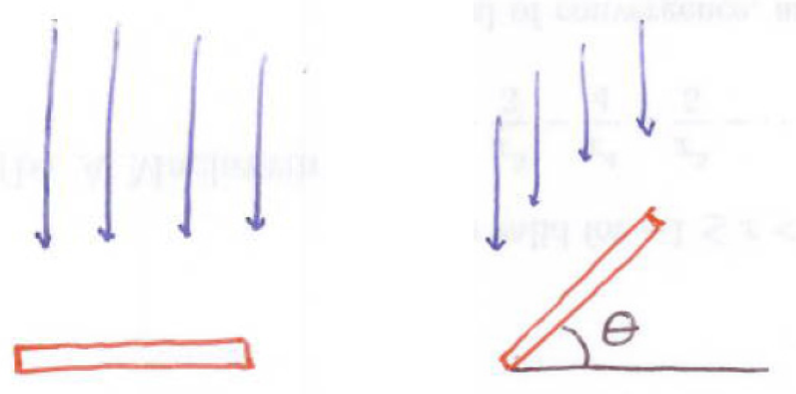
60. *(Leibniz on light)* A ray of light travels from point A to point B in Figure 5.1.23(c) in minimal time. The point A is in one medium, such as air or a vacuum. The point B is in another medium, such as water or glass. In the first medium, light travels at velocity v_1 and in the second at velocity v_2. The media are separated by line L. Show that for the path APB of minimal time,

$$\frac{\sin\alpha}{v_1} = \frac{\sin(\beta)}{v_2}.$$

Leibniz solved this problem with calculus in a paper published in 1684. (The result is called **Snell's law of refraction**.)

Leibniz then wrote, "other very learned men have sought in many devious ways what someone versed in this calculus can accomplish in these lines as by magic." (See C. H. Edwards Jr., *The Historical Development of the Calculus*, p. 259, Springer-Verlag, New York, 1979.)

Exercises 61 to 64 concern the intensity of light.



(a) (b)

Figure 5.1.24

61. Why is it reasonable to assume that the intensity of light from a lamp is inversely proportional to the square of the distance from the lamp? (Imagine the light spreading out in all directions.)

62. A solar panel perpendicular to the sun's rays catches more light than when it is tilted at any other angle, as shown in Figure 5.1.24(a). Let θ be the angle the panel is tilted, as in Figure 5.1.24(b). Show that it then receives $\cos(\theta)$ times the light the panel would receive when perpendicular to the sun's rays.

63. In view of the preceding introduction and exercises, the intensity of light on a small (flat) surface is inversely proportional to the square of the distance from the source and proportional to the angle between the surface and a surface perpendicular to the source.

(a) A person wants to put a light at a horizontal distance of ten feet from his address, which is on a wall (a vertical surface). At what height should the lamp be placed to maximize the intensity of light at the address? (No calculus is needed for this.)

(b) Now the person paints the address on the horizontal surface of the curb. Again the lamp will be placed at a horizontal distance of ten feet from the address. Without doing any calculations sketch what the graph of "intensity of light on the address versus height of lamp" might look like.

(c) Find the height the lamp should have to maximize the light on the address. (Use height as the independent variable.)

64. Solve Exercise 63(c) using an angle as the independent variable.

65. The following calculation occurs in an article concerning the optimum size of new cities: “The net utility to the total client-centered system is

$$U = \frac{RLv}{A} n^{1/2} - nK - \frac{ALc}{v} n^{-1/2}.$$

All symbols except U and n are constant; n is a measure of decentralization. Regarding U as a differentiable function of n, we can determine when $dU/dn = 0$. This occurs when

$$\frac{RLv}{2A} n^{-1/2} - K + \frac{ALc}{2v} n^{-3/2} = 0.$$

This is a cubic equation for $n^{-1/2}$.”

(a) Check that the differentiation is correct.

(b) Of what cubic polynomial is $n^{-1/2}$ a root?

66. Consider the curve $y = x^{-2}$ in the first quadrant. A tangent to this curve, together with axes, determine a triangle.

(a) What is the largest area of such a triangle?

(b) The smallest area?

67. Let f be a differentiable function that is never zero on its domain. Let $g(x) = (f(x))^2$. Show that the functions f and g have the same critical numbers. This is useful for getting rid of square roots.

68. Let f be a differentiable function. Define the function g by $g(x) = \tan(f(x))$. Show that the functions f and g have the same critical numbers.

5.2 Implicit Differentiation

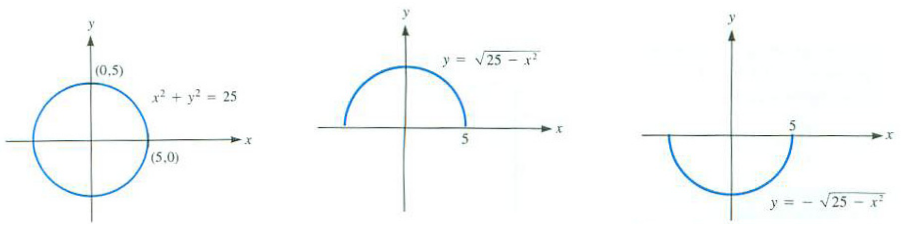
Sometimes a function $y = f(x)$ is given indirectly by an equation that links y and x. This section shows how to differentiate the function without solving for it in terms of x.

A Function Given Implicitly

The equation

$$x^2 + y^2 = 25 \tag{5.2.1}$$

describes a circle of radius 5 and center at the origin, as in Figure 5.2.1(a). This circle is not the graph of a function, since some vertical lines meet the



(a) (b) (c)

Figure 5.2.1

circle in more than one point. However, the top half is the graph of a function and so is the bottom half. To find these functions explicitly, solve (5.2.1) for y:

$$\begin{aligned} y^2 &= 25 - x^2 \\ y &= \pm\sqrt{25 - x^2}. \end{aligned}$$

So either $y = \sqrt{25 - x^2}$ or $y = -\sqrt{25 - x^2}$. The graph of $y = \sqrt{25 - x^2}$ is the top semicircle (see Figure 5.2.1(b)); the graph of $y = -\sqrt{25 - x^2}$ is the bottom semicircle (see Figure 5.2.1(c)). There are two continuous functions that satisfy (5.2.1).

The equation $x^2 + y^2 = 25$ is said to describe the function $y = f(x)$ **implicitly**. The equations

$$y = \sqrt{25 - x^2} \qquad \text{and} \qquad y = -\sqrt{25 - x^2}$$

describe the function $y = f(x)$ **explicitly**.

Differentiating an Implicit Function

It is possible to differentiate a function given implicitly without having to solve for it and express it explicitly. An example will illustrate the method, which is to differentiate both sides of the equation that defines the function implicitly. This procedure is called **implicit differentiation**.

EXAMPLE 1 Let $y = f(x)$ be the continuous function that satisfies the equation

$$x^2 + y^2 = 25$$

such that $y = -4$ when $x = 3$. Find dy/dx when $x = 3$ and $y = -4$.
SOLUTION In this case we can solve for y explicitly, $y = \sqrt{25 - x^2}$ or $y = -\sqrt{25 - x^2}$. Because y equals -4 when x is 3, we are involved with $y = -\sqrt{25 - x^2}$, not $\sqrt{25 - x^2}$. From here we could find the derivative by direct differentiation. However, the square roots do complicate the algebra. Instead we differentiate both sides of the equation

$$x^2 + y^2 = 25$$

with respect to x. This yields

$$\begin{aligned} \frac{d}{dx}(x^2 + y^2) &= \frac{d}{dx}(25), \\ 2x + \frac{d(y^2)}{dx} &= 0. \end{aligned}$$

To differentiate y^2 with respect to x, write $w = y^2$, where y is a function of x.

By the chain rule $$\frac{dw}{dx} = \frac{dw}{dy}\frac{dy}{dx},$$

which gives us $$\frac{d(y^2)}{dx} = 2y\frac{dy}{dx}.$$

Thus $$2x + 2y\frac{dy}{dx} = 0,$$

or $$x + y\frac{dy}{dx} = 0.$$

In particular, when $x = 3$ and $y = -4$, $$3 + (-4)\frac{dy}{dx} = 0,$$

and therefore, $$\frac{dy}{dx} = \frac{3}{4}.$$

Observe that the algebra involves no square roots. ◇

We used implicit differentiation in Section 3.5 when finding the derivatives of $\ln(x)$, $\arcsin(x)$, and $\arctan(x)$. For instance, $y(x) = \ln(x)$ was defined implicitly by the equation $e^{y(x)} = x$. We differentiated both sides of that equation with respect to x to find $y'(x)$.

In the next example implicit differentiation is the only way to find the derivative, for in this case there is no formula expressible in terms of trigonometric and algebraic functions giving y explicitly in terms of x.

EXAMPLE 2 Assume that the equation

$$2xy + \pi \sin(y) = 2\pi$$

defines a function $y = f(x)$. Find dy/dx when $x = 1$ and $y = \pi/2$.

Verify that the point $(1, \pi/2)$ is on the graph of $y = f(x)$ by checking that the equation is satisfied when $x = 1$ and $y = \pi/2$.

SOLUTION Implicit differentiation yields

$$\begin{aligned} \frac{d}{dx}(2xy + \pi \sin y) &= \frac{d(2\pi)}{dx}, \\ \left(2\frac{dx}{dx}y + 2x\frac{dy}{dx}\right) + \pi(\cos y)\frac{dy}{dx} &= 0, \end{aligned}$$

by the formula for the derivative of a product and the chain rule. Hence

$$2y + 2x\frac{dy}{dx} + \pi(\cos y)\frac{dy}{dx} = 0.$$

Solving for the derivative, dy/dx, we get

$$\frac{dy}{dx} = \frac{-2y}{2x + \pi \cos y}.$$

In particular, when $x = 1$ and $y = \pi/2$,

$$\frac{dy}{dx} = -\frac{2 \cdot \frac{\pi}{2}}{2 \cdot 1 + \pi \cos \frac{\pi}{2}} = -\frac{\pi}{2 + \pi \cdot 0} = -\frac{\pi}{2}.$$

◇

Implicit Differentiation and Extrema

Example 3 of Section 5.1 answered the question, "Of all the tin cans that enclose a volume of 100π cubic inches, which requires the least metal?" The radius of the most economical can is $\sqrt[3]{50}$. From this and the fact that its volume is 100π cubic inches, its height was found to be $2\sqrt[3]{50}$, exactly twice the radius. In the next example implicit differentiation is used to answer the same question. Not only will the algebra be simpler but it will provide the shape – the proportion between height and radius – easily.

EXAMPLE 3 Of all the tin cans that enclose a volume of 100π cubic inches, which requires the least metal?

SOLUTION The height h and radius r of any can of volume 100π cubic inches are related by the equation

$$\pi r^2 h = 100\pi. \tag{5.2.2}$$

The surface area S of the can is

$$S = 2\pi r^2 + 2\pi r h \tag{5.2.3}$$

Consider h, and hence S, as functions of r. It is *not* necessary to find h and S explicitly in terms of r. Differentiation of (5.2.2) and (5.2.3) with respect to r yields

$$\pi r^2 \frac{dh}{dr} + 2\pi r h = \frac{d(100\pi)}{dr} = 0 \tag{5.2.4}$$

and

$$\frac{dS}{dr} = 4\pi r + 2\pi r \frac{dh}{dr} + 2\pi h. \tag{5.2.5}$$

When S is a minimum, $dS/dr = 0$, so we have

$$0 = 4\pi r + 2\pi r \frac{dh}{dr} + 2\pi h. \tag{5.2.6}$$

Equations (5.2.4) and (5.2.6) yield, with a little algebra, a relation between h and r, as follows:

Factoring πr out of (5.2.4) and 2π out of (5.2.6) shows that

$$r\frac{dh}{dr} + 2h = 0 \qquad \text{and} \qquad 2r + r\frac{dh}{dr} + h = 0. \tag{5.2.7}$$

Elimination of dh/dr from (5.2.7) yields

$$2r + r\left(\frac{-2h}{r}\right) + h = 0,$$

which simplifies to

$$2r = h. \tag{5.2.8}$$

We have obtained the shape before the specific dimensions. Equation (5.2.8) asserts that the height of the most economical can is the same as its diameter. Moreover, this is the ideal shape, no matter what the prescribed volume happens to be.

The specific dimensions of the most economical can are found by eliminating h from equations (5.2.2) and (5.2.4). This shows that

$$\pi r^2(2r) = 100\pi \qquad \text{or} \qquad r^3 = 50.$$

Hence

$$r = \sqrt[3]{50} \qquad \text{and} \qquad h = 2r = 2\sqrt[3]{50}$$

◇

The procedure illustrated in Example 3 is quite general. It may be of use when maximizing (or minimizing) a quantity that at first is expressed as a function of two variable which are linked by an equation. The equation that links them is called the **constraint**. In Example 3, the constraint is $\pi r^2 h = 100\pi$.

Using Implicit Differentiation in an Extremum Problem

1. Name the various quantities in the problem by letters, such as x, y, h, r, A, V.
2. Identify the quantity to be maximized (or minimized).
3. Express that quantity in terms of other quantities, such as x and y.
4. Obtain an equation relating x and y.
   (This equation is called a constraint.)
5. Differentiate implicitly both the constraint and the quantity to be maximized (or minimized), interpreting all quantities to be functions of a single variable (which you choose).
6. Set the derivative of the quantity to be maximized (or minimized) equal to 0 and combine with the derivative of the constraint to obtain an equation relating x and y at a maximum (or minimum).
7. Step 6 gives only a relation between x and y at an extremum. If the explicit values of x and y are desired, find them by noting that x and y also satisfy the constraint.

Warning: Sometimes an extremum occurs where a derivative, such as dy/dx, is not defined. Exercise 38 illustrates this possibility.

Implicit Differentiation and the Second Derivative

As the next example shows, implicit differentiation also can find the second derivative.

EXAMPLE 4 In Example 2, $y = y(x)$ was given implicitly by $2xy + \pi\sin(y) = 2\pi$. We found that $y' = -2y/(2x + \pi\cos(y))$. Find y''.

SOLUTION

$$\begin{aligned} y'' = (y')' &= \left(\frac{-2y}{2x+\pi\cos(y)}\right)' \\ &= \frac{(2x+\pi\cos(y))\,(-2y') - (-2y)\,(2-\pi\sin(y)y')}{(2x+\pi\cos(y))^2} \\ &= \frac{(2x+\pi\cos(y))\left(-2\left(\frac{-2y}{2x+\pi\cos(y)}\right)\right) - (-2y)\left(2-\pi\sin(y)\left(\frac{-2y}{2x+\pi\cos(y)}\right)\right)}{(2x+\pi\cos(y))^2} \end{aligned}$$

If need be, this can be simplified. The idea is to differentiate y' and whenever y' appears in the computations replace it by its expression in terms of x and y. ◇

Logarithmic Differentiation

If $\ln(f(x))$ is simpler than $f(x)$, there is a technique for finding $f'(x)$ that saves labor. Example 5 illustrates this method, which depends on implicit differentiation.

EXAMPLE 5 Let $y = \frac{\cos(3x)}{\left(\sqrt[3]{x^2+5}\right)^4}$. Find $\frac{dy}{dx}$.

SOLUTION The solution to this problem by **logarithmic differentiation** begins by simplifying $\ln(y)$ using the properties of logarithms:

$$\begin{aligned} \ln(y) &= \ln(\cos(3x)) - \ln\left(\left(\sqrt[3]{x^2+5}\right)^4\right) && [\ln(A/B) = \ln(A) - \ln(B)] \\ &= \ln(\cos(3x)) - \tfrac{4}{3}\ln(x^2+5) && [\ln(A^B) = B\ln(A)]. \end{aligned}$$

Note that y is now given implicitly.

Next, since $\frac{d}{dx}(\ln(y)) = \frac{1}{y}\frac{dy}{dx}$ by the Chain Rule, we have

$$\frac{1}{y}\frac{dy}{dx} = \frac{d}{dx}\left(\ln(\cos(3x)) - \frac{4}{3}\ln\left(x^2+5\right)\right) = \frac{-3\sin(3x)}{\cos(3x)} - \frac{4}{3}\frac{2x}{x^2+5}.$$

Therefore

$$\frac{dy}{dx} = (y)\left(-3\tan(3x) - \frac{4}{3}\frac{2x}{x^2+5}\right).$$

Finally, replace y by its formula, getting

$$\frac{dy}{dx} = \frac{\cos(3x)}{\left(\sqrt[3]{x^2+5}\right)^4}\left(-3\tan(3x) - \frac{4}{3}\frac{2x}{x^2+5}\right).$$

To appreciate logarithmic differentiation, find the derivative directly, as requested in Exercise 26. ◇

If you want to differentiate $\ln(f(x))$ for some function f, first see if you can simplify the expression by using the properties of a logarithm.

Properties of Logarithms

$$\ln(AB) = \ln(A) + \ln(B) \quad \ln\left(\frac{A}{B}\right) = \ln(A) - \ln(B) \quad \ln\left(A^B\right) = B\ln(A)$$

Summary

If a function is given implicitly, differentiate both sides of the equation it satisfies, and solve for the derivative. The derivative is then expressed in terms of the function and its independent variable. To find the second derivative, differentiate the resulting expression, replacing the derivative that appears in it by its formula.

If a function $y(x)$ is a product or a quotient of powers, it may be easier to find $y'(x)$ by differentiating $\ln(y(x))$ implicitly.

EXERCISES for Section 5.2

In Exercises 1 to 4 find dy/dx at the indicated values of x and y in two ways: explicitly (solving for y first) and implicitly.

1. $xy = 4$ at $(1, 4)$

2. $x^2 - y^2 = 3$ at $(2, 1)$

3. $x^2y + xy^2 = 12$ at $(3, 1)$

4. $x^2 + y^2 = 100$ at $(6, -8)$

In Exercises 5 to 8 find dy/dx at the given points by implicit differentiation.

5. $\frac{2xy}{\pi} + \sin y = 2$ at $(1, \pi/2)$

6. $2y^3 + 4xy + x^2 = 7$ at $(1, 1)$

7. $x^5 + y^3x + yx^2 + y^5 = 4$ at $(1, 1)$

8. $x + \tan(xy) = 2$ at $(1, \pi/4)$

9. Solve Example 3 by implicit differentiation, but differentiate (5.2.2) and (5.2.3) with respect to h instead of r.

10. What is the shape of the cylindrical can of largest volume that can be constructed with a given surface area? Do not find the radius and height of the largest can; find the ratio between them. *Suggestion:* Call the surface area S and keep in mind that it is constant.

11. Using implicit differentiation, find $D(\arctan x)$. *Hint:* Start with $x = \tan(y)$.

12. Using implicit differentiation, find $D(\arcsin x)$. *Hint:* Start with $x = \sin(y)$.

In Exercises 13 to 16 find dy/dx at a general point (x, y) on the given curve.

13. $xy^3 + \tan(x + y) = 1$

14. $\sec(x + 2y) + \cos(x - 2y) + y = 2$

15. $-7x^2 + 48xy + 7y^2 = 25$

16. $\sin^3(xy) + \cos(x + y) + x = 1$

In Exercises 17 to 20 implicit differentiation is used to find a second derivative.

17. Assume that $y(x)$ is a differentiable function of x and that $x^3y + y^4 = 2$. Assume that $y(1) = 1$. Find $y''(1)$, following these steps.

(a) Show that $x^3y' + 3x^2y + 4y^3y' = 0$.

(b) Use (a) to find $y'(1)$.

(c) Differentiate the equation in (a) and show that $x^3y'' + 6x^2y' + 6xy + 4y^3y'' + 12y^2(y')^2 = 0$.

(d) Use the equation in (c) to find $y''(1)$. [*Hint:* $y(1)$ and $y'(1)$ are known.]

18. Find $y''(1)$ if $y(1) = 2$ and $x^5 + xy + y^5 = 35$.

19. Find $y'(1)$ and $y''(1)$ if $y(1) = 0$ and $\sin y = x - x^3$.

20. Find $y''(2)$ if $y(2) = 1$ and $x^3 + x^2y - xy^3 = 10$.

21. Use implicit differentiation to find the highest and lowest points on the ellipse $x^2+xy+y^2 = 12$. (What do you know about dy/dx at the highest and lowest points on the graph of a function?)

In Exercises 22 to 25, given information about y', find y''.

22. $y' = (x + y)\sin(x)$

23. $y' = \sin(xy)$

24. $y' = x(y')^3 = (x + 3)y'$

25. $y' \sin(y) + e^{xy} = 0$

26. Differentiate the function in Example 5 directly, without taking logarithms first.

27. Does the tangent line to the curve $x^3 + xy^2 + x^3y^5 = 3$ at the point $(1, 1)$ pass through the point $(-2, 3)$? (Explain.)

In Exercises 28 to 31, find y' two ways: (1) by using the given explicit formula for y and (2) by simplifying $\ln(y)$ and using implicit differentiation.

28. $y = \sqrt{1 + 3x}\,\sqrt[3]{1 + 2x}$

29. $y = (\cos(3x))^{5/2} (\sin(2x))^{1/3}$

30. $y = \frac{\left(1+e^{3x}\right)^4}{(1+e^{2x})^3}$

31. $y = \frac{(\tan(3x))^4\left(x+x^3\right)^5}{\sqrt{x}}$

Exercises 32 and 33 obtain by implicit differentiation the formulas for differentiating $x^{1/n}$ and $x^{m/n}$ with the assumption that they are differentiable functions. Here m and n are integers.

32. Let n be a positive integer. Assume that $y = x^{1/n}$ is a differentiable function of x. From the equation $y^n = x$ deduce by implicit differentiation that $y' = (1/n)x^{1/n-1}$.

33. Let m be a nonzero integer and n a positive integer. Assume that $y = x^{m/n}$ is a differentiable function of x. From the equation $y^n = x^m$ deduce by implicit differentiation that $y' = (m/n)x^{m/n-1}$.

34. Find $D(x^k)$, $x > 0$, by logarithmic differentiation of $y = x^k$.

35. Let $y = x^x$.

(a) Find y' by logarithmic differentiation. That is, first take the logarithm of both sides.

(b) Find y' by first writing the base as $e^{\ln(x)}$. That is, write $y = x^x = \left(e^{\ln(x)}\right)^x = e^{x\ln(x)}$.

36. If $x^3 + y^3 = 1$, find y' and y'' in terms of x and y.

37. Find the first and second derivatives of $y = \sec(x^2)\frac{\sin(x^2)}{x}$.

38.

(a) What difficulty arises when you use implicit differentiation to maximize x^2+y^2 subject to $x^2 + 4y^2 = 16$?

(b) Show that a maximum occurs when dy/dx is not defined. What is the maximum of $x^2 + y^2$ subject to $x^2 + 4y^2 = 16$?

(c) The problem can be viewed geometrically as "Maximize the square of the distance from the origin for points on the ellipse $x^2 + 4y^2 = 16$." Sketch the ellipse and interpret (b) in terms of it.

5.3 Related Rates

The rate at which one quantity changes affects the rate at which another quantity connected to it changes. Implicit differentiation is a convenient tool for finding the relation between the two rates, as the next few examples will illustrate.

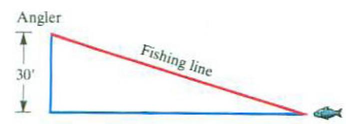


Figure 5.3.1

EXAMPLE 1 An angler has a fish at the end of his line, which is reeled in at 2 feet per second from a bridge 30 feet above the water. At what speed is the fish moving through the water when the amount of line out is 50 feet? 31 feet? Assume the fish is at the surface of the water. (See Figure 5.3.1.)

SOLUTION Our first impression might be that since the line is reeled in at a constant speed, the fish at the end of the line moves through the water at a constant speed. As we will see, this is not the case.

Let s be the length of the line and x the horizontal distance of the fish from the bridge. (See Figure 5.3.2.)

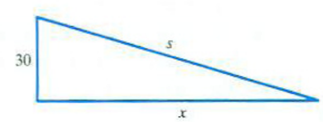


Figure 5.3.2

Since the line is reeled in at the rate of 2 feet per second, s is shrinking, and

$$\frac{ds}{dt} = -2.$$

The rate at which the fish moves through the water is given by the derivative, dx/dt. The problem is to find dx/dt when $s = 50$ and also when $s = 31$.

We need an equation that relates s and x at *any* time, not just when $x = 50$ or $x = 31$. If we consider only $x = 50$ or $x = 31$, there would be no motion, and no chance to use derivatives.

The quantities x and s are related by the Pythagorean Theorem:

This equation is the heart of the example.

$$x^2 + 30^2 = s^2.$$

Both x and s are functions of time t. Thus both sides of the equation may be differentiated with respect to t, yielding

$$\begin{aligned} \frac{d(x^2)}{dt} + \frac{d(30^2)}{dt} &= \frac{d(s^2)}{dt} \\ \text{or} \qquad 2x\frac{dx}{dt} + 0 &= 2s\frac{ds}{dt}. \\ \text{Hence} \qquad x\frac{dx}{dt} &= s\frac{ds}{dt}. \end{aligned}$$

This last equation provides the tool for answering the questions.

Since $ds/dt = -2$,

$$\begin{aligned} & & x\frac{dx}{dt} &= (s)(-2). \\ &\text{Hence} & \frac{dx}{dt} &= \frac{-2s}{x}. \\ &\text{When } s = 50, & x^2 + 30^2 &= 50^2, \end{aligned}$$

so $x = 40$. Thus when 50 feet of line is out, the speed is

$$\left|\frac{dx}{dt}\right| = \frac{2s}{x} = \frac{2 \cdot 50}{40} = 2.5 \text{ feet per second.}$$

$$\begin{aligned} &\text{When } s = 31, & x^2 + 30^2 &= 31^2. \\ &\text{Hence} & x = \sqrt{31^2 - 30^2} &= \sqrt{961 - 900} = \sqrt{61}. \end{aligned}$$

Thus when 31 feet of line is out, the fish is moving at the speed of

$$\frac{dx}{dt} = \frac{2s}{x} = \frac{2 \cdot 31}{\sqrt{61}} = \frac{62}{\sqrt{61}} \approx 7.9 \text{ feet per second.}$$

Let us look at the situation from the fish's point of view. When it is x feet from the point in the water directly below the bridge, its speed is $2s/x$ feet per second. Since s is larger than x, its speed is always greater than 2 feet per second. When x is very large, s/x is near 1 so the fish is moving through the water only a little faster than the line is reeled in. However, when the fish is almost at the point under the bridge, x is very small; then $2s/x$ is huge, and the fish finds itself moving at huge speeds, but according to Einstein, not faster than the speed of light. ◇

In Example 1 it would be a tactical mistake to indicate in Figure 5.3.2 that the hypotenuse of the triangle is 50 feet long, for if one leg is 30 feet and the hypotenuse is 50 feet, the triangle is completely determined; there is nothing left free to vary with time.

In general, label all the lengths or quantities that can change with letters x, y, s, and so on, even if not all are needed in the solution. Only after you finish differentiating do you determine what the rates are at a specified value of the variable.

The General Procedure

The method used in Example 1 applies to many related rate problems. This is the general procedure, broken into steps:

Procedure for Finding a Related Rate

1. Find an equation that relates the varying quantities.
   (If the quantities are geometric, draw a picture and label the varying quantities with letters.)

2. Differentiate both sides of the equation with respect to time, obtaining an equation that relates the various rates of change.

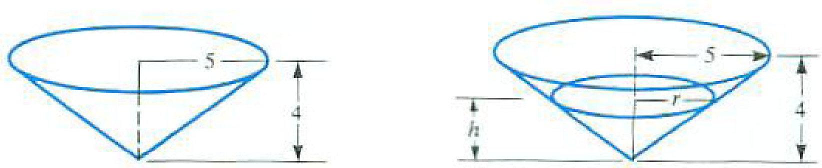
3. Solve the equation obtained in Step 2 for the unknown rate.
   (Only at this step do you substitute constants for variable.)

WARNING Differentiate, then substitute the specific numbers for the variables. If you reversed the order, you would just be differentiating constants.

Finding an Acceleration

The method described in Example 1 for determining unknown rates from known ones extends to finding an unknown acceleration. Just differentiate another time. Example 2 illustrates the procedure.

EXAMPLE 2 Water flows into a conical tank at the constant rate of 3 cubic meters per second. The radius of the cone is 5 meters and its height is 4 meters. Let $h(t)$ represent the height of the water above the bottom of the cone at time t. Find dh/dt (the rate at which the water is rising in the tank) and d^2h/dt^2 (the rate at which that rate changes) when the tank is filled to a height of 2 meters. (See Figure 5.3.3.)



(a) (b)

Figure 5.3.3

SOLUTION Let $V(t)$ be the volume of water in the tank at time t. The fact that water flows into the tank at 3 cubic meters per second is expressed as

$$\frac{dV}{dt} = 3,$$

and, since this rate is constant,

$$\frac{d^2V}{dt^2} = 0.$$

To find dh/dt and d^2h/dt^2, first obtain an equation relating V and h.

When the tank is filled to the height h, the water forms a cone of height h and radius r. (See Figure 5.3.3(b).) By similar triangles,

$$\frac{r}{h} = \frac{5}{4} \qquad \text{or} \qquad r = \frac{5h}{4}.$$

Thus

$$V = \frac{1}{3}\pi r^2 h = \frac{1}{3}\pi\left(\frac{5}{4}h\right)^2 h = \frac{25}{48}\pi h^3.$$

So the equation relating V and h is

$$V = \frac{25\pi}{48}h^3. \tag{5.3.1}$$

From here on, just differentiate as often as needed.

Differentiating both sides of (5.3.1) once (using the chain rule) yields

$$\frac{dV}{dt} = \frac{25\pi}{48}\frac{d(h^3)}{dh}\frac{dh}{dt}$$

or

$$\frac{dV}{dt} = \frac{25\pi}{16}h^2\frac{dh}{dt}.$$

Since $dV/dt = 3$ all the time,

$$3 = \frac{25\pi h^2}{16}\frac{dh}{dt},$$

from which it follows that

$$\frac{dh}{dt} = \frac{48}{25\pi h^2} \text{ meters per second.} \tag{5.3.2}$$

Even though the water enters the tank at a constant rate, it does not rise at a constant rate.

As (5.3.2) shows, the larger h is, the slower the water rises. (Why is this to be expected?)

To find dh/dt when $h = 2$ meters, substitute 2 for h in (5.3.2), obtaining

$$\frac{dh}{dt} = \frac{48}{25\pi 2^2} = \frac{12}{25\pi} \approx 0.15279 \text{ meters per second.}$$

Now we turn to the acceleration, d^2h/dt^2. *We do not* differentiate the equation $dh/dt = 12/(25\pi)$ since this equation holds only when $h = 2$. We must go back to (5.3.2), which holds at any time.

Differentiating (5.3.2) with respect to t yields

$$\frac{d^2h}{dt^2} = \frac{48}{25\pi}\frac{d}{dt}\left(\frac{1}{h^2}\right) = \frac{48}{25\pi}\frac{-2}{h^3}\frac{dh}{dt} = \frac{-96}{25\pi h^3}\frac{dh}{dt}. \tag{5.3.3}$$

The last equation expresses the acceleration in terms of h and dh/dt. Substituting (5.3.2) into (5.3.3) gives

$$\frac{d^2h}{dt^2} = \frac{-96}{25\pi h^3}\frac{48}{25\pi h^2}$$

or

$$\frac{d^2h}{dt^2} = \frac{-(96)(48)}{(25\pi)^2 h^5} \text{ meters per second per second.} \tag{5.3.4}$$

Equation (5.3.4) tells us that, since d^2h/dt^2 is negative, the rate at which the water rises in the tank is decreasing.

The problem also asked for the value of d^2h/dt^2 when $h = 2$. To find it, replace h by 2 in (5.3.4), obtaining

$$\frac{d^2h}{dt^2} = \frac{-(96)(48)}{(25\pi)^2 2^5}$$

or

$$\frac{d^2h}{dt^2} = \frac{-144}{625\pi^2} \approx -0.02334 \text{ meters per second per second.}$$

◇

Summary

If two variables are linked by an equation, the rates at which they change are also related. To find an equation involving those two rates, differentiate both sides of the original equation implicitly.

EXERCISES for Section 5.3

1. How fast is the fish in Example 1 moving through the water when it is 1 foot horizontally from the bridge?

2. The angler in Example 1 decides to let the line out as the fish swims away. The fish swims away at a constant speed of 5 feet per second relative to the water. How fast is the angler playing out his line when the horizontal distance from the bridge to the fish is

(a) 1 foot?

(b) 100 feet?

3. A 10-foot ladder is leaning against a wall. A person pulls the base of the ladder away from the wall at the rate of 1 foot per second.

(a) Draw a neat picture of the situation and label the varying lengths by letters and the fixed lengths by numbers.

(b) Obtain an equation involving the variables in (a).

(c) Differentiate it with respect to time.

(d) How fast is the top going down the wall when the base of the ladder is 6 feet from the wall? 8 feet from the wall? 9 feet from the wall?

4. A kite is flying at a height of 300 feet in a horizontal wind.

(a) Draw a neat picture of the situation of label the varying lengths by letters and the fixed lengths by numbers.

(b) When 500 feet of string is out, the kite is pulling the string out at a rate of 20 feet per second. What is the kite's velocity? (Assume the string remains straight.)

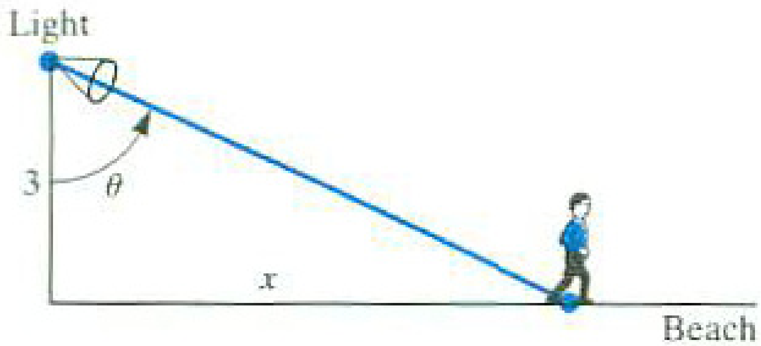


Figure 5.3.4

5. A beachcomber walks 2 miles per hour along the shore as the beam from a rotating light 3 miles offshore follows him. (See Figure 5.3.4.)

(a) Intuitively, what do you think happens to the rate at which the light rotates as the beachcomber walks further and further along the shore away from the lighthouse?

(b) Let x describe the distance of the beachcomber from the point on the shore nearest the light and θ the angle of the light, obtain an equation relating θ and x.

(c) With the aid of (b), show that $d\theta/dt = 6/(9 + x^2)$ (radians per hour).

(d) Does the formula in (c) agree with your guess in (a)?

6. A man 6 feet tall walks at the rate of 5 feet per second away from a street lamp that is 20 feet high. At what rate is his shadow lengthening when he is

(a) 10 feet from the lamp?

(b) 100 feet from the lamp?

7. A large spherical balloon is being inflated at the rate of 100 cubic feet per minute. At what rate is the radius increasing when the radius is

(a) 10 feet?

(b) 20 feet?

(The volume of a sphere of radius r is $V = 4\pi r^3/3$.)

8. A shrinking spherical balloon loses air at the rate of 1 cubic inch per second. At what rate is its radius changing when the radius is

(a) 2 inches

(b) 1 inch?

9. Bulldozers are moving earth at the rate of $1,000$ cubic yards per hour onto a conically shaped hill whose height of the hill increasing when the hill is

(a) 20 yards high?

(b) 100 yards high?

(The volume of a cone of radius r and height h is $V = \pi r^2 h/3$.)

10. The lengths of the two legs of a right triangle depend on time. One leg, whose length is x, increases at the rate of 5 feet per second, while the other, of length y, decreases at the rate of 6 feet per second. At what rate is the hypotenuse changing when $x = 3$ feet and $y = 4$ feet? Is the hypotenuse increasing or decreasing then?

11. Two sides of a triangle and their included angle are changing with respect to time. The angle increases at the rate of 1 radian per second, one side increases at the rate of 3 feet per second, and the other side decrease at the rate of 2 feet per second. Find the rate at which the area is changing when the angle is $\pi/4$, the first side is 4 feet long, and the second side is 5 long. Is the area decreasing or increasing then?

12. The length of a rectangle is increasing at the rate of 7 feet per second, and the width is decreasing at the rate of 3 feet per second. When the length is 12 feet and the width is 5 feet, find the rate of change of

(a) the area,

(b) the perimeter

(c) the length of the diagonal.

Exercises 13 to 17 concern acceleration.

13. What is the acceleration of the fish described in Example 1 when the length of line is

(a) 300 feet?

(b) 31 feet?

The notation $\dot{x}$ for dx/dt, $\dot{\theta}$ for $d\theta/dt$, $\ddot{x}$ for d^2x/dt^2, and $\ddot{\theta}$ for $d^2\theta/dt^2$ was introduced by Newton and is still common in physics.

14. A woman on the ground is watching a jet through a telescope as it approaches at a speed of 10 miles per minute at an altitude of 7 miles. At what rate (in radians per minute) is the angle of the telescope changing when the horizontal distance of the jet from the woman is 24 miles? When the jet is directly above the woman?

15. Find $\ddot{\theta}$ in Example 14 when the horizontal distance from the jet is

(a) 7 miles,

(b) 1 mile.

16. A particle moves on the parabola $y = x^2$ in such a way that $\dot{x} = 3$ throughout the journey. Find the formulas for (a) $\dot{y}$ and (b) $\ddot{y}$.

17. Call one acute angle of a right triangle θ. The adjacent leg has length x and the opposite leg has length y.

(a) Obtain an equation relating x, y and θ.

(b) Obtain an equation involving $\dot{x}$, $\dot{y}$, and $\dot{\theta}$ (and other variables).

(c) Obtain an equation involving $\ddot{x}$, $\ddot{y}$, and $\ddot{\theta}$ (and other variables).

18. A two-piece extension ladder leaning against a wall is collapsing at the rate of 2 feet per second and the base of the ladder is moving away from the wall at the rate of 3 feet per second. How fast is the top of the ladder moving down the wall when it is 8 feet from the ground and the foot is 6 feet from the wall? (See Figure 5.3.5.)

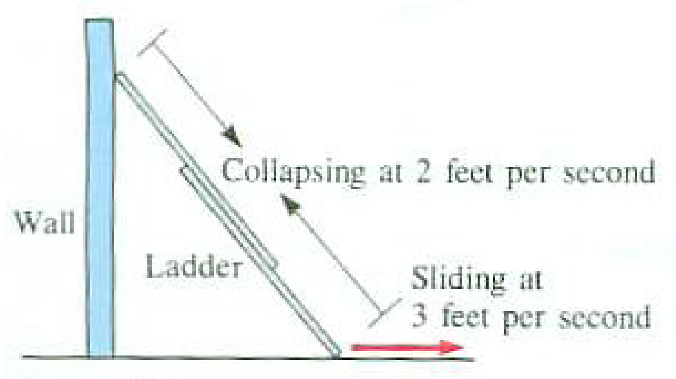


Figure 5.3.5

19. At an altitude of x kilometers, the atmospheric pressure decreases at a rate of $128(0.88)^x$ millibars per kilometer. A rocket is rising at the rate of 5 kilometers per second vertically. At what rate is the atmospheric pressure changing (in millibars per second) when the altitude of the rocket is (a) 1 kilometer? (b) 50 kilometers?

20. A woman is walking on a bridge that is 20 feet above a river as a boat passes directly under the center of the bridge (at a right angle to the bridge) at 10 feet per second. At that moment the woman is 50 feet from the center and approaching it at the rate of 5 feet per second.

(a) At what rate is the distance between the boat and woman changing at that moment?

(b) Is the rate at which they are approaching or separating increasing or is it decreasing?

21. A spherical raindrop evaporates at a rate proportional to its surface area. Show that the radius shrinks at a constant rate.

22. A couple is on a Ferris wheel when the sun is directly overhead. The diameter of the wheel is 50 feet, and its speed is 0.01 revolution per second.

(a) What is the speed of their shadows on the ground when they are at a two-o'clock position?

(b) A one-o'clock position?

(c) Show that the shadow is moving its fastest when they are at the top or bottom, and its slowest when they are at the three-o'clock or nine-o'clock position.

23. Water is flowing into a hemispherical bowl of radius 5 feet at the constant rate of 1 cubic foot per minute.

(a) At what rate is the top surface of the water rising when it height above the bottom of the bowl is 3 feet? 4 feet? 5 feet?

(b) If $h(t)$ is the depth in feet at time t, find $\ddot{h}$ when $h = 3, 4$, and 5.

24. A detective is aiming a flashlight at a door. The axis of the conical beam is perpendicular to the door. Let $x(t)$ be the distance between the detective and the door at time t. Let $A(t)$ be the area of the illuminated disk on the door at time t. The detective is walking towards the door.

(a) Is dx/dt positive or negative?

(b) Is dA/dt positive or negative?

(c) Is there a constant k such that $\frac{dA}{dt} = k\frac{dx}{dt}$?

Explain each answer.

25. The rate at which the variable $B(t)$ changes is proportional to the square of the rate at which the variable $C(t)$ changes. Does it follow that the acceleration of $A(t)$, $A''(t)$, is proportional to the square of $B''(t)$? Explain.

5.4 Higher Derivatives and the Growth of a Function

The only higher derivative we've used so far is the second derivative. In the study of motion, if y denotes position then y'' is acceleration. In the study of graphs, the second derivative determines whether the graph is concave up ($y'' > 0$) or down ($y'' < 0$). Later, in Section 9.6, the second derivative will appear in a formula that measures the curviness of a curve.

Now we will see how the higher derivatives (including the second derivative) influence the growth of a function. In the next section this will be applied to estimate the error in approximating a function by a polynomial.

Introduction

Imagine that you are in a car motionless at the origin of the x-axis. Then you put your foot to the gas pedal and accelerate. The greater the acceleration, the faster the speed increases; the greater the speed, the further you travel in a given time. So the acceleration, which is the second derivative of the position function, influences the function itself. This illustrates how a higher derivative of a function influences the growth of a function. In this section we examine this influence in more detail.

The following lemma is the basis for our analysis. In terms of daily life, it says, "The faster runner wins the race."

Lemma 5.4.1. *Let $f(x)$ and $g(x)$ be differentiable functions on an interval I. Let a be a number in I where $f(a) = g(a)$. Assume that $f'(x) \leq g'(x)$ for x in I. Then $f(x) \leq g(x)$ for all x in I to the right of a and $f(x) \geq g(x)$ for all x in I to the left of a.*

Figure 5.4.1 makes this plausible, when the graphs of f and g are straight lines. To the right of $x = a$ the steeper line lies above the other line. To the left of $x = a$ the steeper line lies below the other line.

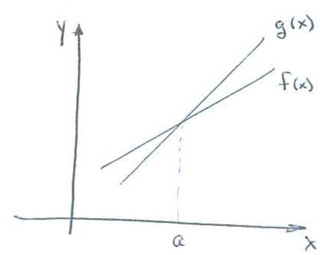


Figure 5.4.1

Proof of Lemma 5.4.1

Consider the case when $x > a$. Let $h(x) = g(x) - f(x)$. Then $h(a) = 0$ and $h'(x) = g'(x) - f'(x) \geq 0$. Thus, h is a non-decreasing function. Since $h(a) = 0$, it follows that $h(x) \geq 0$ for $x \geq a$. That is, $g(x) - f(x) \geq 0$, hence $f(x) \leq g(x)$ for $x > a$. •

Repeated application of Lemma 5.4.1 will enable us to establish a connection between higher derivatives and the function itself.

Higher Derivatives and the Growth of a Function

$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120.$

In the following theorem we name the function $R(x)$ because that will be the notation in the next section when $R(x)$ is the "remainder" function. The notation $n!$ (read: "n factorial") for a positive integer n is shorthand for the product of all integers from 1 through n: $n! = n(n-1)\cdots 3\cdot 2\cdot 1$. The symbol $0!$ is usually defined to be 1.

Theorem 5.4.2 (Growth Theorem)**.** *Assume that at a the function R and its first n derivatives are zero,*

$$R(a) = R'(a) = R''(a) = R^{(3)}(a) = \cdots = R^{(n)}(a) = 0.$$

Assume also that $R(x)$ has continuous derivatives up through the derivative of order $n+1$ in some open interval I containing the numbers a and x. Then, assuming $x > a$, there is a number c_n in the interval $[a, x]$ such that

$$R(x) = R^{(n+1)}(c_n)\frac{(x-a)^{n+1}}{(n+1)!}. \tag{5.4.1}$$

Before giving the straightforward proof, we illustrate the theorem by several examples.

The Growth Theorem with $n = 1$ and $a = 0$ describes the position of an accelerating car. Let $R(x)$ be the position of the car on the y-axis at time x. One has $R(0) = 0$ (at time 0 the car is at position 0), $R'(0) = 0$ (at time 0 the car is not moving) and R'' describes the acceleration. If that acceleration is constant, equal to k, then (5.4.1) gives the car's position at time x as $R(x) = k\frac{x^2}{2!}$. If the acceleration is not constant, it says that $R(x)$ equals the acceleration at some time multiplied by $x^2/2$.

EXAMPLE 1 Show that $|e^x - 1 - x| \leq \frac{e}{2}x^2$ for x in $(-1, 1)$.

SOLUTION Let $R(x) = e^x - 1 - x$. Then $R(0) = e^0 - 1 - 0 = 0$. And, since $R'(x) = e^x - 1$, $R'(0) = e^0 - 1 = 0$ also. $R''(x) = e^x$. By the Growth Theorem, with $a = 0$ and $n = 1$, there is a number c_1 in $(-1, 1)$ such that

$$e^x - 1 - x = e^{c_1}\frac{(x-0)^2}{2!}.$$

We do not know c_1, but, since it is less than 1, $e^{c_1} < e$. Thus

$$|e^x - 1 - x| \leq e\frac{x^2}{2}. \tag{5.4.2}$$

◇

The inequality (5.4.2) in the preceding example provides a way to estimate e^x when x is small. For instance, $|e^{0.1} - 1 - 0.1| \leq \frac{e}{2}(0.1)^2 = e/200$. The estimate 1.1 for $e^{0.1}$ is off by at most $e/200 \approx 0.013591$.

EXAMPLE 2 Let $R(x) = \cos(x) - 1 + \frac{x^2}{2}$. Show that $|R(x)| \leq \frac{|x^3|}{6}$.
SOLUTION As in Example 1 we use the Growth Theorem with $a = 0$, $n = 2$, and $x > 0$.

$$\begin{array}{lrcll} & R(x) & = & \cos(x) - 1 - \frac{x^2}{2}, & \text{so R(0) = 1-1+0 = 0;} \\ & R'(x) & = & -\sin(x) + x, & \text{so } R'(0) = 0 + 0 = 0; \\ & R''(x) & = & -\cos(x) + 1, & \text{so } R''(0) = -1 + 1 = 0; \\ \text{nd} & R^{(3)}(x) & = & \sin(x). & \end{array}$$

By the Growth Theorem, with $a = 0$ and $n = 2$,

$$R(x) = \sin(c_2)\frac{x^3}{3!} \qquad \text{for some number } c_2 \text{ between 0 and } x.$$

Because $|\sin(x)| \leq 1$,

$$|R(x)| \leq \left|(1)\frac{x^3}{6}\right| = \frac{|x|^3}{6}.$$

◇

Example 2 provides a good estimate for values of the cosine function for small angles. For instance, if $x = 0.1$ radians, we have

0.1 radians $= 0.1\frac{180°}{\pi}° \approx 5.7°$

$$\left|\cos(0.1) - 1 + \frac{0.1^2}{2}\right| \leq \frac{0.1^3}{6} = 0.00016667 = 1.6667 \times 10^{-4}.$$

Thus, $1 - \frac{0.1^2}{2} = 1 - 0.005 = 0.995$ is an estimate of $\cos(0.1)$ with an error less than $\frac{1}{6} \times 10^{-3} \approx 0.00016667$. Incidentally, $\cos(0.1) \approx 0.9950041653$ so the error is only 0.0000041653.

> **Remark:** An even better bound on the growth of $R(x)$ in Example 2 is possible. In addition to $R(0) = R'(0) = R''(0) = 0$, notice that $R^{(3)}(0) = \sin(0) = 0$. This means that $|R(x)| \leq \left|M_4\frac{(x-0)^4}{4!}\right|$ where M_4 is the maximum value of $R^{(4)}(t) = \cos(t)$ in the interval $[0, x]$. As in Example 2, $M \leq 1$. Thus,
>
> $$|R(x)| \leq \left|(1)\frac{x^4}{4!}\right| = \frac{x^4}{24}.$$
>
> This means the difference between the exact value of $\cos(0.1)$ and the estimate $1 - \frac{0.1^2}{2} = 0.995$ is no more than $\frac{0.1^4}{24} = 4.16667 \times 10^{-6}$. This shows the estimate in Example 2 is accurate to five decimal places.

In fact, $|\cos(0.1) - 0.995| \approx 4.16528 \times 10^{-6}$.

In any case, $1 - \frac{x^2}{2}$ is a good estimate of $\cos(x)$ for small values of x. The next section describes how to find polynomials that provide good estimates of functions.

A Refinement of the Growth Theorem

When proving the Growth theorem we will establish something stronger:

Theorem 5.4.3 (Refined Growth Theorem)**.** *If $m \leq R^{(n+1)}(t) \leq M$ and all earlier derivatives of R are 0 at a, then*

$$R(x) \text{ is between } m\frac{(x-a)^{n+1}}{(n+1)!} \text{ and } M\frac{(x-a)^{n+1}}{(n+1)!}. \tag{5.4.3}$$

This statement holds even if x is less than a and $(x-a)$ is negative.

EXAMPLE 3 Let $R(x) = e^x - (1 + x + \frac{x^2}{2!} + \frac{x^3}{3!})$. Show that $\frac{1}{1152} \leq R(\frac{1}{2}) \leq \frac{1}{128}$. Use this estimate to obtain approximations, with error bounds, for $\sqrt{e} = e^{1/2}$.
SOLUTION

$$\begin{aligned} R(0) &= e^0 - 1 - 0. \\ R'(x) &= e^x - (1 + x + \frac{x^2}{2!}), && \text{so} && R'(0) = 0. \\ R''(x) &= e^x - (1+x), && \text{so} && R''(0) = 0. \\ R^{(3)}(x) &= e^x - 1, && \text{so} && R^{(3)}(0) = 0. \\ R^{(4)}(x) &= e^x, && \text{and} && R^{(4)}(0) = 1 \neq 0. \end{aligned}$$

But, for x in $I = (-1, 1)$, $\frac{1}{3} \leq e^{-1} \leq e^x \leq e^1 < 3$. Theorem 5.4.3, with $a = 0$, $n = 3$, $m = \frac{1}{3}$, $M = 3$, and $x = \frac{1}{2}$ gives

$$\begin{array}{lrcl} & \frac{1}{3}\frac{(1/2)^4}{4!} & \leq R(1/2) \leq & 3\frac{(1/2)^4}{4!} \\ \text{Then,} & \frac{1}{1152} & \leq \sqrt{e} - \left(1 + \frac{1}{2} + \frac{(1/2)^2}{2!} + \frac{(1/2)^3}{3!}\right) \leq & \frac{1}{128} \\ \text{or} & \frac{79}{48} + \frac{1}{1152} & \leq \sqrt{e} \leq & \frac{79}{48} + \frac{1}{128} \\ \text{so} & 1.64670 & \leq \sqrt{e} \leq & 1.65365. \end{array}$$

As you can check with your calculator, $\sqrt{e} \approx 1.64872$ to five decimal places. ⋄

As Example 3 shows, the Growth Theorem provides not only upper bounds on the error in approximating a function by certain polynomials, but lower bounds on that error as well.

Proof of the Growth Theorem

Proof of the Growth Theorem

We illustrate the proof in the case $n = 2$. For convenience, we take the case $x > a$. The case with $x < a$ is complicated by the fact that $x - a$ is then negative and the sign of $(x - a)^n$ depends on whether n is odd or even.

Assume $R(a) = R'(a) = R''(a) = 0$ and $R^{(3)}(x)$ is continuous in the interval $[a, x]$. We want to show there is a number c_2 in $[a, x]$ such that

$$R(x) = R^{(3)}(c_2)\frac{(x-a)^3}{3!}.$$

Let M be the maximum of $R^{(3)}(t)$ and m be the minimum of $R^{(3)}(t)$ on the closed interval $[a, x]$. Thus

$$m \le R^{(3)}(t) \le M \qquad \text{for all } t \text{ in } [a, x].$$

We will see first what the inequality $R^{(3)}(t) \le M$ implies about $R(x)$.

We rewrite that inequality as

$$\frac{d}{dt}\left(R^{(2)}(t)\right) \le \frac{d}{dt}\left(M(t-a)\right). \tag{5.4.4}$$

Now apply Lemma 5.4.1 with $f(t) = R^{(2)}(t)$ and $g(t) = M(t - a)$. Note that $f(a) = 0$ and $g(a) = M(a - a) = 0$. (That is why we used the antiderivative $M(t - a)$ rather than the expected Mt.) Also $f''(a) = 0 = g''(a)$. By the lemma

$$R^{(2)}(t) \le M(t-a). \tag{5.4.5}$$

Next, rewrite (5.4.5) as

$$\frac{d}{dt}\left(R'(t)\right) \le \frac{d}{dt}\left(M\frac{(t-a)^2}{2}\right).$$

Applying the lemma again shows that

$$R'(t) \le M\frac{(t-a)^2}{2}. \tag{5.4.6}$$

Finally, rewrite (5.4.6) as

$$\frac{d}{dt}\left(R(t)\right) \le \frac{d}{dt}\left(M\frac{(t-a)^3}{3\cdot 2}\right).$$

The lemma asserts that

$$R(t) \le M\frac{(t-a)^3}{3!}. \tag{5.4.7}$$

Similar reasoning, starting with $m \leq R^{(3)}(t)$ shows that

$$m\frac{(t-a)^3}{3!} \leq R(t). \tag{5.4.8}$$

Combining (5.4.7) and (5.4.8) gives two bounds on $R(t)$; in particular on $R(x)$:

$$m\frac{(x-a)^3}{3!} \leq R(x) \leq M\frac{(x-a)^3}{3!}.$$

Because $R^{(3)}$ is continuous on $[a, x]$ it assumes all values between m and M. Thus there is a number c_2 in $[a, x]$ such that

$$R(x) = R^{(3)}(c_2)\frac{(x-a)^3}{3!}.$$

•

Summary

The bound on the size of the derivative of a function limits the growth of the function itself. This observation applied repeatedly shows that if a function $R(x)$ and its first n derivatives are all zero at a, then

$$R(x) = R^{(n+1)}(c_n)\frac{(x-a)^{n+1}}{(n+1)!} \qquad \text{for some } c_n \text{ between } a \text{ and } x.$$

The number c_n depends on n, not just on a, x, and the function $R(x)$.

EXERCISES for Section 5.4

1. If $f'(x) \geq 3$ for all $x \in (-\infty, \infty)$ and $f(0) = 0$, what can be said about $f(2)$? about $f(-2)$?

2. If $f'(x) \geq 2$ for all $x \in (-\infty, \infty)$ and $f(1) = 0$, what can be said about $f(3)$? about $f(-3)$?

3. What can be said about $f(2)$ if $f(1) = 0$, $f'(1) = 0$, and $2.5 \leq f''(x) \leq 2.6$ for all x?

4. What can be said about $f(4)$ if $f(1) = 0$, $f'(1) = 0$, and $2.9 \leq f''(x) \leq 3.1$ for all x?

5. A car starts from rest and travels for 4 hours. Its acceleration is always at least 5 miles per hour per hour, but never exceeds 12 miles per hour per hour. What can you say about the distance traveled during those 4 hours?

6. A car starts from rest and travels for 6 hours. Its acceleration is always at least 4.1 miles per hour per hour, but never exceeds 15.5 miles per hour per hour. What can you say about the distance traveled during those 6 hours?

7. State the Growth Theorem for $x \geq a$ in the case when R has at least five continuous derivatives and $R(a) = R'(a) = R''(a) = R^{(3)}(a) = R^{(4)}(a) = 0$.

8. State the Growth Theorem in words, using as little math notation as possible.

9. If $R(1) = R'(1) = R''(1) = 0$ and $R^{(3)}(x)$ is continuous on an interval that includes 1 and $R^{(3)}(x) \leq 2$, what can be said about $R(4)$?

10. If $R(3) = R'(3) = R''(3) = R^{(3)}(3) = R^{(4)}(3) = 0$ and $R^{(5)}(x) \leq 6$, what can be said about $R(3.5)$?

11. Let $R(x) = \sin(x) - \left(x - \frac{x^3}{6}\right)$. Show that

(a) $R(0) = R'(0) = R''(0) = R^{(3)}(0) = 0$.

(b) $R^{(4)}(x) = \sin(x)$.

(c) $|R(x)| \leq \frac{x^4}{24}$.

(d) Use $x - \frac{x^3}{6}$ to approximate $\sin(x)$ for $x = 1/2$.

(e) Use (c) to estimate the difference between the exact value for $\sin\left(\frac{1}{2}\right)$ and the approximation obtained in (d).

(f) Explain why $|R(x)| \leq \frac{|x|^5}{120}$. How can this be used to obtain a better estimate of the difference between the exact value for $\sin\left(\frac{1}{2}\right)$ and the approximation obtained in (d)?

(g) By how much does the estimate in (d) differ from $\sin\left(\frac{1}{2}\right)$?

Incidentally, an angle of $\frac{1}{2}$ radian is about 29°.

12. Let $R(x) = \cos(x) - \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!}\right)$. Show that

(a) $R(0) = R'(0) = R''(0) = R^{(3)}(0) = R^{(4)}(0) = R^{(5)}(0) = 0$.

(b) $R^{(6)}(x) = -\cos(x)$.

(c) $|R(x)| \leq \frac{x^6}{6!}$.

(d) Use $1 - \frac{x^2}{2!} + \frac{x^4}{4!}$ to estimate $\cos(x)$ for $x = 1$.

(e) By how much does the estimate in (d) differ from $\cos(1)$?

Incidentally, an angle of 1 radian is about 57°.

13. Let $R(x) = (1+x)^5 - (1 + 5x + 10x^2)$. Show that

(a) $R(0) = R'(0) = R''(0) = 0$.

(b) $R^{(3)}(x) = 60(1+x)^2$.

(c) $|R(x)| \leq 80x^3$ on $[-1, 1]$.

(d) Use $1 + 5x + 10x^2$ to estimate $(1+x)^5$ for $x = 0.2$.

(e) By how much does the estimate in (d) differ from $(1.2)^5$?

14. If $f(3) = 0$ and $f'(x) \geq 2$ for all $x \in (-\infty, \infty)$, what can be said about $f(1)$? Explain.

15. If $f(0) = 3$ and $f'(x) \geq -1$ for all $x \in (-\infty, \infty)$, what can be said about $f(2)$ and about $f(-2)$? Explain.

16. Use the polynomial in Example 3 to estimate e. Provide two numbers p and q, such that $p < e < q$ and $|p - q|$ is "small."

In Example 2 the polynomial $1 - \frac{x^2}{2}$ was shown to be a good approximation to $\cos(x)$ for x near 0. You may wonder how that polynomial was chosen. Exercise 17 shows how.

17. Let $P(x) = a_0 + a_1x + a_2x^2$ be an arbitrary quadratic polynomial. For which values of a_0, a_1, and a_2 is:

(a) $\cos(0) - P(0) = 0$?

(b) $\cos'(0) - P'(0) = 0$?

(c) $\cos''(0) - P''(0) = 0$?

(d) Let $R(x) = \cos(x) - P(x)$. For which $P(x)$ is $R(0) = R'(0) = R''(0) = 0$?

18. Find constants a_0, a_1, a_2, and a_3 such that if $R(x) = \tan(x) - \left(a_0 + a_1x + a_2x^2 + a_3x^3\right)$ then $R(0) = R'(0) = R''(0) = R^{(3)}(0) = 0$.

19. Find constants a_0, a_1, a_2, and a_3 such that if $R(x) = \sqrt{1+x} - \left(a_0 + a_1x + a_2x^2 + a_3x^3\right)$ then $R(0) = R'(0) = R''(0) = R^{(3)}(0) = 0$.

20. Find constants a_0, a_1, a_2, and a_3 such that if

$$R(x) \quad = \quad \sin x - \left(a_0 + a_1\left(x - \frac{\pi}{6}\right) + a_2\left(x - \frac{\pi}{6}\right)^2 + a_3\left(x - \frac{\pi}{6}\right)^3\right)$$

then $R\left(\frac{\pi}{6}\right) = R'\left(\frac{\pi}{6}\right) = R''\left(\frac{\pi}{6}\right) = R^{(3)}\left(\frac{\pi}{6}\right) = 0$. (Consider derivatives evaluated at $\pi/6$.)

Exercises 21 to 25 are related.

21. Because $e > 1$, it is known that $e^x \geq 1$ for every $x \geq 0$.

(a) Use Lemma 5.4.1 to deduce that $e^x > 1 + x$, for $x > 0$.

(b) Use (a) and Lemma 5.4.1 to deduce that, for $x > 0$, $e^x > 1 + x + \frac{x^2}{2!}$.

(c) Use (b) and Lemma 5.4.1 to deduce that, for $x > 0$, $e^x > 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!}$.

(d) In view of (a), (b), and (c), what is the general inequality that can be proved by this approach?

22. Let k be a fixed positive number. For x in $[0, k]$, $e^x \leq e^k$.

(a) Deduce that $e^x \leq 1 + e^k x$ for x in $[0, k]$.

(b) Deduce that $e^x \leq 1 + x + e^k \frac{x^2}{2!}$ for x in $[0, k]$.

(c) Deduce that $e^x \leq 1 + x + \frac{x^2}{2!} + e^k \frac{x^3}{3!}$ for x in $[0, k]$.

(d) In view of (a), (b), and (c), what is the general inequality that can be proved by this approach?

23. Combine the results of Exercises 21 and 22 to estimate $e = e^1$ to two decimal places. Assume $e \leq 3$.

24. What properties of e^x did you use in Exercises 21 and 22?

25. Let $E(x)$ be a function such that $E(0) = 1$ and $E'(x) = E(x)$ for all x.

(a) Show that $E(x) \geq 1$ for all $x \geq 0$.

(b) Use (a) to show that $E(x)$ is an increasing function for all $x \geq 0$. (Show that $E'(x) \geq 1$, for all $x \geq 0$.)

(c) Show $E(x) \geq 1 + x + \frac{x^2}{2}$ for all $x \geq 0$.

Exercises 26 to 32 show that $\lim_{x\to\infty} \frac{x}{e^x}$, $\lim_{x\to\infty} \frac{\ln(y)}{y}$, $\lim_{x\to 0^+} x\ln(x)$, $\lim_{x\to\infty} \frac{x^k}{b^x}$ $(b > 1)$, and $\lim_{x\to 0^+} x^x$ are closely connected. (If you know one of them you can deduce the other three.)

Exercises 26 and 27 use the inequalityt $e^x > 1 + x + \frac{x^2}{2}$ for all $x > 0$ (see Exercise 21).

26. Evaluate $\lim_{x\to\infty} \frac{x}{e^x}$.

27. Evaluate $\lim_{y\to\infty} \frac{\ln(y)}{y}$. (Let $y = e^x$ and compare with Exercise 26.)

Exercises 28 and 29 provide proofs of the fact that the exponential function grows faster than any power of x. **28.**

(a) Let n be a positive integer. Write $\frac{x^n}{e^x} = \left(\frac{x}{e^{x/n}}\right)\left(\frac{x}{e^{x/n}}\right)\cdots\left(\frac{x}{e^{x/n}}\right)$. Let $y = x/n$ so that $\frac{x}{e^{x/n}} = \frac{ny}{e^y}$. Use Exercise 26 ($n$ times) to show that $\lim_{x\to\infty} \frac{x^n}{e^x} = 0$.

(b) Deduce that for any fixed number k, $\lim_{x\to\infty} \frac{x^k}{e^x} = 0$.

29. (See Exercise 28.) Show that for any positive integer n, $lim_{x\to\infty} x^n/e^x = 0$, using Exercise 21(d).

30. Evaluate $\lim_{x\to 0^+} x\ln(x)$ as follows: Let $x = 1/t$, where $t \to \infty$. Then $x\ln(x) = \frac{1}{t}\ln\left(\frac{1}{t}\right) = \frac{-\ln(t)}{t}$. and refer to Exercise 27.

31. Evaluate $\lim_{x\to 0^+} x^x$ as follows: Let $y = x^x$. Then $\ln(y) = x\ln(x)$, a limit that was evaluated in Exercise 30. Explain why $\ln(y) \to 0$ implies $y \to 1$.

32. Evaluate $\lim_{x\to\infty} \frac{x^k}{b^x}$ for any $b > 1$ and k is a positive integer, (Use the result obtained in Exercises 28 or 29.)

33. Explain why $f(a) = g(a)$ and $f'(x) \leq g'(x)$ on $[a, b]$ with $a > b$ implies $f(x) \geq g(x)$ for all x in $[a, b]$.

34. In Example 1 it is shown that $|e^x - 1 - x| \leq \frac{e}{2}x^2$ for all x in $(-1, 1)$. Find a bound for

(a) $R(x) = e^x - 1 - x - \frac{x^2}{2}$ on $(-1, 1)$.

(b) $R(x) = e^x - 1 - x$ on $(-2, 1)$.

(c) $R(x) = e^x - 1 - x$ on $(-1, 2)$.

(d) $R(x) = e^x - 1 - x - \frac{x^2}{2}$ on $(-2, 1)$.

(e) $R(x) = e^x - 1 - x - \frac{x^2}{2}$ on $(-1, 2)$.

35. Apply Lemma 5.4.1 for $x > a$ to the case when $R(a) = R''(a) = 0$, $R^{(3)}(t) \leq M$, (for all t in $[a, x]$) but $R'(a) = 5$.

36. Consider the following proposal by Sam: "As usual, I can do things more simply than the text. For instance, say $R(a) = R'(a) = R''(a) = 0$ and $R^{(3)}(x) \leq M$. I'll show how M affects the size of $R(x)$, for $x > a$.
By the Mean-Value Theorem, $R(x) = R(x) - R(a) = R'(c_1)(x - a)$ for some c_1 in $[a, x]$. Then I just use the MVT again, this time finding $R'(c_1) = R'(c_1) - R'(a) = R''(c_2)(c_1 - a)$ for some c_2 in $[a, c_1]$. One more application of this idea then gives $R''(c_2) = R''(c_2) - R''(a) = R^{(3)}(c_3)(c_3 - a)$.
Then I put these all together, getting

$$R(x) \leq M(x - a)(c_2 - a)(c_3 - a).$$

Since c_1, c_2, and c_3 are in $[a, x]$, I can certainly say that

$$R(x) \leq M(x - a)^3.$$

I didn't need that lemma about two functions."
Is Sam correct? Is this a valid substitute for the text's treatment? Explain.

37. The proof of the Growth Theorem when x is less than a is slightly different than the proof when x is greater than a. Prove it for the case $n = 4$. Note that in this case $(x - a)^3$ and $(x - a)$ are negative $x < a$.

5.5 Taylor Polynomials and Their Errors

We spend years learning how to add, subtract, multiply, and divide. These same operations are built into any calculator or computer. Both we and machines can evaluate a polynomial, such as

$$a_0 + a_1x + a_2x^2 + \cdots + a_nx^n,$$

when x and the coefficients a_0, a_1, a_2, ..., a_n are given. Only multiplication and addition are needed. But how do we evaluate e^x? We resort to our calculators or look in a table that lists values of e^x. If e^x were a polynomial in disguise, then it would be easy to evaluate it by finding the polynomial and evaluating it instead. But e^x cannot be a polynomial, as the reasons in the margin show.

Three different reasons:

1. Because e^x equals its own derivative and no polynomial equals its own derivative (other than the polynomial that has constant value 0).
2. When you differentiate a non-constant polynomial, you get a polynomial with a lower degree.
3. Also, $e^x \to 0$ as $x \to -\infty$ and no non-constant polynomial has this property.

Since we cannot write e^x as a polynomial, we settle for the next best thing. Let's look for a polynomial that closely *approximates* e^x. However, no polynomial can be a good approximation of e^x for *all* x, since e^x grows too fast as $x \to \infty$. We search, instead, for a polynomial that is close to e^x for x in some short interval.

In this section we develop a method to construct polynomial approximations to functions. The accuracy of these approximations can be determined using the Growth Theorem from the previous section. Higher derivatives play a pivotal role.

Fitting a Polynomial, Near 0

Suppose we want to find a polynomial that closely approximates a function $y = f(x)$ for x near the input 0. For instance, what polynomial $p(x)$ of the form $a_0 + a_1x + a_2x^2 + a_3x^3$ might produce a good fit?

First we insist that

$$p(0) = f(0) \tag{5.5.1}$$

so the approximation is exact when $x = 0$.

Second, we would like the slope of the graph of $p(x)$ to be the same as that of $f(x)$ when x is 0. Therefore, we require

$$p'(0) = f'(0). \tag{5.5.2}$$

There are many polynomials that satisfy these two conditions. To find the best choices for the four numbers a_0, a_1, a_2, and a_3 we need four equations. To get them we continue the pattern started by (5.5.1) and (5.5.2). So we also insist that

$$p''(0) = f''(0) \tag{5.5.3}$$

and

$$p^{(3)})(0) = f^{(3)}(0). \tag{5.5.4}$$

Equation (5.5.3) forces the polynomial $p(x)$ to have the same sense of concavity as the function $f(x)$ at $x = 0$. We expect the graphs of $f(x)$ and such a polynomial $p(x)$ to resemble each other for x close to a.

To find the unknowns a_0, a_1, a_2, and a_3 we first compute $p(x)$, $p'(x)$, $p''(x)$, and $p^{(3)}(x)$ at 0. Table 5.5.1 displays the computations that yield formulas for the unknowns, a_0, a_1, a_2, and a_3, in terms of $f(x)$ and its derivatives. For example, note how we compute $p''(x) = 2a_2 + 3 \cdot 2a_3 x$ and evaluate it at 0 to obtain $p''(0) = 2a_2 + 3 \cdot 2a_3 \cdot 0 = 2a_2$. Then we obtain an equation for a_2 by equating $p''(0)$ and $f''(0)$; that is, $2a_2 = f''(0)$, so $a_2 = \frac{1}{2}f''(0)$.

$p(x)$ and its derivatives	Their values at 0	Equation for a_k	Formula for a_k
$p(x) = a_0 + a_1x + a_2x^2 + a_3x^3$	$p(0) = a_0$	$a_0 = f(0)$	$a_0 = f(0)$
$p^{(1)}(x) = a_1 + 2a_2x + 3a_3x^2$	$p^{(1)}(0) = a_1$	$a_1 = f^{(1)}(0)$	$a_1 = f^{(1)}(0)$
$p^{(2)}(x) = 2a_2 + 3 \cdot 2a_3x$	$p^{(2)}(0) = 2a_2$	$2a_2 = f^{(2)}(0)$	$a_2 = \frac{1}{2}f^{(2)}(0)$
$p^{(3)}(x) = 3 \cdot 2a_3$	$p^{(3)}(0) = 3 \cdot 2a_3$	$3 \cdot 2a_3 = f^{(3)}(0)$	$a_3 = \frac{1}{3 \cdot 2}f^{(3)}(0)$

Table 5.5.1

We can write a general formula for a_k if we let $f^{(0)}(x)$ denote $f(x)$ and recall that $0! = 1$ (by definition), $1! = 1$, $2! = 2 \cdot 1 = 2$, and $3! = 3 \cdot 2$. According to Table 5.5.1,

Factorials appear in the denominators.

$$a_k = \frac{f^{(k)}(0)}{k!}, \qquad k = 0, 1, 2, 3.$$

Therefore

$$p(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f^{(3)})(0)}{3!}x^3.$$

The coefficient of x^k is *completely determined* by the k^{th} derivative of f evaluated at 0. It equals the k^{th} derivative of f at 0 divided by $k!$. This suggests the following definition.

DEFINITION *(Taylor Polynomials at 0)* Let n be a non-negative integer and let f be a function with derivatives at 0 of all orders through n. Then the polynomial

The n^{th}-order Taylor polynomial has degree at most n.

$$f(0) + f^{(1)}(0)x + \frac{f^{(2)}(0)}{2!}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n \tag{5.5.5}$$

is called the **n^{th}-order Taylor polynomial of f centered at 0** and is denoted $P_n(x; 0)$. It is also called a **Maclaurin polynomial**.

Whether $P_n(x;0)$ approximates $f(x)$ for x near 0 is not obvious. We will show that the Macaurin polynomials for e^x do provide good approximations of the function when x is not too large.

EXAMPLE 1 Find the Maclaurin polynomial $P_4(x;0)$ associated with $1/(1-x)$.

SOLUTION The first step is to compute $1/(1-x)$ and its first four derivatives, then evaluate them at $x=0$. Dividing them by suitable factorials gives the coefficients of the Maclaurin polynomial. Table 5.5.2 records the computations.

		at x	at 0
$f(x)$	$=$	$\frac{1}{1-x}$	1
$f'(x)$	$=$	$\frac{1}{(1-x)^2}$	1
$f''(x)$	$=$	$\frac{2}{(1-x)^3}$	2
$f^{(3)}(x)$	$=$	$\frac{3\cdot 2}{(1-x)^4}$	$3\cdot 2$
$f^{(4)}(x)$	$=$	$\frac{4\cdot 3\cdot 2}{(1-x)^5}$	$4\cdot 3\cdot 2$

Table 5.5.2

So the fourth-degree Maclaurin polynomial is

$$P_4(x;0) = 1 + \frac{1}{1!}x + \frac{2}{2!}x^2 + \frac{3\cdot 2}{3!}x^3 + \frac{4\cdot 3\cdot 2}{4!}x^4,$$

which simplifies to

$$P_4(x;0) = 1 + x + x^2 + x^3 + x^4.$$

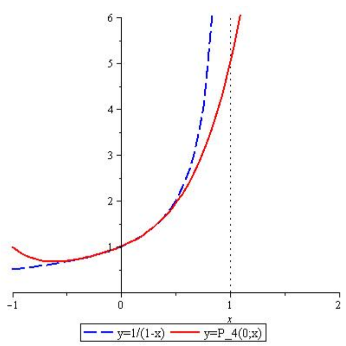


Figure 5.5.1

Figure 5.5.1 suggests that $P_4(x;0)$ does a fairly good job of approximating $1/(1-x)$ for x near 0. $\diamond$

The calculations in Example 1 suggest that

> The Maclaurin polynomial $P_n(x;0)$ associated with $1/(1-x)$ is
> $$1 + x + x^2 + x^3 + \cdots + x^{n-1}.$$

Because all the derivatives of e^x at 0 are 1,

> The Maclaurin polynomial $P_n(x;0)$ associated with e^x is
> $$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^{n-1}}{(n-1)!}.$$

EXAMPLE 2 Find the Maclaurin polynomial $P_5(x;0)$ for $f(x)=\sin(x)$.

SOLUTION Again we make a table for computing the coefficients of the Taylor polynomial centered at 0. (See Table 5.5.3.)

at x			at 0		
$f^{(0)}(x)$	$=$	$\sin(x)$	$f^{(0)}(0)$	$=$	$\sin(0) = 0$
$f^{(1)}(x)$	$=$	$\cos(x)$	$f^{(1)}(0)$	$=$	$\cos(0) = 1$
$f^{(2)}(x)$	$=$	$-\sin(x)$	$f^{(2)}(0)$	$=$	$-\sin(0) = 0$
$f^{(3)}(x)$	$=$	$-\cos(x)$	$f^{(3)}(0)$	$=$	$-\cos(0) = -1$
$f^{(4)}(x)$	$=$	$\sin(x)$	$f^{(4)}(0)$	$=$	$\sin(0) = 0$
$f^{(5)}(x)$	$=$	$\cos(x)$	$f^{(5)}(0)$	$=$	$\cos(0) = 1$

Table 5.5.3

Thus

$$\begin{aligned} P_5(x;0) &= f^{(0)}(0) + f^{(1)}(0)x + \frac{f^{(2)}(0)}{2!}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \frac{f^{(4)}(0)}{4!}x^4 + \frac{f^{(5)}(0)}{5!}x^5 \\ &= 0 + (1)x + \frac{0}{2!}x^2 + \frac{-1}{3!}x^3 + \frac{0}{4!}x^4 + \frac{1}{5!}x^5 \\ &= x - \frac{x^3}{3!} + \frac{x^5}{5!}. \end{aligned}$$

x	$\sin(x)$	$P_5(x;0)$
0.0	0.000000	0.000000
0.1	0.099833	0.099833
0.5	0.479426	0.479427
1.0	0.841471	0.841667
2.0	0.909297	0.933333
π	0.000000	0.524044
2π	0.000000	46.546732

Table 5.5.4

Figure 5.5.2 illustrates the graphs of $P_5(x;0)$ and $\sin(x)$ near 0. ◇

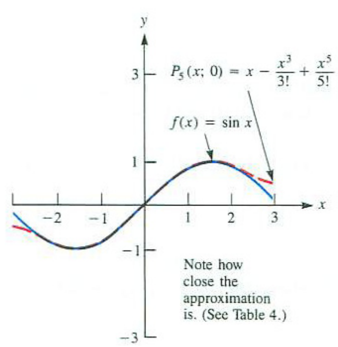


Figure 5.5.2

Having found the fifth-order Maclaruin polynomial for $\sin(x)$, let us see how good an approximation it is of $\sin(x)$. Table 5.5.4 compares their values to six-decimal-place accuracy for inputs both near 0 and far from 0. As we see, the closer x is to 0, the better the Taylor approximation is. When x is large, $P_5(x;0)$ gets very large, but the value of $\sin(x)$ stays between -1 and 1.

A Shorthand Notation

The Maclaurin polynomials associated with $\sin(x)$ have only odd powers and its terms alternate in sign:

$$P_m(x;0) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots \pm \frac{x^m}{m!}.$$

The $\pm$ in front of $x^m/m!$ indicates the coefficient is either positive or negative. For the terms involving x, x^5, x^9, ..., the coefficient is $+1$. For x^3, x^7, x^{11}, ...it is -1. If m is odd, it can be written as $2n+1$ for some integer n. If n is even, the coefficient of x^{2n+1} is $+1$. If n is odd, the coefficient of x^{2n+1} is -1. The shorthand notation to write the typical summand is

$$(-1)^n \frac{x^{2n+1}}{(2n+1)!}.$$

So we may write

$$P_{2n+1}(x;0) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1_!}.$$

Taylor Polynomials Centered at a

We may be interested in estimating a function $f(x)$ near a number a, not just near 0. In that case, we express the approximating polynomial in terms of powers of $x-a$ instead of powers of $x = x-0$ and make the derivatives of the approximating polynomial, evaluated at a, coincide with the derivatives of the function at a. Calculuations similar to those that gave us the polynomial (5.5.5) produce the polynomial called a "Taylor polynomial centered at a". (If a is not 0, it is not called a Maclaurin polynomial.)

The n^{th}-order Taylor polynomial of f centered at a is denoted $P_n(x;a)$. It's degree is at most n.

DEFINITION *(Taylor Polynomials of n^{th} order, $P_n(x;a)$)* If the function f has derivatives through order n at a, then the **n^{th}-order Taylor polynomial of f centered at a** is defined as

$$f(a) + f^{(1)}(a)(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

and is denoted $P_n(x;a)$.

EXAMPLE 3 Find the n^{th}-order Taylor polynomial centered at a for $f(x) = e^x$.

SOLUTION All the derivatives of e^x evaluated at a are e^a. Thus

$$P_n(x;a) = e^a + e^a(x-a) + \frac{e^a}{2!}(x-a)^2 + \frac{e^a}{3!}(x-a)^3 + \cdots + \frac{e^a}{n!}(x-a)^n.$$

◇

The Error in Using A Taylor Polynomial

There is no point using $P_n(x;a)$ to estimate a function $f(x)$ if we have no idea how large the difference between $f(x)$ and $P_n(x;a)$ may be. So let us take a look at the difference.

Define the **remainder** to be the difference between the function, $f(x)$, and the Taylor polynomial, $P_n(x;a)$. Denote the remainder as $R_n(x;a)$. Then,

$$f(x) = P_n(x;a) + R_n(x;a).$$

We will be interested in the absolute value of the remainder. We call $|R_n(x;a)|$ the **error** in using $P_n(x;a)$ to approximate $f(x)$. We do not care whether $P_n(x;a)$ is larger or smaller than the exact value.

Theorem 5.5.1 (The Lagrange Form of the Remainder)**.** *Assume that a function $f(x)$ has continuous derivatives of orders through $n+1$ in an interval that includes the numbers a and x. Let $P_n(x;a)$ be the n^{th}-order Taylor polynomial associated with $f(x)$ in powers of $x-a$. Then there is a number c_n between a and x such that*

$$R_n(x;a) = \frac{f^{(n+1)}(c_n)}{(n+1)!}(x-a)^{n+1}.$$

Proof of Theorem 5.5.1

For simplicity, we denote the remainder $R_n(x;a) = f(x) - P_n(x;a)$ by $R(x)$.

Since $P_n(a;a) = f(a)$,

$$R(a) = f(a) - P_n(a;a) = f(a) - f(a) = 0.$$

Similarly, repeated differentiation of $R(x)$, leads to

$$R^{(k)}(x) = f^{(k)}(x) - P_n^{(k)}(x;a), \tag{5.5.6}$$

for each integer k, $1 \le k \le n$. From the definition of $P_n(x;a)$,

$$R^{(k)}(a) = f^{(k)}(a) - P_n^{(k)}(a;a) = 0.$$

Since $P_n(x;a)$ is a polynomial of degree at most n, its $(n+1)^{\text{st}}$ derivative is 0. As a result, the $(n+1)^{\text{st}}$ derivative of $R(x)$ is the same as the $(n+1)^{\text{st}}$ derivative of $f(x)$. Thus, $R(x)$ satisfies all the assumptions of the Growth Theorem. Recalling (5.4.1) from Section 5.4, we see

$R^{(n+1)}(x) = f^{(n+1)}(x)$

Lagrange Form of the Remainder

There is a number c_n between a and x such that

$$R_n(x;a) = \frac{f^{(n+1)}(c_n)}{(n+1)!}(x-a)^{n+1}.$$

•

EXAMPLE 4 Discuss the error in using $x - \frac{x^3}{3!} + \frac{x^5}{5!}$ to estimate $\sin(x)$ for $x > 0$.

SOLUTION Example 2 showed that $x - \frac{x^3}{3!} + \frac{x^5}{5!}$ is the fifth-order Maclaurin polynomial, $P_5(x;0)$, associated with $\sin(x)$. In this case $f(x) = \sin(x)$ and each derivative of $f(x)$ is either $\pm\sin(x)$ or $\pm\cos(x)$. Therefore, $|f^{n+1}(c_n)|$ is at most 1, and we have

$$\frac{|f^{5+1}(c_5)|}{6!}x^6 \leq \frac{x^6}{6!}.$$

Then

$$\left|\sin(x) - \left(x - \frac{x^3}{6} + \frac{x^5}{120}\right)\right| \leq \frac{|x|^6|}{6!} = \frac{x^6}{720}.$$

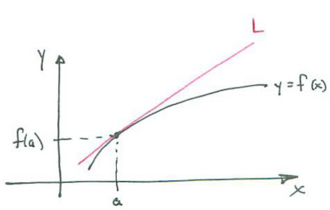
For instance, with $x = 1/2$,

$$\begin{aligned} \left|\sin\left(\frac{1}{2}\right) - \left(\left(\frac{1}{2}\right) - \frac{\left(\frac{1}{2}\right)^3}{6} + \frac{\left(\frac{1}{2}\right)^5}{120}\right)\right| \quad &\leq \quad \frac{\left(\frac{1}{2}\right)^6}{720} \\ &= \quad \frac{1}{(64)(720)} = \frac{1}{46,080} \approx 0.0000217 = 2.17 \times \end{aligned}$$

So the approximation

$$P_5(\frac{1}{2};0) = \frac{1}{2} - \frac{1}{3!}\left(\frac{1}{2}\right)^3 + \frac{1}{5!}\left(\frac{1}{2}\right)^5 = \frac{1}{2} - \frac{1}{48} + \frac{1}{3840} = \frac{1841}{3840} \approx 0.4794271$$

differs from $\sin(1/2)$ (the sine of half a radian) by less than 2.17×10^{-5}; this means at least the first four decimal places are correct. The exact value of $\sin(1/2)$, to ten decimal places is 0.4794255386 and our estimate is correct to five decimal places. By comparison, a calculator gives $\sin(1/2) \approx 0.479426$, which is also correct to five decimal places. ◇



The Linear Approximation $P_1(x;a)$

The graph of the Taylor polynomial $P_1(x;a) = f(a) + f'(a)(x-a)$ is a line that passes through the point $(a, f(a))$ and has the same slope as f does at a. That means that the graph of $P_1(x;a)$ is the *tangent line* to the graph of f at $(a, f(a))$. It is customary to call $P_1(x;a) = f(a) + f'(a)(x-a)$ the **linear approximation** to $f(x)$ for x near a. It is often denoted $L(x)$. Figure 5.5.3 shows the graphs of f and L near the point $(a, f(a))$.

Let x be a number close to a and define $\Delta x = x - a$ and $\Delta y = f(a+\Delta x) - f(a)$, quantities used in the definition of the derivative: $f'(a) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$.

Often Δx is denoted by dx and $f'(a)dx$ is defined to be "dy", as shown in Figure 5.5.4. Note that dy is an approximation to Δy, and $f(a) + dy$ is an approximation to $f(a + \Delta x) = f(a) + \Delta y$.

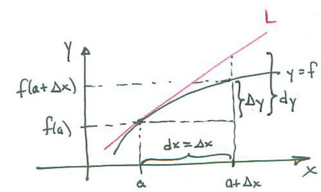


Figure 5.5.4

In Section 8.2 we will use $dy = f'(x)dx$ and dx as bookkeeping tools to simplify the search for antiderivatives.

The expressions "dx" and "dy" are called **differentials**. In the seventeenth century, dx and dy referred to "infinitesimals", infinitely small numbers. Leibniz viewed the derivative as the quotient $\frac{dy}{dx}$, and that notation for the derivative persists more than three centuries later.

> **WARNING** *(The derivative is not a quotient.)* The derivative is the limit of a quotient.

The next example uses the linear approximation to estimate $\sqrt{x}$ near $x = 1$.

EXAMPLE 5 Use $P_1(x;1)$ to estimate $\sqrt{x}$ for x near 1. Then discuss the error.

SOLUTION In this case $f(x) = \sqrt{x}$, $f'(x) = \frac{1}{2\sqrt{x}}$, and $f'(1) = 1/2$. The linear approximation of $f(x)$ near $a = 1$ is

$$P_1(x;1) = f(1) + f'(1)(x-1) = 1 + \frac{1}{2}(x-1)$$

and the remainder is

$$R_1(x;1) = \sqrt{x} - \left(1 + \frac{1}{2}(x-1)\right).$$

Table 5.5.5 shows how rapidly $R_1(x;1)$ approaches 0 as $x \to 1$ and compares this difference with $(x-1)^2$.

The final column in Table 5.5.5 shows that $\frac{R_1(x;1)}{(x-1)^2}$ is nearly constant. Because $(x-1)^2 \to 0$ as $x \to 0$, this means $R_1(x;1)$ approaches 0 at the same rate as the square of $(x-1)$.

Since the Lagrange form for $R_1(x;1)$ is approximately $\frac{1}{2}f''(1)(x-1)^2$ when x is near 1, $\frac{R_1(x;1)}{(x-1)^2}$ should be near $\frac{1}{2}f''(1)$ when x is near 1. Just as a check, we compute $\frac{1}{2}f''(1)$. We have $f''(x) = \frac{-1}{4}x^{-3/2}$. Thus $\frac{1}{2}f''(1) = \frac{1}{2}\left(\frac{-1}{4}\right) = \frac{-1}{8} = -0.125$. This is consistent with the final column of Table 5.5.5. ◇

x	$R_1(x;1)$	$(x-1)^2$	$R_1(x;1)/(x-1)^2$
2.0	$\sqrt{2} - \left(1+\frac{1}{2}(2-1)\right) \approx -0.08578643$	1	-0.08579
1.5	$\sqrt{1.5} - \left(1+\frac{1}{2}(1.5-1)\right) \approx -0.02525512$	0.25	-0.10102
1.1	$\sqrt{1.1} - \left(1+\frac{1}{2}(1.1-1)\right) \approx -0.00119115$	0.01	-0.11912
1.01	$\sqrt{1.01} - \left(1+\frac{1}{2}(1.01-1)\right) \approx -0.00001243$	0.0001	-0.12438

Table 5.5.5

Summary

We define the "zeroth derivative" of a function to be the function itself *and* start counting from 0. This allows us to say simply that the derivatives $P_n^{(k)}(x;a)$ coincide with $f^{(k)}(a)$ for $k = 0, 1, \ldots, n$.

Given a function f with n derivatives on an interval that contains the number a we defined the n^{th}-order Taylor polynomial at a, $P_n(x;a)$. The first n derivatives of the Taylor polynomial of degree n coincide with the first n derivatives of the given function f at a. Also, $P_n(x;a)$ has the same function value at a that f does.

$$P_n(x;a) = f(a) + f^{(1)}(a)(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

If $a = 0$, $P_n(x;0)$ is call a Maclaurin polynomial. The general Maclaurin polynomial associated with

$$\begin{array}{ccl} e^x & \text{is} & 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} \\ \sin(x) & \text{is} & x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^n \frac{x^{2n+1}}{(n+1)!} \\ \cos(x) & \text{is} & 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + (-1)^n \frac{x^{2n}}{(2n)!} \\ 1/(1-x) & \text{is} & 1 + x + x^2 + x^3 + \cdots + x^n \end{array}$$

The remainder in using the Taylor polynomial of degree n to estimate a function involves the $(n+1)^{\text{st}}$ derivative of the function:

$$R_n(x;a) = f(x) - P_n(x;a) = \frac{f^{(n+1)}(c_n)}{(n+1)!}(x-a)^{n+1}$$

where c_n is a number between a and x . The error is the absolute value of the remainder, $|R_n(x;a)|$.

The linear approximation to a function near a is

$$L(x) = P_1(x;a) = f(a) + f'(a)(x-a).$$

The differentials are $dx = x - a$ and $dy = f'(a)dx$. While $dx = \Delta x$, $dy \approx \Delta y = f(x + \Delta x) - f(x)$.

EXERCISES for Section 5.5

Use a graphing calculator or computer algebra computer algebra system to assist with the computations and with the graphing.

1. Give at least three reasons $\sin(x)$ cannot be a polynomial.

In Exercises 2 to 13 compute the Taylor polynomials. Graph $f(x)$ and $P_n(x;a)$ on the same axes on a domain centered at a. Keep in mind that the graph of $P_1(x;a)$ is the tangent line at the point $(a, f(a))$.

2. $f(x) = 1/(1+x)$, $P_1(x;0)$ and $P_2(x;0)$

3. $f(x) = 1/(1+x)$, $P_1(x;1)$ and $P_2(x;1)$

4. $f(x) = \ln(1+x)$, $P_1(x;0)$, $P_2(x;0)$ and $P_3(x;0)$

5. $f(x) = \ln(1+x)$, $P_1(x;1)$, $P_2(x;1)$ and $P_3(x;1)$

6. $f(x) = e^x$, $P_1(x;0)$, $P_2(x;0)$, $P_3(x;0)$, and $P_4(x;0)$

7. $f(x) = e^x$, $P_1(x;2)$, $P_2(x;2)$, $P_3(x;2)$, and $P_4(x;2)$

8. $f(x) = \arctan(x)$, $P_1(x;0)$, $P_2(x;0)$, and $P_3(x;0)$

9. $f(x) = \arctan(x)$, $P_1(x;-1)$, $P_2(x;-1)$, and $P_3(x;-1)$

10. $f(x) = \cos(x)$, $P_2(x;0)$ and $P_4(x;0)$

11. $f(x) = \sin(x)$, $P_7(x;0)$

12. $f(x) = \cos(x)$, $P_6(x;\pi/4)$

13. $f(x) = \sin(x)$, $P_7(x;\pi/4)$

14. Can there be a polynomial $p(x)$ such that $\sin(x) = p(x)$ for all x in the interval $[1, 1.0001]$? Explain.

15. Can there be a polynomial $p(x)$ such that $\ln(x) = p(x)$ for all x in the interval $[1, 1.0001]$? Explain.

16. State the Lagrange formula for the error in using a Taylor polynomial as an estimate of the value of a function. Use as little mathematical notation as you can.

In Exercises 17 to 22 obtain the Maclaurin polynomial of order n associated with the given function.

17. $1/(1-x)$

18. e^x

19. e^{-x}

20. $\sin(x)$

21. $\cos(x)$

22. $1/(1+x)$

23. Let $f(x) = \sqrt{x}$.

(a) What is the linear approximation, $P_1(x;4)$, to $\sqrt{x}$ at $x = 4$?

(b) Fill in the following table.

x	$R_1(x;4) = f(x) - P_1(x;4)$	$(x-4)^2$	$\frac{R_1(x;4)}{(x-4)^2}$
5.0			
4.1			
4.01			
3.99			

(c) Compute $f''(4)/2$. Explain the relationship between this number and the entries in the fourth column of the table in (b).

24. Repeat Exercise 23 for the linear approximation to $\sqrt{x}$ at $a = 3$. Use $x = 4$, 3.1, 3.01, and 2.99.

25. Assume $f(x)$ has continuous first and second derivatives and that $4 \leq f''(x) \leq 5$ for all x.

(a) What can be said in general about the error in using $f(2) + f'(2)(x-2)$ to approximate $f(x)$?

(b) How small should $x - 2$ be to be sure that the error — the absolute value of the remainder — is less than or equal to 0.005? This ensures the approximate value is correct to 2 decimal places.

26. Let $f(x) = 2 + 3x + 4x^2$.

(a) Find $P_2(x;0)$.

(b) Find $P_3(x;0)$.

(c) Find $P_2(x;5)$.

(d) Find $P_3(x;5)$.

27.

(a) What can be said about the degree of the polynomial $P_n(x;0)$?

(b) When is the degree of $P_n(x;0)$ less than n?

(c) When is the degree of $P_n(x; a)$ less than n? ($a \neq 0$)

28. In the case of $f(x) = 1/(1-x)$ the error $R_n(x; 0)$ in using a Maclaurin polynomial $P_n(x; 0)$ to estimate the function can be calculated exactly. Show that it equals $\left|x^{n+1}/(1-x)\right|$.

Exercises 29 to 32 are related.

29. Let $f(x) = (1+x)^3$.

(a) Find $P_3(x; 0)$ and $R_3(x; 0)$.

(b) Check that your answer to (a) is correct by multiplying out $(1+x)^3$.

30. Let $f(x) = (1+x)^4$.

(a) Find $P_4(x; 0)$ and $R_4(x; 0)$.

(b) Check that your answer to (a) is correct by multiplying out $(1+x)^4$.

31. Let $f(x) = (1+x)^5$. Using $P_5(x; 0)$, show that

$$(1+x)^5 = 1 + 5x + \frac{5 \cdot 4}{1 \cdot 2}x^2 + \frac{5 \cdot 4 \cdot 3}{1 \cdot 2 \cdot 3}x^3 + \frac{5 \cdot 4 \cdot 3 \cdot 2}{1 \cdot 2 \cdot 3 \cdot 4}x^4 + \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5}x^5.$$

For a positive integer n and a non-negative integer k, with $k \leq n$, the symbol $\binom{n}{k}$ denotes the **binomial coefficient**:

$$\binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{1 \cdot 2 \cdot 3 \cdots k} = \frac{n!}{k!(n-k)!}.$$

Thus

$$(1+x)^5 = \binom{5}{0} + \binom{5}{1}x + \binom{5}{2}x^2 + \binom{5}{3}x^3 + \binom{5}{4}x^4 + \binom{5}{5}x^5.$$

Using $P_n(x; 0)$ one can show that, for any positive integer n,

$$(1+x)^n = \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{n-1}x^{n-1} + \binom{n}{n}x^n = \sum_{k=0}^{n}\binom{n}{k}x^k.$$

This is the basis for the **Binomial Theorem**,

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Recall that $\binom{n}{0} = \frac{n!}{0!n!} = 1$ and $\binom{n}{n} = \frac{n!}{n!0!} = 1$.

32.

(a) Using algebra (no calculus) derive the binomial theorem for $(a+b)^3$ from the binomial theorem for $(1+x)^3$.

(b) Obtain the binomial theorem for $(a+b)^{12}$ from the special case $(1+x)^{12} = \sum_{k=0}^{12} \binom{12}{k} x^k$.

In Exercises 33 and 34, use a calculator or computer to help evaluate the Taylor polynomials

33. Let $f(x) = e^x$.

(a) Find $P_{10}(x;0)$.

(b) Compute $f(x)$ and $P_{10}(x;0)$ at $x=1$, $x=2$, and $x=4$.

34. Let $f(x) = \ln(x)$.

(a) Find $P_{10}(x;1)$.

(b) Compute $f(x)$ and $P_{10}(x;1)$ at $x=1$, $x=2$, and $x=4$.

35.

(a) Find $P_5(x;0)$ for $f(x) = \ln(1+x)$.

(b) What is $P_n(x;0)$?

(c) Estimate $\ln(1.05)$ using $P_5(x;0)$ and put a bound on the error.

Exercises 36 to 39 involve even and odd functions Recall, from Section 2.6, that a function is even if $f(-x) = f(x)$ and is odd if $f(-x) = -f(x)$.

36. Show that if f is an odd function, f' is an even function.

37. Show that if f is an even function, f' is an odd function.

38.

(a) Which polynomials are even functions?

(b) If f is an even function, are its associated Maclaurin polynomials necessarily even functions? Explain.

39.

(a) Which polynomials are odd functions?

(b) If f is an odd function, are its associated Maclaurin polynomials necessarily odd functions? Explain.

40. This exercise constructs Maclaurin polynomials that do not approximate the associated function. Let $f(x) = e^{-1/x^2}$ if $x \neq 0$ and $f(0) = 0$.

(a) Find $f'(0)$.

(b) Find $f''(0)$.

(c) Find $P_2(x; 0)$.

(d) What is $P_{100}(x; 0)$.

(Recall the definition of the derivative.)

41. Show that in an open interval in which f''' is positive, that $f(x) > f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2$. (Treat the cases $a < x$ and $x > a$ separately.) See also Exercise 17 in Section 4.4.

42.

(a) Show that in an open interval in which $f^{(n+1)}$ is positive (n a positive integer), that $f(x)$ is greater than $P_n(x; 0)$.

(b) What additional information is needed to make this a true statement for $x < a$?

See also Exercise 41.

43. The quantity $\sqrt{1 - v^2/c^2}$ occurs often in the theory of relativity. Here v is the velocity of an object and c the velocity of light. Justify the following approximations that physicists use:

(a) $\sqrt{1-\frac{v^2}{c^2}} \approx 1-\frac{1}{2}\frac{v^2}{c^2}$.

(b) $\frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \approx 1+\frac{1}{2}\frac{v^2}{c^2}$.

Even for a rocket v/c is very small.

44. Using the formula for the sum of a finite geometric series, justify the factorization used in Section 2.2. (See Exercise 42, Section 2.2. on page 93.)

$$x^n - a^n = (x-a)(x^{n-1} + x^{n-2}a + x^{n-3}a^2 + \cdots + a^{n-1}).$$

45. If $P_n(x;0)$ is the Maclaurin polynomial associated with $f(x)$, is $P_n(-x;0)$ the Maclaurin polynomial associated with $f(-x)$? Explain.

46. Let $P(x)$ be the Maclaurin polynomial of the second-order associated with $f(x)$. Let $Q(x)$ be the Maclaurin polynomial of the second-order associated with $g(x)$. What part, if any, of $P(x)Q(x)$ is a Maclaurin polynomial associated with $f(x)g(x)$? Explain.

5.6 L'Hôpital's Rule for Finding Certain Limits

There are two types of limits in calculus: those that you can evaluate at a glance, and those that require some work to evaluate. In Section 2.4 we called the limits that can be evaluated easily **determinate** and those that require some work **indeterminate**.

For instance $\lim_{x\to\pi/2} \frac{\sin(x)}{x}$ is clearly $1/(\pi/2) = 2/\pi$. That's easy. But $\lim_{x\to 0}(\sin(x))/x$ is not obvious. Back in Section 2.2 we used a diagram of circles, sectors, and triangles, to show that this limit is 1.

In this section we describe a technique for evaluating more indeterminate limits, for instance

$$\lim_{x\to a} \frac{f(x)}{g(x)}$$

when both $f(x)$ and $g(x)$ approach 0 as x approaches a. The numerator is trying to drag $f(x)/g(x)$ toward 0, at the same time as the denominator is trying to make the quotient large. L'Hôpital's rule helps determine which term wins or whether there is a compromise.

L'Hôpital is pronounced lope - ee - tall.

Indeterminate Limits

The following limits are called **indeterminate** because you can't determine them without knowing more about the functions f and g.

$$\lim_{x\to a} \frac{f(x)}{g(x)}, \text{ where } \lim_{x\to a} f(x) = 0 \text{ and } \lim_{x\to a} g(x) = 0$$

$$\lim_{x\to a} \frac{f(x)}{g(x)}, \text{ where } \lim_{x\to a} f(x) = \infty \text{ and } \lim_{x\to a} g(x) = \infty$$

L'Hôpital's Rule provides a way for dealing with these limits (and limits that can be transformed to those forms.) In short, l'Hôpital's rule applies only when you need it.

Theorem 5.6.1 (L'Hôpital's Rule (zero-over-zero case))**.** *Let a be a number and let f and g be differentiable over some open interval that contains a. Assume also that $g'(x)$ is not 0 for any x in that interval except perhaps at a. If*

$$\lim_{x\to a} f(x) = 0, \ \lim_{x\to a} g(x) = 0, \text{ and } \lim_{x\to a} \frac{f'(x)}{g'(x)} = L,$$

then

$$\lim_{x\to a} \frac{f(x)}{g(x)} = L.$$

In short, "to evaluate the limit of a quotient that is indeterminant, evaluate the limit of the quotient of their derivatives." You evaluate the limit of the quotient of the derivatives, <u>not</u> the derivative of the quotient. We will discuss the proof after some examples.

EXAMPLE 1 Find $\lim_{x\to 1}(x^5-1)/(x^3-1)$.
SOLUTION In this case,

$$a = 1,\ f(x) = x^5 - 1,\ \text{and } g(x) = x^3 - 1.$$

Remember to check that the hypotheses of l'Hôpital's Rule are satisfied.

All the assumptions of l'Hôpital's rule are satisfied. In particular,

$$\lim_{x\to 1}(x^5-1) = 0 \text{ and } \lim_{x\to 1}(x^3-1) = 0.$$

According to l'Hôpital's rule,

$$\lim_{x\to 1}\frac{x^5-1}{x^3-1} \overset{\text{l'H}}{=} \lim_{x\to 1}\frac{(x^5-1)'}{(x^3-1)'}$$

if the latter limit exists. Now,

$\lim_{x\to 1}\frac{(x^5-1)'}{(x^3-1)'}$	$=$	$\lim_{x\to 1}\frac{5x^4}{3x^2}$	differentiation of numerator and differentiation of denominator
	$=$	$\lim_{x\to 1}\frac{5}{3}x^2$	algebra
	$=$	$\frac{5}{3}$.	

Thus

$$\lim_{x\to 1}\frac{x^5-1}{x^3-1} = \frac{5}{3}.$$

◇

Sometimes it may be necessary to apply l'Hôpital's Rule more than once, as in the next example.

EXAMPLE 2 Find $\lim_{x\to 0}(\sin(x)-x)/x^3$.
SOLUTION As $x \to 0$, both numerator and denominator approach 0. By l'Hôpital's Rule,

$$\lim_{x\to 0}\frac{\sin(x)-x}{x^3} \overset{\text{l'H}}{=} \lim_{x\to 0}\frac{(\sin(x)-x)'}{(x^3)'} = \lim_{x\to 0}\frac{\cos(x)-1}{3x^2}.$$

But as $x \to 0$, both $\cos(x) - 1 \to 0$ and $3x^2 \to 0$. So use l'Hôpital's Rule again:

$$\lim_{x\to 0}\frac{\cos(x)-1}{3x^2} \overset{\text{l'H}}{=} \lim_{x\to 0}\frac{(\cos(x)-1)'}{(3x^2)'} = \lim_{x\to 0}\frac{-\sin(x)}{6x}.$$

Both $\sin(x)$ and $6x$ approach 0 as $x \to 0$. Use l'Hôpital's Rule yet another time:

Or recall from Section 2.2 that $\lim_{x\to 0}\frac{\sin x}{x} = 1$.

$$\lim_{x\to 0}\frac{-\sin(x)}{6x} \overset{\text{l'H}}{=} \lim_{x\to 0}\frac{(-\sin(x))'}{(6x)'} = \lim_{x\to 0}\frac{-\cos(x)}{6} = \frac{-1}{6}.$$

So after three applications of l'Hôpital's Rule we find that

$$\lim_{x\to 0}\frac{\sin(x)-x}{x^3} = -\frac{1}{6}.$$

◇

Sometimes a limit may be simplified before l'Hôpital's Rule is applied. For instance, consider

$$\lim_{x\to 0}\frac{(\sin(x)-x)\cos^5(x)}{x^3}.$$

Since $\lim_{x\to 0}\cos^5(x) = 1$, we have

$$\lim_{x\to 0}\frac{(\sin(x)-x)\cos^5(x)}{x^3} = \left(\lim_{x\to 0}\frac{\sin(x)-x}{x^3}\right)\cdot 1,$$

which, by Example 2, is $-\frac{1}{6}$. This shortcut saves a lot of work, as may be checked by finding the limit using l'Hôpital's Rule without separating $\cos^5(x)$.

Theorem 5.6.1 concerns limits as $x \to a$. L'Hôpital's Rule also applies if $x \to \infty$, $x \to -\infty$, $x \to a^+$, or $x \to a^-$. In the first case, we would assume that $f(x)$ and $g(x)$ are differentiable in some interval (c, ∞) and $g'(x)$ is not zero there. In the case of $x \to a^+$, assume that $f(x)$ and $g(x)$ are differentiable in some open interval (a, b) and $g'(x)$ is not 0 there.

Infinity-over-Infinity Limits

Theorem 5.6.1 concerns the limit of $f(x)/g(x)$ when both $f(x)$ and $g(x)$ approach 0. But a similar problem arises when both $f(x)$ and $g(x)$ get arbitrarily large as $x \to a$ or as $x \to \infty$. The behavior of the quotient $f(x)/g(x)$ will be influenced by how rapidly $f(x)$ and $g(x)$ become large.

"Infinity-over-infinity" is indeterminate.

In short, if $\lim_{x\to a} f(x) = \infty$ and $\lim_{x\to a} g(x) = \infty$, then $\lim_{x\to a}(f(x)/g(x))$ is an indeterminate form. The next theorem presents a l'Hôpital Rule for this case.

Theorem 5.6.2 (L'Hôpital's Rule (infinity-over-infinity case)**.** *Let f and g be defined and differentiable for all x larger than some number c. Then, if $g'(x)$ is not zero for all $x > c$,*

$$\lim_{x\to\infty} f(x) = \infty,\ \lim_{x\to\infty} g(x) = \infty,\ \textit{and}\ \lim_{x\to\infty}\frac{f'(x)}{g'(x)} = L,$$

it follows that

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = L.$$

A similar result holds for $x \to a$, $x \to a^-$, $x \to a^+$, *or* $x \to -\infty$. *Moreover,* $\lim_{x\to\infty} f(x)$ *and* $\lim_{x\to\infty} g(x)$ *could both be* $-\infty$, *or one could be* ∞ *and the other* $-\infty$.

EXAMPLE 3 Find $\lim_{x\to\infty} \frac{\ln(x)}{x^2}$.

SOLUTION Since $\ln(x) \to \infty$ and $x^2 \to \infty$ as $x \to \infty$, we may use l'Hôpital's Rule in the "infinity-over-infinity" form.

We have

$$\lim_{x\to\infty} \frac{\ln(x)}{x^2} \overset{\text{l'H}}{=} \lim_{x\to\infty} \frac{(\ln(x))'}{(x^2)'} = \lim_{x\to\infty} \frac{1/x}{2x} = \lim_{x\to\infty} \frac{1}{2x^2} = 0.$$

Hence $\lim_{x\to\infty}((\ln(x))/x^2) = 0$. This says that $\ln(x)$ grows much more slowly than x^2 does as x gets large. ◇

EXAMPLE 4 Find

$$\lim_{x\to\infty} \frac{x - \cos(x)}{x}. \tag{5.6.1}$$

SOLUTION Both numerator and denominator approach ∞ and $x \to \infty$. Trying l'Hôpital's Rule, we obtain

$$\lim_{x\to\infty} \frac{x - \cos(x)}{x} \overset{\text{l'H}}{=} \lim_{x\to\infty} \frac{(x - \cos(x))'}{x'} = \lim_{x\to\infty} \frac{1 + \sin(x)}{1}.$$

L'Hôpital's Rule may fail to provide an answer.

But $\lim_{x\to\infty}(1+\sin(x))$ does not exist, since $\sin(x)$ oscillates back and forth from -1 to 1 as $x \to \infty$

What can we conclude about the limit in (5.6.1)? Nothing at all. L'Hôpital's Rule says that if $\lim_{x\to\infty} f'(x)/g'(x)$ exists, then $\lim_{x\to\infty} f(x)/g(x)$ exists and has the same value. It say nothing about the case when $\lim_{x\to\infty} f'(x)/g'(x)$ does not exist.

It is not difficult to evaluate (5.6.1) directly, as follows:

$$\begin{aligned}
\lim_{x\to\infty} \frac{x - \cos(x)}{x} &= \lim_{x\to\infty} \left(1 - \frac{\cos(x)}{x}\right) && \text{algebra} \\
&= 1 - 0 && \text{since } |\cos(x)| \le 1 \\
&= 1.
\end{aligned}$$

◇

Moral: Look carefully at a limit before you decide to use l'Hôpital's Rule.

Two cars can help make Theorem 5.6.2 plausible. Imagine that $f(t)$ and $g(t)$ describe the locations on the x-axis of two cars at time t. Call the cars the f-car and the g-car. See Figure 5.6.1. Their velocities are therefore $f'(t)$ and $g'(t)$. These two cars are on endless journeys. But assume that as time $t \to \infty$ the f-car tends to travel at a speed closer and closer to L times the speed of the g-car. That is, assume that

$$\lim_{t\to\infty} \frac{f'(t)}{g'(t)} = L.$$

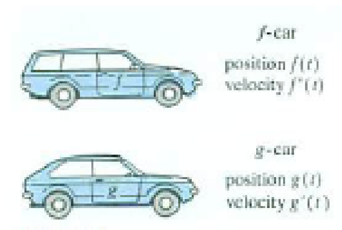


Figure 5.6.1 ARTIST: Update car profiles.

No matter how the two cars move in the short run, it seems reasonable that in the long run the f-car will tend to travel about L times as far as the g-car; that is,

$$\lim_{t\to\infty} \frac{f(t)}{g(t)} = L.$$

Transforming Limits So You Can Use l'Hôpital's Rule

Many limits can be transformed to limits to which l'Hôpital's Rule applies. For instance, the problem of finding

"zero-times-infinity" is indeterminate

$$\lim_{x\to 0^+} x\ln(x)$$

does not fit into l'Hôpital's Rule, since it does not involve the quotient of two functions. As $x \to 0^+$, one factor, x, approaches 0 and the other factor $\ln(x)$, approaches $-\infty$. So this is another type of indeterminate limit, involving a small number times a large number ("zero-times-infinity"). It is not obvious how this product, $x\ln(x)$, behaves as $x \to 0^+$. (Such a limit can turn out to be "zero, medium, large, or infinite"). A little algebra transforms the zero-times-infinity case into a problem to which l'Hôpital's Rule applies, as the next example illustrates.

EXAMPLE 5 Find $\lim_{x\to 0^+} x\ln(x)$.
SOLUTION Rewrite $x\ln(x)$ as a quotient, $\frac{\ln(x)}{(1/x)}$. Note that

$$\lim_{x\to 0^+} \ln(x) = -\infty \text{ and } \lim_{x\to 0^+} \frac{1}{x} = \infty.$$

By l'Hôpital's Rule,

$$\lim_{x\to 0^+} \frac{\ln(x)}{1/x} \overset{\text{l'H}}{=} \lim_{x\to 0^+} \frac{1/x}{-1/x^2} = \lim_{x\to 0^+}(-x) = 0.$$

Thus

$$\lim_{x\to 0^+} \frac{\ln(x)}{1/x} = 0,$$

The factor x, which approaches 0, dominates the factor $\ln(x)$ which "slowly grows towards $-\infty$."

from which it follows that $\lim_{x\to 0^+} x\ln(x) = 0$. ◇

The final example illustrates another type of limit that can be found by first relating it to limits to which l'Hôpital's Rule applies.

Try this on your calculator first.

EXAMPLE 6 $\lim_{x\to 0^+} x^x$.

SOLUTION Since this limit involves an exponential, not a quotient, it does not fit directly into l'Hôpital's Rule. But a little algebra changes the problem to one covered by l'Hôpital's Rule.

Let	y	$=$	x^x.
Then	$\ln(y)$	$=$	$\ln(x^x) = x\ln(x)$
By Example 5,	$\lim_{x\to 0^+} x\ln(x)$	$=$	0.

Therefore, $\lim_{x\to 0^+} \ln(y) = 0$. By the definition of $\ln(y)$ and the continuity of $e^x = \exp(x)$,

$$\lim_{x\to 0^+} y = \lim_{x\to 0^+} \exp(\ln(y)) = \exp\left(\lim_{x\to 0^+} (\ln(y))\right) = e^0 = 1.$$

Hence $x^x \to 1$ as $x \to 0^+$. ◇

Concerning the Proof

A complete proof of Theorem 5.6.1 may be found in Exercises 71 to 73. The following argument is intended to make the theorem plausible. To do so, consider the *special case* where f, f', g, and g' are all continuous throughout an open interval containing a — in particular, all four functions are defined at a. Assume that $g'(x) \neq 0$ throughout the interval. Since we have $\lim_{x\to a} f(x) = 0$ and $\lim_{x\to a} g(x) = 0$, it follows by continuity that $f(a) = 0$ and $g(a) = 0$.

Assume that $\lim_{x\to a} \frac{f'(x)}{g'(x)} = L$. Then

$$\begin{aligned}
\lim_{x\to a} \frac{f(x)}{g(x)} &= \lim_{x\to a} \frac{f(x)-f(a)}{g(x)-g(a)} && \text{since } f(a)=0 \text{ and } g(a)=0 \\
&= \lim_{x\to a} \frac{\frac{f(x)-f(a)}{x-a}}{\frac{g(x)-g(a)}{x-a}} && \text{algebra} \\
&= \frac{\lim_{x\to a} \frac{f(x)-f(a)}{x-a}}{\lim_{x\to a} \frac{g(x)-g(a)}{x-a}} && \text{limit of quotient} = \text{quotient of limits} \\
&= \frac{f'(a)}{g'(a)} && \text{definitions of } f'(a) \text{ and } g'(a) \\
&= \frac{\lim_{x\to a} f'(x)}{\lim_{x\to a} g'(x)} && f',\ g' \text{ continuous, by assumption} \\
&= \lim_{x\to a} \frac{f'(x)}{g'(x)} && \text{quotient of limits} = \text{limit of quotients} \\
&= L && \text{by assumption.}
\end{aligned}$$

Consequently,

$$\lim_{x\to a} \frac{f(x)}{g(x)} = L.$$

Summary

Indeterminate Forms	Name	Conversion Method	New Form
$f(x)g(x)$; $f(x) \to 0$, $g(x) \to \infty$	Zero-times-infinity $(0 \cdot \infty)$	Write as $\frac{f(x)}{1/g(x)}$ or $\frac{g(x)}{1/f(x)}$	$\frac{0}{0}$ or $\frac{\infty}{\infty}$
$f(x)^{g(x)}$; $f(x) \to 1$, $g(x) \to \infty$	One-to-infinity (1^∞)	Let $y = f(x)^{g(x)}$; take $\ln(y)$, find limit of $\ln(y)$, and then find limit of $y = e^{\ln(y)}$	$\ln(y)$ has form $\infty \cdot 0$
$f(x)^{g(x)}$; $f(x) \to 0$, $g(x) \to 0$	Zero-to-zero (0^0)	Same as for 1^∞	$\ln(y)$ has form $0 \cdot \infty$.

Table 5.6.1

We described l'Hôpital's Rule, which is a technique for dealing with limits of the indeterminate form "zero-over-zero" and "infinity-over-infinity" . In both of these cases

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)},$$

if the latter limit exists. Note that it concerns the quotient of two derivatives, *not* the derivative of the quotient.

Table 5.6.1 shows how some limits of other indeterminate forms can be converted into either of these two forms.

L'Hôpital's rule comes in handy during our study of a uniform sprinkler in the Calculus is Everywhere section at the end of this chapter.

EXERCISES for Section 5.6

In Exercises 1 to 16 check that l'Hôpital's Rule applies and use it to find the limits. *Identify all uses of l'Hôpital's Rule, including the type of indeterminant form.*

1. $\lim\limits_{x\to 2} \dfrac{x^3-8}{x^2-4}$

2. $\lim\limits_{x\to 1} \dfrac{x^7-1}{x^3-1}$

3. $\lim\limits_{x\to 0} \dfrac{\sin(3x)}{\sin(2x)}$

4. $\lim\limits_{x\to 0} \dfrac{\sin(x^2)}{(\sin(x))^2}$

5. $\lim\limits_{x\to 0} \dfrac{\sin(5x)\cos(3x)}{x}$

6. $\lim\limits_{x\to 0} \dfrac{\sin(5x)\cos(3x)}{x-\frac{\pi}{2}}$

7. $\lim\limits_{x\to \frac{\pi}{2}} \dfrac{\sin(5x)\cos(3x)}{x}$

8. $\lim\limits_{x\to \frac{\pi}{2}} \dfrac{\sin(5x)\cos(3x)}{x-\frac{\pi}{2}}$

9. $\lim\limits_{x\to \infty} \dfrac{x^3}{e^x}$

10. $\lim\limits_{x\to \infty} \dfrac{x^5}{3^x}$

11. $\lim\limits_{x\to 0} \dfrac{1-\cos(x)}{x^2}$

12. $\lim\limits_{x\to 0} \dfrac{\sin(x)-x}{(\sin(x))^3}$

13. $\lim\limits_{x\to 0} \dfrac{\tan(3x)}{\ln(1+x)}$

14. $\lim\limits_{x\to 1} \dfrac{\cos(\pi x/2)}{\ln(x)}$

15. $\lim\limits_{x\to 2} \dfrac{(\ln(x))^2}{x}$

16. $\lim\limits_{x\to 0} \dfrac{\arcsin(x)}{e^{2x}-1}$

In each of Exercises 17 to 22 transform the problem into one to which l'Hôpital's Rule applies; then find the limit. *Identify all uses of l'Hôpital's Rule, including the type of indeterminant form.*

17. $\lim\limits_{x\to 0} (1-2x)^{1/x}$

18. $\lim\limits_{x\to 0} (1+\sin(2x))^{\csc(x)}$

19. $\lim_{x\to 0^+} (\sin(x))^{(e^x-1)}$

20. $\lim_{x\to 0^+} x^2 \ln(x)$

21. $\lim_{x\to 0^+} (\tan(x))^{\tan(2x)}$

22. $\lim_{x\to 0^+} (e^x - 1)\ln(x)$

WARNING *(Do Not Overuse l'Hôpital's Rule)* Remember that l'Hôpital's Rule, carelessly applied, may give a wrong answer or no answer.

In Exercises 23 to 51 find the limits. Use l'Hôpital's Rule only if it applies. *Identify all uses of l'Hôpital's Rule, including the type of indeterminant form.*

23. $\lim_{x\to\infty} \frac{2^x}{3^x}$

24. $\lim_{x\to\infty} \frac{2^x + x}{3^x}$

25. $\lim_{x\to\infty} \frac{\log_2(x)}{\log_3(x)}$

26. $\lim_{x\to 1} \frac{\log_2(x)}{\log_3(x)}$

27. $\lim_{x\to\infty} \left(\frac{1}{x} - \frac{1}{\sin(x)}\right)$

28. $\lim_{x\to\infty} \left(\sqrt{x^2+3} - \sqrt{x^2+4x}\right)$

29. $\lim_{x\to\infty} \frac{x^2 + 3\cos(5x)}{x^2 - 2\sin(4x)}$

30. $\lim_{x\to\infty} \frac{e^x - 1/x}{e^x - 1/x}$

31. $\lim_{x\to 0} \frac{3x^3 + x^2 - x}{5x^3 + x^2 + x}$

32. $\lim_{x\to\infty} \frac{3x^3 + x^2 - x}{5x^3 + x^2 + x}$

33. $\lim_{x\to\infty} \frac{\sin(x)}{4 + \sin(x)}$

34. $\lim_{x\to\infty} x\sin(3x)$

35. $\lim_{x \to 1^+} (x-1)\ln(x-1)$

36. $\lim_{x \to \pi/2} \frac{\tan(x)}{x - (\pi/2)}$

37. $\lim_{x \to 0} (\cos(x))^{1/x}$

38. $\lim_{x \to 0^+} x^{1/x}$

39. $\lim_{x \to 0} (1+x)^{1/x}$

40. $\lim_{x \to 0} \left(1+x^2\right)^x$

41. $\lim_{x \to 1} \frac{x^2-1}{x^3-1}$

42. $\lim_{x \to 0} \frac{xe^x(1+x)^3}{e^x-1}$

43. $\lim_{x \to 0} \frac{xe^x \cos^2(6x)}{e^{2x}-1}$

44. $\lim_{x \to 0} (\csc(x) - \cot(x))$

45. $\lim_{x \to 0} \frac{\csc(x) - \cot(x)}{\sin(x)}$

46. $\lim_{x \to 0} \frac{5^x - 3^x}{\sin(x)}$

47. $\lim_{x \to 0} \frac{(\tan(x))^5 - (\tan(x))^3}{1 - \cos(x)}$

48. $\lim_{x \to 2} \frac{x^3+8}{x^2+5}$

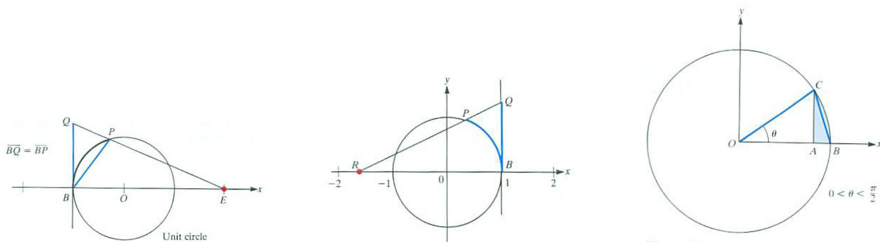
49. $\lim_{x \to \pi/4} \frac{\sin(5x)}{\sin(3x)}$

50. $\lim_{x \to 0} \left(\frac{1}{1-\cos(x)} - \frac{2}{x^2}\right)$

51. $\lim_{x \to 0} \frac{\arcsin(x)}{\arctan(2x)}$

52. In Figure 5.6.2(a) the unit circle is centered at O, BQ is a vertical tangent line, and the length of BP is the same as the length of BQ. What happens to the point E as $Q \to B$?

53. In Figure 5.6.2(b) the unit circle is centered at the origin, BQ is a vertical tangent line, and the length of BQ is the same as the arc length $\widehat{BP}$. Show that the x-coordinate of R approaches -2 as $P \to B$.



(a) (b) (c)

Figure 5.6.2

54. Exercise 44 of Section 2.2 asked you to guess a certain limit. Now that limit will be computed.

> **WARNING** *(Common Sense)* As Albert Einstein observed, "Common sense is the deposit of prejudice laid down in the mind before the age of 18."

In Figure 5.6.2(c), which shows a circle, let $f(\theta)$ be the area of triangle ABC and $g(\theta)$ be the area of the shaded region formed by deleting triangle OAC from sector OBC.

(a) Why is $f(\theta)$ smaller than $g(\theta)$?

(b) What would you guess is the value of $\lim_{\theta\to 0} f(\theta)/g(\theta)$?

(c) Find $\lim_{\theta\to 0} f(\theta)/g(\theta)$.

55. The following argument appears in an economics text:
"Consider the production function

$$y = k\left(\alpha x_1^{-\rho} + (1-\alpha)x_2^{-\rho}\right)^{-1/\rho},$$

where k, α, x_1, and x_2 are positive constants and $\alpha < 1$. Taking the limit as $\rho \to 0^+$, we find that

$$\lim_{\rho\to 0^+} y = k x_1^{\alpha} x_2^{1-\alpha},$$

which is the Cobb-Douglas function, as expected."
Fill in the details.

56. Sam proposes the following proof for Theorem 5.6.1: "Since

$$\lim_{x\to a^+} f(x) = 0 \qquad \text{and} \qquad \lim_{x\to a+} g(x) = 0,$$

I will define $f(a) = 0$ and $g(a) = 0$. Next I consider $x > a$ but near a. I now have continuous functions f and g defined on the closed interval $[a, x]$ and differentiable on the open interval (a, x). So, using the Mean-Value Theorem, I conclude that there is a number c, $a < c < x$, such that

$$\frac{f(x) - f(a)}{x - a} = f'(c) \qquad \text{and} \qquad \frac{g(x) - g(a)}{x - a} = g'(c).$$

Since $f(a) = 0$ and $g(a) = 0$, these equations tell me that

$$f(x) = (x - a)f'(c) \qquad \text{and} \qquad g(x) = (x - a)g'(c)$$

$$\begin{array}{lrcl} \text{Thus} & \frac{f(x)}{g(x)} & = & \frac{f'(c)}{g'(c)} \\ \text{Hence} & \lim\limits_{x\to a^+} \dfrac{f(x)}{g(x)} & = & \lim\limits_{x\to a^+} \dfrac{f'(c)}{g'(c)}. \end{array}$$

Sam made one error. What is it?

57. Find $\lim_{x\to 0} \left(\frac{1+2^x}{x}\right)^{1/x}$.

58. R. P. Feynman, in *Lectures in Physics*, wrote: "Here is the quantitative answer of what is right instead of kT. This expression

$$\frac{\hbar\omega}{e^{\hbar\omega/kT} - 1}$$

should, of course, approach kT as $\omega \to 0$. .... See if you can prove that it does — learn how to do the mathematics."
Do the mathematics. All symbols, except ω, denote constants.

59. Graph $y = x^x$ for $0 < x \leq 1$, showing its minimum point.

In Exercises 60 to 62 graph the specified function, being sure to show (a) where the function is increasing and decreasing, (b) where the function has any asymptotes, and (c) how the function behaves for x near 0.
60. $f(x) = (1 + x)^{1/x}$ for $x > -1$, $x \neq 0$

61. $y = x\ln(x)$

62. $y = x^2\ln(x)$

63. In which cases below is it possible to determine $\lim_{x\to a} f(x)^{g(x)}$ without further information about the functions?

(a) $\lim_{x\to a} f(x) = 0$; $\lim_{x\to a} g(x) = 7$

(b) $\lim_{x\to a} f(x) = 2$; $\lim_{x\to a} g(x) = 0$

(c) $\lim_{x\to a} f(x) = 0$; $\lim_{x\to a} g(x) = 0$

(d) $\lim_{x\to a} f(x) = 0$; $\lim_{x\to a} g(x) = \infty$

(e) $\lim_{x\to a} f(x) = \infty$; $\lim_{x\to a} g(x) = 0$

(f) $\lim_{x\to a} f(x) = \infty$; $\lim_{x\to a} g(x) = -\infty$

64. In which cases below is it possible to determine $\lim_{x\to a} f(x)/g(x)$ without further information about the functions?

(a) $\lim_{x\to a} f(x) = 0$; $\lim_{x\to a} g(x) = \infty$

(b) $\lim_{x\to a} f(x) = 0$; $\lim_{x\to a} g(x) = 1$

(c) $\lim_{x\to a} f(x) = 0$; $\lim_{x\to a} g(x) = 0$

(d) $\lim_{x\to a} f(x) = \infty$; $\lim_{x\to a} g(x) = -\infty$

65. Sam is angry. "Now I know why calculus books are so long. They spend all of page 82 showing that $\lim_{x\to 0} \frac{\sin(x)}{x}$ is 1. They could have saved space (and me a lot of trouble) if they had just used l'Hôpital's approach."
Is Sam right, for once?

66. Jane says, "I can get $\lim_{x\to 0} \frac{e^x-1}{x}$ easily. It's just the derivative of e^x evaluated at 0. I don't need l'Hôpital's Rule." Is Jane right, or has Sam's influence affected her ability to reason?

67.

$$\begin{array}{lrcl} \text{If} & \lim_{t\to\infty} f(t) = & \infty & = \lim_{t\to\infty} g(t) \\ \text{and} & \lim_{t\to\infty} \frac{f(t)}{g(t)} & = & 3, \end{array}$$

what can be said about

$$\lim_{t\to\infty} \frac{\ln(f(t))}{\ln(g(t))}?$$

Do *not* assume f and g are differentiable.

68. Give an example of a pair of functions f and g such that we have $\lim_{x\to 0} f(x) = 1$, $\lim_{x\to 0} g(x) = \infty$, and $\lim_{x\to 0} f(x)^{g(x)} = 2$.

69. Obtain l'Hôpital's Rule for $\lim_{x\to\infty} \frac{f(x)}{g(x)}$ from the case $\lim_{t\to 0^+} \frac{f(t)}{g(t)}$. (Let $t = 1/x$.)

70. Find the limit of $(1^x + 2^x + 3^x)^{1/x}$ as

(a) $x \to 0$

(b) $x \to \infty$

(c) $x \to -\infty$.

The proof of Theorem 5.6.1, to be outlined in Exercise 73, depends on the following generalized mean-value theorem.

Generalized Mean-Value Theorem. Let f and g be two functions that are continuous on $[a, b]$ and differentiable on (a, b). Furthermore, assume that $g'(x)$ is never 0 for x in (a, b). Then there is a number c in (a, b) such that

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}.$$

The proof of this is in Exercise 72.

71. During a given time interval one car travels twice as far as another car. Use the Generalized Mean-Value Theorem to show that there is at least one instant when the first car is traveling exactly twice as fast as the second car.

72. To prove the Generalized Mean-Value Theorem, introduce a function h defined by

$$h(x) = f(x) - f(a) - \frac{f(b) - f(a)}{g(b) - g(a)}(g(x) - g(a)). \qquad (5.6.2)$$

Show that $h(b) = 0$ and $h(a) = 0$. Then apply Rolle's Theorem to h on (a, b). Rolle's Theorem is Theorem 4.1.2 in Section 4.1.

> **Remark:** The function h in (5.6.2) is similar to the function h used in the proof of the Mean-Value Theorem (Theorem 4.1.3 in Section 4.1). Check that $h(x)$ is the vertical distance between the point $(g(x), f(x))$ and the line through $(g(a), f(a))$ and $(g(b), f(b))$.

73. Assume the hypotheses of Theorem 5.6.1. Define $f(a) = 0$ and $g(a) = 0$, so that f and g are continuous at a. Note that

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{g(x) - g(a)},$$

and apply the Generalized Mean-Value Theorem from Exercise 71. This Exercise proves Theorem 5.6.1, l'Hôpital's Rule in the zero-over-zero case.

74.

$$\begin{array}{lrcl} \text{If} & \lim_{t\to\infty} f(t) = & \infty & = \lim_{t\to\infty} g(t) \\ \text{and} & \lim_{t\to\infty} \frac{\ln(f(t))}{\ln(g(t))} & = & 1, \\ \text{must} & \lim_{t\to\infty} \frac{f(t)}{g(t)} & = & 1? \end{array}$$

Explain.

75. Assume that f, f', and f'' are defined in $[-1,1]$ and are continuous. Also, $f(0) = 0$, $f'(0) = 0$, and $f''(0) > 0$.

(a) Sketch what the graph of f may look like for x in $[0, a]$, where a is a small positive number.

(b) Interpret the quotient

$$Q(a) = \frac{\int_0^a f(x)\ dx}{af(a) - \int_0^a f(x)\ dx}$$

in terms of the graph in (a).

(c) What do you think happens to $Q(a)$ as $a \to 0$?

(d) Find $\lim_{a\to 0} Q(a)$.

(Because f''' might not be continuous at 0, you need to use $\lim_{a\to 0} \frac{f'(a)}{a} = f''(0)$.)

76.

Sam: I bet I can find $\lim_{x\to 0} \frac{e^x - 1 - x - \frac{x^2}{2}}{x^3}$ by using the Maclaurin polynomial $P_2(x; 0)$ for e^x and paying attention to the error.

Is Sam right?

5.7 Natural Growth and Decay

In 2009 the population of the United States was about 306 million and growing at a rate of about 1% (roughly 3 million people) a year. The world population was about 6.79 billion and growing at a rate of about 1.5% (roughly 100 million people) a year. Both are examples of natural growth.

Natural Growth

If t denotes time in years and $P(t)$ is the US population, $k = 0.01$.

Let $P(t)$ be the size of a population at time t. If its rate of growth is proportional to its size, there is a positive constant k such that

$$\frac{dP(t)}{dt} = kP(t). \tag{5.7.1}$$

To find an explicit formula for $P(t)$ as a function of t, rewrite (5.7.1) as

$$\frac{\frac{dP(t)}{dt}}{P(t)} = k. \tag{5.7.2}$$

The left-hand side can be rewritten as the derivative of $\ln(P(t))$ and so (5.7.2) can be rewritten as

$$\frac{d(\ln(P(t))}{dt} = \frac{d(kt)}{dt}.$$

Therefore there is a constant C such that

$$\ln(P(t)) = kt + C. \tag{5.7.3}$$

From (5.7.3) it follows, by the definition of a logarithm, that

$$P(t) = e^{kt+C},$$

hence

$$P(t) = e^C e^{kt}.$$

Since C is a constant, so is e^C, which we call A. We have the following simple explicit formula for $P(t)$:

The equation for **natural growth** is

$$P(t) = Ae^{kt}$$

where k is a positive constant. Because $P(0) = Ae^{k(0)} = A$, the coefficient A is the **initial population**.

Because of the presence of the exponential e^{kt}, natural growth is also called **exponential growth**.

EXAMPLE 1 The size of the world population at the beginning of 1988 was approximately 5.14 billion. At the beginning of 1989 it was 5.23 billion. Assume that the growth rate remains constant.

(a) What is the growth constant k?

(b) What would the population be in 2009?

(c) When will the population double is size?

SOLUTION Let $P(t)$ be the population in billions at time t. For convenience, measure time starting in the year 1988; that is, $t = 0$ corresponds to 1988 and $t = 1$ to 1989. Thus $P(0) = 5.14$ and $P(1) = 5.23$. The natural growth equation describing the population in billions at time is

$$P(t) = 5.14e^{kt}. \tag{5.7.4}$$

(a) To find k, we note that

$$P(1) = 5.14e^{k\cdot 1},$$

so

$$\begin{aligned} 5.14e^k &= 5.23 \\ e^k &= \frac{5.23}{5.14} \\ k &= \ln\left(\frac{5.23}{5.14}\right) \approx 0.174. \end{aligned}$$

Hence (5.7.4) takes the form

$$P(t) = 5.14e^{0.174t}.$$

This equation is all that we need to answer the remaining questions.

(b) The year 2009 corresponds to $t = 21$, so in the year 2009 the population, in billions, would be

$$P(21) = 5.14e^{0.174\cdot 21} = 5.14e^{0.3654} \approx 5.14(1.441) \approx 7.41.$$

The population would be approximately 7.41 billion in 2009. (Recall from the introduction of this section that the actual estimate of the world population in 2009 is about 6.79 billion. This suggests that the actual growth rate has not been constant; it has increased during the past 21 years.)

(c) The population will double when it reaches $2(5.14) = 10.28$ billion. We need to solve for t in the equation $P(t) = 10.28$. We have

$$\begin{aligned} 5.14e^{kt} &= 10.28 \\ e^{kt} &= 2 \\ kt &= \ln(2) \\ t &= \frac{\ln(2)}{k} \approx \frac{0.6931}{0.0174} \approx 39.8360. \end{aligned}$$

The world population will double approximately 40 years after 1988, which corresponds to the year 2028.

◇

The time it takes for a population to double is called the **doubling time** and is denoted t_2. Exponential growth is often described by its doubling time t_2 rather than by its growth constant k. However, if you know either t_2 or k you can figure out the other, as they are related by the equation

$$t_2 = \frac{\ln(2)}{k}$$

which appeared during part (c) of the solution to Example 1.

Exponential growth may also be described in terms of an annual percentage increase, such as "The population is growing 6 percent per year." That is, each year the population is multiplied by the factor 1.06: $P(t+1) = P(t)(1.06)$.

On the other hand, from the exponential growth function, we see that

$$P(t+1) = P(0)e^{k(t+1)} = P(0)e^{kt}e^k = P(t)e^k.$$

That is, during each unit of time the population is "magnified" by a factor of e^k. Now, when k is small, $e^k \approx 1 + k$. Consequently we can approximate 6 percent annual growth by letting $k = 0.06$. This approximation is valid whenever the growth rate is only a few percent. Since population figures are themselves only an approximation, setting the growth constant k equal to the annual percentage rate is a reasonable tactic.

EXAMPLE 2 Find the doubling time if the growth rate is 2 percent per year.

SOLUTION The growth rate is 2 percent, so we set $k = 0.02$. Then

$$t_2 = \frac{\ln(2)}{k} \approx \frac{0.693}{0.02} = 34.65 \text{ years}.$$

◇

The Mathematics of Natural Decay

As Glen Seaberg observes in the conversation reported on page 433, some radioactive elements decay at a rate proportional to the amount present. The time it takes for half the initial amount to decay is denoted $t_{1/2}$ and is called the element's **half-life**.

Similarly, in medicine one speaks of the half-life of a drug administered to a patient: the time required for half the drug to be removed from the body. This half-life depends both on the drug and the patient, and can be from 20 minutes for penicillin to 2 weeks for quinacrine, an antimalarial drug. This half-life is critical to determining how frequently a drug can be administered. Some elderly patients have died from overdoses before it was realized that the half-life of some drugs is longer in the elderly than in the young.

Letting $P(t)$ again represent the amount present at time t, we have

Now k is negative.

$$P'(t) = kP(t) \qquad k < 0$$

where k is the **decay constant**. This is the same equation as (5.7.1), so

$$P(t) = P(0)e^{kt},$$

as before, except now k is a *negative number*. Since k is negative, the factor e^{kt} is a decreasing function of t.

Just as the doubling time is related to (positive) k by the equation $t_2 = (\ln(2))/k$, the half-life is related to (negative) k by the equation $t_{1/2} = (\ln(1/2))/k$, which can be rewritten as $t_{1/2} = -(\ln(2))/k$.

EXAMPLE 3 The Chernobyl nuclear reactor accident, in April 1986, released radioactive cesium 137 into the air. The half-life of ^{137}Cs is 27.9 years.

(a) Find the decay constant k of ^{137}Cs.

(b) When will only one-fourth of an initial amount remain?

(c) When will only 20 percent of an initial amount remain?

SOLUTION

(a) The formula for the half-life can be solved for k to give:

$$k = \frac{-\ln(2)}{t_{1/2}} \approx \frac{-0.693}{27.9} \approx -0.0248.$$

(b) This can be done without the aid of any formulas. Since $\frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2}$, in two half-lives only one-quarter of an initial amount remains. The answer is $2(27.9) = 55.8$ years.

(c) We want to find t such that only 20 percent remains. While we know the answer is greater than 55.8 years (since 20% is less than 25%), finding the exact time requires using the formula for $P(t)$.

We want
$$P(t) = 0.20P(0).$$
That is, we want to solve
$$\begin{array}{rrcl} & P(0)e^{kt} & = & 0.20P(0). \\ \text{Then} & e^{kt} & = & 0.20 \\ & kt & = & \ln(0.20) \\ & t & = & \dfrac{\ln(0.20)}{k}. \end{array}$$
Since $k \approx -0.0248$, this gives
$$t \approx \frac{-1.609}{-0.0248} = 64.9 \text{ years.}$$
After 64.9 years (that is, in 2051) only 20% of the original amount remains.

◇

Summary

We developed the mathematics of growth or decay that is proportional to the amount present. This required solving the differential equation
$$\frac{dP}{dt} = kP$$
where k is a constant, positive in the case of growth and negative in the case of decay. The solution is
$$P(t) = Ae^{kt}$$
where A is $P(0)$, the amount of the substance present when $t = 0$.

In the case of growth, the time for the quantity to double (the "doubling time") is denoted t_2. In the case of decay, the time when only half the original amount survives is denoted $t_{1/2}$, the "half-life." One has
$$t_2 = \frac{\ln(2)}{k} \qquad \text{and} \qquad t_{1/2} = \frac{\ln(1/2)}{k} = -\frac{\ln(2)}{k}.$$

The Scientist, The Senator, and Half-Life

During the hearings in 1963 before the Senate Foreign Relations Committee on the nuclear test ban treaty, this exchange took place between Glen Solborg, winner of the Nobel prize for chemistry in 1951, and Senator James W. Fulbright.

Seaborg: Tritium is used in a weapon, and it decays with a half-life of about 12 years. But the plutonium and uranium have such long half-lives that there is no detectable change in a human lifetime.

Fulbright: I am sure this seems to be a very naive question, but why do you refer to half-life rather than whole life? Why do you measure by half-lives?

Seaborg: Here is something that I could go into a very long discussion on.

Fulbright: I probably wouldn't benefit adequately from a long discussion. It seems rather odd that you should call it a half-life rather than its whole life.

Seaborg: Well, I will try. If we have, let us say, one million atoms of a material like tritium, in 12 years half of those will be transformed into a decay product and you will have 500,000 atoms.

Then, in another 12 years, half of what remains transforms, so you have 250,000 atoms left. And so forth.

On that basis it never all decays, because half is always left, but of course you finally get down to where your last atom is gone.

EXERCISES for Section 5.7

1.

(a) Show that exponential growth can be expressed as $P = Ab^t$ for some constants A and b.

(b) What can be said about b?

2.

(a) Show that exponential decay can be expressed as $P = Ab^t$ for some constants A and b.

(b) What can be said about b?

3. If $P(t) = 30e^{0.2t}$ what are the initial size and the doubling time?

4. If $P(t) = 30e^{-0.2t}$ what are the initial size and the half life?

5. What is the doubling time for a population always growing at 1% a year?

6. What is the half life for a population always shrinking at 1% a year?

7. A quantity is increasing according to the law of natural growth. The amount present at time $t = 0$ is A. It will double when $t = 10$.

(a) Express the amount at time t in the form Ae^{kt} for a suitable k.

(b) Express the amount at time t in the form Ab^t for a suitable b.

8. The mass of a certain bacterial culture after t hours is $10 \cdot 3^t$ grams.

(a) What is the initial amount?

(b) What is the growth constant k?

(c) What is the percent increase in any period of 1 hour?

9. Let $f(t) = 3 \cdot 2^t$.

(a) Solve the equation $f(t) = 12$.

(b) Solve the equation $f(t) = 5$.

(c) Find k such that $f(t) = 3e^{kt}$.

10. In 1988 the world population was about 5.1 billion and was increasing at the rate of 1.7 percent per year. If it continues to grow at that rate, when will it (a) double? (b) quadruple? (c) reach 100 billion?

11. The population of Latin America has a doubling time of 27 years. Estimate the percent it grows per year.

12. At 1:00 P.M. a bacterial culture weighed 100 grams. At 4:30 P.M. it weighed 250 grams. Assuming that it grows at a rate proportional to the amount present, find (a) at what time it will grow to 400 grams, (b) its growth constant.

13. A bacterial culture grows from 100 to 400 grams in 10 hours according to the law of natural growth.

(a) How much was present after 3 hours?

(b) How long will it take the mass to double? quadruple? triple?

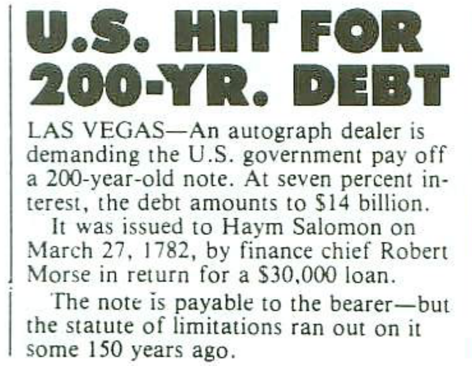
14. A radioactive substance disintegrates at the rate of 0.05 grams per day when its mass is 10 grams.

(a) How much of the substance will remain after t days if the initial amount is A?

(b) What is its half-life?

15. In 2009 the population of Mexico was 111 million and of the United States 308 million. If the population of Mexico increases at 1.15% per year and the population of the United States at 1.0% per year, when would the two nations have the same size population?

16. The size of the population in India was 689 million in 1980 and 1,027 million in 2007. What is its doubling time t_2?



U.S. HIT FOR 200-YR. DEBT

LAS VEGAS—An autograph dealer is demanding the U.S. government pay off a 200-year-old note. At seven percent interest, the debt amounts to $14 billion.

It was issued to Haym Salomon on March 27, 1782, by finance chief Robert Morse in return for a $30,000 loan.

The note is payable to the bearer—but the statute of limitations ran out on it some 150 years ago.

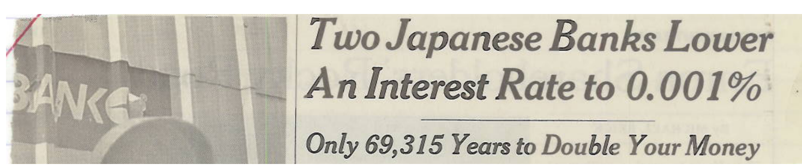
Figure 5.7.1

17. The newspaper article shown in Figure 5.7.1 illustrates the rapidity of exponential growth.

(a) Is the figure of $14 billion correct? Assume that the interest is compounded annually.

(b) What interest rate would be required to produce an account of $14 billion if interest were compounded once a year?

(c) Answer (b) for "continuous compounding," which is another term for natural growth (a bank account increases at a rate proportional to the amount in the account at any instant).



Two Japanese Banks Lower An Interest Rate to 0.001%

Only 69,315 Years to Double Your Money

Figure 5.7.2

18. The headline shown in Figure 5.7.2 appeared in 2002. Is the number 69,315 correct? Explain.

19. Carbon 14 (chemical symbol ^{14}C), an isotope of carbon, is radioactive and has a half-life of approximately 5,730 years. If the ^{14}C concentration in a piece of wood of unknown age is half of the concentration in a present-day live specimen, then it is about 5,730 years old. (This assumes that ^{14}C concentrations in living objects remain about the same.) This gives a way of estimating the age of an undated specimen. Show that if A_C is the concentration of ^{14}C in a live (contemporary) specimen and A_u is the concentration of ${}^{14}C$ in a specimen of unknown age, then the age of the undated material is about $8,300\ln(A_C/A_u)$ years. This method, called **radiocarbon dating** is reliable up to about 70,000 years.

20. From a letter to an editor in a newspaper:

> **I've been hearing bankers and investment advisers talk about something called the "rule of 72." Could you explain what it means?**
>
> How quickly would you like to double your money? That's what the "rule of 72" will tell you. To find out how fast your money will double at any given interest rate or yield, simply divide that yield into 72. This will tell you how many years doubling will take.
>
> Let's say you have a long-term certificate of deposit paying 12 percent [annually]. At that rate your money would double in six years. A money-market fund paying 10 percent would take 7.2 years to double your investment.

(a) Explain the rule of 72 and what number should be used instead of 72.

(b) Why do you think 72 is used?

21. Benjamin Franklin conjectured that the population of the United States would double every 20 years, beginning in 1751, when the population was 1.3 million.

(a) If Franklin's conjecture were right, what would the population of the United States be in 2010?

(b) In 2010 the population was 310 million. Assuming natural growth, what would the doubling time be?

22. (Doomsday equation) A differential equation of the form $dP/dt = kP^{1.01}$ is called a **doomsday equation**. The rate of growth is just slightly higher than that for natural growth. Solve the differential equation to find $P(t)$. How does $P(t)$ behave as t increases? Does $P(t)$ increase forever?

23. The following situations are all mathematically the same:

1. A drug is administered in a dose of A grams to a patient and gradually leaves the system through excretion.
2. Initially there is an amount A of smoke in a room. The air conditioner is turned on and gradually the smoke is removed.

3. Initially there is an amount A of some pollutant in a lake, when further dumping of toxic materials is prohibited. The rate at which water enters the lake equals the rate at which it leaves. (Assume the pollution is thoroughly mixed.)

In each case, let $P(t)$ be the amount present at time t (whether drug, smoke, or pollution).

(a) Why is it reasonable to assume that there is a constant k such that for small intervals of time, Δt, $\Delta P \approx kP(t)\Delta t$?

(b) From (a) deduce that $P(t) = Ae^{kt}$.

(c) Is k positive or negative?

24. **Newton's law of cooling** assumes that an object cools at a rate proportional to the difference between its temperature and the room temperature. Denote the room temperature as A. The differential equation for Newton's law of cooling is $dy/dt = k(y - A)$ where k and A are constants.

(a) Explain why k is negative.

(b) Draw the slope field for this differential equation when $k = -1/2$.

(c) Use (b) to conjecture the behavior of $y(t)$ as $t \to \infty$.

(d) Solve for y as a function of t.

(e) Draw the graph of $y(t)$ on the slope field produced in (b).

(f) Find $\lim_{t\to\infty} y(t)$.

25. Let $I(x)$ be the intensity of sunlight at a depth of x meters in the ocean. As x increases, $I(x)$ decreases.

(a) Why is it reasonable to assume that there is a constant k (negative) such that $\Delta I \approx kI(x)\Delta x$ for small Δx?

(b) Deduce that $I(x) = I(0)e^{kx}$, where $I(0)$ is the intensity of sunlight at the surface. Incidentally, sunlight at a depth of 1 meter is only one-fourth as intense as at the surface.

26. A particle moving through a liquid meets a "drag" force proportional to the velocity; that is, its acceleration is proportional to its velocity. Let x denote its position and v its velocity at time t. Assume $v > 0$.

(a) Show that there is a positive constant k such that $dv/dt = -kv$.

(b) Show that there is a constant A such that $v = Ae^{-kt}$.

(c) Show that there is a constant B such that $x = -\frac{1}{k}Ae^{-kt} + B$.

(d) How far does the particle travel as t goes from 0 to ∞? (Is this a finite or infinite distance?)

27.

(a) Show that the natural growth function $P(t) = Ae^{kt}$ can be written in terms of A and t_2 as $P(t) = A \cdot 2^{t/t_2}$.

(b) Check that the function found in (a) is correct when $t = 0$ and $t = t_2$.

28.

(a) Express the natural decay function $P(t) = Ae^{kt}$ in terms of A and $t_{1/2}$.

(b) Check that the function found in (a) is correct when $t = 0$ and $t = t_{1/2}$.

29. A population is growing exponentially. Initially, at time 0, it is P_0. Later, at time u it is P_u.

(a) Show that at time t it is $P_0(P_u/P_0)^{t/u}$.

(b) Check that the formula in (a) gives the correct population when $t = 0$ and $t = u$.

30. Let $P(t) = Ae^{kt}$. Then $\frac{P(t+1)-P(t)}{P(t)} = e^k - 1$. Show that when k is small, $e^k - 1 \approx k$. That means the relative change in one unit of time is approximately k.

31. A certain fish population increases in number at a rate proportional to the size of the population. In addition, it is being harvested at a constant rate. Let $P(t)$ be the size of the fish population at time t.

(a) Show that there are positive constants h and k such that for small Δt, $\Delta P \approx kP\Delta t - h\Delta t$.

(b) Find a formula for $P(t)$ in terms of $P(0)$, h, and k. (First divide by Δt in (a) and then take limits as $\Delta t \to 0$.)

(c) Describe the behavior of $P(t)$ in the three cases $h = kP(0)$, $h > kP(0)$, and $h < kP(0)$

32. The half-life of a drug administered to a certain patient is 8 hours. It is given in a 1-gram dose every 8 hours.

(a) How much is there in the patient just after the second dose is administered?

(b) How much is there in the patient just after the third dose? The fourth dose?

(c) Let $P(t)$ be the amount in the patient at t hours after the first dose. Graph $P(t)$ for a period of 48 hours. $P(t)$ has meaning for all values of t, not just at the integers.

(d) Does the amount in the patient get arbitrarily large as time goes on?

33. The half-life of the drug in Exercise 32 is 16 hours when administered to a different patient. Answer, for this patient, the questions in Exercise 32.

34. The half-life of a drug in a certain patient is $t_{1/2}$ hours. It is administered every h hours. Can it happen that the concentration of the drug gets arbitrarily high? Explain your answer.

Exercises 35 to 37 introduce and analyze the **inhibited or logistic growth** model. This model will be encountered in CIE 13 about petroleum at the end of Chapter 10.

35. In many cases of growth there is obviously a finite upper bound M which the population cannot exceed. Why is it reasonable to assume (or to take as a model) that

$$\frac{dP}{dt} = kP(t)(M - P(t)) \qquad 0 < P(t) < M \tag{5.7.5}$$

for some constant k?

36.

(a) Solve the differential equation in Exercise 35. (You will need the partial fraction identity

$$\frac{1}{P(M-P)} = \frac{1}{M}\left(\frac{1}{P} + \frac{1}{M-P}\right)$$

and the property of logarithms: $\ln(A) - \ln(B) = \ln\left(\frac{A}{B}\right)$.) After simplification, your answer should have the form

$$P(t) = \frac{M}{1 + ae^{-Mkt}}$$

for a suitable constant a.

(b) Find $\lim_{t\to\infty} P(t)$. Is this reasonable?

(c) Express a in terms of $P(0)$, M, and k.

37. By considering (5.7.5) in Exercise 35 directly (not the explicit formula in Exercise 36), show that

(a) P is an increasing function.

(b) The maximum rate of change of P occurs when $P(t) = M/2$.

(c) The graph of $P(t)$ has an inflection point.

38. A salesman, trying to persuade a tycoon to invest in Standard Coagulated Mutual Fund, shows him the accompanying graph which records the value of a similar investment made in the fund in 1965. "Look! In the first 5 years the investment increased \$1,000," the salesman observed, "but in the past 5 years it increased by \$2,000. It's really improving. Look at the graph of the graph from 1985 to 1990, which you can see clearly in Figure 5.7.3."

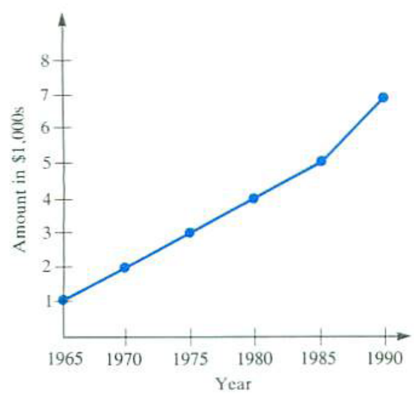


Figure 5.7.3

The tycoon replied, "Hogwash. Though your graph is steeper from 1985 to 1990, in fact, the rate of return is less than from 1965 to 1970. Indeed, that was your best period."

(a) If the percentage return on the accumulated investment remains the same over each 5-year period as the first 5-year period, sketch the graph.

(b) Explain the tycoon's reasoning.

39. Each of two countries is growing exponentially but at different rates. One is describe by the function $A_1e^{k_1t}$, the other by $A_2e^{k_2t}$, and k_1 is not equal to k_2. Is their total population growing exponentially? That is, are there constants A and k such that the formula describing their total population has the form Ae^{kt}. Explain your answer.

40. Assume c_1, c_2, and c_3 are distinct constants. Can there be constants A_1, A_2, and A_3, not all 0, such that $A_1e^{c_1x} + A_2e^{c_2x} + A_3e^{c_3x} = 0$ for all x?

41. If each of two functions describes natural growth does their (a) product? (b) quotient? (c) sum?

5.8 The Hyperbolic Functions and Their Inverses

Certain combinations of the exponential functions e^x and e^{-x} occur often enough in differential equations and engineering — for instance, in the study of the shape of electrical transmission or suspension cables — to be given names. This section defines these **hyperbolic functions** and obtains their basic properties. Since the letter x will be needed later for another purpose, we will use the letter t when writing the two preceding exponentials, namely, e^t and e^{-t}.

The Hyperbolic Functions

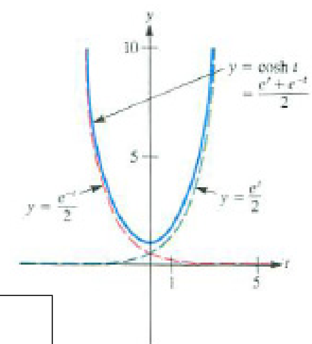


Figure 5.8.1

DEFINITION *(The hyperbolic cosine.)* Let t be a real number. The **hyperbolic cosine** of t, denoted $\cosh(t)$, is given by the formula

$$\cosh(t) = \frac{e^t + e^{-t}}{2}.$$

To graph $\cosh(t)$, note first that

$$\cosh(-t) = \frac{e^{-t} + e^{-(-t)}}{2} = \frac{e^t + e^{-t}}{2} = \cosh(t).$$

Since $\cosh(-t) = \cosh(t)$, the cosh function is even, and so its graph is symmetric with respect to the vertical axis. Furthermore, $\cosh(t)$ is the sum of two terms

$$\cosh(t) = \frac{e^t}{2} + \frac{e^{-t}}{2}.$$

Pronounced as written, "cosh," rhyming with "gosh."

As $t \to \infty$, the second term, $e^{-t}/2$, is positive and approaches 0. Thus, for $t > 0$ and large, the graph of $\cosh(t)$ is just a little above the graph of $e^t/2$. This information, together with the fact that $\cosh(0) = (e^0 + e^{-0})/2 = 1$, is the basis for Figure 5.8.1.

For $|t| \to \infty$, the graph of $y = \cosh(t)$ is asymptotic to the graph of $y = e^t/2$ or $y = e^{-t}/2$.

The curve $y = \cosh(t)$ in Figure 5.8.1 is called a **catenary** (from the Latin *catena* meaning "chain"). It describes the shape of a free-hanging chain. (See the CIE on the Suspension Bridge and the Hanging Cable for Chapter 15.)

"sinh" is pronounced "sinch," rhyming with "pinch."

DEFINITION *(The hyperbolic sine.)* Let t be a real number. The **hyperbolic sine** of t, denoted $\sinh(t)$, is given by the formula

$$\sinh(t) = \frac{e^t - e^{-t}}{2}.$$

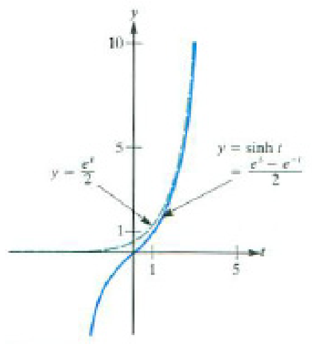


Figure 5.8.2

It is a simple matter to check that $\sinh(0) = 0$ and $\sinh(-t) = -\sinh(t)$, so that the graph of $\sinh(t)$ is symmetric with respect to the origin. Moreover, it lies below the graph of $e^t/2$. However, the graphs of $\sinh(t)$ and $e^t/2$ approach each other since $e^{-t}/2 \to 0$ as $t \to \infty$. Figure 5.8.2 shows the graph of $\sinh(t)$.

Note the contrast between $\sinh(t)$ and $\sin(t)$. As $|t|$ becomes large, the hyperbolic sine becomes large, $\lim_{t\to\infty} \sinh(t) = \infty$ and $\lim_{t\to-\infty} \sinh(t) = -\infty$. There is a similar contrast between $\cosh(t)$ and $\cos(t)$. While the trigonometric functions are periodic, the hyperbolic functions are not.

Example 1 shows why the functions $(e^t + e^{-t})/2$ and $(e^t - e^{-t})/2$ are called **hyperbolic**.

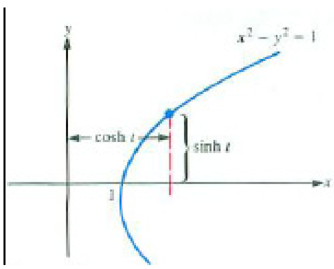


Figure 5.8.3

EXAMPLE 1 Show that for any real number t the point with coordinates

$$x = \cosh(t), \qquad y = \sinh(t)$$

lies on the hyperbola $x^2 - y^2 = 1$.

SOLUTION Compute $x^2 - y^2 = \cosh^2(t) - \sinh^2(t)$ and see whether it simplifies to 1. We have

$$\begin{aligned}
\cosh^2(t) - \sinh^2(t) &= \left(\tfrac{e^t+e^{-t}}{2}\right)^2 - \left(\tfrac{e^t-e^{-t}}{2}\right)^2 \\
&= \tfrac{e^{2t}+2e^te^{-t}+e^{-2t}}{4} - \tfrac{e^{2t}-2e^te^{-t}+e^{-2t}}{4} \\
&= \tfrac{2+2}{4} && \text{cancellation} \\
&= 1.
\end{aligned}$$

Observe that since $\cosh(t) \geq 1$, the point $(\cosh(t), \sinh(t))$ is on the right half of the hyperbola $x^2 - y^2 = 1$, as shown in Figure 5.8.3. ◇

Example 1 proves a fundamental identity for hyperbolic functions: $\cosh^2(t) - \sinh^2(t) = 1$

By contrast, $(\cos(\theta), \sin(\theta))$ lies on the circle $x^2 + y^2 = 1$, so the trigonometric functions are also called **circular functions**.

There are four more hyperbolic functions: the hyperbolic tangent, hyperbolic secant, hyperbolic cotangent, and hyperbolic cosecant. They are defined as follows:

$$\tanh(t) = \frac{\sinh(t)}{\cosh(t)} \qquad \operatorname{sech}(t) = \frac{1}{\cosh(t)} \qquad \coth(t) = \frac{\cosh(t)}{\sinh(t)} \qquad \operatorname{csch}(t) = \frac{1}{\sinh(t)}.$$

Each can be expressed in terms of exponentials. For instance,

$$\tanh(t) = \frac{(e^t - e^{-t})/2}{(e^t + e^{-t})/2} = \frac{e^t - e^{-t}}{e^t + e^{-t}}.$$

As $t \to \infty$, $e^t \to \infty$ and $e^{-t} \to 0$. Thus $\lim_{t\to\infty} \tanh(t) = 1$. Similarly, $\lim_{t\to-\infty} \tanh(t) = -1$. Figure 5.8.4 is a graph of $y = \tanh(t)$.

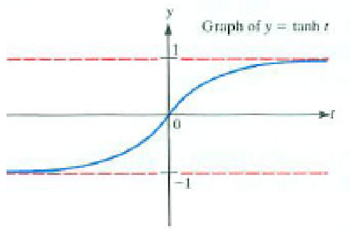


Figure 5.8.4

The Derivatives of the Hyperbolic Functions

The derivatives of the six hyperbolic functions can be computed directly. For instance,

$$(\cosh(t))' = \left(\frac{e^t + e^{-t}}{2}\right)' = \frac{e^t - e^{-t}}{2} = \sinh(t).$$

Table 5.8.1 lists the derivatives of the six hyperbolic functions. Notice that the formulas, except for the signs, are like those for the derivatives of the trigonometric functions.

Function	Derivative
$\cosh(t)$	$\sinh(t)$
$\sinh(t)$	$\cosh(t)$
$\tanh(t)$	$\operatorname{sech}^2(t)$
$\coth(t)$	$-\operatorname{csch}^2(t)$
$\operatorname{sech}(t)$	$-\operatorname{sech}(t)\tanh(t)$
$\operatorname{csch}(t)$	$-\operatorname{csch}(t)\coth(t)$

Table 5.8.1

The Inverses of the Hyperbolic Functions

Inverse hyperbolic functions appear on some calculators and in tables of mathematical functions. Just as the hyperbolic functions are expressed in terms of the exponential function, each inverse hyperbolic function can be expressed in terms of a logarithm. They provide useful antiderivatives as well as solutions to some differential equations.

Consider the inverse of $\sinh(t)$ first. Since $\sinh(t)$ is increasing, it is one-to-one; there is no need to restrict its domain. To find its inverse, it is necessary to solve the equation

$$x = \sinh(t)$$

for t as a function of x. The steps are straightforward:

Finding the inverse of the hyperbolic sine

$$\begin{array}{rrcll} & x & = & \frac{e^t - e^{-t}}{2}, & \text{definition of } \sinh(t) \\ & 2x & = & e^t - \frac{1}{e^t}, & e^{-t} = 1/e^t \\ & 2xe^t & = & (e^t)^2 - 1, & \text{multiply by } e^t \\ \text{or} & (e^t)^2 - 2xe^t - 1 & = & 0. & \end{array}$$

Equation (5.8) is quadratic in the unknown e^t. By the quadratic formula,

$$e^t = \frac{2x \pm \sqrt{(-2x)^2 + 4}}{2} = x \pm \sqrt{x^2 + 1}.$$

Since $e^t > 0$ and $\sqrt{x^2+1} > x$, the plus sign is kept and the minus sign is rejected. Thus

$$e^t = x + \sqrt{x^2 + 1} \qquad \text{and} \qquad t = \ln\left(x + \sqrt{x^2 + 1}\right).$$

Consequently, the inverse of the function $\sinh(t)$ is given by the formula

Formula for $\operatorname{arcsinh}(x)$

$$\operatorname{arcsinh}(x) = \sinh^{-1}(x) = \ln\left(x + \sqrt{x^2+1}\right).$$

Computation of $\operatorname{arctanh}(x)$ is a little different. Since the derivative of $\tanh(t)$ is $\operatorname{sech}^2(t)$, the function $\tanh(t)$ is increasing and has an inverse. However, $|\tanh(t)| < 1$, and so the inverse function will be defined only for $|x| < 1$. Computations similar to those for $\operatorname{arcsinh}(x)$ show that

Formula for $\operatorname{arctanh}(x)$

$$\operatorname{arctanh}(x) = \tanh^{-1}(x) = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right) \qquad |x| < 1.$$

Inverses of the other four hyperbolic functions are computed similarly. The functions $\operatorname{arccosh}(x)$ and $\operatorname{arcsech}(x)$ are chosen to be positive. Their formulas are included in Table 5.8.2.

The derivatives are found by differentiating the formulas in the second column.

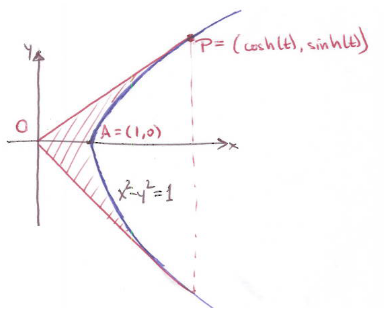
Function	Formula	Derivative	Domain
$\operatorname{arccosh}(x)$	$\ln(x+\sqrt{x^2-1})$	$\frac{1}{\sqrt{x^2-1}}$	$x \geq 1$
$\operatorname{arcsinh}(x)$	$\ln(x+\sqrt{x^2+1})$	$\frac{1}{\sqrt{x^2+1}}$	x-axis
$\operatorname{arctanh}(x)$	$\frac{1}{2}\ln\left(\frac{1+x}{1-x}\right)$	$\frac{1}{1-x^2}$	$\|x\| < 1$
$\operatorname{arccoth}(x)$	$\frac{1}{2}\ln\left(\frac{x+1}{x-1}\right)$	$\frac{1}{1-x^2}$	$\|x\| > 1$
$\operatorname{arcsech}(x)$	$\ln\left(\frac{1+\sqrt{1-x^2}}{x}\right)$	$\frac{-1}{x\sqrt{1-x^2}}$	$0 < x \leq 1$
$\operatorname{arccsch}(x)$	$\ln\left(\frac{1}{x}+\sqrt{1+\frac{1}{x^2}}\right)$	$\frac{-1}{\|x\|\sqrt{1+x^2}}$	$x \neq 0$

Table 5.8.2

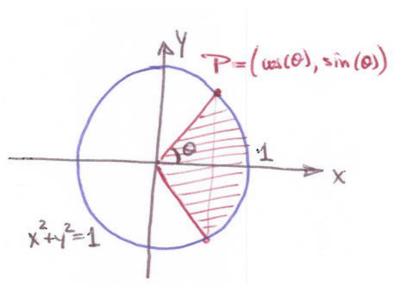
Summary

We introduced the six hyperbolic functions and their inverses, including $\sinh(x)$ (pronounced *sinch*), $\cosh(x)$ (pronounced *cōsh*), $\tanh(x)$ (pronounced *tanch* to rhyme with "ranch") and their inverses $\operatorname{arcsinh}(x)$, arccosh, and arctanh. Because they are all expressible in terms of exponentials, square roots, and logarithms, they do not add to the collection of elementary functions. However, some of them are especially convenient.

The point $(\cosh(t), \sinh(t))$ lies on the graph of the hyperbola $x^2 - y^2 = 1$. (See Example 1.) The parameter t, which can be any number, has a geometric interpretation: it is the area of the shaded region in Figure 5.8.5(a). This corresponds to the fact that a sector of the unit circle with angle 2θ has area θ, as shown in Figure 5.8.5(b). (See Exercise 64 in the Chapter 6 Summary.)



(a)



(b)

Figure 5.8.5

EXERCISES for Section 5.8

1.

(a) Compute $\cosh(t)$ and $e^t/2$ for $t = 0$, 1, 2, 3, and 4.

(b) Using the data in (a), graph $y = \cosh(t)$ and $y = e^t/2$ relative to the same axes.

2.

(a) Compute $\tanh(t)$ for $t = 0$, 1, 2, and 3.

(b) Using the data in (a), and the fact that $\tanh(-t) = \tanh(t)$, graph $y = \tanh(t)$.

In Exercises 3 to 5 obtain the derivatives of the given functions and express them in terms of hyperbolic functions.

3. $\tanh(x)$

4. $\sinh(x)$

5. $\cosh(x)$

6.

(a) Compute $\sinh(t)$ and $\cosh(t)$ for $t = -3$, -2, -1, 0, 1, 2, and 3.

(b) Plot the seven points $(x, y) = (\cosh(t), \sinh(t))$ found in (a).

(c) Explain why the point plotted in (b) lie on the hyperbola $x^2 - y^2 = 1$.

7.

(a) Show that $\operatorname{sech}^2(x) + \tanh^2(x) = 1$.

(b) What equation links $\sec(\theta)$ and $\tan(\theta)$?

In Exercises 8 to 16 use the definitions of the hyperbolic functions to verify the given identities. Notice how they differ from the corresponding identities for the trigonometric functions. In Section 12.6, it is shown that the hyperbolic functions are simply the trigonometric functions evaluated at complex numbers.

8. $\cosh(x + y) = \cosh(x)\cosh(y) + \sinh(x)\sinh(y)$

9. $\sinh(x+y) = \sinh(x)\cosh(y) + \cosh(x)\sinh(y)$

10. $\tanh(x+y) = \frac{\tanh(x)+\tanh(y)}{1+\tanh(x)\tanh(y)}$

11. $\cosh(x-y) = \cosh(x)\cosh(y) - \sinh(x)\sinh(y)$

12. $\sinh(x-y) = \sinh(x)\cosh(y) - \cosh(x)\sinh(y)$

13. $\cosh(2x) = \cosh^2(x) + \sinh^2(x)$

14. $\sinh(2x) = 2\sinh(x)\cosh(x)$

15. $2\sinh^2(x/2) = \cosh(x) - 1$

16. $2\cosh^2(x/2) = \cosh(x) + 1$

In Exercises 17 to 19 obtain a formula for the given function.

17. $\operatorname{arctanh}(x)$

18. $\operatorname{arcsech}(x)$

19. $\operatorname{arccosh}(x)$

In Exercises 20 to 23 show that the derivative of the first function is the second function.

20. $\operatorname{arccosh}(x)$; $1/\sqrt{x^2-1}$

21. $\operatorname{arcsinh}(x)$; $1/\sqrt{x^2+1}$

22. $\operatorname{arcsech}(x)$; $1/(x\sqrt{1-x^2})$

23. $\operatorname{arccsch}(x)$; $1/(x\sqrt{1+x^2})$

24. Find the inflection points on the curve $y = \tanh(x)$.

25. Graph $y = \sinh(x)$ and $y = \operatorname{arcsinh}(x)$ relative to the same axes. Show any inflection points.

26. One of the applications of hyperbolic functions is to the study of motion in which the resistance of the medium is proportional to the square of the velocity. Suppose that a body starts from rest and falls x meters in t seconds. Let g (a constant) be the acceleration due to gravity. It can be shown that there is a constant $V > 0$ such that $x = \frac{V^2}{g}\ln\left(\cosh\left(\frac{gt}{V}\right)\right)$.

(a) Find the velocity $v(t) = dx/dt$ as a function of t.

(b) Show that $\lim_{t\to\infty} v(t) = V$.

(c) Compute the acceleration $a(t) = dv/dt$ as a function of t.

(d) Show that the acceleration equals $g - g(v/V)^2$.

(e) What is the limit of the acceleration as $t \to \infty$?

27. In this exercise you will discover two different formulas for an antiderivative of $f(x) = \frac{1}{\sqrt{ax+b}\sqrt{cx+d}}$. The correct formula to use depends on the signs of a and c.

(a) Show that $\frac{2}{\sqrt{-ac}} \arctan \sqrt{\frac{-c(ax+b)}{a(cx+d)}}$ is an antiderivative of $f(x)$ when $a > 0$ and $c < 0$.

(b) Show that $\frac{2}{\sqrt{ac}} \operatorname{arctanh} \sqrt{\frac{c(ax+b)}{a(cx+d)}}$ is an antiderivative of $f(x)$ when $a > 0$ and $c > 0$.

5.S Chapter Summary

This chapter shows the derivative at work; applying it to practical problems, estimating errors, and evaluating some limits.

To determine the extrema of some quantity one must find a function that tells how the quantity depends on other quantities. Then, finding the extrema is like finding the highest or lowest points on the graph of the function.

When two varying quantities are related by an equation, the derivative can tell the relation between the rates at which they change: just differentiate both sides of the equation that relates them. That differentiation depends on the chain rule and is called implicit differentiation because one differentiates a function without having an explicit formula for it.

The next two sections form a unit that presents one of the main uses of higher derivatives: to estimate errors when approximating a function by a polynomial and later, in Section 6.5, to estimate errors in approximating area under a curve by trapezoids and parabolas.

The key to the Growth Theorem is that if R is a function such that

$$0 = R(a) = R'(a) = R''(a) = \cdots = R^{(n)}(a)$$

and in some interval around a we know $R^{(n+1)}(x)$ is continuous, then there is a number c_n in $[a, x]$ such that

$$|R(x)| \leq R^{(n+1)}(c_n)\frac{|x-a|^{n+1}}{(n+1)!} \qquad \text{for all } x \text{ in that interval.}$$

That means we have information on how rapidly $R(x)$ can grow for x near a. This information was used to control the error when using a polynomial to approximate a function.

A likely candidate for the polynomial of degree n that closely resembles a given function f near $x = a$ is the one whose derivatives at a, up through order n, agree with those of f there. That polynomial is

$$P(x) = P_n(x; a) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

Because the polynomial was chosen so that $P^{(k)}(a) = f^{(k)}(a)$ for all k up through n, the remainder function function $R(x) = f(x) - P(x)$ has all its derivatives up through order n at a equal to 0. Moreover, since the $(n+1)^{\text{st}}$ derivative of any polynomial of degree at most n is identically 0, $R^{(n+1)}(x) = f^{(n+1)}(x)$. Thus the error $|f(x) - P(x)|$ is at most $M\frac{|x-a|^{n+1}}{(n+1)!}$, if $|f^{(n+1)}(t)|$ stays less than or equal to M for t between a and x. A similar conclusion holds if $|f^{n+1}(t)|$ stays larger than a fixed number. Using these facts we obtained

Lagrange's formula for the error:

$$\frac{f^{(n+1)}(c_n)}{(n+1)!}(x-a)^{n+1} \qquad \text{for some } c_n \text{ between } a \text{ and } x.$$

The case $n = 1$ reduces to the linear approximation of a curve by the tangent line at $(a, f(a))$. In this case the error is controlled by the second derivative.

We return to Taylor polynomials in Chapter 12, where we express e^x, $\sin(x)$, and $\cos(x)$ as "polynomials of infinite degree," and use them and complex numbers to express $\sin(x)$ and $\cos(x)$ in terms of exponential functions.

Section 5.6 concerns l'Hôpital's rule, a tool for computing certain limits, such as the limit of a quotient whose numerator and denominator both approach zero.

The final two sections, on natural growth and decay and the hyperbolic functions, conclude the chapter. While these sections are not needed in future chapters of this book, they are important applications in a wide variety of disciplines, including biology and engineering.

EXERCISES for 5.S

1. Arrange the following numbers by order of increasing size as $x \to \infty$.

(a) $1000x$

(b) $\log_2(x)$

(c) $\sqrt{x}$

(d) $(1.0001)^x$

(e) $\log_{1000}(x)$

(f) $0.01x^3$

In Exercises 2 to 28 find the limits, if they exist.

2. $\displaystyle\lim_{u\to\infty}\left(\frac{u+1}{u}\right)^{u+1}\frac{1}{\sqrt{u}}$

3. $\displaystyle\lim_{x\to\infty}\left(\frac{x+2}{x+1}\right)^{x+3}$

4. $\displaystyle\lim_{x\to\infty}\left(\frac{x}{x+1}\right)^{x+1}$

5. $\displaystyle\lim_{x\to 3}\frac{x-2}{\cos(\pi x)}$

6. $\displaystyle\lim_{x\to 3}\frac{x-2}{\sin(\pi x)}$

7. $\lim\limits_{x\to\infty} \dfrac{\sqrt{1+x^2}}{x}$

8. $\lim\limits_{x\to\infty} \dfrac{\sqrt{1+x^2}}{\sqrt{2+x^2}}$

9. $\lim\limits_{x\to\infty} \dfrac{(1+x^2)^{1/2}}{(2+x^2)^{1/3}}$

10. $\lim\limits_{x\to\infty} \dfrac{1+x+x^2}{2+3x+4x^2}$

11. $\lim\limits_{x\to 1} \dfrac{\ln(x)\tan\left(\frac{\pi x}{4}\right)}{\cos\left(\frac{\pi x}{2}\right)}$

12. $\lim\limits_{x\to 0} \dfrac{f(3+x)-f(3)}{x}$ where $f(x)=(x^2+5)\sin^2(3x)$.

13. $\lim\limits_{x\to\infty} \dfrac{\ln(6x)-\ln(5x)}{\ln(7x)-\ln(6x)}$

14. $\lim\limits_{x\to\infty} \dfrac{\ln(6x)-\ln(5x)}{x\ln(7x)-x\ln(6x)}$

15. $\lim\limits_{x\to\pi} \dfrac{e^{-x^2}\sin(x)}{x^2-\pi^2}$

16. $\lim\limits_{x\to\pi} \dfrac{\ln(x^3-\sin(x))-3\ln(\pi)}{x-\pi}$

17. $\lim\limits_{x\to 0} \dfrac{(x+2)}{(x+3)}\dfrac{(\cos(5x)-1)}{\cos(7x)-1)}$

18. $\lim\limits_{x\to\infty} \left(\dfrac{x+2}{x+1}\right)^{2x}$

19. $\lim\limits_{x\to\pi} \dfrac{\sin^4(x)}{(\pi^4-x^4)^2}$

20. $\lim\limits_{x\to\infty} \dfrac{\sec^4(x)\tan(3x)}{\sin(2x)}$

21. $\lim\limits_{x\to 1} \dfrac{e^{3x}(x^2-1)}{\cos(\sqrt{2}x)\tan(3x-3)}$

22. $\lim\limits_{x\to 0}(1+0.005x)^{20x}$

23. $\lim\limits_{t\to 0} \dfrac{e^{3(x+t)}-e^{3x}}{5t}$

24. $\lim\limits_{t\to 0} \dfrac{e^{3(x+t)}-e^{3x}}{5t}$

25. $\lim\limits_{x\to 0} \left(\dfrac{1+2^x}{2}\right)^{1/x}$

26. $\lim\limits_{x\to 0} \left(\dfrac{1+2^x}{1+3^x}\right)^{1/x}$

27. $\lim\limits_{x\to\infty}(1+0.003x)^{20/x}$

28. $\lim_{x\to\infty}(1+0.003x)^{20/x}$

In Exercises 29 to 36 find the derivative of the given function.

29. $(\cos(x))^{1/x^2}$

30. $\ln\left(\sec^2(3x)\sqrt{1+x^2}\right)$

31. $\ln\left(\sqrt{e^{x^3}}\right)$

32. $\frac{5+3x+7x^2}{58-4x+x^2}$

33. $\frac{\tan^2(2x)}{(1+\cos(2x))^4}$

34. $(\cos^2(3x))^{\cos^2(2x)}$

35. $f(x) = \begin{cases} x^2\sin(\pi/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$ (Use the definition of the derivative to find $f'(0)$.)

36. $f(x) = \begin{cases} \frac{\sin(\pi x)}{x} & \text{if } x \neq 0 \\ \pi & \text{if } x = 0 \end{cases}$

37.

(a) Find $P_1(x;64)$ for $f(x)=\sqrt{x}$.

(b) Use $P_1(x;64)$ to estimate $\sqrt{67}$.

(c) Put bounds on the error in the estimate in (b).

38.

(a) Show that when x is small $\sqrt[3]{1+x}$ is approximately $1+x/3$.

(b) Use (a) to estimate $\sqrt[3]{0.94}$ and $\sqrt[3]{1.06}$.

39.

(a) Show that when x is small $1/\sqrt[3]{1+x}$ is approximately $1-x/3$.

(b) Use (a) to estimate $1/\sqrt[3]{0.94}$ and $1/\sqrt[3]{1.06}$.

40.

(a) Find the Maclaurin polynomial of degree 6 associated with $\cos(x)$.

(b) Use (a) to estimate $\cos(\pi/4)$.

(c) What is the error between the estimate found in (b) and the exact value, $\sqrt{2}/2$.

(d) What is the Lagrange bound for the error?

In Exercises 41 to 52, examine the limit, determine whether it exists, and, if it does exist, find its value.

41. $\lim_{x\to 1} \frac{1-e^x}{1-e^{2x}}$

42. $\lim_{x\to 0} \frac{x}{\sqrt{1+x^2}}$

43. $\lim_{x\to 0} \frac{1-e^x}{1-e^{2x}}$

44. $\lim_{x\to\infty} \frac{x^2}{(1+x^3)^{2/3}}$

45. $\lim_{x\to\infty} x^2 \sin(x)$

46. $\lim_{x\to 8} \frac{2^x - 2^8}{x-8}$

47. $\lim_{x\to 1} \frac{e^{x^2} - e^x}{x-1}$

48. $\lim_{x\to 4} \frac{2^x + 2^4}{x+4}$

49. $\lim_{x\to 0} \frac{\sin(x) - e^{2x}}{x}$

50. $\lim_{x\to 0} \frac{e^{3x}\sin(2x)}{\tan(3x)}$

51. $\lim_{x\to 0} \frac{\sqrt{1+x^2}-1}{\sqrt[3]{1+x^2}-1}$

52. $\lim_{x\to\pi/2} \frac{\sin 9x)\cos(x)}{x-\pi/2}$

53. If $\lim_{x\to\infty} f'(x) = 3$ and $\lim_{x\to\infty} g'(x) = 3$, what, if anything, can be said about

(a) $\lim_{x\to\infty} \frac{f(x)}{3x}$

(b) $\lim_{x\to\infty}(g(x) - f(x))$

(c) $\lim_{x\to\infty} \frac{f(x)}{g(x)}$

(d) $\lim_{x\to\infty}(f(x)-3x)$

(e) $\lim_{x\to\infty}\frac{(f(x))^3}{(g(x))^3}$

54. Let $f(x) = (5x^3 + x + 2)^{20}$. Find (a) $f^{(60)}(4)$ and (b) $f^{(61)}(2)$.

55. The point $P = (c, d)$ lies in the first quadrant. Each line through P of negative slope determines a triangle whose vertices are the intercepts of the line on the axes, and the origin.

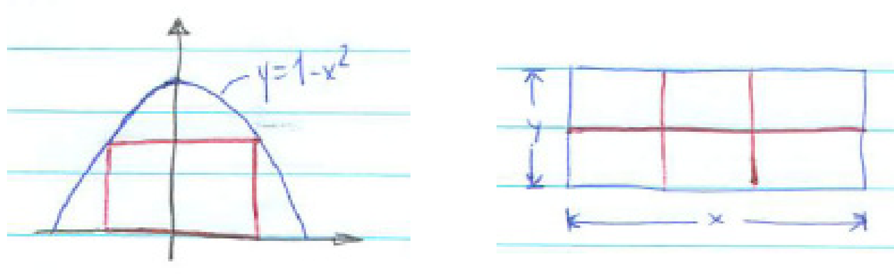
(a) Find the slope of the line that minimizes the area.

(b) Find the minimum area.

56. Figure 5.S.1(a) shows a typical rectangle whose base is the x-axis, inscribed in the parabola $y = 1 - x^2$.

(a) Find the rectangle of largest perimeter.

(b) Find the rectangle of largest area.



(a) (b)

Figure 5.S.1

57. A rectangle of perimeter 12 inches is spun around one of its edges to produce a circular cylinder.

(a) For which rectangle is the area of the curved surface of the cylinder a maximum?

(b) For which rectangle is the volume of the cylinder a maximum?

58. Consider isosceles triangles whose equal sides have length a and the angle where these two sides meet is θ. For which angle θ is the area of the triangle a maximum?

(a) Solve this problem using calculus.

(b) Solve the same problem without calculus.

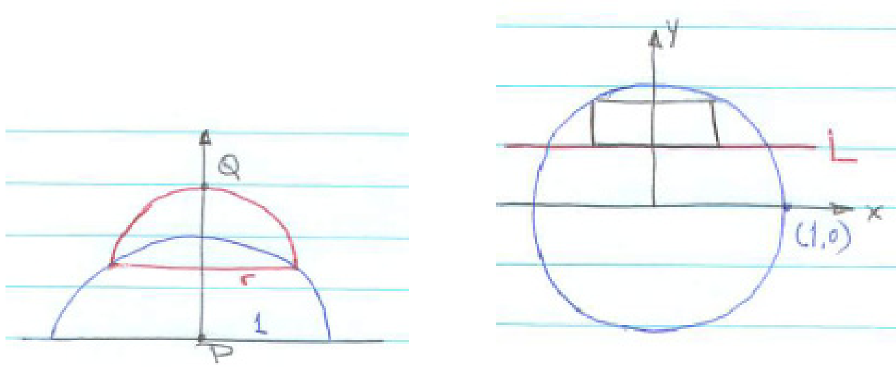
59. A farmer has 200 feet of fence which he wants to use to enclose a rectangle divided into six congruent rectangles, as shown in Figure 5.S.1(b). He wishes to enclose a maximum area.

(a) If x is near 0, what is the area, approximately?

(b) How large can x be?

(c) In the case that produces the maximum area, which do you think will be larger x or y? Why?

(d) Find the dimensions x and y that maximizes the area.



(a) (b)

Figure 5.S.2

60. A semicircle of radius $a < r \leq 1$ rests upon a semicircle of radius 1, as shown in Figure 5.S.2(a). The length of PQ, the segment from the origin of the lower circle to the top of the upper circle is a function of r, $f(r)$.

(a) Find $f(0)$ and $f(1)$.

(b) Find $f(r)$.

(c) Maximize $f(r)$, testing the maximum by the second derivative.

Exercises 61 to 63 are independent, but related. They contain a surprise.

61. Figure 5.S.2(b) shows the unit circle $x^2 + y^2 = 1$, the line L whose equation is $y = 1/3$, and a typical rectangle with base on L, inscribed in the circle. Find the rectangle with base on L that has (a) minimum perimeter and (b) maximum perimeter.

62. Like Exercise 61 but this time the line L has the equation $y = 1/2$.

63. The analyses in Exercises 61 to 62 are different. Let the line L have the equation $y = c$, $0 < c < 1$. For which values of c is the analysis like that for (a) Exercise 61? (b) Exercise 62?

64. A. Bellemans, in "Power Demand in Walking and Pace Optimization," Amer. J. Physics 49(1981) pp. 25–27, modeling the work spent on walking, writes "$H = L(1 - \cos(\gamma)$ or, to a sufficient approximation for the present purpose, $H = L\gamma^2/2$." Justify this approximation.

65. Two houses, A and B, are a distance p apart. They are distances q and r, respectively, from a straight road, and on the same side of the road. Find the length of the shortest path that goes from A to the road, and then on to the other house B.

(a) Use calculus.

(b) Use only elementary geometry. *Hint:* Introduce an imaginary house C such that the midpoint of B and C is on the road and the segment BC is perpendicular to the road; that is, "reflect" B across the road to become C.

66. Let k be a constant. Determine $\lim_{x\to\infty} x\left(e^{-k} - \left(1 - \frac{k}{x}\right)^x\right)$.

67. Let k be a constant. Determine $\lim_{x\to\infty} x\left(e^{k} - \left(1 + \frac{k}{x}\right)^x\right)$.

68. Let $p_n(x)$ be the Maclaurin polynomial of degree n associated with e^x. Because $e^x \cdot e^{-x} = 1$, we might expect that $p_n(x)p_n(-x)$ would also be 1. But that cannot be because the degree of the product is $2n$.

(a) Compute $p_2(x)p_2(-x)$ and $p_3(x)p_3(-x)$.

(b) Make a conjecture about $p_n(x)p_n(-x)$ based on (a).

69. Let $p_n(x)$ be the Maclaurin polynomial of degree n associated with e^x. Because $e^{2x} = e^x \cdot e^x$, we might expect that $p_{2n}(x) = p_n(x)p_n(x)$.

(a) Why is that equation false for $n \geq 1$?

(b) To what extent does $p_2(x)p_2(x)$ resemble $p_2(2x)$ and $p_3(x)p_3(x)$ resemble $p_3(2x)$?

(c) Make a conjecture based on (a) and (b).

70. Let $p_n(x)$ be the Maclaurin polynomial of degree n associated with e^x. The equation $e^{x+y} = e^x \cdot e^y$ suggests that $p_n(x+y)$ might equal $p_n(x)p_n(y)$.

(a) Why is that hope not realistic?

(b) To what extent does $p_2(x)p_2(y)$ resemble $p_2(x+y)$?

71. What can be said about $f(10)$ if $f(1) = 5$, $f'(1) = 3$ and $2, f''(x) < 4$ for x in $(-10, 20)$?

72. The demand for a product is influenced by its price. In one example an economics text links the amount sold (x) to the price (P) by the equation $x = b - aP$, where b and a are positive constants. As the price increases the sales go down. The cost of producing x items is an increasing function $C(x) = c + kx$, where c and k are positive constants.

(a) Express P in terms of x.

(b) Express the total revenue $R(x)$ in terms of x.

(c) Note that $C(0) = c$. So what is the economic significance of c?

(d) What is the economic significance of k?

(e) Let $E(x)$ be the profit, that is, the revenue minus the cost. Express $E(x)$ as a function of x.

(f) Which value of x produces the maximum profit?

(g) The marginal revenue is defined as dR/dx and the marginal cost as dC/dx. Show that for the value of x that produces the maximum profit, $dR/dx = dC/dx$.

(h) What is the economic significance of $dR/dx = dC/dx$ in (g)?

73. This exercise concerns a function used to describe the consumption of a finite resource, such as petroleum. Let Q be the amount initially available. Let a be a positive constant and b be a negative constant. Let $y(t)$ be the amount used up by the time t. The function $Q/(1 + ae^{bt})$ is often used to represent $y(t)$.

(a) Show that $\lim_{t\to\infty} y(t) = Q$ and $\lim_{t\to-\infty} y(t) = 0$. Why are these realistic?

(b) Show that $y(t)$ has an inflection point when $t = -\ln(a)/b$.

(c) Show that at the inflection point, $y(t) = Q/2$, that is, half the resource has been used up.

(d) Sketch the graph of $y(t)$.

(e) Where is $y'(t)$, the rate of using the resource, greatest?

The same function describes limited growth that is bounded by Q, so called **logistic growth**.

74. About 100 cubic yards are added to a land fill every day. The operator decides to pile the debris up in the form of a cone whose base angle is $\pi/4$. (He hopes to make a ski run where it never snows.) At what rate is the height of the cone increasing when the height is (a) 10 yards? (b) 20 yards? (c) 100 yards? (d) How long will it take to make a cone 100 yards high? 300 yards high? The volume of a circular cone is one third the product of its height and the area of its base.

75. A wine dealer has a case of wine that he could sell today for \$100. Or, he could decide to store it, letting it mellow, and sell later for a higher price. Assume he could sell in t years for \$ $100e^{\sqrt{t}}$. In order to decide which option to choose he computes the present value of the sale. If the interest rate is r, the present value of one dollar t years hence is e^{-rt}. When should he sell the wine?

76. A differentiable function is defined throughout $(-\infty, \infty)$. Its derivative is 0 at exactly two inputs.

(a) Can there be exactly one relative extremum?

(b) Could it have two relative maxima?

(c) What is the maximum number of relative extrema possible?

(d) What is the minimum number?

(Sketch graphs, then explain.)

77. A differentiable function is defined throughout $(-\infty, \infty)$. Its derivative is 0 at exactly three inputs. and the function approaches 0 as x approaches ∞

(a) Can there be exactly two relative extremum?

(b) Could it have three relative maxima?

(c) What is the maximum number of relative extrema possible?

(d) What is the minimum number?

(Sketch graphs, then explain.)

78. A differentiable function is defined throughout $(-\infty, \infty)$. Its derivative is 0 at exactly two inputs. and the function approaches the same finite limit as x approaches ∞ and $-\infty$.

(a) Can there be exactly one relative extremum?

(b) Could it have two relative maxima?

(c) What is the greatest number of relative extrema possible?

(d) What is the least number?

(Sketch graphs, then explain.)

79. In the paper cited in the Exercise 64, Bellemans writes "The total mechanical power required for walking is $P(v, a) = \alpha M v^3/a + (\beta M g v)/L)a$. Enlarging the pace, a, at a constant speed v, lowers the first term and increases the second one so that the formula predicts an optimal pace $a^*(v)$, minimizing $P(v, a)$." In the formula, α, M, v, β, g, and L are constants.

(a) Show that $a^*(v) = \left(\frac{\alpha}{\beta}\right)^{1/2} \left(\frac{L}{g}\right)^{1/2} v$

(b) Verify that the "corresponding minimum power" is

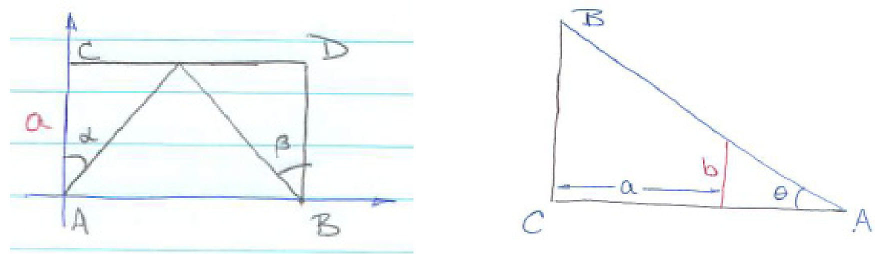
$$P(v, a^*(v)) = 2(\alpha\beta)^{1/2} \left(\frac{g}{L}\right)^{1/2} M v^2.$$

"One would therefore expect that, when walking naturally on the flat at a fixed velocity, a subject will adjust its pace automatically to the optimum value corresponding to the minimum work expenditure. This has indeed been verified experimentally."

80. Figure 5.S.3(a) shows two points A and B a mile apart and both at a distance a from the river CD. Sam is at A. He will walk in a straight line to the river at 4 mph, fill a pail, then continue on to B at 3 mph. He wishes to do this in the shortest time.

(a) For the fastest route which angle in Figure 5.S.3 do you expect to be larger, α or β?

(b) Show that for the fastest route $\sin(\alpha)/\sin(\beta)$ equals 4/3.



(a) (b)

Figure 5.S.3

81. A fence b feet high is a feet from a tall building, whose wall contains BC. as shown in Figure 5.S.3(b). Find the angle θ that minimizes the length of AB. (That angle produces the shortest ladder to reach the building and stay above the fence.)

82.

(a) Show that if a differentiable function f is even, then f' is odd, by differentiating both sides of the equation $f(-x) = f(x)$.

(b) Explain why the conclusion in (a) is to be expected by interpreting it in terms of the graph of f.

83. Show that if a differentiable function is odd, then its derivative is even.

84. What do the previous two exercises imply about a Maclaurin polynomial associated with an odd function? associated with an even function?

85. Show that

(a) If $p_n(x)$ is a Maclaurin polynomial associated with $f(x)$, then $p_n'(x)$ is a Maclaurin polynomial associated with $f'(x)$.

(b) Use (a) to find the 6^{th}-order Maclaurin polynomial for $1/(1-x)^2$.

86. (Assume $e < 3$.) Let $P_1(x)$ be the Maclaurin polynomial associated with e^x. For how large an x can you be sure that

(a) $|e^x - P_1(x)| < 0.01$?

(b) $|e^x - P_2(x)| < 0.01$?

(c) $|e^x - P_3(x)| < 0.01$?

87. A number b is **algebraic** if there is a non-zero polynomial $\sum_{i=0}^{n} a_i x^i = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$, with coefficients a_i that are rational numbers, such that $\sum_{i=0}^{n} a_i b^i = 0$. In other words, b is algebraic if there is a function f that satisfies (a) $f(b) = 0$, (b) all derivatives of f at 0 are rational, but not all zero, and (c) there is a positive integer m such that $D^m(f) = 0$. (Recall that D is the differentiation operator.)
We call a number b **almost algebraic** if (a) b is not algebraic and there is a function f with (b) $f(b) = 0$, (c) all derivatives of f at 0 are rational, but not all zero, and (d) there is a non-zero polynomial $p(D)$ such that $p(D)(f) = 0$. For example, if $p(x) = x^2 + 1$ then $p(D)(f) = D^2(f) + f = f'' + f$.
Show that π is almost algebraic. (Assume it is not algebraic.)

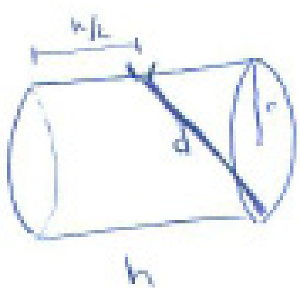


Figure 5.S.4 ARTIST: Show wine level inside the barrel.

88. Kepler, the astrologer and astronomer, to celebrate his wedding in 1613, ordered some wine, which was available in cylindrical barrels of various shapes. He was surprised by the way the merchant measure the volume of a barrel. A ruler was pushed through the opening in the side of the barrel (used to fill the barrel) until it came to a stop at the edge of a circular base. The merchant used the length of the part of the ruler inside the barrel to determine the volume of the barrel. Figure 5.S.4 shows the method.
The barrel in Figure 5.S.4 has radius r, height h, and volume V. The length of the ruler inside the barrel is d.

(a) Using common sense, show that d does not determine V.

(b) How small can V be for a given value of d?

(c) Using calculus, show that the maximum volume for a given d occurs when $h = 2\sqrt{2}d/\sqrt{6}$ and $r = d/\sqrt{6}$.

(d) Show that to maximize the volume the height must be $\sqrt{2}$ times the diameter. (This is what Kepler showed.)

Try to solve this problem two different ways. One without implicit differentiation and the other with implicit differentiation.

89. Let m and n be positive integers. Let $f(x) = \sin^m(x)\cos^n(x)$ for x in $[0, \pi/2]$.

(a) For which x is $f(x)$ a minimum?

(b) For which x is $f(x)$ a maximum?

(c) What is the maximum value of $f(x)$?

90.

(a) Let $P(x)$ be a polynomial such that $D^2(x^2P(x)) = 0$. Show that $P(x) = 0$.

(b) Does the same conclusion follow if instead we assume $D^2(xP(x)) = 0$?

(If $P(x)$ has degree n, what are the degrees of $xP(x)$ and $x^2P(x)$?)

91. Translate this news item into the language of calculus: "The one positive sign during the quarter was a slowing in the rate of increase in home foreclosures."

92. In May 2009 it was reported that "the nation's industrial production fell in April by the smallest amount in six months, fresh evidence that the pace of the economy's decline is slowing."
Let $P(t)$ denote the total production up to time t with t representing the number of months since January 2000 ($t = 0$).

(a) Translate the above statement into the language of calculus, that is, in terms of $P(t)$ and its derivatives (evaluated at appropriate values of t).

(b) Sketch a possible graph of $P(t)$ for November 2008 through April 2009.

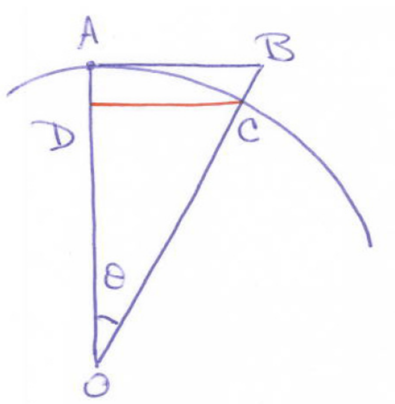


Figure 5.S.5

93. (A challenge to your intuition.) In Figure 5.S.5 AB is tangent to an arc of a circle, OA is a radius and DC is parallel to AB.

(a) What do you think happens to the ratio of the area of ABC to the area of ADC as $\theta \to 0$?

(b) Using calculus, find the limit of that ratio as $\theta \to 0$.

(c) In view of (b), which provides a better estimate of the area of a disk, the circumscribed regular n-gon or the inscribed regular n-gon?

(d) In view of the limit in (b), what combination of the estimates by the inscribed regular n-gon and the circumscribed regular n-gon would likely provide a very good estimate of the area of the disk?

94. Let f and g be differentiable.

(a) If $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 3$, must $\lim_{x\to\infty} \frac{f'(x)}{g'(x)}$ exist and be 3?

(b) If the second limit in (a) exists, can it have a value other than 3?

95. Evaluate each limit, indicating the indeterminate form each time l'Hôpital's Rule is applied.

(a) $\displaystyle\lim_{x\to 0} \left(\frac{1+2^x}{2}\right)^{1/x}$

(b) $\displaystyle\lim_{x\to 0} \left(\frac{1+2^x}{1+3^x}\right)^{1/x}$

96. Evaluate $\lim_{x\to\infty} \frac{x(x+1)^x}{x^{x+1}}$.

97. Evaluate $\lim_{x\to\infty} \frac{(x+1)^x}{x^{x+1}}$.

98.

Jane: I wonder which is bigger, 2001^{2000} or 2000^{2001}?

Sam: Obviously the one with the bigger base.

Jane: But its exponent is smaller than the exponent of the other.

Sam: I think the base has more influence.

Jane: And I think the exponent has more impact.

Settle the dispute by examining the ratio 20012000/(20002001).

99.

Sam: I can use Taylor polynomials to get l'Hôpital's theorem.

Jane: How so?

Sam: I write $f(x) = f(0)+f'(0)x+f''(c)x^2/2$ and $g(x) = g(0)+g'(0)x+g''(d)x^2/2$.

Jane: O.K.

Sam: Since $\lim_{x\to 0} f(x)$ and $\lim_{x\to 0} g(x)$ are both zero I have $f(0) = g(0) = 0$. I can write, after canceling some x's

$$\frac{f(x)}{g(x)} = \frac{f'(0) + f''(c)x/2}{g'(0) + g''(d)x/2}.$$

Jane: But you don't know the second derivatives.

Sam: It doesn't matter. I just take limits and get

$$\lim_{x\to 0} \frac{f(x)}{g(x)} = \lim_{x\to 0} \frac{f'(0) + f''(c)x/2}{g'(0) + g''(d)x/2}.$$

So

$$\lim_{x\to 0} \frac{f(x)}{g(x)} = \frac{f'(0)}{g'(0)} = \lim_{x\to 0} \frac{f'(x)}{g'(x)}.$$

There you have it.

Jane: Let me check your steps.

Check the steps and comment on Sam's proof.

When you throw a fair six-sided die many times, you would expect a 5 to show about 1/6 of the times. That is, if you throw it n times and get k 5's, you would expect k/n to be near 1/6.
More generally, if a certain trial has probability p of success and $q = 1-p$ of failure, and is repeated n times, with k successes, you would expect k/n to be near p. That means that if n is large you would expect $(k/n) - p$ to be small. In other words, let $\epsilon = (k/n) - p$, where ϵ approaches 0 as $n \to \infty$. This means that in most cases $k = np + \epsilon n$, or $k = np + z$, where $z/n \to 0$ as $n \to \infty$.
The probability of exactly k successes (and $n-k$ failures) in n trials is

$$\frac{n!}{k!(n-k)!}p^k q^{n-k}. \tag{5.S.1}$$

Exercises 100 to 104 show that for large n (and k) (5.S.1) is approximately

$$\frac{1}{\sqrt{2\pi npq}} \exp\left(\frac{-z^2}{2npq}\right). \tag{5.S.2}$$

Note that (5.S.2) involves $\exp(-x^2)$, whose graph has the shape of the famous bell curve associated with the **normal (or Gaussian) distribution** in probability and statistics. (See also Exercises 23 and 24 in Section 10.4 on page 912.)
100. In Exercise 28 in Section 11.6 we will derive Stirling's formula for an approximation to $n!$:

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

Use Stirling's formula to show that (5.S.1) is approximately

$$\left(\frac{n}{2\pi k(n-k)}\right)^{1/2}\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k} \tag{5.S.3}$$

in the sense that (5.S.2) divided by (5.S.3) approaches 1 as $n \to \infty$.

101. Show that as $n \to \infty$, the first factor in (5.S.3) is asymptotic to

$$\left(\frac{1}{2\pi pqn}\right)^{1/2} \tag{5.S.4}$$

in the sense that the ratio between it and (5.S.4) approaches 1 as $n \to \infty$.

102. To relate the rest of (5.S.3) to the exponential function, $\exp(x)$, we take its logarithm. Show that

$$\begin{aligned}&\ln\left(\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k}\right)\\ &\quad = \quad -(np+z)\ln\left(1+\frac{z}{np}\right) - (nq-z)\ln\left(1-\frac{z}{nq}\right).\end{aligned}$$

103. Using the Maclaurin polynomial of degree two to approximate $\ln(1+t)$, show that for large n, (5.S.5) is approximately

$$\frac{-z^2}{2pqn}.$$

104. Conclude that for large n, (5.S.1) is approximately (5.S.2).

105. If $P(x)$ is a Maclaurin polynomial associated with $f(x)$, what is the Maclaurin polynomial of the same degree associated with $f(2x)$?

106. Find the Maclaurin polynomial of degree 6 associated with $1/e^x$.

107. Find the Maclaurin polynomial of degree 6 associated with $\sin(x)\cos(x)$. (There is a short way and a long way to find it.)

108. The center $(x, 0)$ of a circle C_1 of radius 1 is at a distance $x < 3$ from the center $(0, 0)$ of a circle C_2 of radius 2. AB is the chord joining their two points in common. Let A_1 be the area within C_1 to the left of that chord and A_2 the area within C_2 to the right of that chord.

(a) Which is larger, A_1 or A_2? (Sketch a diagram of these circles and the chord.)

(b) If $\lim_{x\to 3^-} A_2/A_1$ exists, what do you think it is?

(c) Determine whether the limit in (b) exists. If it does, find it.

109. In the set-up of Exercise 108, let O_1 be the center of C_1 and O_2 the center of C_2. What happens to the ratio of the area common to the two disks and the area of the quadrilateral AO_1BO_2 as $x \to 3^-$?

110. Let $g(x) = f(x^2)$.

(a) Express the Maclaurin polynomial for $g(x)$ up through the term of degree 4 in terms of f and its derivatives.

(b) How is the answer in (a) related to a Maclaurin polynomial associated with f?

111. Find $\lim_{x\to\pi/2^-}(\sec(x) - \tan(x))$

(a) Using l'Hôpital's rule

(b) Without using l'Hôpital's rule

112. Assume that $\lim_{x\to\infty} f(x) = \infty$ and $\lim_{x\to\infty} g(x) = \infty$.

(a) If $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$, what, if anything, can be said about $\lim_{x\to\infty} \frac{ln(f(x))}{ln(g(x))}$?

(b) If $\lim_{x\to\infty} \frac{ln(f(x))}{ln(g(x))} = 1$, what, if anything, can be said about $\lim_{x\to\infty} \frac{f(x)}{g(x)}$?

113. Assume that the function $f(x)$ is defined on $[0, \infty)$, has a continuous positive second derivative and $\lim_{x\to\infty} f(x) = 0$.

(a) Can $f(x)$ ever be negative?

(b) Can $f'(x)$ ever be positive?

(c) What are the possible general shapes for the graph of f?

(d) Give an explicit formula for an example of such a function.

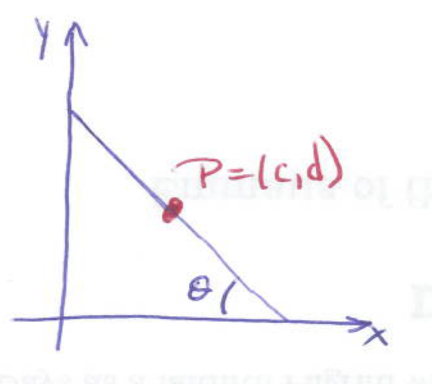


Figure 5.S.6

114. Let c and d be fixed positive numbers. Consider line segments through $P = (c, d)$ whose ends are on the positive x- and y-axes, as in Figure 5.S.6. Let θ be the acute angle between the line and the x-axis. Show that the angle α that produces the shortest line segment through P has $\tan^3(\alpha) = d/c$.

115. (See Exercise 114.)

(a) Show that for the angle β such that the the area of the triangle determined by the line segment and the two axes is a minimum, $\tan(\beta) = d/c$.

(b) Show that for β as in (a), OP bisects the line into two parts of equal length.

116. An adventurous bank decides to compound interest twice a year, at time x ($0 < x < 1$) and at time 1 (instead of at the usual 1/2 and 1). Assume that the annual interest rate is r. Is there a time, x, such that the account grows to more than if the interest was computed at 1/2 and 1?

117. Every six hours a patient takes an amount A of a medicine. Once in the patient, the medicine decays exponentially. In six hours the amount declines from A to kA, where k is less than 1 (and positive). Thus, in 12 hours, the amount in the system is $kA + k^2A$. At exactly 12 hours, the patient takes another pill and the amount in her system is $A + kA + k^2A$.

(a) Graph the general shape of the sketch showing the amount of medicine in the patient as a function of time.

(b) When a pill is taken at the end of n six-hour periods how much is in the system?

(c) Does the amount in the system become arbitrarily large? (If so, this could be dangerous.)

The constant k depends on many factors, such as the age of the patient. For this reason, a dosage tested on a 20-year old may be lethal on a 70-year. (See also Exercise 30 in Section 11.2.)

Skill Drill: Derivatives

The remaining exercises offer an opportunity to practice differentiating. In each case show that the derivative of the first function is the second function.

118. $\arctan\left(\frac{x}{a}\right)$; $\frac{a}{x^2+a^2}$.

119. $\frac{2(3ax-2b)}{15a^2}\sqrt{(ax+b)^3}$; $x\sqrt{ax+b}$.

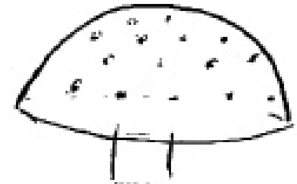
120. $\sin(ax) - \frac{1}{3}\sin^3(ax)$; $a\cos^3(ax)$.

121. $e^{ax}(a\cos(bx) + b\sin(bx))$; $(a^2+b^2)e^{ax}\cos(bx)$.

Calculus is Everywhere # 7
The Uniform Sprinkler

One day one of the authors (S.S.) realized that the sprinkler did not water his lawn evenly. Placing empty cans throughout the lawn, he discovered that some places received as much as nine times as much water as other places. That meant some parts of the lawn were getting too much water or not enough water.

The sprinkler, which had no moving parts, consisted of a hemisphere, with holes distributed uniformly on its surface, as in Figure C.7.1. Even though the holes were uniformly spaced, the water was not supplied uniformly to the lawn. Why not?



A little calculus answered that question and advised how the holes should be placed to have an equitable distribution. For convenience, it was assumed that the radius of the spherical head was 1, that the speed of the water as it left the head was the same at any hole, and air resistance was disregarded.

Consider the water contributed to the lawn by the uniformly spaced holes in a narrow band of width $d\phi$ near the angle ϕ, as shown in Figure C.7.2. To be sure the jet was not blocked by the grass, the angle ϕ is assumed to be no more than $\pi/4$.

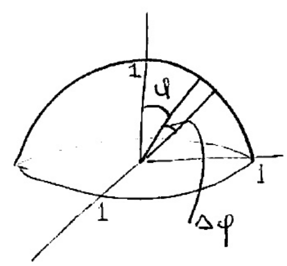


Figure C.7.2

Water from this band wets the ring shown in Figure C.7.3.

The area of the band on the sprinkler is roughly $2\pi\sin(\phi)\,d\phi$. As shown in Section 9.3, see Exercises 24 and 25, water from this band lands at a distance from the sprinkler of about

$$x = kv^2\sin(2\phi).$$

Here k is a constant and v is the speed of the water as it leaves the sprinkler. The width of the corresponding ring on the lawn is roughly

$$dx = 2kv^2\cos(2\phi)d\phi.$$

Since its radius is approximately $kv^2\sin(2\phi)$, its area is approximately

$$2\pi\left(kv^2\sin(2\phi)\right)\left(2kv^2\cos(2\phi)\,d\phi\right),$$

which is proportional to $\sin(2\phi)\cos(2\phi)$, hence to $\sin(4\phi)$.

Thus the water supplied by the band was proportional to $\sin(\phi)$ but the area watered by that band was proportional to $\sin(4\phi)$. The ratio

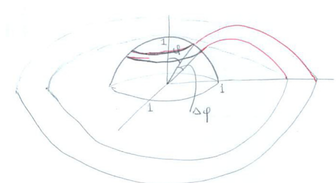


Figure C.7.3

$$\frac{\sin(4\phi)}{\sin(\phi)} = \frac{\text{Area watered on lawn}}{\text{Area of supply on sprinkler}}$$

is the key to understanding both why the distribution was not uniform and to finding out how the holes should be placed to water the lawn uniformly.

By l'Hôpital's rule, this fraction approaches 4 as ϕ approaches zero:

$$\lim_{\phi \to 0} \frac{\sin(4\phi)}{\sin(\phi)} = 4. \tag{C.7.1}$$

This means that for angles ϕ near 0 that ratio is near 4. When ϕ is $\pi/4$, that ratio is $\frac{\sin(\pi)}{\sin(\pi/4)} = 0$, and water was supplied much more heavily far from the sprinkler than near it. To compensate for this bias the number of holes in the band should be proportional to $\sin(4\phi)/\sin(\phi)$. Then the amount of water is proportional to the area watered, and watering is therefore uniform.

Professor Anthony Wexler of the Mechanical Engineering Department of UC-Davis calculated where to drill the holes and made a prototype, which produced a beautiful fountain and a much more even supply of water. Moreover, if some of the holes were removed, it would water a rectangular lawn.

We offered the idea to the firm that made the biased sprinkler. After keeping the prototype for half a year, it turned it down because "it would compete with the product we have."

Recently, we mentioned our sprinkler to another manufacturer, who showed great interest. In fact, the State of California has introduced criteria on how uniformly a sprinkler must distribute water. Perhaps our uniform sprinkler may eventually water many a lawn.

EXERCISES

1. Show that the limit (C.7.1) is 4

(a) using only trigonometric identities.

(b) using l'Hôpital's rule.

2. Show that $\sin(4x)/\sin(x)$ is a decreasing function for x in the interval $[0, \pi/4]$. (Use trigonometric identities and no calculus. (However, you may be amused if you also do this by calculus.))

3. An oscillating sprinkler goes back and forth at a fixed angular speed.

(a) Does it water a lawn uniformly?

(b) If not, how would you modify it to provide more uniform coverage?

Chapter 6

The Definite Integral

Up to this point we have been concerned with the derivative, which provides local information, such as the slope at a particular point on a curve or the velocity at a particular time. Now we introduce the second major concept of calculus, the definite integral. In contrast to the derivative, the definite integral provides global information, such as the area under a curve.

Section 6.1 motivates the definite integral through three of its applications. Section 6.2 defines the definite integral and Section 6.3 presents ways to estimate it. Sections 6.4 and 6.5 develop the connection between the derivative and the definite integral, which culminates in the Fundamental Theorems of Calculus. The derivative turns out to be essential for evaluating many definite integrals.

Chapters 2 to 6 form the core of calculus. Later chapters are mostly variations or applications of the key ideas in those chapters.

6.1 Three Problems That Are One Problem

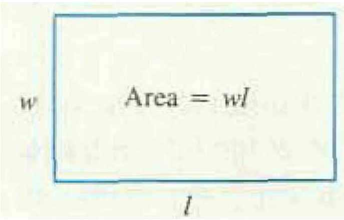


Figure 6.1.1

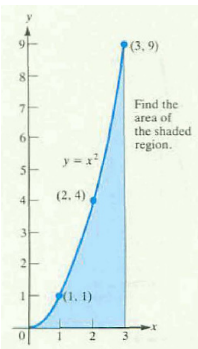


Figure 6.1.2

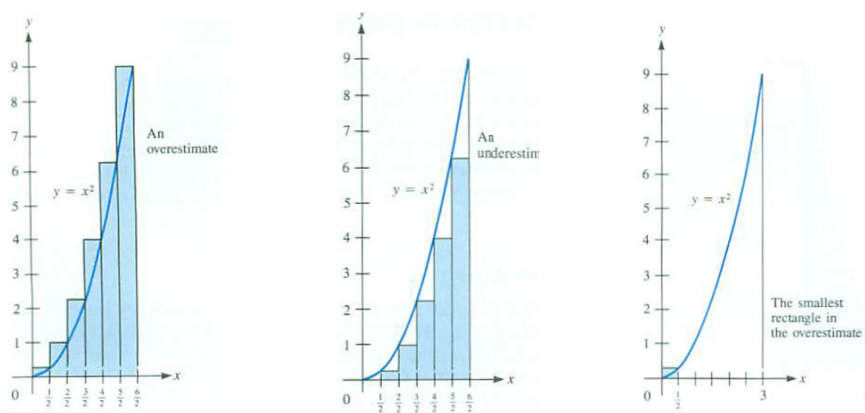
The definite integral is introduced with three problems. At first glance these problems may seem unrelated, but by the end of the section it will be clear that they represent one basic problem in various guises. They lead up to the concept of the definite integral, defined in the next section.

Estimating an Area

It is easy to find the exact area of a rectangle: multiply its length by its width (see Figure 6.1.1). But how do you find the area of the region in Figure 6.1.2? In this section we will show how to make accurate *estimates* of that area. The technique we use will lead up in the next section to the definition of the definite integral of a function.

PROBLEM 1 Estimate the area of the region bounded by the curve $y = x^2$, the x-axis, and the vertical line $x = 3$, as shown in Figure 6.1.2.

Since we know how to find the area of a rectangle, we will use rectangles to approximate the region. Figure 6.1.3(a) shows an approximation by six rectangles whose total area is more than the area under the parabola. Figure 6.1.3(b) shows a similar approximation whose area is less than the area under the parabola.



(a) (b) (c)

Figure 6.1.3

In each case we break the interval $[0, 3]$ into six short intervals, all of width $\frac{1}{2}$. In order to find the areas of the overestimate and of the underestimate, we must find the height of each rectangle. That height is determined by the curve $y = x^2$. Let us examine only the overestimate, leaving the underestimate for the Exercises.

There are six rectangles in the overestimate shown in Figure 6.1.3(a). The smallest rectangle is shown in Figure 6.1.3(c). The height of this rectangle is equal to the value of x^2 when $x = \frac{1}{2}$. Its height is therefore $\left(\frac{1}{2}\right)^2$ and its area is $\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)$, the product of its height and its width. The areas of the other five rectangles can be found similarly. In each case evaluate x^2 at the right end of the rectangle's base in order to find the height. The total area of the six rectangles is

$$\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{2}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{3}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{4}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{5}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{6}{2}\right)^2\left(\frac{1}{2}\right).$$

This equals

$$\frac{1}{8}\left(1^2+2^2+3^2+4^2+5^2+6^2\right)=\frac{91}{8}=11.375. \tag{6.1.1}$$

The area under the parabola is therefore less than 11.375.

To get a closer estimate we should use more rectangles. Figure 6.1.4 shows an overestimate in which there are 12 rectangles. Each has width $\frac{3}{12}=\frac{1}{4}$. The total area of the overestimate is

$$\left(\frac{1}{4}\right)^2\left(\frac{1}{4}\right)+\left(\frac{2}{4}\right)^2\left(\frac{1}{4}\right)+\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)+\cdots+\left(\frac{12}{4}\right)^2\left(\frac{1}{4}\right).$$

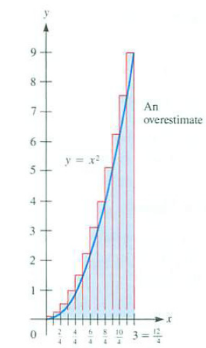


Figure 6.1.4

This equals

$$\frac{1}{4^3}\left(1^2+2^2+3^2+\cdots+12^2\right)=\frac{650}{64}=10.15625. \tag{6.1.2}$$

Now we know the area under the parabola is less than 10.15625.

To get closer estimates we would cut the interval $[0,3]$ into more sections, maybe 100 or 10,000 or more, and calculate the total area of the corresponding rectangles. (This is an easy computation on a computer.)

Figure 6.1.5

In general, we would divide $[0,3]$ into n sections of equal length. The length of each section is then $\frac{3}{n}$. Their endpoints are shown in Figure 6.1.5.

Then, for each integer $i = 1,\ 2,\ \ldots, n$, the i^{th} section from the left has endpoints $(i-1)\left(\frac{3}{n}\right)$ and $i\left(\frac{3}{n}\right)$, as shown in Figure 6.1.6.

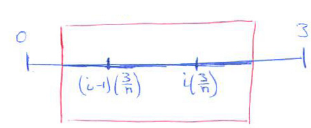


Figure 6.1.6 [ARTIST: Redraw Figure 6.1.6 to give effect of zooming in on ith interval]

To make an overestimate, observe that x^2 is increasing for $x > 0$ and evaluate x^2 at the right endpoint of each interval. Then multiply the result by the width of the interval, getting

$$\left(i\left(\frac{3}{n}\right)\right)^2\frac{3}{n}=3^3\frac{i^2}{n^3}.$$

Then, sum these overestimates for all n intervals:

$$3^3\frac{1^2}{n^3}+3^3\frac{2^2}{n^3}+3^3\frac{3^2}{n^3}+\cdots+3^3\frac{(n-1)^2}{n^3}+3^3\frac{n^2}{n^3}$$

which simplifies to

$$3^3\left(\frac{1^2+2^2+3^2+\cdots+(n-1)^2+n^2}{n^3}\right). \tag{6.1.3}$$

In summation notation, this equals

$$\frac{3^3}{n^3}\sum_{i=1}^{n} i^2.$$

We have already seen that these overestimates become more and more accurate as the number of intervals increases. We would like to know what happens to the overestimate as n gets larger and larger. More specifically, does

$$\lim_{n\to\infty}\frac{1^2+2^2+3^2+\cdots+(n-1)^2+n^2}{n^3} \tag{6.1.4}$$

exist? If it does exist, call it L. (Then the area would be 3^3L.)

The numerator gets large, tending to make the fraction large. But the denominator also gets large, which tends to make the fraction small. Once again we encounter one of the "limit battles" that occurs in the foundation of calculus.

Archimedes, some 2200 years ago, found a short formula for the numerator in (6.1.3), enabling him to find the limit in (6.1.4). See, for instance, S. Stein, "Archimedes: What did he do besides cry Eureka?".

To estimate L, use, say, $n = 6$. Then we have

$$\frac{1}{6^3}\left(1^2+2^2+3^2+4^2+5^2+6^2\right)=\frac{91}{216}\approx 0.42130.$$

Try a larger value of n to get a closer estimate of L.

If we knew L we would know the area under the parabola and above the interval $[0,3]$, for the area is 3^3L. Since we do not know L, we don't know the area. Be patient. We will find L indirectly in this section. You may want to compute the quotient in (6.1.4) for some n and guess what L is. For example, with $n = 12$, the estimate is $\frac{650}{12^3}=\frac{650}{1728}\approx 0.37616$.

Estimating a Distance Traveled

The units simplify: $\frac{\text{mi}}{\text{hr}}\times\text{hr}=\text{mi}$.

If you drive at a constant speed of v miles per hour for a period of t hours, you travel vt miles:

$$\text{Distance} = \text{Speed}\times\text{Time} = vt \text{ miles}.$$

But how would you compute the total distance traveled if your speed were not constant? (Imagine that your odometer, which records distance traveled, was broken. However, your speedometer is still working fine, so you know your speed at any instant.) The next problem illustrates how you could make accurate estimates of the total distance traveled.

PROBLEM 2 A snail is crawling about for three minutes. This remarkable snail knows that she is traveling at the rate of t^2 feet per minute at time t minutes. For instance, after half a minute, she is slowly moving at the rate of $\left(\frac{1}{2}\right)^2$ feet per minute. At the end of her journey she is moving along at 3^2 feet per minute. Estimate how far she travels during the three minutes.

The speed during the three-minute trip increases from 0 to 9 feet per minute. During shorter time intervals, such a wide fluctuation does not occur. As in Problem 1, cut the three minutes of the trip into six equal intervals each 1/2 minute long, and use them to estimate the total distance covered. Represent time by a line segment cut into six parts of equal length, as in Figure 6.1.7.

Figure 6.1.7

Consider the distance she travels during one of the six half-minute intervals, say during the interval $[\frac{3}{2}, \frac{4}{2}]$. At the beginning of this time interval her speed was $\left(\frac{3}{2}\right)^2$ feet per minute; at the end she was going $\left(\frac{4}{2}\right)^2$ feet per minute. The highest speed during this half hour was $\left(\frac{4}{2}\right)^2$ feet per minute. Therefore, she traveled at most $\left(\frac{4}{2}\right)^2\left(\frac{1}{2}\right)$ feet during the time interval $[3/2, 4/2]$. Similar reasoning applies to the other five half-minute periods. Adding up these upper estimates for the distance traveled during each interval of time, we get the total distance traveled is less than

Speed increases as t increases.

$$\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{2}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{3}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{4}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{5}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{6}{2}\right)^2\left(\frac{1}{2}\right).$$

If we divide the time interval into n equal sections of duration $\frac{3}{n}$, the right endpoint of the i^{th} interval is $i\left(\frac{3}{n}\right)$. At that time the speed is $(3i/n)^2$ feet per minute. So the distance covered during the i^{th} interval of time is less than

$$\underbrace{\left(\frac{3i}{n}\right)^2}_{\text{max speed}}\underbrace{\frac{3}{n}}_{\text{time}} = \frac{3^3 i^2}{n^3}.$$

The total overestimate is then

$$3^3\frac{1^2}{n^3} + 3^3\frac{2^2}{n^3} + 3^3\frac{3^2}{n^3} + \cdots + 3^3\frac{(n-1)^2}{n^3} + 3^3\frac{n^2}{n^3}$$

or

$$3^3\left(\frac{1^2+2^2+3^2+\cdots+(n-1)^2+n^2}{n^3}\right). \tag{6.1.5}$$

The calculations in the area problem, (6.1.3), and in the distance problem, (6.1.5), are the same. Thus, the area and distance have the same upper estimates. Their lower estimates are also the same, as you may check. The limit of (6.1.5) is 3^3L. The two problems are really the same problem.

Estimating a Volume

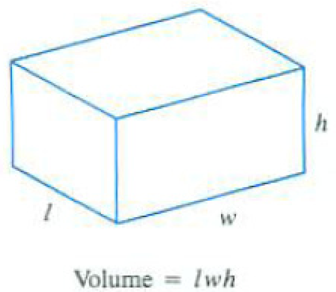
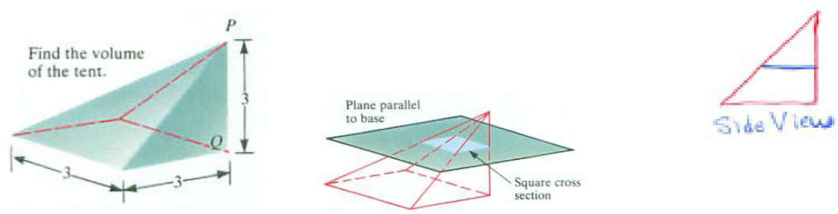


Figure 6.1.8

The volume of a rectangular box is easy to compute; it is the product of its length, width, and height. See Figure 6.1.8. But finding the volume of a pyramid or ball requires more work. The next example illustrates how we can estimate the volume inside a certain tent.

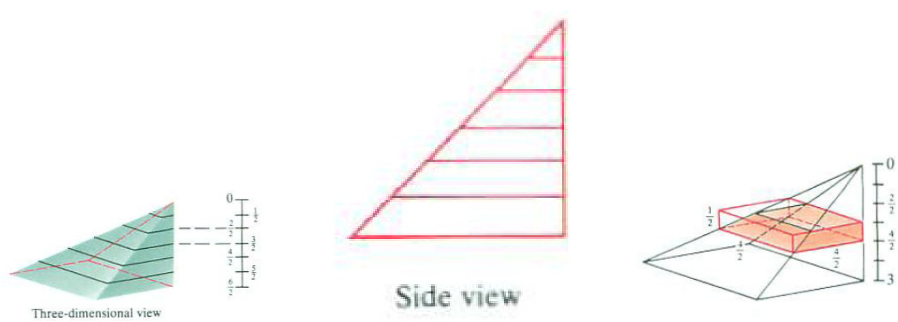
PROBLEM 3 Estimate the volume inside a tent with a square floor of side 3 feet, whose vertical pole, 3 feet long, is located *above one corner* of the floor. The tent is shown in Figure 6.1.9(a).



(a) (b) (c)

Figure 6.1.9

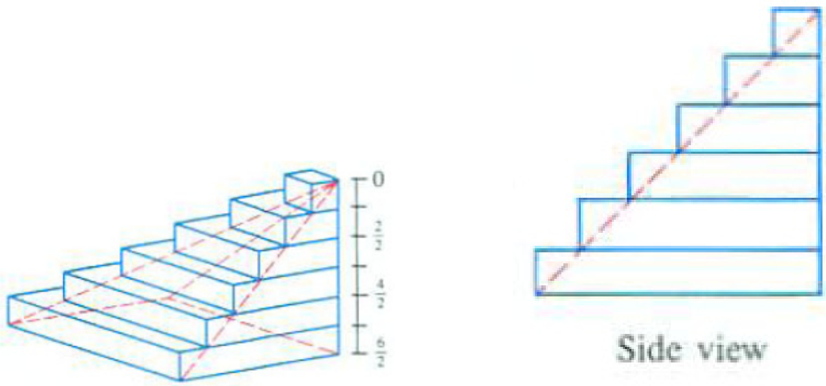
The cross section of the tent made by any plane parallel to the base is a square, as shown in Figure 6.1.9(b). The width of the square equals its distance from the top of the pole, as shown in Figure 6.1.9(c). Using this fact, we can approximate the volume inside the tent with rectangular boxes with square cross sections. Begin by cutting a vertical line, representing the pole, into six



(a) (b) (c)

Figure 6.1.10

sections of equal length, each $\frac{1}{2}$ foot long. Draw the corresponding square cross section of the tent, as in Figure 6.1.10(a). Use these square cross sections to



(a) (b)

Figure 6.1.11

form rectangular boxes. Consider the part of the tent corresponding to the interval $[\frac{3}{2}, \frac{4}{2}]$ on the pole. The base of this section is a square with sides $\frac{4}{2}$ feet. The box with this square as a base and height $\frac{1}{2}$ foot encloses completely the part of the tent corresponding to $[\frac{3}{2}, \frac{4}{2}]$. (See Figure 6.1.10(c).) The volume of this box is $\left(\frac{4}{2}\right)^2\left(\frac{1}{2}\right)$ cubic feet. Figures Figure 6.1.11(a) and (b) show six such boxes, whose total volume is greater than the volume of the tent.

Since the volume of each box is the area of its base times its height, the total volume of the six boxes is

$$\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{2}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{3}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{4}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{5}{2}\right)^2\left(\frac{1}{2}\right)+\left(\frac{6}{2}\right)^2\left(\frac{1}{2}\right) \quad \text{cubic feet.}$$

This sum, which we have encountered twice before, equals 11.375. It is an *overestimate* of the volume of the tent. Better (over)estimates can be obtained by cutting the pole into shorter pieces. Evidently, the arithmetic for the tent volume is the same as for the previous two problems.

We now know that the number describing the volume of the tent is the same as the number describing the area under the parabola and also the length of the snail's journey. That number is 3^3L. The arithmetic of the estimates is the same in all three cases.

A Neat Bit of Geometry

If we knew the limit L in (6.1.3), we would then find the answers to all three problems. But we haven't found L. Luckily, there is a way to find the volume of the tent without knowing L.

The key is that *three identical copies of the tent* fill up a cube of side 3

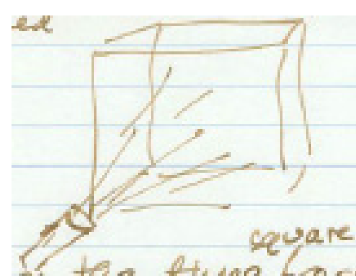
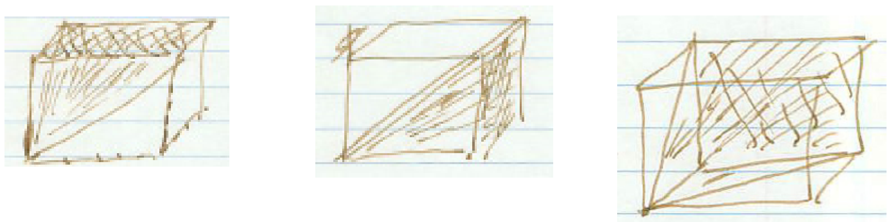


Figure 6.1.12

feet. To see why, imagine a flashlight at one corner of the cube, aimed into the cube, as in Figure 6.1.12.



(a) (b) (c)

Figure 6.1.13

This trick is like the way the area of a right triangle is found by arranging two copies to form a rectangle.

The flashlight illuminates the three square faces not meeting the corner at the flashlight. The rays from the flashlight to the top, side, and back, as shown in Figure 6.1.13(a), (b), and (c), respectively, fill out a copy of the tent.

Since three copies of the tent fill a cube of volume $3^3 = 27$ cubic feet, the tent has volume 9 cubic feet. From this, we see that the area under the parabola above $[0, 3]$ is 9 and the snail travels 9 feet. Incidentally, the limit L must be $\frac{1}{3}$, since the area under the parabola is both 9 and 3^3L. In short,

$$\lim_{n\to\infty} \frac{1^2 + 2^2 + 3^2 + \cdots + (n-1)^2 + n^2}{n^3} = \lim_{n\to\infty} \frac{1}{n^3}\sum_{i=1}^{n} i^2 = \frac{1}{3}$$

Summary

Using upper estimates, we showed that problems concerning area, distance traveled, and volume were the same problem in various disguises. We were really studying a problem concerning a particular function, x^2, over a particular interval $[0, 3]$. We solved this problem by cutting a cube into three congruent pieces. By the end of this chapter you will learn general techniques that will make such a special device unnecessary.

EXERCISES for Section 6.1

Exercises 1 to 21 concern estimates of areas under curves.

1. In Problem 1 we broke the interval $[0,3]$ into six sections. Instead, break $[0,3]$ into four sections of equal lengths and estimate the area under $y = x^2$ and above $[0,3]$ as follows.

(a) Draw the four rectangles whose total area is larger than the area under the curve. The value of x^2 at the right endpoint of each section determines the height of each rectangle.

(b) On the diagram in (a), show the height and width of each rectangle.

(c) Find the total area of the four rectangles.

2. Like Exercise 1, but this time obtain an underestimate of the area by using the value of x^2 at the left endpoint of each section to determine the height of the rectangle.

3. Estimate the area under $y = x^2$ and above $[1,2]$ using the five rectangles with equal widths shown in Figure 6.1.14(a).

4. Repeat Exercise 3 with the five rectangles in Figure 6.1.14(b).

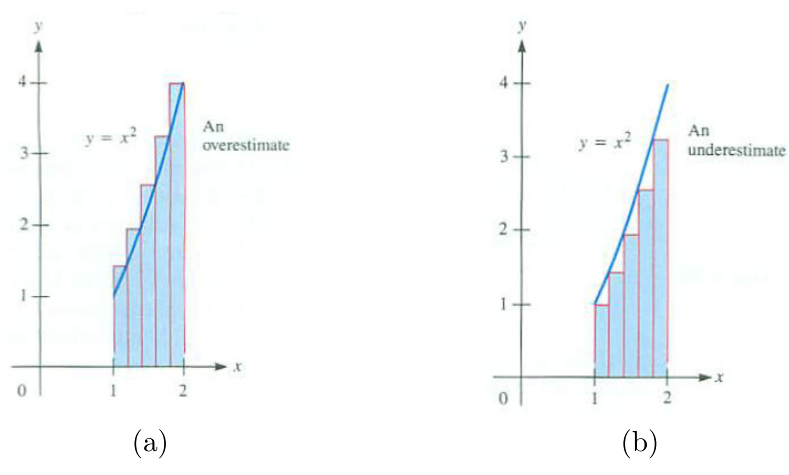


Figure 6.1.14

5. Evaluate

(a) $\sum_{i=1}^{4} i^2$

(b) $\sum_{i=1}^{4} 2^i$

(c) $\sum_{n=3}^{4} (n-3)$

6. Evaluate

(a) $\sum_{i=1}^{4} i^3$

(b) $\sum_{i=2}^{5} 2^i$

(c) $\sum_{k=1}^{4}(k^3 - k^2)$

7. Figure 6.1.15(a) shows the curve $y = \frac{1}{x}$ above the interval $[1, 2]$ and an approximation to the area under the curve by five rectangles of equal width.

(a) Make a large copy of Figure 6.1.15(a).

(b) On your diagram show the height and width of each rectangle.

(c) Find the total area of the five rectangles.

(d) Find the total area of the five rectangles in Figure 6.1.15(b).

(e) On the basis of (c) and (d), what can you say about the area under the curve $y = 1/x$ and above $[1, 2]$?

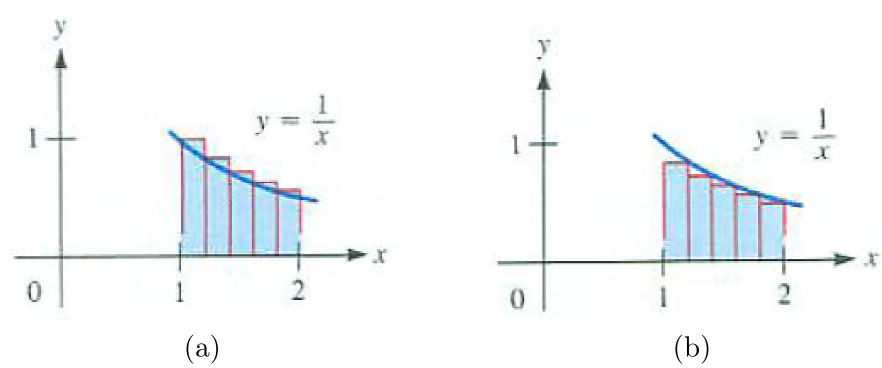


Figure 6.1.15

Exercises 8 and 9 develop underestimates for each of the problems considered in this section.

8. In Problem 1 we found overestimates for the area under the parabola x^2 over the interval $[0, 3]$. Here we obtain underestimates for this area as follows.

(a) Break $[0, 3]$ into six sections of equal lengths and draw the six rectangles whose total area is smaller than the area under the curve.

(b) Because x^2 is increasing on $[0, 3]$, the left endpoint of each section determines the height of each rectangle. Show the height and width of each rectangle you drew in (a).

(c) Find the total area of the six rectangles.

9. Repeat Exercise 8 with 12 sections of equal lengths.

10. Consider the area under $y = 2^x$ and above $[-1, 1]$.

(a) Graph the curve and estimate the area by eye.

(b) Make an overestimate of the area, using four sections of equal width.

(c) Make an underestimate of the area, using four sections of equal width.

11. Use the information found in Exercises 3 and 4 to complete this sentence: The area in Problem 1 is certainly less than ______ but larger than ______.

12. Estimate the area in Problem 1, using the division of $[0, 3]$ into four sections with endpoints 0, 1, $\frac{5}{3}$, $\frac{11}{4}$, and 3 (see Figure 6.1.16(a)).

(a) Estimate the area when the right-hand endpoints of each section are used to find the heights of the rectangles.

(b) Repeat (a), using the left-hand endpoints of each section to find the heights of the rectangles.

(c) Repeat (a) computing the heights of the rectangles at the points $\frac{1}{2}$, $\frac{3}{2}$, 2, and $\frac{14}{5}$.

Figure 6.1.16

In each of Exercises 13 to 18

(a) Draw the region.

(b) Draw six rectangles of equal widths whose total area overestimates the area of the region.

(c) On your diagram indicate the height and width of each rectangle.

(d) Find the total area of the six rectangles. (Give this answer accurate to two decimal places.)

13. Under $y = x^2$, above $[2, 3]$.

14. Under $y = \frac{1}{x}$, above $[2, 3]$.

15. Under $y = x^3$, above $[0, 1]$.

16. Under $y = \sqrt{x}$, above $[1, 4]$.

17. Under $y = \sin(x)$, above $[0, \pi/2]$.

18. Under $y = \ln(x)$, above $[1, e]$.

19. Estimate the area under $y = x^2$ and above $[-1, 2]$ by dividing the interval into six sections of equal lengths.

(a) Draw the six rectangles that form an overestimate for the area under the curve. Note that you cannot do this using only left-endpoints or only right-endpoints.

(b) Find the total area of all six rectangles.

(c) Repeat (a) and (b) to find an underestimate for this area.

20. Estimate the area between the curve $y = x^3$, the x-axis, and the vertical line $x = 6$ using a division into

(a) six sections of equal lengths with left endpoints;

(b) six sections of equal lengths with right endpoints;

(c) three sections of equal lengths with midpoints;

(d) six sections of equal lengths with midpoints.

21. Estimate the area below the curve $y = \frac{1}{x^2}$ and above $[1, 7]$ following the directions in Exercise 20.

22. To estimate the area in Problem 1 you divide the interval $[0, 3]$ into n sections of equal lengths. Using the right-hand endpoint of each of the n sections you then obtain an overestimate. Using the left-hand endpoint, you obtain an underestimate.

(a) Show that these two estimates differ by $\frac{27}{n}$.

(b) How large should n be chosen in order to be sure the difference between the upper estimate and the area under the parabola is less than 0.01?

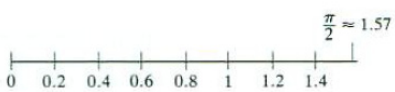
23. Estimate the area of the region under the curve $y = \sin(x)$ and above the interval $[0, \frac{\pi}{2}]$, cutting the interval as shown in Figure 6.1.17(a) and using

(a) left endpoints

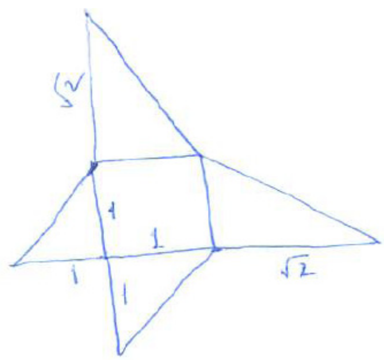
(b) right endpoints

(c) midpoints.

(All but the last section are of the same length.)



(a)



(b)

Figure 6.1.17

24. Make three copies of the tent in Problem 3 by folding a pattern as shown in Figure 6.1.17(b). Check that they fill up a cube.

25. An electron is being accelerated in such a way that its velocity is t^3 kilometers per second after t seconds. Estimate how far it travels in the first 4 seconds, as follows:

(a) Draw the interval $[0, 4]$ as the time axis and cut it into eight sections of equal length.

(b) Using the sections in (a), make an estimate that is too large.

(c) Using the sections in (a), make an estimate that is too small.

26. A business which now shows no profit is to increase its profit flow gradually in the next 3 years until it reaches a rate of 9 million dollars per year. At the end of the first half year the rate is to be $\frac{1}{4}$ million dollars per year; at the end of 2 years, 4 million dollars per year. In general, at the end of t years, where t is any number between 0 and 3, the rate of profit is to be t^2 million dollars per year. Estimate the total profit during its first 3 years if the plan is successful using

(a) using six intervals and left endpoints;

(b) using six intervals and right endpoints;

(c) using six intervals and midpoints.

27. Oil is leaking out of a tank at the rate of 2^{-t} gallons per minute after t minutes. Describe how you would estimate how much oil leaks out during the first 10 minutes. Illustrate your procedure by computing one estimate.

28. Archimedes showed that $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$. You can prove this as follows:

(a) Check that the formula is correct for $n = 1$.

(b) Show that if the formula is correct for the integer n, it is also correct for the next integer, $n + 1$.

(c) Why do (a) and (b) together show that Archimedes' formula holds for all positive integers n?

This type of proof is known as **mathematical induction**.

29.

(a) Explain why the area of the region under the curve $y = x^2$ and above the interval $[0, b]$ is given by
$$\lim_{n\to\infty} \sum_{i=1}^{n} \left(\frac{bi}{n}\right)^2 \frac{b}{n}.$$

(b) Use Exercise 28 to find this limit.

(c) Give an explicit formula for the area of the region under $y = x^2$ and above $[0, b]$.

(d) For $0 < a < b$, what is the area under the curve $y = x^2$ and above the interval $[a, b]$?

30. The function $f(x)$ is increasing for x in the interval $[a, b]$ and is positive. To estimate the area under the graph of $y = f(x)$ and above $[a, b]$ you divide the interval $[a, b]$ into n sections of equal lengths. You then form an overestimate B (for "big") using right-hand endpoints of the sections and an underestimate S (for "small") using left-hand endpoints. Express the difference between the two estimates, $B - S$, as simply as possible.

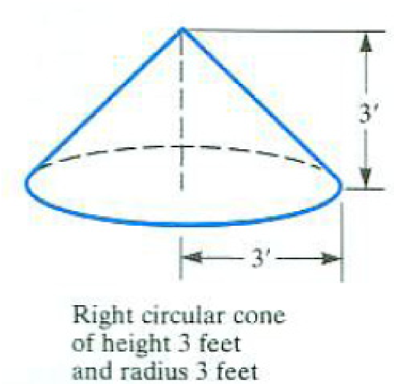
31. A right circular cone has a height of 3 feet and a radius of 3 feet, as shown in Figure 6.1.18(a). Estimate its volume by the sum of the volumes of six cylindrical slabs, just as we estimated the volume of the tent with the aid of six rectangular slabs.

(a) Make a large and neat diagram that shows the six cylinders used in making an overestimate.

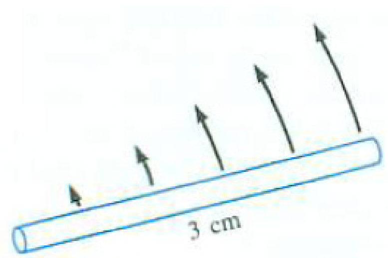
(b) Compute the total volume of the six cylinders in (a).

(c) Make a separate diagram showing a corresponding underestimate.

(d) Compute the total volume of the six cylinders in (c). (Note: One of the cylinders has radius 0.)



(a)



(b)

Figure 6.1.18

32. The kinetic energy of an object, for example, a baseball or car, of mass m grams and speed v centimeters per second is defined as $\frac{1}{2}mv^2$ ergs. Now, in a certain machine a uniform rod 3 centimeters long and weighing 32 grams rotates once per second around one of its ends as shown in Figure 6.1.18(b). Estimate the kinetic energy of this rod by cutting it into six sections, each $\frac{1}{2}$ centimeter long, and taking as the "speed of a section" the speed of its midpoint.

33. Express the sum $\displaystyle\sum_{i=1}^{n} \ln\left(\frac{i+1}{i}\right)$ as simply as possible. (So that you could compute the sum in the fewest steps.)

Skill Drill

In Exercises 34 to 39 differentiate the expression.

34. $(1+x^2)^{4/3}$

35. $\frac{(1+x^3)\sin(3x)}{\sqrt[3]{5x}}$

36. $\frac{3x}{8}+\frac{3x\sin(4x)}{32}+\frac{\cos^3(2x)\sin(2x)}{8}$

37. $\frac{3}{8(2x+3)^2}-\frac{1}{4(2x+3)}$

38. $\frac{\cos^3(2x)}{6}-\frac{\cos(2x)}{2}$

39. $x^3\sqrt{x^2-1}\tan(5x)$

In Exercises 40 to 50 give an antiderivative of the expression.

40. $(x+2)^3$

41. $(x^2+1)^2$

42. $x\sin(x^2)$

43. $x^3+\frac{1}{x^3}$

44. $\frac{1}{\sqrt{x}}$

45. $\frac{3}{x}$

46. e^{3x}

47. $\frac{1}{1+x^2}$

48. $\frac{1}{x^2}$

49. 2^x

50. $\frac{4}{\sqrt{1-x^2}}$

6.2 The Definite Integral

We now introduce the other main concept in calculus, the "definite integral of a function over an interval."

The preceding section was not really about area under a parabola, distance a snail traveled, or volume of a tent. The common theme of all three was a procedure we carried out with the function x^2 and the interval $[0, 3]$: Cut the interval into small pieces, evaluate the function somewhere in each section, form certain sums, and then see how those sums behave as we choose the sections smaller and smaller.

Here is the general procedure. We have a function f defined at least on an interval $[a, b]$. We cut, or "partition," the interval into n sections by the numbers $x_0 = a$, x_1, x_2, …, x_{n-1}, $x_n = b$, as in Figure 6.2.1. They need not

The sections $[a, x_1]$, $[x_1, x_2]$, …, $[x_{n-1}, b]$ form a **partition** of $[a, b]$.

Figure 6.2.1

all be of the same length, though usually, for convenience, they will be.

Then we pick a **sampling number** in each interval, c_1 in $[x_0, x_1]$, c_2 in $[x_1, x_2]$, …, c_i in $[x_{i-1}, x_i]$, …, c_n in $[x_{n-1}, x_n]$ (as in Figure 6.2.1). In Section 6.1, the c_i's were mostly either right-hand or left-hand endpoints or midpoints. However, they can be anywhere in each section.

Next we bring in the particular function f. (In Section 6.1 the function was x^2.) We evaluate that function at each c_i and form the sum

$$f(c_1)(x_1 - x_0) + f(c_2)(x_2 - x_1) + \cdots + f(c_i)(x_i - x_{i-1}) \qquad (6.2.1)$$
$$+ \cdots + f(c_{n-1})(x_{n-1} - x_{n-2}) + f(c_n)(x_n - x_{n-1}). \qquad (6.2.2)$$

Rather than continue to write out such a long expression, we choose to take advantage of the fact that each term in (6.2.1) follows the same general pattern: for each of the n sections, multiply the function value at the sampling number by the length of the section. This pattern is easily expressed in the shorthand Σ-notation as:

$$\sum_{i=1}^{n} f(c_i)(x_i - x_{i-1}). \qquad (6.2.3)$$

If the length of section i is written as $\Delta x_i = x_i - x_{i-1}$, the expression for the sum becomes even shorter:

$$\sum_{i=1}^{n} f(c_i)\Delta x_i. \tag{6.2.4}$$

If all the sections have the same length, each Δx_i equals $(b-a)/n$, since the length of $[a,b]$ is $b-a$. Let Δx denote $\frac{b-a}{n}$. We can write (6.2.3) and (6.2.4) also as

$$\sum_{i=1}^{n} f(c_i)\left(\frac{b-a}{n}\right) \qquad \text{or as} \qquad \sum_{i=1}^{n} f(c_i)\Delta x \tag{6.2.5}$$

where $\Delta x = \frac{b-a}{n}$.

The final step is to investigate what happens to the sums of the form (6.2.4) (or (6.2.5)) as the lengths of all the sections approach 0. That is, we try to find

$$\lim_{\text{all } \Delta x_i \text{ approach } 0} \sum_{i=1}^{n} f(c_i)\Delta x_i. \tag{6.2.6}$$

The sums in (6.2.1)–(6.2.5) are called **Riemann sums** in honor of the nineteenth century mathematician, Bernhard Riemann.

Bernhard Riemann, 1826–1866, http://en.wikipedia.org/wiki/Bernhard_Riemann.

In advanced mathematics it is proved that if f is continuous on $[a,b]$ then the sums in (6.2.6) do approach a single number. This brings us to the definition of the definite integral.

The Definite Integral

DEFINITION *(Definite Integral of a function f over an interval $[a,b]$)* Let f be a continuous function defined at least on the interval $[a,b]$. The limit of sums of the form $\sum_{i=1}^{n} f(c_i)\Delta x_i$, for partitions of $[a,b]$ where every Δx_i approaches 0, exists (no matter how the sampling numbers c_i are chosen). The limiting value is called the **definite integral of f over the interval** $[a,b]$ and is denoted

$$\int_a^b f(x)\,dx.$$

The symbol $\int$ comes from "S," for "sum". The "dx," strictly speaking, is not needed. Both symbols were introduced by Liebniz.

Gottfried Liebniz, 1646–1716, http://en.wikipedia.org/wiki/Gottfried_Leibniz.

The limit in this definition is a little unusual. It requires the length of every segment within the partition to approach 0. It is not sufficient to simply consider partitions of $[a, b]$ with more and more segments as this does not prevent segments with lengths that do not approach 0. Another way of stating this requirement is that the length of the largest segment in the partion must approach zero.

EXAMPLE 1 Express the area under $y = x^2$ and above $[0, 3]$ as a definite integral.
SOLUTION Here the function is $f(x) = x^2$ and the interval is $[0, 3]$. As we saw in the previous section, the area equals the limit of Riemann sums

$$\lim_{\Delta x \to 0^+} \sum_{i=1}^{n} c_i^2 \Delta x = \int_0^3 x^2 \, dx. \tag{6.2.7}$$

◇

The dx traditionally suggests the length of a small section of the x-axis and denotes the **variable of integration** (usually x, as in this case). The function $f(x)$ is called the **integrand**, while the numbers a and b are called the **limits of integration**; a is the **lower limit of integration** and b is the **upper limit of integration**.

The symbol $\int_a^b x^2 \, dx$ is read as "the integral from a to b of x^2." Freeing ourselves from the variable x, we could say, "the integral from a to b of the squaring function". There is nothing special about the symbol x in "x^2." We could just as well have used the letter t — or any other letter. (We would typically pick a letter near the end of the alphabet, since letters near the beginning are customarily used to denote constants.) The notations

$$\int_a^b x^2 \, dx, \qquad \int_a^b t^2 \, dt, \qquad \int_a^b z^2 \, dz, \qquad \int_a^b u^2 \, du, \qquad \int_a^b \theta^2 \, d\theta$$

all denote the same number, that is, "the definite integral of the squaring function from a to b". Taken to the extreme, we could express (6.2.7) as

$$\int_a^b (\ )^2 \, d(\ ).$$

Usually, however, we find it more convenient to use some letter to name the independent variable. Since the letter chosen to represent the variable has no significance of its own, it is called a **dummy variable**. Later in this chapter there will be cases where the interval of integration is $[a, x]$ instead of $[a, b]$. Were we to write $\int_a^x x^2 \, dx$, it would be easy to think there is some relation between the x in x^2 and the x in the upper limit of integration. To avoid possible confusion, we prefer to use a different dummy variable and write, for example, $\int_a^x t^2\,dt$ in such cases.

It is important to realize that area, distance traveled, and volume are merely applications of the definite integral. (It is a mistake to link the definite integral too closely with one of its applications, just as it narrows our understanding of the number 2 to link it always with the idea of two fingers.) The definite integral $\int_a^b f(x)\,dx$ is also call the **Riemann integral**.

Slope and velocity are particular interpretations or applications of the derivative, which is a purely mathematical concept defined as a limit:

derivatives are limits

$$\text{derivative of } f \text{ at } x = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}.$$

Similarly, area, total distance, and volume are just particular interpretations of the definite integral, which is also defined as a limit:

definite integrals are also limits

$$\text{definite integral of } f \text{ over } [a,b] = \lim_{\text{as all } \Delta x_i \,\to\, 0^+} \sum_{i=1}^{n} f(c_i)(x_i - x_{i-1}).$$

The Definite Integral of a Constant Function

To bring the definition down to earth, let us use it to evaluate the definite integral of a constant function.

EXAMPLE 2 Let f be the function whose value at any number x is 4; that is, f is the constant function given by the formula $f(x) = 4$. *Use only the definition* of the definite integral to compute

$$\int_1^3 f(x)\,dx.$$

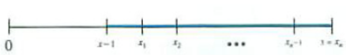

A typical partition of [1, 3]

Figure 6.2.2

SOLUTION In this case, every partition of the interval $[1,3]$ has $x_0 = 1$ and $x_n = 3$. See Figure 6.2.2. Since, no matter how the sampling number c_i is chosen, $f(c_i) = 4$, the approximating sum equals

$$\sum_{i=1}^{n} f(c_i)\Delta x = \sum_{i=1}^{n} f(c_i)(x_i - x_{i-1}) = \sum_{i=1}^{n} 4(x_i - x_{i-1})$$

Now

$$\sum_{i=1}^{n} 4(x_i - x_{i-1}) = 4\sum_{i=1}^{n}(x_i - x_{i-1}) = 4\cdot(x_n - x_0) = 4\cdot 2 = 8.$$

This is true because the sum of the widths of the sections is the width of the interval $[1, 3]$, namely 2. All approximating sums have the same value, namely, 8. For every partition,

$$\sum_{i=1}^{n} f(c_i)\Delta x = \sum_{i=1}^{n} f(c_i)(x_i - x_{i-1}) = 8.$$

Thus, as all sections are chosen smaller, the values of the sums are always 8. This number must be the limit:

$$\int_1^3 4\ dx = 8.$$

◇

We could have guessed the value of $\int_1^3 4\ dx$ by interpreting the definite integral as as area. To do so, draw a rectangle of height 4 and base coinciding with the interval $[1, 3]$. (See Figure 6.2.3.) Since the area of a rectangle is its base times its height, it follows again that $\int_1^3 4\ dx = 8$.

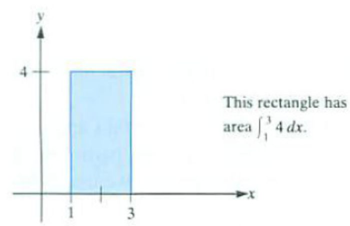


Figure 6.2.3

Similar reasoning shows that for any constant function that has the fixed value c,

$$\int_a^b c\ dx = c(b-a) \qquad (c \text{ is a constant function})$$

The Definite Integral of x

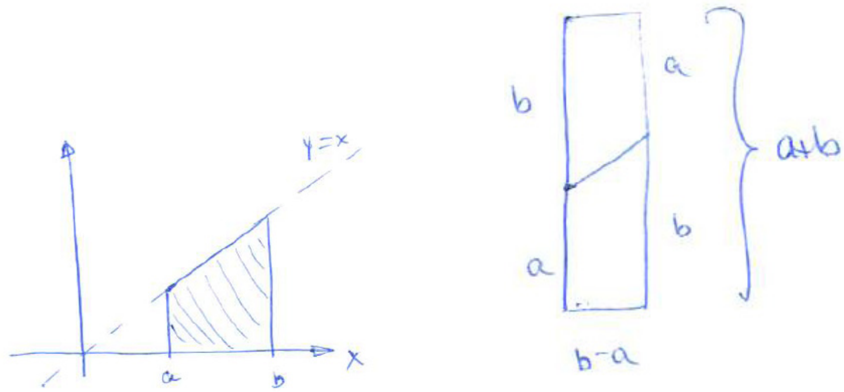
Exercise 34 shows us how to find $\int_a^b x\ dx$ directly from the definition. Alternatively, let us use the "area" interpretation of the definite integral to predict the value of $\int_a^b x\ dx$.

When the integrand is positive, that is, $0 < a < b$, the area in question then lies above the x-axis, as shown in Figure 6.2.4(a). Two copies of this region form a rectangle of width $b - a$ and height $a + b$, as shown in Figure 6.2.4(b). Thus, the area shown in Figure 6.2.4(a) is half of $(b - a)(b + a) = b^2 - a^2$. Hence,

$$\int_a^b x\ dx = \frac{b^2}{2} - \frac{a^2}{2}.$$

The Definite Integral of x^2

We will find $\int_0^b x^2\ dx$ by examining the approximating sums when all the sections have the same length, as they did in Section 6.1.



(a) (b)

Figure 6.2.4

Pick a positive integer n and cut the interval $[0, b]$ into n sections of length $\Delta x = b/n$ as in Figure 6.2.5. Then the points of subdivision are 0, Δx, $2\Delta x$, $\dots$, $(n-1)\Delta x$, and $n\Delta x = b$.

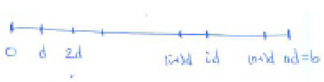


Figure 6.2.5

In the typical section $[(i-1)\Delta x, i\Delta x]$ we pick the right-hand endpoint as the sampling number. Thus the approximating sum is

$$\sum_{i=1}^{n} (i\Delta x)^2 \ (\Delta x) = (\Delta x)^3 \sum_{i=1}^{n} i^2.$$

Since $\Delta x = b/n$, these overestimates can be written as

$$\frac{b^3}{n^3} \sum_{i=1}^{n} i^2. \qquad (6.2.8)$$

Or, see Exercise 29 in Section 6.1.

In Section 6.1 we used geometry to find that

$$\lim_{n\to\infty} \frac{1}{n^3} \sum_{i=1}^{n} i^2 = \frac{1}{3}.$$

Thus, (6.2.8) approaches $b^3/3$ as n increases, and we conclude that

$$\int_0^b x^2 \, dx = \frac{b^3}{3}.$$

Note that when $b = 3$, we have $b^3/3 = 9$, agreeing with the three problems in Section 6.1.

A little geometry suggests the value of $\int_a^b x^2\,dx$, for $0 \le a < b$. Interpret $\int_a^b x^2\,dx$ as the area under $y = x^2$ and above $[a, b]$. This area is equal to the area under $y = x^2$ and above $[0, b]$ minus the area under $y = x^2$ and above $[0, a]$, as shown in Figure 6.2.6. Then

$$\int_a^b x^2\,dx = \frac{b^3}{3} - \frac{a^3}{3}.$$

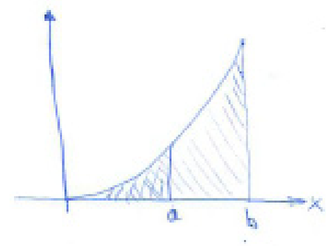


Figure 6.2.6

The Definite Integral of 2^x

EXAMPLE 3 Use the definition of the definite integral to evaluate $\int_0^b 2^x\,dx$. (Assume $b > 0$.)

SOLUTION Divide the interval $[0, b]$ into n sections of equal length, $d = b/n$. This time let's evaluate the integrand at the left-hand endpoint of each section. Call this number c_i, $c_i = (i-1)d$. The approximating sum has one term for each section. The contribution from the i^{th} section is

d = width of section

$$2^{c_i}d = 2^{(i-1)d}d.$$

The total estimate is the sum

$$2^0 d + 2^d d + 2^{2d} d + \cdots + 2^{(i-1)d} d + \cdots + 2^{(n-1)d} d.$$

This equals

$$d\left(1 + 2^d + (2^d)^2 + \cdots + (2^d)^i + \cdots + (2^d)^{n-1}\right). \tag{6.2.9}$$

The terms inside the large parentheses in (6.2.9) form a geometric progression with n terms, whose first term is 1 and whose ratio is 2^d. Thus, its sum is

Sum of geometric series: $a + ar + ar^2 + \cdots + ar^{n-1} = a\frac{1-r^n}{1-r}$.

$$\frac{1-(2^d)^n}{1-2^d}.$$

Therefore this typical underestimate is

$$\frac{d(1-(2^d)^n)}{1-2^d} = \frac{d(1-2^{dn})}{1-2^d} = \frac{d(1-2^b)}{1-2^d}. \tag{6.2.10}$$

In the last step we used the fact that $dn = b$. We can rewrite (6.2.10) as

$$\frac{d}{2^d-1}\left(2^b - 1\right). \tag{6.2.11}$$

It still remains to take the limit as n increases without bound. To find what happens to (6.2.11) as $n \to \infty$, we must investigate how $\frac{d}{2^d-1}$ behaves as

d approaches 0 (from the right). Though we haven't met this quotient before, we have met its reciprocal, $\frac{2^d-1}{d}$. This quotient occurs in the definition of the derivative of 2^x at $x = 0$:

$$\lim_{x\to 0} \frac{2^x - 2^0}{x} = \lim_{x\to 0} \frac{2^x - 1}{x}.$$

As we saw in Section 3.5, the derivative of 2^x is $2^x \ln(2)$. Thus $D(2^x)$ at $x = 0$ is $\ln(2)$. Hence

$$\lim_{d\to 0^+} \frac{d}{2^d - 1}(2^b - 1) = \lim_{d\to 0^+} \frac{1}{\left(\frac{2^d-1}{d}\right)} \left(2^b - 1\right) = \frac{2^b - 1}{\ln(2)}.$$

Incidentally, $\frac{1}{\ln(2)} \approx 1.443$.

We conclude that

$$\int_0^b 2^x \, dx = \frac{1}{\ln(2)} \left(2^b - 1\right).$$

$\diamond$

To evaluate $\int_a^b 2^x \, dx$ with $b > a \geq 0$, we reason as we did when we generalized $\int_0^b x^2 \, dx$ to $\int_a^b x^2 \, dx$. Namely,

$$\int_a^b 2^x \, dx = \int_0^b 2^x \, dx - \int_0^a 2^x \, dx = \frac{2^b - 1}{\ln(2)} - \frac{2^a - 1}{\ln(2)} = \frac{2^b}{\ln(2)} - \frac{2^a}{\ln(2)}.$$

Summary

We defined the definite integral of a function $f(x)$ over an interval $[a, b]$. It is the limit of sums of the form $\sum_{i=1}^{n} f(c_i)\Delta x_i$ created from partitions of $[a, b]$. It is a purely mathematical idea. You could estimate $\int_a^b f(x) \, dx$ with your calculator – even without having any application in mind. However, the definite integral has many applications: three of them are "area under a curve," "distance traveled," and "volume."

The following table contains a great deal of information. Compare the first three cases with the fourth, which describes the fundamental definition of integral calculus. In this table, all the functions, whether cross-sectional length, velocity, or cross-sectional area, are denoted by the same symbol $f(x)$.

Underlying these three applications is one purely mathematical concept, the definite integral, $\int_a^b f(x) \, dx$. The definite integral is defined as a certain limit; it is a number. It is essential to keep the definition of the number $\int_a^b f(x) \, dx$ clear. *It is a limit of certain sums.*

$f(x)$	$\sum_{i=1}^{n} f(c_i)(x_i - x_{i-1})$	$\int_a^b f(x)\ dx$
Variable length of cross section of set in plane	Approximate area of set in the plane	The area of set in the plane
Variable velocity	Approximation to the distance traveled	The distance traveled
Variable cross section of a solid	Approximate of volume	The volume of a solid
A function	Just a certain sum	The limit of the sums as the $\Delta x_i \to 0$

Spend some time studying this table. The concepts it summarizes will be used often.

EXERCISES for Section 6.2

1. Using the formula for $\int_a^b x^2\,dx$, find the area under the curve $y = x^2$ and above the interval

(a) $[0, 5]$

(b) $[0, 4]$

(c) $[4, 5]$

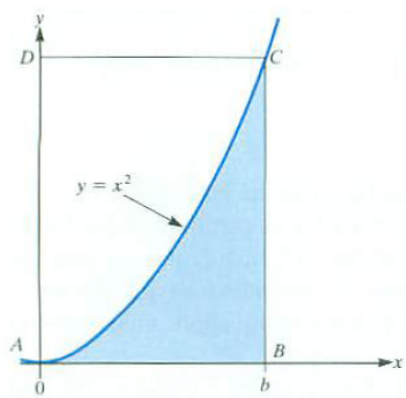


Figure 6.2.7

2. Figure 6.2.7 shows the curve $y = x^2$. What is the ratio between the shaded area under the curve and the area of the rectangle $ABCD$?

3.

(a) Define "the definite integral of $f(x)$ from a to b, $\int_a^b f(x)\,dx$."

(b) Define the definite integral, using as few mathematical symbols as you can.

(c) Give three applications of the definite integral.

4. Assume $f(x)$ is decreasing for x in $[a, b]$. When you form an approximating sum for $\int_a^b f(x)\,dx$ with left-hand endpoints as sampling points, is your estimate too large or too small? Explain (in one or more complete sentneces).

In Exercises 5 to 8 evaluate the sum

5.

(a) $\sum_{i=1}^{3} i$

(b) $\sum_{i=3}^{7}(2i + 3)$

(c) $\sum_{d=1}^{3} d^2$

6.

(a) $\sum_{i=2}^{4} i^2$

(b) $\sum_{j=2}^{4} j^2$

(c) $\sum_{i=1}^{3}(i^2+i)$

7.

(a) $\sum_{i=1}^{4} 1^i$

(b) $\sum_{k=2}^{6}(-1)^k$

(c) $\sum_{j=1}^{150} 3$

8.

(a) $\sum_{i=3}^{5} \frac{1}{i}$

(b) $\sum_{i=0}^{4} \cos(2\pi i)$

(c) $\sum_{i=1}^{3} 2^{-i}$

In Exercises 9 to 12 write each sum in Σ-notation. (Do not evaluate the sum.)

9.

(a) $1+2+2^2+2^3+\cdots+2^{100}$

(b) $x^3+x^4+x^5+x^6+x^7$

(c) $\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\cdots+\frac{1}{102}+\frac{1}{103}$

10.

(a) $\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{100}$

(b) $\frac{1}{4}+\frac{1}{6}+\frac{1}{8}+\frac{1}{10}+\frac{1}{12}+\frac{1}{14}$

(c) $\frac{1}{1^2}+\frac{1}{3^2}+\frac{1}{5^2}+\cdots+\frac{1}{101^2}$

11.

(a) $x_0^2(x_1 - x_0) + x_1^2(x_2 - x_1) + x_2^2(x_3 - x_2)$

(b) $x_1^2(x_1 - x_0) + x_2^2(x_2 - x_1) + x_3^2(x_3 - x_2)$

12.

(a) $8t_0^2(t_1 - t_0) + 8t_1^2(t_2 - t_1) + \cdots + 8t_{99}^2(t_{100} - t_{99})$

(b) $8t_1^2(t_1 - t_0) + 8t_2^2(t_2 - t_1) + \cdots + 8t_n^2(t_n - t_{n-1})$

13.

(a) Use the definition of definite integral to evaluate $\int_0^b e^x \, dx$. (See Example 3.)

(b) From (a), deduce that, for $0 \le a < b$, $\int_a^b e^x \, dx = e^b - e^a$.

14.

(a) Use the definition of definite integral to evaluate $\int_0^b 3^x \, dx$.

(b) From (a), deduce that, for $0 \le a < b$, $\int_a^b 3^x \, dx = (3^b - 3^a)/\ln(3)$.

15. The fact that $\int_a^b f(x) \, dx = \lim_{n\to\infty} \sum_{i=1}^n f(c_i)\Delta x$ provides another way to evaluate some limits of sums that would otherwise be very challenging to evaluate. Use this idea to write each of the following limits as a definite integral. (Do not evaluate the definite integrals.)

(a) $\displaystyle\lim_{n\to\infty} \sum_{i=1}^{n} e^{i/n} \frac{1}{n}$

(b) $\displaystyle\lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{1 + \left(1 + \frac{2i}{n}\right)^2} \frac{2}{n}$

(c) $\displaystyle\lim_{n\to\infty} \sum_{i=1}^{n} \sin\left(\frac{i\pi}{n}\right) \frac{\pi}{n}$

(d) $\displaystyle\lim_{n\to\infty} \sum_{i=1}^{n} \left(2 + \frac{3i}{n}\right)^4 \frac{3}{n}$

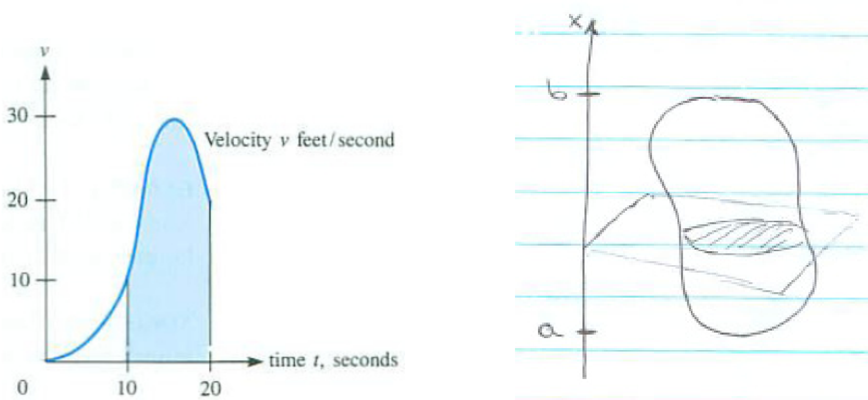
In Exercises 16 to 18 evaluate $\sum_{i=1}^{n} f(c_i)(x_i - x_{i-1})$ for the given function, partition, and sampling numbers.

16. $f(x) = \sqrt{x}$, $x_0 = 1$, $x_1 = 3$, $x_2 = 5$, $c_1 = 1$, $c_2 = 4$ $(n = 2)$

17. $f(x) = \sqrt[3]{x}$, $x_0 = 0$, $x_1 = 1$, $x_2 = 4$, $x_3 = 10$, $c_1 = 0$, $c_2 = 1$, $c_3 = 8$ $(n = 3)$

18. $f(x) = 1/x$, $x_0 = 1$, $x_1 = 1.25$, $x_2 = 1.5$, $x_3 = 1.75$, $x_4 = 2$, $c_1 = 1$, $c_2 = 1.25$, $c_3 = 1.6$, $c_4 = 2$ $(n = 4)$

19. The velocity of an automobile at time t is $v(t)$ feet per second. [Assume $v(t) \geq 0$.] The graph of v for t in $[0, 20]$ is shown in Figure 6.2.8(a). Explain, in complete sentences, why the shaded area under the curve equals the change in position.



(a) (b)

Figure 6.2.8

In Exercises 20 to 23 partition the interval into 4 sections of equal lengths. Estimate the definite integral using sampling numbers chosen to be (a) the left endpoints and (b) the right endpoints.

20. $\int_1^2 (1/x^2)\ dx$.

21. $\int_1^5 \ln(x)\ dx$.

22. $\int_1^5 \frac{2^x}{x}\ dx$.

23. $\int_0^1 \sqrt{1 + x^3}\ dx$.

24. Write the following expression using summation notation.

$$c^{n-1} + c^{n-2}d + c^{n-3}d^2 + \cdots + cd^{n-2} + d^{n-1}.$$

25. Assume that $f(x) \leq -3$ for all x in $[1, 5]$. What can be said about the value of $\int_1^5 f(x)\, dx$? Explain, in detail, using the definition of the definite integral.

26. A rocket's speed if $f(t)$ miles per second at time t seconds. Let $t_0, \ldots, t_n$ be a partition of $[a, b]$, and let $T_1, \ldots, T_n$ be sampling numbers. What is the physical interpretation of each of the following quantities?

(a) $t_i - t_{i-1}$

(b) $f(T_i)$

(c) $f(T_i)(t_i - t_{i-1})$

(d) $\sum_{i=1}^{n} f(T_i)(t_i - t_{i-1})$

(e) $\int_a^b f(t)\, dt$

27.

(a) Sketch $y = \cos(x)$, for x in $[0, \pi/2]$.

(b) Estimate, by eye, the area under the curve and above $[0, \pi/2]$.

(c) Partition $[0, \pi/2]$ into three equal sections and use them to provide an overestimate of the area under the curve.

(d) Use the same partition to provide an underestimate of the area under the curve.

28. Repeat Exercise 27 for the area under the curve $y = e^{-x}$ above $[0, 3]$.

29. For x in $[a, b]$, let $A(x)$ be the area of the cross section of a solid perpendicular to the x-axis at x (think of slicing a potato). Let $x_0, x_1, \ldots, x_n$ be a partition of $[a, b]$. Let $c_1, \ldots, c_n$ be the corresponding sampling numbers. What is the geometric interpretation of each of the following quantities? (Refer to Figure 6.2.8(b).)

(a) $x_i - x_{i-1}$

(b) $A(c_i)$

(c) $A(c_i)(x_i - x_{i-1})$

(d) $\sum_{i=1}^{n} A(c_i)(x_i - x_{i-1})$

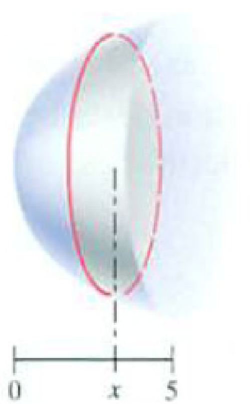
(e) $\int_a^b A(x)\, dx$

30. Show that the volume of a right circular cone of radius a and height h is $\frac{\pi a^3 h}{3}$. (First show that a cross section by a plane perpendicular to the axis of the cone and a distance x from the vertex is a circle of radius ax/h.) See Exercise 29.

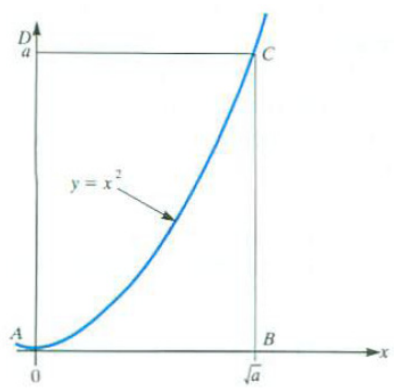
31.

(a) Set up an appropriate definite integral $\int_a^b f(x)\,dx$ which equals the volume of the headlight in Figure 6.2.9(a) whose cross section by a typical plane perpendicular to the x-axis at x is a disk whose radius is $\sqrt{x/\pi}$. A circle is a curve and a disk is the flat region inside a circle.

(b) Evaluate the definite integral found in (a).



(a)



(b)

Figure 6.2.9

32.

(a) By considering Figure 6.2.9(b), in particular the area of region ACD, show that $\int_0^a \sqrt{x}\,dx = \frac{2}{3}a^{3/2}$.

(b) Use (a) to evaluate $\int_a^b \sqrt{x}\,dx$ when $0 < a < b$.

Exercises 33 to 36 involve "telescoping sums". Let f be a function defined at least for positive integers. A sum of the form $\sum_{i=1}^{n}(f(i+1) - f(i))$ is called telescoping. To show why, write the sum out in longhand:

$$(f(2)-f(1))+(f(3)-f(2))+(f(4)-f(3))+\cdots+(f(n)-f(n-1))+(f(n+1)-f(n)).$$

Everything cancels except $-f(1)$ and $f(n+1)$. The whole sum shrinks like a collapsible telescope, to $f(n+1) - f(1)$.

33.

(a) Show that $\sum_{i=1}^{n} \left((i+1)^2 - i^2\right) = (n+1)^2 - 1$. (This is a telescoping sum.)

(b) From (a), show that $\sum_{i=1}^{n}(2i+1) = (n+1)^2 - 1$.

(c) From (b), show that $n + 2\sum_{i=1}^{n} i = (n+1)^2 - 1$.

(d) From (c), show that $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$.

34. Exercise 33 showed that $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$. Use this information to find $\int_0^b x\, dx$ directly from the definition of the definite integral (not by interpreting it as an area). No picture is needed.

35.

(a) Starting with the telescoping sum $\sum_{i=1}^{n} \left((i+1)^3 - i^3\right)$ show that

$$n + 3\sum_{i=1}^{n} i^2 + 3\sum_{i=1}^{n} i = (n+1)^3 - 1.$$

(b) Use (a) to show that $\sum_{i=1}^{n} i^2 = \frac{1}{6}n(n+1)(2n+1)$.

(c) Use (b) to show that $\int_0^b x^2\, dx = \frac{b^3}{3}$.

See Exercise 34.

36.

(a) Using the techniques of Exercises 33 to 35, find a short formula for the sum $\sum_{i=1}^{n} i^3$.

(b) Use the formula found in (a) to show that $\int_0^b x^3\, dx = \frac{b^4}{4}$.

37. The function $f(x) = 1/x$ has a remarkable property, namely, for a and b greater than 1,

$$\int_1^a \frac{1}{x}\, dx = \int_b^{ab} \frac{1}{x}\, dx.$$

In other words, "magnifying the interval $[1, a]$ by a positive number b does not change the value of the definite integral." The following steps show why this is so.

(a) Let $x_0 = 1$, x_1, x_2, $\ldots$, $x_n = a$ divide the interval $[1, a]$ into n sections. Using left endpoints write out an approximating sum for $\int_1^a \frac{1}{x}\, dx$.

(b) Let $bx_0 = b$, bx_1, bx_2, $\ldots$, $bx_n = ab$ divide the interval $[b, ab]$ into n sections. Using left endpoints write out an approximating sum for $\int_b^{ab} \frac{1}{x}\, dx$.

(c) Explain why $\int_1^a \frac{1}{x}\, dx = \int_b^{ab} \frac{1}{x}\, dx$.

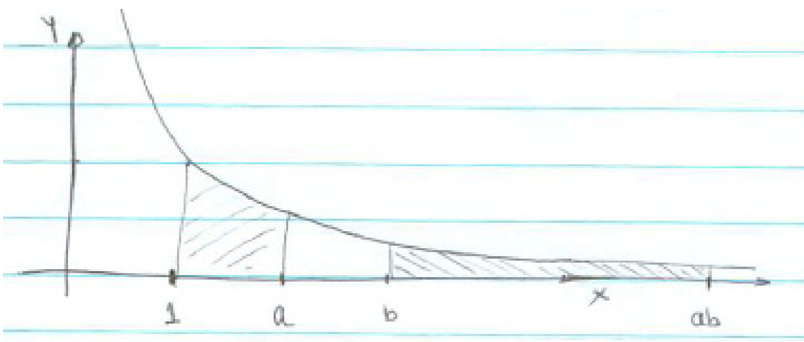


Figure 6.2.10

38. Let $L(t) = \int_1^t \frac{1}{x}\, dx$, $t > 1$.

(a) Show that $L(a) = L(ab) - L(b)$.

(b) By (a), conclude that $L(ab) = L(a) + L(b)$.

(c) What familiar function has the property listed in (b)?

Gregory St. Vincent noticed the property (a) in 1647, and his friend A.A. de Sarasa saw that (b) followed. Euler, in the 18th century, recognized that $L(x)$ is the logarithm of x to the base e. In short, the area under the hyperbola $y = 1/x$ and above $[1, a]$, $a > 1$, is $\ln(a)$. It can be shown that for a in $(0, 1)$, the negative of the area below that curve and above $[a, 1]$ is $\ln(a)$. (See C. H. Edwards Jr., The Historical development of the Calculus, pp. 154–158.)

39. In Exercise 13 it was shown that for $0 \le a \le b$, $\int_a^b e^x\, dx = e^b - e^a$.

(a) Use this information and a diagram to show that $\int_{e^a}^{e^b} \ln(x)\, dx = e^b(b-1) - e^a(a-1)$.

(b) From (a), deduce that for $1 \le c \le d$, $\int_c^d \ln(x)\, dx = (d\ln(d) - d) - (c\ln(c) - c)$.

(c) By differentiating $x\ln(x) - x$, show that it is an antiderivative of $\ln(x)$.

40.

(a) To estimate $\int_1^2 \frac{1}{x}\,dx$ divide $[1,2]$ into n sections of equal lengths and use right endpoints as the sampling points.

(b) Deduce from (a) that

$$\lim_{n\to\infty}\sum_{i=n+1}^{2n}\frac{1}{i}=\lim_{n\to\infty}\left(\frac{1}{n+1}+\frac{1}{n+2}+\cdots+\frac{1}{2n}\right)=\text{area under } y=1/x \text{ and above } [$$

(c) Let $g(n)=\frac{1}{n+1}+\frac{1}{n+2}+\cdots+\frac{1}{2n}$. Show that $\frac{1}{2}\le g(n)<1$ and $g(n+1)<g(n)$.

41. (This Exercise is used in Exercise 42.) Consider $b>1$ and n a positive integer. Define $r(n)$ by the equation $(r(n))^n=b$.

(a) In the case $b=5$, find $r(n)$ for $n=1$, 2, 3, and 10. (Note that $r=b^{1/n}$, so you could use the x^y key on a calculator.)

(b) The calculations in (a) suggest that $\lim_{n\to\infty} r(n)=1$. Show that this conjecture is correct. (Start by taking ln of both sides of the equation $(r(n))^n=b$.)

42. For $b>1$ and k any number, Pierre Fermat (1601–1665) found the area under $y=x^k$ and above $[1,b]$ by using approximating sums. However, he did not cut the interval $[1,b]$ into n sections of equal widths. Instead, for a given positive integer n, he introduced the numbers r such that $r^n=b$. As n increases, r approaches 1, as Exercise 41 shows. Then he divided the interval $[0,b]$ into sections using the number r, r^2, r^3, ..., r^{n-1}, as shown in Figure 6.2.11. The n sections are $[1,r]$, $[r,r^2]$, ..., $[r^{n-1},r^n]=[r^{n-1},b]$.

(a) Show that the width of the i^{th} section, $[r^{i-1},r^i]$, is $r^{i-1}(r-1)$.

(b) Using the left endpoints of each section, obtain an underestimate of $\int_1^b x^2\,dx$.

(c) Show that the estimate in (b) is equal to

$$\frac{b^3-1}{1+r+r^2}.$$

(d) Find $\lim_{n\to\infty}\frac{b^3-1}{1+r+r^2}$. (Remember that r depends on n.)

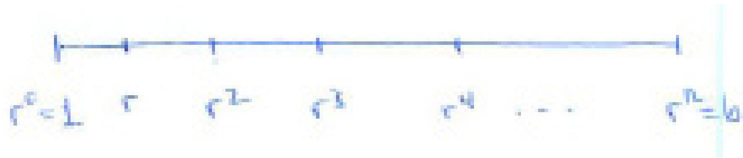


Figure 6.2.11

43. Use Fermat's method (see Exercise 42) to find $\int_1^b x^3\ dx$.

44. Use Fermat's method (see Exercise 42) to find $\int_1^b x^2\ dx$.

45. Use Fermat's approach outlined in Exercise 42, but with right endpoints as the sampling points, to obtain an overestimate of the area under x^2, above $[1, b]$, and then find its limit as $n \to \infty$.

46.

(a) Obtain an underestimate and an overestimate of $\int_0^{\pi/2} \cos(x)\ dx$ that differ by at most 0.1. Remember that the angles are measured in radians.

(b) Average the two estimates in (a).

(c) If $\int_0^{\pi/2} \cos(x)\ dx$ is a famous number, what do you think it is?

47. By considering the approximating sums in the definition of a definite integral, show that $\int_3^4 \frac{dx}{(x+5)^3}$ equals $\int_2^3 \frac{dx}{(x+6)^3}$.

48. For a continuous function f defined for all x, is $\int_a^b f(x+1)\ dx$ equal to $\int_{a+1}^{b+1} f(x)\ dx$?

49. For continuous functions f and g defined for all x, is $\int_a^b f(x)g(x)\ dx$ equal to the product of $\int_a^b f(x)\ dx$ and $\int_a^b g(x)\ dx$?

50. If f is an increasing function such that $f(1) = 3$ and $f(6) = 7$, what can be said about $\int_2^4 f(x)\ dx$? Explain.

51.

(a) Using formulas already developed, evaluate $G(x) = \int_1^x t^2\,dt$.

(b) Find $G'(x)$.

(c) Repeat (a) and (b) for $G(x) = \int_1^x 2^t\,dt$.

(d) Do you notice what appears to be a coincidence in both (b) and (c)?

Skill Drill

In Exercises 52 to 59 give two antiderivatives for the given functions.

52. x^2

53. $1/x^3$

54. e^{-4x}

55. $1/(2x+1)$

56. 2^x

57. $\sin(3x)$

58. $\frac{3}{1+9x^2}$

59. $\frac{4}{\sqrt{1-x^2}}$

6.3 Properties of the Antiderivative and the Definite Integral

In Section 3.6 we defined an antiderivative of a function $f(x)$. It is any function $F(x)$ whose derivative is $f(x)$. For instance, x^3 is an antiderivative of $3x^2$. So is $x^3 + 2011$. Keep in mind that *an antiderivative is a function.*

F is an antiderivative of f when $F'(x) = f(x)$

In this section we discuss various properties of antiderivatives and definite integrals. These properties will be needed in Section 6.4 where we obtain a relation between antiderivatives and definite integrals. That relation will be a great time-saver in evaluating many (but not all) definite integrals.

We have not yet introduced a symbol for an antiderivative of a function. We will adopt the following standard notation:

> **Notation**: Any antiderivative of f is denoted $\int f(x)\ dx$.

Warning: If a function has an antiderivative, then it has lots of antiderivatives.

For instance, $x^3 = \int 3x^2\ dx$. This equation is read "x^3 is an antiderivative of $3x^2$". That means simply that "the derivative of x^3 is $3x^2$". It is true that $x^3 + 2011 = \int 3x^2\ dx$, since $x^3 + 2011$ is also an antiderivative of $3x^2$. That does *not* mean that the functions x^3 and $x^3 + 2011$ are equal. All it means is that these two functions both have the same derivative, $3x^2$. The symbol $\int 3x^2\ dx$ refers to *any function* whose derivative is $3x^2$.

If $F'(x) = f(x)$ we write $F(x) = \int f(x)\ dx$. The function $f(x)$ is called the **integrand**. The function $F(x)$ is called an antiderivative of $f(x)$. The symbol for an antiderivative, $\int f(x)\ dx$, is similar to the symbol for a definite integral, $\int_a^b f(x)\ dx$, but they denote vastly different concepts. An antiderivative is often called an "integral" or "indefinite integral," but should not be confused with a definite integral. The symbol $\int f(x)dx$ denotes a *function* — any function whose derivative is $f(x)$. The symbol $\int_a^b f(x)\ dx$ denotes a *number* — one that is defined by a limit of certain sums. The value of the definite integral may vary as the interval $[a, b]$ changes.

$\int f(x)\ dx$ is a function
$\int_a^b f(x)\ dx$ is a number.

We apologize for the use of such similar notations, $\int f(x)\ dx$ and $\int_a^b f(x)\ dx$, for such distinct concepts. However, it is not for us to undo over three centuries of custom. Rather, it is up to you to read the symbols $\int f(x)\ dx$ and $\int_a^b f(x)\ dx$ carefully. You distinguish between such similar-looking words as "density" and "destiny" or "nuclear" and "unclear". Be just as careful when reading mathematics.

Properties of Antiderivatives

The tables inside the covers of this book list many antiderivatives. One example is $\int \sin(x)\ dx = -\cos(x)$. Of course, $-\cos(x) + 17$ also is an antiderivative of $\sin(x)$. In Section 4.1 it was shown that if F and G have the same derivative on an interval, they differ by a constant, C. So $F(x) - G(x) = C$ or $F(x) = G(x) + C$. For emphasis, we state this as a theorem.

The following theorem asserts that if you find an antiderivative $F(x)$ for a function $f(x)$, then any other antiderivative of $f(x)$ is of the form $F(x) + C$ for some constant C.

This result was anticipated back in Section 3.6.

Theorem 6.3.1. *If F and G are both antiderivatives of f on some interval, then there is a constant C such that*

$$F(x) = G(x) + C.$$

Many tables of integrals, including the ones in the cover of this book, omit the $+C$.

When using an antiderivative, it is best to include the constant C. For example,

$$\begin{aligned} \int 5\ dx &= 5x + C \\ \int e^x\ dx &= e^x + C \\ \text{and} \qquad \int \sin(2x)\ dx &= \frac{-1}{2}\cos(2x) + C. \end{aligned}$$

Observe that

$$\frac{d}{dx}\left(\int x^3\ dx\right) = x^3 \qquad \text{and} \qquad \frac{d}{dx}\left(\int \sin(2x)\ dx\right) = \sin(2x). \qquad (6.3.1)$$

Are these two equations profound or trivial? Read them aloud and decide.

We know that the square of the square root of 7 is 7 and that $e^{\ln(3)} = 3$, both by the definition of inverse functions.

The first says, "The derivative of an antiderivative of x^3 is x^3." *It is true simply because that is how we defined the antiderivative.* We know that

$$\frac{d}{dx}\left(\int \frac{\ln(1+x^2)}{(\sin(x))^2}\ dx\right) = \frac{\ln(1+x^2)}{(\sin(x))^2}$$

even though we cannot write out a formula for an antiderivative of $\frac{\ln(1+x^2)}{(\sin(x))^2}$. In other words, by the very definition of the antiderivative,

$$\frac{d}{dx}\left(\int f(x)\ dx\right) = f(x).$$

Any property of derivatives gives us a corresponding property of antiderivatives. Three of the most important properties of antiderivatives are recorded in the next theorem.

Theorem 6.3.2 (Properties of Antiderivatives)**.** *Assume that f and g are functions with antiderivatives $\int f(x)\ dx$ and $\int g(x)\ dx$. Then the following hold:*

Properties of antiderivatives

A. $\int cf(x)\ dx = c\int f(x)\ dx$ *for any constant c.*

B. $\int (f(x) + g(x))\ dx = \int f(x)\ dx + \int g(x)\ dx$.

C. $\int (f(x) - g(x))\ dx = \int f(x)\ dx - \int g(x)\ dx$.

Proof

(A) Before we prove that $\int cf(x)\ dx = c\int f(x)\ dx$, we stop to see what it means. This equation says that "c times an antiderivative of $f(x)$ is an antiderivative of $cf(x)$". Let $F(x)$ be an antiderivative of $f(x)$. Then the equation says "c times $F(x)$ is an antiderivative of $cf(x)$". To determine if this statement is true we must differentiate $cF(x)$ and check that we get $cf(x)$. So, we compute $(cF(x))'$:

$$\begin{aligned} (cF(x))' &= cF'(x) && [c \text{ is a constant}] \\ &= cf(x). && [F \text{ is antiderivative of } f] \end{aligned}$$

Thus $cF(x)$ is indeed an antiderivative of $cf(x)$. Therefore, we may write

$$cF(x) = \int cf(x)\ dx.$$

Since $F(x) = \int f(x)\ dx$, we conclude that

$$c\int f(x)\ dx = \int cf(x)\ dx.$$

(B) The proof is similar. We show that $\int f(x)\ dx + \int g(x)\ dx$ is an antiderivative of $f(x) + g(x)$. To do this we compute the derivative of the sum $\int f(x)\ dx + \int g(x)\ dx$:

$$\begin{aligned} \frac{d}{dx}\left(\int f(x)\ dx + \int g(x)\ dx\right) &= \frac{d}{dx}\left(\int f(x)\ dx\right) + \frac{d}{dx}\left(\int g(x)\ dx\right) && [\text{derivative of a sum}] \\ &= f(x) + g(x). && [\text{definition of antiderivatives}] \end{aligned}$$

(C) The proof is similar to the one for (b). •

EXAMPLE 1 Find (a) $\int 6\cos(x)\,dx$, (b) $\int(6\cos(x)+3x^2)\,dx$, and (c) $\int(6\cos(x)-\frac{5}{1+x^2}\,dx$.

SOLUTION (a) Part (a) of the theorem is used to move the "6" (a constant) past the integral sign, "$\int$". We then have:

$$\int 6\cos(x)\ dx = 6\int \cos(x)\ dx = 6\sin(x) + C.$$

Notice that the "C" is added as the last step in finding an antiderivative.

(b)

$$\begin{aligned}\int(6\cos(x)+3x^2)\ dx &= \int 6\cos(x)\ dx + \int 3x^2\ dx && \text{[part (b) of the theorem]}\\ &= 6\sin(x) + x^3 + C.\end{aligned}$$

Here, notice that separate constants are not needed for each antiderivative; again only one "C" is needed for the overall antiderivative.

(c)

$$\begin{aligned}\int\left(6\cos(x)-\tfrac{5}{1+x^2}\right)\ dx &= \int 6\cos(x)\ dx - \int \tfrac{5}{1+x^2}\ dx && \text{[part (c) of the theorem]}\\ &= 6\sin(x) - 5\int \tfrac{1}{1+x^2}\ dx && \text{[part (a) of the theorem]}\\ &= 6\sin(x) - 5\arctan(x) + C && [(\arctan(x))' = \tfrac{1}{1+x^2}]\end{aligned}$$

◇

The last two parts of Theorem 6.3.2 extend to any finite number of functions. For instance,

$$\int(f(x)-g(x)+h(x))\ dx = \int f(x)\ dx - \int g(x)\ dx + \int h(x)\ dx.$$

Theorem 6.3.3. *Let a be a number other than -1. Then*

$$\int x^a\ dx = \frac{x^{a+1}}{a+1} + C.$$

Proof

$$\left(\frac{x^{a+1}}{a+1}\right)' = \frac{(a+1)x^{(a+1)-1}}{a+1} = x^a.$$

•

EXAMPLE 2 Find $\int \left(\frac{3}{\sqrt{1-x^2}} - \frac{2}{x} + \frac{1}{x^3}\right)\, dx$, $0 < x < 1$.

If $-1 < x < 0$, we would write the antiderivative of $1/x$ as $\ln|x|$.

SOLUTION

$$\begin{aligned}\int \left(\frac{3}{\sqrt{1-x^2}} - \frac{2}{x} + \frac{1}{x^3}\right)\, dx &= 3\int \frac{1}{\sqrt{1-x^2}}\, dx - 2\int \frac{1}{x}\, dx + \int x^{-3}\, dx \\ &= 3\arcsin(x) - 2\ln(x) + \frac{x^{-2}}{-2} + C \\ &= 3\arcsin(x) - 2\ln(x) - \frac{1}{2x^2} + C.\end{aligned}$$

◇

Properties of Definite Integrals

Some of the properties of definite integrals look like properties of antiderivatives. However, they are assertions about numbers, not about functions. In the notation for the definite integral, $\int_a^b f(x)\, dx$, b is larger than a. It will be useful to be able to speak about "the definite integral from a to b" even if b is less than or equal to a. The following two definitions meet this need and we will use them in the proofs of the two fundamental theorems of calculus in the next section.

DEFINITION *(Integral from a to b, where $b < a$.)* If b is less than a, then

$$\int_a^b f(x)\, dx = -\int_b^a f(x)\, dx.$$

EXAMPLE 3 Compute $\int_3^0 x^2\, dx$, the integral from 3 to 0 of x^2.
SOLUTION The symbol $\int_3^0 x^2\, dx$ is defined as $-\int_0^3 x^2\, dx$. As was shown in Section 6.2, $\int_0^3 x^2\, dx = 9$. Thus

$$\int_3^0 x^2\, dx = -9.$$

◇

DEFINITION *(Integral from a to a.)*

$$\int_a^a f(x)\,dx = 0$$

Remark: The definite integral is defined with the aid of partitions of an interval. Rather than permit partitions to have sections of length 0, it is simpler just to make this definition.

The point of making these two definitions is that now the symbol $\int_a^b f(x)\,dx$ is defined for any numbers a and b and any continuous function f, assuming $f(x)$ is defined for x in $[a, b]$. It is no longer necessary that a be less than b.

The definite integral has several properties, some of which we will be using in this section and some in later chapters. Justifications of these properties are provided immediately after the following table.

Properties of antiderivatives

Theorem 6.3.4 (Properties of the Definite Integral). *Let f and g be continuous functions, and let c be a constant. Then*

1. ***Moving a Constant Past*** $\int_a^b$
   $\int_a^b cf(x)\ dx = c\int_a^b f(x)\ dx$
2. ***Definite Integral of a Sum***
   $\int_a^b (f(x)+g(x))\ dx = \int_a^b f(x)\ dx + \int_a^b g(x)\ dx$
3. ***Definite Integral of a Difference***
   $\int_a^b (f(x)-g(x))\ dx = \int_a^b f(x)\ dx - \int_a^b g(x)\ dx$
4. ***Definite Integral of a Non-Negative Function***
   If $f(x) \geq 0$ *for all* x *in* $[a,b]$, $a<b$, *then* $\displaystyle\int_a^b f(x)\ dx \geq 0.$
5. ***Definite Integrals Preserve Order***
   If $f(x) \geq g(x)$ *for all* x *in* $[a,b]$, $a<b$, *then*
   $$\int_a^b f(x)\ dx \geq \int_a^b g(x)\ dx.$$
6. ***Sum of Definite Integrals Over Adjoining Intervals***
   If a, b, *and* c *are numbers, then*
   $$\int_a^c f(x)\ dx + \int_c^b f(x)\ dx = \int_a^b f(x)\ dx.$$
7. ***Bounds on Definite Integrals***
   If m *and* M *are numbers and* $m \leq f(x) \leq M$ *for all* x *between* a *and* b, *then*
   $$m(b-a) \quad \leq \quad \int_a^b f(x)\ dx \leq M(b-a) \qquad \text{if } a<b$$
   and
   $$m(b-a) \quad \geq \quad \int_a^b f(x)\ dx \geq M(b-a) \qquad \text{if } a>b$$

Proof of Property 1

Take the case $a < b$. The equation $\int_a^b cf(x)\ dx = c\int_a^b f(x)\ dx$ resembles part (a) of Theorem 6.3.2 about antiderivatives: $\int cf(x)\ dx = c\int f(x)\ dx$. However, its proof is quite different, since $\int_a^b cf(x)\ dx$ is defined as a limit of sums.

We have

$$\begin{aligned}
\int_a^b cf(x)\ dx &= \lim_{\text{all } \Delta x_i \to 0} \sum_{i=1}^{n} cf(c_i)\Delta x_i && \text{definition of definite integral}\\
&= \lim_{\text{all } \Delta x_i \to 0} c\sum_{i=1}^{n} f(c_i)\Delta x_i && \text{algebra (distributive law)}\\
&= c\lim_{\text{all } \Delta x_i \to 0} \sum_{i=1}^{n} f(c_i)\Delta x_i && \text{property of limits}\\
&= c\int_a^b f(x)\ dx. && \text{definition of definite integral}
\end{aligned}$$

•

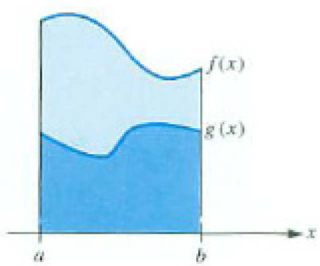


Figure 6.3.1

Similar approaches can be used to justify each of the other properties. However, we pause only to make them plausible by giving an intuitive interpretation of each property in terms of area for positive functions.

Plausibility of Argument for Property 5

This amounts to the assertion that when the graph of $y = f(x)$ is always at least as high as the graph of $y = g(x)$, then the area of a region under the curve $y = f(x)$ is greater than or equal to the area under the curve $y = g(x)$ above a given interval. (See Figure 6.3.) •

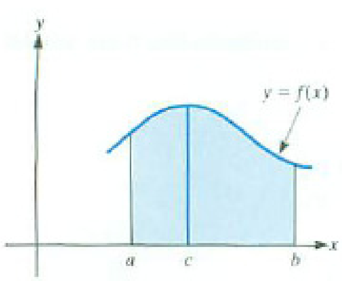


Figure 6.3.2

Plausibility of Argument for Property 6

In the case that $a < c < b$ and $f(x)$ assumes only positive values, this property asserts that the area of the region below the graph of $y = f(x)$ and above the interval $[a, b]$ is the sum of the areas of the regions below the graph and above the smaller intervals $[a, c]$ and $[c, b]$. Figure 6.3.2 shows that this is certainly plausible. •

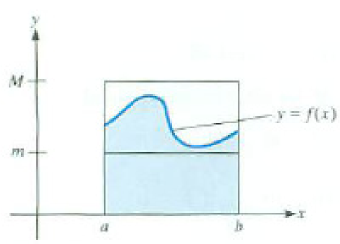


Figure 6.3.3

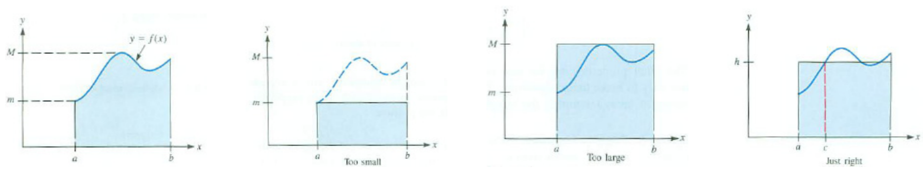
Plausibility of Argument for Property 7

The inequalities in this property compare the area under the graph of $y = f(x)$ with the areas of two rectangles, one of height M and one of height m. (See Figure 6.3.3.) In the case $a < b$, the area of the larger rectangle is $M(b - a)$ and the area of the smaller rectangle is $m(b - a)$. •

The Mean-Value Theorem for Definite Integrals

The mean-value theorem for *derivatives* says that (under suitable hypotheses) $f(b)-f(a)=f'(c)(b-a)$ for some number c in $[a,b]$. The mean-value theorem for *definite integrals* has a similar flavor. First, we state it geometrically.

If $f(x)$ is positive and $a<b$, then $\int_a^b f(x)\,dx$ can be interpreted as the area of the shaded region in Figure 6.3.4(a).



(a) (b) (c) (d)

Figure 6.3.4

Let m be the minimum and M the maximum values of $f(x)$ for x in $[a,b]$. We assume that $m<M$. The area of the rectangle of height M is larger than the shaded area; the area of the rectangle of height m is smaller than the shaded area. (See Figures 6.3.4(b) and (c).) Therefore, there is a rectangle whose height h is somewhere between m and M, whose area is the same as the shaded area under the curve $y=f(x)$. (See Figure 6.3.4(d).) Hence $\int_a^b f(x)\,dx=(b-a)h$.

What can you say about the case when $m=M$?

Now, h is a number between m and M. By the Intermediate-Value Property for continuous functions, in Section 2.5 there is a number c in $[a,b]$ such that $f(c)=h$. (See Figure 6.3.4(d).) Hence,

$$\text{Area of shaded region under curve} = f(c)(b-a).$$

This suggests the mean-value theorem for definite integrals.

Mean-Value Theorem for Definite Integrals

Theorem 6.3.5 (Mean-Value Theorem for Definite Integrals)**.** *Let a and b be numbers, and let f be a continuous function defined at least on the interval $[a,b]$. Then there is a number c in $[a,b]$ such that*

$$\int_a^b f(x)\,dx = f(c)(b-a).$$

Proof of the Mean-Value Theorem for Definite Integrals, using only properties of the definite integral
Consider the case when $a < b$. Let M be the maximum and m the minimum of $f(x)$ on $[a, b]$. Property 7, combined with division by $b - a$, gives

$$m \leq \frac{\int_a^b f(x)\, dx}{b-a} \leq M,$$

Because f is continuous on $[a, b]$, by the Intermediate-Value Property of Section 2.5 there is a number c in $[a, b]$ such that

$$f(c) = \frac{\int_a^b f(x)\, dx}{b-a},$$

The case $b < a$ can be obtained from the case $a < b$. (see Exercise 37).

and the theorem is proved (without depending on a picture). •

EXAMPLE 4 Verify the mean-value theorem for definite integrals when $f(x) = x^2$ and $[a, b] = [0, 3]$.
SOLUTION In Section 6.2 it was shown that $\int_0^3 x^2\, dx = 9$. Since $f(x) = x^2$, we are looking for c in $[0, 3]$ such that

$$\int_0^3 x^2\, dx = 9 = c^2(3 - 0)$$

$-\sqrt{3}$ is not in $[0, 3]$.

That is, $9 = 3c^2$, so $c^2 = \frac{9}{3} = 3$, $c = \sqrt{3}$. (See Figure 6.3.5.) The rectangle with height $f(\sqrt{3}) = (\sqrt{3})^2 = 3$ and base $[0, 3]$ has the same area as the region under the curve $y = x^2$ and above $[0, 3]$. ◇

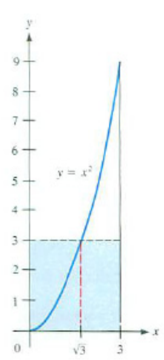


Figure 6.3.5

The Average Value of a Function

Let $f(x)$ be a continuous function defined on $[a, b]$. What shall we mean by the "average value of $f(x)$ over $[a, b]$"? We cannot add up all the values of $f(x)$ for all x's in $[a, b]$ and divide by the number of x's, since there are an infinite number of such x's. However, we can work with the average (or mean) of n numbers $a_1, a_2, \ldots, a_n$, which is their sum divided by n: $\frac{1}{n}\sum_{i=1}^n a_i$. For example, the average of 1, 2, and 6 is $\frac{1}{3}(1 + 2 + 6) = \frac{9}{3} = 3$.

This suggests how to define the "average value of $f(x)$ over $[a, b]$". Choose a large integer n and partition $[a, b]$ into n sections of equal length, $\Delta x = (b - a)/n$. Let the sampling points c_i be the left endpoint of each section,

$c_1 = a$, $c_2 = a + \Delta x$, ..., $c_n = a + (n-1)\Delta x = b - \Delta x$. Then an estimate of the "average" would be

$$\frac{1}{n}(f(c_1) + f(c_2) + \cdots + f(c_n)). \tag{6.3.2}$$

Since $\Delta x = (b-a)/n$, it follows that $\frac{1}{n} = \frac{\Delta x}{b-a}$. Therefore, (6.3.2) can be rewritten as

$$\frac{1}{b-a}\sum_{i=1}^{n} f(c_i)\Delta x.$$

But, $\sum_{i=1}^{n} f(c_i)\Delta x$ is an estimate of $\int_a^b f(x)\ dx$. It follows that, as $n \to \infty$, this average of the n function values approaches $\frac{1}{b-a}\int_a^b f(x)\ dx$. This motivates the following definition:

DEFINITION *(Average Value of a Function over an Interval)* Let $f(x)$ be defined on the interval $[a, b]$. Assume that $\int_a^b f(x)\ dx$ exists. The **average value** or **mean value** of f on $[a, b]$ is defined to be

$$\frac{1}{b-a}\int_a^b f(x)\ dx.$$

Geometrically speaking (if $f(x)$ is positive), this average value is the height of the rectangle that has the base $[a, b]$ and the same area as the area of the region under the curve $y = f(x)$, above $[a, b]$. (See Figure 6.3.6.) Observe that the average value of $f(x)$ over $[a, b]$ is between its maximum and minimum values for x in $[a, b]$. However, it is not necessarily the average of these two numbers.

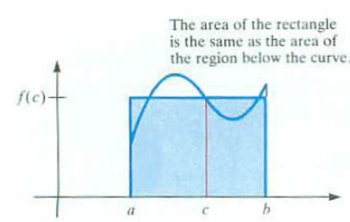


Figure 6.3.6

EXAMPLE 5 Find the average value of 2^x over the interval $[1, 3]$.
SOLUTION The average value of 2^x over $[1, 3]$ by definition equals

$$\frac{1}{3-1}\int_1^3 2^x\ dx.$$

First, by Example 3 in Section 6.2,

$$\int_1^3 2^x\ dx = \frac{1}{\ln(2)}(2^3 - 2^1) = \frac{6}{\ln(2)}.$$

The average of the maximum and minimum values of 2^x on $[1,3]$ is $\frac{1}{2}(2^3+2^1)=5$. It's not the same as the average value.

Hence,

$$\text{average value of } 2^x \text{ over } [1,3] = \frac{1}{3-1}\frac{6}{\ln(2)} = \frac{3}{\ln(2)} \approx 4.2381.$$

$\diamond$

The Zero-Integral Principle

Let f be a continuous function on the interval $[a,b]$. Suppose *for every subinterval* $[c,d]$ *of* $[a,b]$ that $\int_c^d f(x)\,dx$ is zero. For example, the constant function $f(x)=0$ has this property. We now show that this is the only such function with this property.

Let $f(x)$ be any continuous function on $[a,b]$ that is not the constant function 0. Then there is a number q in $[a,b]$ such that $f(q)=p$ is not zero. We consider the case when p is positive. (The case when p is negative can be treated the same way. See Exercise 46.)

By the Permanence Property (see Theorem 2.5.4 in Section 2.5), there is a subinterval $[c,d]$ of $[a,b]$, where the function values remain larger than $p/2$. The integral of f over $[c,d]$ is at least $p/2$ times the length of the interval $[c,d]$, hence not 0. This contradicts the assumption that $\int_c^d f(x)\,dx=0$ for all subintervals $[c,d]$ of the domain of f. As a result, the hypothesis must also be false and so f is zero on $[a,b]$.

This interesting result will be applied in Section 18.5 and 18.8.

Zero-Integral Principle

Let f be a continuous function on an interval $[a,b]$. If f has the property that $\int_c^d f(x)\,dx=0$ for every subinterval $[c,d]$ of $[a,b]$, then $f(x)=0$ for all x on $[a,b]$.

WARNING *(Antiderivative Terminology)* As mentioned earlier, in the real world an antiderivative is most often called an "integral" or "indefinite integral". If you stay alert, the context will always reveal whether the word "integral" refers to an antiderivative (a function) or to a definite integral (a number). They are two wildly different beasts. Even so, the next section will show that there is a very close connection between them. This connection ties the two halves of calculus — differential calculus and integral calculus — into one neat package.

Summary

We introduced the notation $\int f(x)\ dx$ for an **antiderivative** of $f(x)$. Using this notation we stated several properties of antiderivatives.

We defined the symbol $\int_a^b f(x)\ dx$ in the special case when $b \le a$, and stated various properties of definite integrals.

The mean-value theorem for definite integrals asserts that for a continuous function $f(x)$, $\int_a^b f(x)\ dx$ equals $f(c)$ times $(b-a)$ for at least one value of c in $[a,b]$.

The quantity $\frac{1}{b-a}\int_a^b f(x)\ dx$ is called the **average value** (or **mean value**) of $f(x)$ over $[a,b]$. It can be thought of as the height of the rectangle whose area is the same as the area of the region under the curve $y = f(x)$.

We justified the zero-integral principle, which says that if the integral of a continuous function is 0 over each interval, then it must be the constant function with valued 0.

EXERCISES for Section 6.3

In Exercises 1 to 12 evaluate each antiderivative. Remember to add a constant to each answer. Check each answer by differentiating it.

1. $\int 5x^2\ dx$

2. $\int \left(7/x^2\right)\ dx$

3. $\int (2x - x^3 + x^5)\ dx$

4. $\int \left(6x^2 + 2x^{-1} + \frac{1}{\sqrt{x}}\right)\ dx$

5.

(a) $\int e^x\ dx$

(b) $\int e^{x/3}\ dx$

6.

(a) $\int \frac{1}{1+x^2}\ dx$

(b) $\int \frac{1}{\sqrt{1-x^2}}\ dx$

7.

(a) $\int \cos(x)\ dx$

(b) $\int \cos(2x)\ dx$

8.

(a) $\int \sin(x)\ dx$

(b) $\int \sin(3x)\ dx$

9.

(a) $\int (2\sin(x) + 3\cos(x))\ dx$

(b) $\int (\sin(2x) + \cos(3x))\ dx$

10. $\int \sec(x)\tan(x)\ dx$

11. $\int (\sec(x))^2\ dx$

12. $\int (\csc(x))^2\ dx$

13. State the mean-value theorem for definite integrals in words, using no mathematical symbols.

14. Define the average value of a function over an interval, using no mathematical symbols.

15. Evaluate

(a) $\int_2^5 x^2\ dx$

(b) $\int_5^2 x^2\ dx$

(c) $\int_5^5 x^2\ dx$

16. Evaluate

(a) $\int_1^2 x\ dx$

(b) $\int_2^1 x\ dx$

(c) $\int_3^3 x\ dx$

17. Find

(a) $\int x\ dx$

(b) $\int_3^4 x\ dx$

18. Find

(a) $\int 3x^2\ dx$

(b) $\int_1^4 3x^2\ dx$

19. If $2 \leq f(x) \leq 3$, what can be said about $\int_1^6 f(x)\ dx$?

20. If $-1 \leq f(x) \leq 4$, what can be said about $\int_{-2}^7 f(x)\ dx$?

21. Write a sentence or two, in your own words, that tells what the symbols $\int f(x)\ dx$ and $\int_a^b f(x)\ dx$ mean. Include examples. Use as few mathematical symbols as possible.

22. Let $f(x)$ be a differentiable function. In this exercise you will determine if the following equation is true or false:

$$f(x) = \int \frac{df}{dx}(x)\ dx.$$

(a) Pick several functions of your choice and test if the equation is true.

(b) Determine if the equation is always true. Write a brief justification for your answer. (Read the equation out loud.)

The mean-value theorem for definite integrals asserts that if $f(x)$ is continuous throughout the interval with endpoints a and b, then $\int_a^b f(x)\ dx = f(c)(b-a)$ for some number c in $[a,b]$. In each of Exercises 23 to 26 find $f(c)$ and at least one value of c in $[a,b]$.

23. $f(x) = 2x$; $[a,b] = [1,5]$

24. $f(x) = 5x+2$; $[a,b] = [1,2]$

25. $f(x) = x^2$; $[a,b] = [0,4]$

26. $f(x) = x^2 + x$; $[a,b] = [1,4]$

27. If $\int_1^2 f(x)\ dx = 3$ and $\int_1^5 f(x)\ dx = 7$, find

(a) $\int_2^1 f(x)\ dx$

(b) $\int_2^5 f(x)\ dx$

28. If $\int_1^3 f(x)\ dx = 4$ and $\int_1^3 g(x)\ dx = 5$, find

(a) $\int_1^3 (2f(x) + 6g(x))\ dx$

(b) $\int_3^1 (f(x) - g(x))\ dx$

29. If the maximum value of $f(x)$ on $[a,b]$ is 7 and the minimum value on $[a,b]$ is 4, what can be said about

(a) $\int_a^b f(x)\,dx$?

(b) the mean value of $f(x)$ on $[a,b]$?

30. Let $f(x) = c$ (constant) for all x in $[a,b]$. Find the average value of $f(x)$ on $[a,b]$.

Exercises 31 to 34 concern the average of a function over an interval. In each case, find the minimum, maximum, and average value of the function over the given interval.

31. $f(x) = x^2$, $[2,3]$

32. $f(x) = x^2$, $[0,5]$

33. $f(x) = 2^x$, $[0,4]$

34. $f(x) = 2^x$, $[2,4]$

35. Let a, b, and c be constants. Assume that the integral of $\left(ax^2+bx+c\right)^2$ over any interval is zero. Find a, b, and c.

36. Let a and b be constants. Assume that the integral of $ae^{x^3} + b\cos^{10}(x)$ over every interval is zero. Find a and b.

37. Prove the mean-value theorem for definite integrals in the case when $b < a$. (Use the definition of $\int_a^b f(x)\,dx$ when $b < a$.)

38. Is $\int f(x)g(x)\,dx$ always equal to $\int f(x)\,dx \int g(x)\,dx$? Are they ever equal? (Explain.)

39.

(a) Show that $\frac{1}{3}(\sin(x))^3$ is *not* an antiderivative of $\sin(x))^2$.

(b) Use the identity $(\sin(x))^2 = \frac{1}{2}(1-\cos(2x))$ to find an antiderivative of $\sin(x))^2$.

(c) Verify your answer in (b) by differentiation.

In Exercises 40 and 41 verify the equations quoted from a table of antiderivatives (integrals). Just differentiate each of the alleged antiderivatives and see whether you obtain the quoted integrand. (The number a is a constant in each case.)

40. $\int x^2 \sin(ax)\ dx = \frac{2x}{a^2}\sin(ax) + \frac{2}{a^3}\cos(ax) - \frac{x^2}{a}\cos(ax) + C$

41. $\int x(\sin(ax))^2\ dx = \frac{x^2}{4} - \frac{x}{4a}\sin(2ax) - \frac{1}{8a^2}\cos(2ax) + C$

42. Define $f(x) = \begin{cases} -x & 0 < x \leq 1 \\ -1 & 1 < x \leq 2 \\ 1 & 2 < x \leq 3 \\ 4-x & 3 < x \leq 4 \end{cases}$.

(a) Sketch the graphs of $y = f(x)$ and $y = (f(x))^2$ on the interval $[0, 4]$.

(b) Find the average value of f on the interval $[0, 4]$.

(c) The **root mean square** (RMS) of a function f on $[a, b]$ is defined as $\sqrt{\frac{1}{b-a}\int_a^b f(x)^2\ dx}$ (The voltage, e.g., 110 volts, for an alternating electric current is the root mean square of a varying voltage.) Find the "root mean square" value of f on the interval $[0, 4]$. That is, compute $\sqrt{\frac{1}{4-0}\int_0^4 (f(x))^2\ dx}$.

(d) Why is it not surprising that your answer in (b) is zero and your answer in (c) is positive?

43.

Sam: The text makes the average value of a function on $[a, b]$ too hard.

Jane: How so?

Sam: It's easy. Just average $f(a)$ and $f(b)$.

Jane: That sure is easier.

(a) Show that Sam is correct when $f(x)$ is any polynomial of degree 0 or 1.

(b) Is Sam always correct? Explain.

Exercise 44 describes the famous **Buffon neeedle** problem, now over 200 years old. Exercise 47 is related, but not nearly as famous.

44. On the floor there are parallel lines a distance d from each other, such as the edges of slats. You throw a straight wire of length d on the floor at random. Sometimes it ends up crossing a line, sometimes it avoids a line.

(a) Perform the experiment at least 20 times and use the results to estimate the percentage of times the wire crosses a line.

(b) If the wire makes an angle θ with a line perpendicular to the lines, show that the probability that it crosses a line is $\cos(\theta)$.

(c) Find the average value of that probability. That average is the probability that the wire crosses a line.

(d) How close is the experimental value in (a) to the theoretical value in (c)?

45. Assume that f and g are continuous functions and that $\int_a^b f(x)\,dx$ equals $\int_a^b g(x)\,dx$ for every interval $[a, b]$. Show that $f(x)$ equals $g(x)$ for all x.

46. Provide the details for the proof of the Zero-Integral Principle in the case when p is negative.

47. An infinite floor is composed of congruent square tiles arranged as in a checkerboard. You have a straight wire whose length is the same as the length of a side of a square. The edges of the squares form lines in perpendicular directions. What is the probability that when you throw the wire at random it crosses two lines, one in each of the two perpendicular directions? (This is related to Exercise 44, the classic Buffon needle problem.) You can check if your answer is reasonable by carrying out the experiment.

48. The average value of a certain function $f(x)$ on $[1, 3]$ is 4. On $[3, 6]$ the average value of the same function is 5. What is its average value on $[1, 6]$? (Explain your answer.)

49. This exercise evaluates two definite integrals that appear often in applications.

(a) Draw the graphs of $y = (\cos(x))^2$ and $y = (\sin(x))^2$. On the basis of your picture, decide how $\int_0^{\pi/2} (\cos(x))^2\,dx$ and $\int_0^{\pi/2} (\sin(x))^2\,dx$ compare.

(b) Using (a) and a trigonometric identity, show that

$$\int_0^{\pi/2} (\cos(x))^2\,dx = \frac{\pi}{4} = \int_0^{\pi/2} (\sin(x))^2\,dx.$$

(c) Evaluate $\int_0^{\pi} (\cos(x))^2\,dx$.

6.4 The Fundamental Theorem of Calculus

Introduction and Motivation

This is the most important section of the entire book.

FTC I gives a shortcut to evaluating $\int_a^b f(x)\ dx$

FTC II gives a way to evaluate $\frac{d}{dx}\left(\int_a^x f(t)\ dt\right)$

In this section we obtain two closely related theorems. They are called the Fundamental Theorems of Calculus I and II, or simply The Fundamental Theorem of Calculus (FTC). The first part of the FTC provides a way to evaluate a definite integral if you are lucky enough to know an antiderivative of the integrand. That means that the derivative, developed in Chapter 3, has yet another application.

The second fundamental theorem tells how rapidly the value of a definite integral changes as you change the interval $[a, b]$ over which you are integrating. This part of the Fundamental Theorem is used to prove the first part of the FTC.

Motivation for the Fundamental Theorem of Calculus I

In Section 6.2 we found that $\int_a^b c\ dx = cb - ca$ and $\int_a^b x\ dx = \frac{b^2}{2} - \frac{a^2}{2}$. In the same section we found that $\int_a^b x^2\ dx = \frac{b^3}{3} - \frac{a^3}{3}$; in this case our reasoning was based on the fact that congruent lopsided tents fill a cube. Finally, using the formula for the sum of a geometric series, we showed that $\int_a^b 2^x\ dx = \frac{2^b}{\ln(2)} - \frac{2^a}{\ln(2)}$.

Notice that all four results follow a similar pattern:

$$\begin{array}{rclcrcl} \int_a^b c\ dx &=& cb - ca & \qquad & \int_a^b x\ dx &=& \frac{b^2}{2} - \frac{a^2}{2} \\ \int_a^b x^2\ dx &=& \frac{b^3}{3} - \frac{a^3}{3} & \qquad & \int_a^b 2^x\ dx &=& \frac{2^b}{\ln(2)} - \frac{2^a}{\ln(2)} \end{array}$$

To describe the similarity in detail, compute an antiderivative of each of the four integrands:

We omit "$+C$" since only one antiderivative is needed here. See Exercises 40 and 41.

$$\begin{array}{rclcrcl} \int c\ dx &=& cx & \qquad & \int x\ dx &=& \frac{x^2}{2} \\ \int x^2\ dx &=& \frac{x^3}{3} & \qquad & \int 2^x\ dx &=& \frac{2^x}{\ln(2)}. \end{array}$$

In each case *the definite integral equals the difference between the values of an antiderivative of the integrand evaluated at b and at a*, the endpoints of the interval.

This suggests that *maybe* for any integrand $f(x)$, the following may be true: If $F(x)$ is an antiderivative of $f(x)$, then

$$\int_a^b f(x)\ dx = F(b) - F(a). \qquad (6.4.1)$$

If this is correct, then, instead of resorting to special tricks to evaluate a definite integral, such as cutting up a cube or summing a geometric series, we should look for an antiderivative of the integrand.

We may reason using "velocity and distance" to provide further evidence for (6.4.1). Picture a particle moving upwards on the y-axis. At time t it is at position $F(t)$ on that line. The velocity at time t is $F'(t)$.

But we saw that the definite integral of the velocity from time a to time b tells the change in position, that is,

"the definite integral of the velocity = the final position − the initial position"

In symbols,

$$\int_a^b F'(t)\ dt = F(b) - F(a). \tag{6.4.2}$$

If we give $F'(t)$ the name $f(t)$, then we can restate (6.4.2) as:

$$\text{If } f(t) = F'(t), \text{ then } \int_a^b f(t)\ dt = F(b) - F(a).$$

In other words,

$$\text{If } F \text{ is an antiderivative of } f, \text{ then } \int_a^b f(t)\ dt = F(b) - F(a).$$

Formulas we found for the integrands c, x, x^2, and 2^x and reasoning about motion are all consistent with

Theorem 6.4.1 (Fundamental Theorem of Calculus I)**.**
If f is continuous on $[a, b]$ and if F is an antiderivative of f then

$$\int_a^b f(x)\ dx = F|_b^a = F(b) - F(a).$$

FTC I

WARNING *(Notation)* When applying FTC I, the difference $F(b) - F(a)$ is usually written as $F|_a^b$.

In practical terms this theorem says, "To evaluate the definite integral of f from a to b, look for an antiderivative of f. Evaluate the antiderivative at b and subtract its value at a. This difference is the value of the definite integral you are seeking". The success of this approach hinges on finding an antiderivative of the integrand f. For many functions, it is easy to find an antiderivative. For some it is hard, but they can be found. For others, the antiderivatives cannot be expressed in terms of the functions met in Chapters 2 and 3, such as polynomials, quotients of polynomials, and functions built up from trigonometric, exponential, and logarithm functions and their inverses.

Some techniques for finding antiderivatives are discussed in Chapter 7.

Example 1 shows the power of FTC I.

EXAMPLE 1 Use the Fundamental Theorem of Calculus to evaluate $\int_0^{\pi/2} \cos(x)\, dx$

SOLUTION Since $(\sin(x))' = \cos(x)$, $\sin(x)$ is an antiderivative of $\cos(x)$. By FTC I,

$$\int_0^{\pi/2} \cos(x)\, dx \quad = \quad \sin(x)|_0^{\pi/2} = \sin(\frac{\pi}{2}) - \sin(0) = 1 - 0 = 1.$$

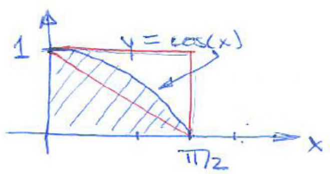


Figure 6.4.1

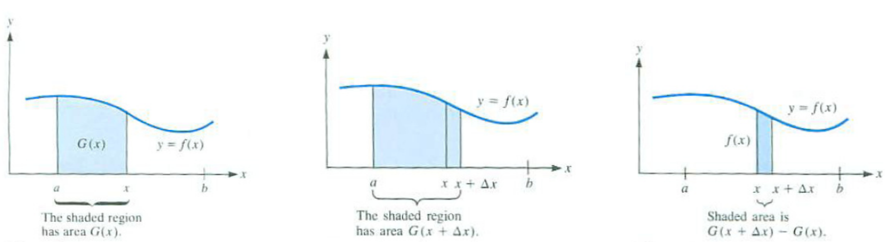
This tells us that the area under the curve $y = \cos(x)$ and above $[0, \pi/2]$, shown in Figure 6.4.1, is 1.

This result is reasonable since the area lies inside a rectangle of area $1 \times \frac{\pi}{2} = \frac{\pi}{2} \approx 1.5708$ and contains a triangle of area $\frac{1}{2}\left(\frac{\pi}{2}\right)1 = \frac{\pi}{4} \approx 0.7854$. ◇

How would the evaluation be different if we used $\sin(x) + 5$ as the antiderivative of $\cos(x)$?

Motivation for the Fundamental Theorem of Calculus II

Let f be a continuous function such that $f(x)$ is positive for x in $[a, b]$. For x in $[a, b]$, let $G(x)$ be the area of the region under the graph of f and above the interval $[a, x]$, as shown in Figure 6.4.2(a). In particular, $G(a) = 0$.



(a) (b) (c)

Figure 6.4.2

We will compute the derivative of $G(x)$, that is,

$$G'(x) = \lim_{\Delta x \to 0} \frac{\Delta G}{\Delta x} = \lim_{\Delta x \to 0} \frac{G(x + \Delta x) - G(x)}{\Delta x}.$$

(This is one of several occasions when we must go back to the definition of the derivative as a limit.) For simplicity, keep Δx positive. Then $G(x+\Delta x)$ is the area under the curve $y = f(x)$ above the interval $[a, x+\Delta x]$. If Δx is small, $G(x+\Delta x)$ is only slightly larger than $G(x)$, as shown in Figure 6.4.2(b). Then $\Delta G = G(x+\Delta x) - G(x)$ is the area of the thin shaded strip in Figure 6.4.2(c).

When Δx is small, the narrow shaded strip above $[x, x+\Delta x]$ resembles a rectangle of base Δx and height $f(x)$, with area $f(x)\Delta x$. Therefore, it seems reasonable that when Δx is small,

$$\frac{\Delta G}{\Delta x} \approx \frac{f(x)\Delta x}{\Delta x} = f(x).$$

In short, it seems plausible that

$$\lim_{\Delta x \to 0} \frac{\Delta G}{\Delta x} = f(x).$$

Briefly,

$$G'(x) = f(x).$$

In words, "the derivative of the area of the region under the graph of f and above $[a, x]$ with respect to x is the value of f at x".

Now we state these observations in terms of definite integrals.

Let f be a continuous function. Let $G(x) = \int_a^x f(t)\ dt$. Then we expect that

$$\frac{d}{dx}\left(\int_a^x f(t)\ dt\right) = f(x).$$

We use t in the integrand to avoid using x to denote both an end of the interval and a variable that takes values between a and x.

This equation says that "the derivative of the definite integral of f with respect to the right end of the interval is simply f evaluated at that end". This is the substance of the Fundamental Theorem of Calculus II. It tells how rapidly the definite integral changes as we change the upper limit of integration.

FTC II

Theorem 6.4.2 (Fundamental Theorem of Calculus II)**.**
Let f be continuous on the interval $[a, b]$. Define

$$G(x) = \int_a^x f(t)\ dt \qquad \text{for all } a \le x \le b.$$

Then G is differentiable on $[a, b]$ and its derivative is f; that is,

$$G'(x) = f(x).$$

As a consequence of FTC II, *every continuous function is the derivative of some function.*

There is a similar theorem for $H(x) = \int_x^b f(t)\,dt$: $H'(x) = -f(x)$. A glance at Figure 6.4.3 shows why there is a minus sign: the area in this figure shrinks as x increases.

See Exercise 63.

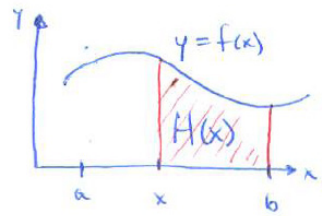
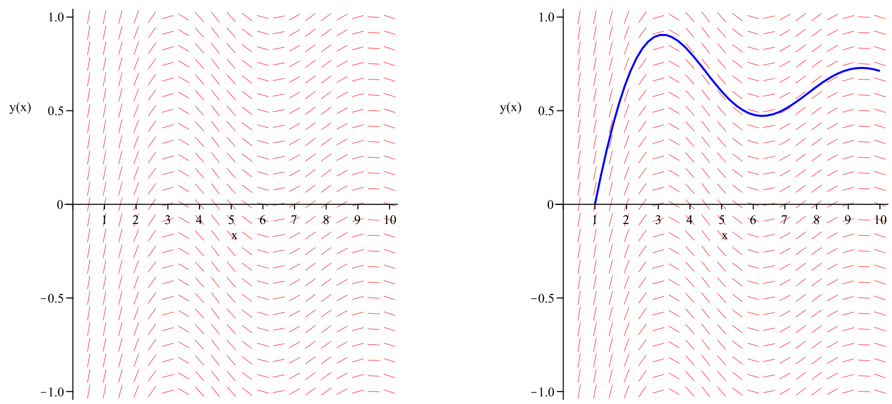


Figure 6.4.3

EXAMPLE 2 Give an example of an antiderivative of $\frac{\sin(x)}{x}$.

SOLUTION There are many antiderivatives of $\frac{\sin(x)}{x}$. Any two antiderivatives differ by a constant. These curves can be seen in the slope field for $y' = \frac{\sin(x)}{x}$ shown in Figure 6.4.4 (a).



(a) (b)

Figure 6.4.4 (a) slope field for $y' = \frac{\sin(x)}{x}$ and (b) same slope field with solution with $y'(1) = \sin(1)$

Let $G(x) = \int_1^x \frac{\sin(t)}{t}\,dt$. By FTC II, $G'(x) = \frac{\sin(x)}{x}$. The graph of $y = G(x)$ is shown in Figure 6.4.4 (b). Notice that $G(1) = 0$. ◇

You probably expected the answer in Example 2 to be an explicit formula for the antiderivative expressed in terms of the familiar functions discussed in Chapters 2 and 3. Recall, from Section 3.6, that the derivative of every elementary function is an elementary function. Liouville proved that there are (many) elementary functions that do not have elementary antiderivatives. Nobody will ever find an explicit formula in terms of elementary functions for an antiderivative of $\frac{\sin(x)}{x}$. (The proof is reserved for a graduate course.)

Joseph Liouville (1809–1882) `http://en.wikipedia.org/wiki/Joseph_Liouville`

EXAMPLE 3 Give an example of an antiderivative of $\frac{\sin(\sqrt{x})}{\sqrt{x}}$.

SOLUTION This integrand appears more terrifying than $\frac{\sin(x)}{x}$, yet it does have an elementary antiderivative, namely $-2\cos(\sqrt{x})$. To check, we differentiate $y = -2\cos(\sqrt{x})$ by the Chain Rule. We have $y = -2\cos(u)$ where

$u = \sqrt{x}$. Therefore,

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = -2(-\sin(u))\frac{1}{2\sqrt{x}} = \frac{\sin(\sqrt{x})}{\sqrt{x}}.$$

◇

Because the antiderivatives of $\frac{\sin(\sqrt{x})}{\sqrt{x}}$ are elementary functions, it would be easy to calculate $\int_1^2 \frac{\sin(\sqrt{x})}{\sqrt{x}}\,dx$.

Any antiderivative of e^x is of the form $e^x + C$, an elementary function. However, *no* antiderivative of e^{-x^2} is elementary. Statisticians define the **error function** to be $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2/2}\,dt$. Except that $\operatorname{erf}(0) = 0$, there is no easy way to evaluate $\operatorname{erf}(x)$. Since $\operatorname{erf}(x)$ is not elementary, it is customary to collect approximate values of it for various values of x in a table. Approximate values of special functions such as the error function can also be obtained from mathematical software and even a few calculators.

Net Area

When we evaluate $\int_0^\pi \cos(x)\,dx$, we obtain $\sin(\pi) - \sin(0) = 0 - 0 = 0$. What does this say about areas? Inspection of Figure 6.4.5 shows what is happening.

For x in $[\pi/2, \pi]$, $\cos(x)$ is negative and the curve $y = \cos(x)$ lies *below* the x-axis. If we interpret the corresponding area as negative, then we see that it cancels with the area from 0 to $\pi/2$. Let us agree that when we say "$\int_a^b f(x)\,dx$ represents the area under the curve $y = f(x)$", we mean that it represents the area between the curve and the x-axis, *with area below the x-axis taken as negative*. This is the **net area** under $y = f(x)$ on the interval $[a, b]$. Note that the net area can be positive, zero, or negative.

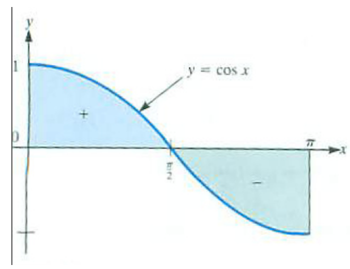


Figure 6.4.5 The area of a region below the x-axis is negative.

EXAMPLE 4 Evaluate $\int_1^2 \frac{1}{x^2}\,dx$ by the Fundamental Theorem of Calculus I.

SOLUTION In order to apply FTC I we have to find an antiderivative of $\frac{1}{x^2}$. In Section 6.3 it was observed that

$$\int x^a\,dx = \frac{1}{a+1}x^{a+1} + C \qquad a \neq -1.$$

In particular, with $a = -2$,

$$\int \frac{1}{x^2}\,dx = \int x^{-2}\,dx = \frac{1}{(-2)+1}x^{(-2)+1} + C = \frac{1}{-1}x^{-1} + C = \frac{-1}{x} + C$$

By FTC I

$$\int_1^2 \frac{1}{x^2}\,dx = \left(\frac{-1}{x} + C\right)\Bigg|_1^2 = \left(\frac{-1}{2} + C\right) - \left(\frac{-1}{1} + C\right) = \frac{-1}{2} - (-1) = \frac{1}{2}.$$

Note that the C's cancel. We do not need the C when applying FTC I.

◇

The First Fundamental Theorem of Calculus asserts that

$$\underbrace{\int_1^2 \frac{1}{x^2}\,dx}_{\text{The definite integral: a limit of sums}} = \underbrace{\int \frac{1}{x^2}\,dx\bigg|_1^2}_{\text{The difference between an antiderivative evaluated at 2 and at 1}}$$

The symbols on the right and left of the equal sign are so similar that it is tempting to think that the equation is obvious or says nothing whatsoever.

> **WARNING** *(Notation)* This equation is a special instance of the First Fundamental Theorem of Calculus, FTC I.

> **Remark:** Often we write $\int \frac{1}{x^2}\,dx$ as $\int \frac{dx}{x^2}$, merging the 1 with the dx. More generally, $\int \frac{f(x)}{g(x)}\,dx$ may be written as $\int \frac{f(x)\,dx}{g(x)}$.

Some Terms and Notation

The related processes of computing $\int_a^b f(x)\,dx$ and of finding an antiderivative $\int f(x)\,dx$ are both called **integrating** $f(x)$. Thus integration refers to two separate but related problems: computing a number $\int_a^b f(x)\,dx$ or finding a function $\int f(x)\,dx$.

In practice, both FTC I and FTC II are called "the Fundamental Theorem of Calculus." The context always makes it clear which one is meant.

Proofs of the Two Fundamental Theorems of Calculus

We now prove both parts of the Fundamental Theorem of Calculus — without referring to motion, area, or concrete examples. The proofs use only the mathematics of functions and limits. We prove FTC II first; then we will use it to prove FTC I.

Proof of the Second Fundamental Theorem of Calculus

The Second Fundamental Theorem of Calculus asserts that the derivative of $G(x) = \int_a^x f(t)\,dt$ is $f(x)$. We gave a convincing argument using areas of regions. However, since definite integrals are defined in terms of approximating sums, not areas, we include a proof that uses only properties of definite integrals.

Proof of Fundamental Theorem of Calculus II

We wish to show that $G'(x) = f(x)$. To do this we must make use of the definition of the derivative of a function.

We have

$$
\begin{aligned}
G'(x) &= \lim_{\Delta x \to 0} \frac{G(x+\Delta x) - G(x)}{\Delta x} && \text{(definition of derivative)} \\
&= \lim_{\Delta x \to 0} \frac{\int_a^{x+\Delta x} f(t)\ dt - \int_a^x f(t)\ dt}{\Delta x} && \text{(definition of } G\text{)} \\
&= \lim_{\Delta x \to 0} \frac{\int_a^x f(t)\ dt + \int_x^{x+\Delta x} f(t)\ dt - \int_a^x f(t)\ dt}{\Delta x} && \text{(property 6 in Section 6.3)} \\
&= \lim_{\Delta x \to 0} \frac{\int_x^{x+\Delta x} f(t)\ dt}{\Delta x} && \text{(canceling)} \\
&= \lim_{\Delta x \to 0} \frac{f(c)\Delta x}{\Delta x} && \text{(MVT for Definite Integrals; } c \text{ between } x \text{ and } x+\Delta x\text{)} \\
&= \lim_{\Delta x \to 0} f(c) && \text{(canceling)} \\
&= f(x). && \text{(continuity of } f\text{; } c \to x \text{ as } \Delta x \to 0\text{)}
\end{aligned}
$$

Hence

$$G'(x) = f(x),$$

which is what we set out to prove. •

A similar argument shows that

$$\frac{d}{dx} \int_x^b f(t)\ dt = -f(x).$$

For integrands whose values are positive, the minus sign is to be expected. As x increases, the interval shrinks, and so the (positive) area under the curve shrinks as well.

Proof of the First Fundamental Theorem of Calculus

The First Fundamental Theorem of Calculus asserts that if $F' = f$, then $\int_a^b f(x)\,dx = F(b) - F(a)$. We persuaded ourselves that this is true by thinking of f as "velocity" and F as "position", and also by four special cases ($f(x) = c$, $f(x) = x$, $f(x) = x^2$, and $f(x) = 2^x$). We now prove the theorem, which is an immediate consequence of the Second Fundamental Theorem of Calculus and the fact that two antiderivatives of the same function differ by a constant.

Proof of the Fundamental Theorem of Calculus I

We are assuming that $F' = f$ and wish to show that $F(b) - F(a) = \int_a^b f(x)\ dx$. Define $G(x)$ to be $\int_a^x f(t)\ dt$. By FTC II, G is an antiderivative of f. Since F and G are both antiderivatives of f, they differ by a constant, say C. That is,

$$F(x) = G(x) + C.$$

Thus,

$$\begin{aligned}
F(b) - F(a) &= (G(b) + C) - (G(a) + C) & \\
&= G(b) - G(a) & (C\text{'s cancel}) \\
&= \int_a^b f(t)\ dt - \int_a^a f(t)\ dt & (\text{definition of } G) \\
&= \int_a^b f(t)\ dt & \left(\int_a^a f(t)\ dt = 0\right)
\end{aligned}$$

•

Summary

This section links the two basic ideas of calculus, the derivative (more precisely, the antiderivative) and the definite integral.

FTC I says that if you can find a formula for an antiderivative F of f, then you can evaluate $\int_a^b f(x)\ dx$:

$$\boxed{\int_a^b f(x)\ dx = F(b) - F(a).}$$

FTC II says that if f is continuous then it has an antiderivative, namely $G(x) = \int_a^x f(t)\ dt$; that is, $G'(x) = f(x)$. Unfortunately, G might not be an elementary function. However, a reasonable graph of an antiderivative of f can be obtained from the slope field for $\frac{dy}{dx} = f(x)$.

EXERCISES for Section 6.4

1. State (a) FTC I and (b) FTC II.

2. Using only words, no mathematical symbols, state the First Fundamental Theorem of Calculus.

3. Using only words, no mathematical symbols, state the Second Fundamental Theorem of Calculus.

In Exercises 4 and 5 evaluate the given expressions.

4.

(a) $x^3\big|_1^2$

(b) $x^2\big|_{-1}^2$

(c) $\cos(x)|_0^\pi$

5.

(a) $(x + \sec(x))|_0^{\pi/4}$

(b) $\left.\dfrac{1}{x}\right|_2^3$

(c) $\sqrt{x-1}\big|_5^{10}$

In Exercises 6 to 19 use FTC I to evaluate the given definite integrals.

6. $\displaystyle\int_1^2 5x^3\ dx$

7. $\displaystyle\int_{-1}^3 2x^4\ dx$

8. $\displaystyle\int_1^4 (x + 5x^2)\ dx$

9. $\displaystyle\int_1^2 (6x - 3x^2)\ dx$

10. $\displaystyle\int_{\pi/6}^{\pi/3} 5\cos(x)\ dx$

11. $\displaystyle\int_{\pi/4}^{3\pi/4} 3\sin(x)\ dx$

12. $\displaystyle\int_{0}^{\pi/2} \sin(2x)\ dx$

13. $\displaystyle\int_{0}^{\pi/6} \cos(3x)\ dx$

14. $\displaystyle\int_{4}^{9} 5\sqrt{x}\ dx$

15. $\displaystyle\int_{1}^{9} \frac{1}{\sqrt{x}}\ dx$

16. $\displaystyle\int_{1}^{8} \sqrt[3]{x^2}\ dx$

17. $\displaystyle\int_{2}^{4} \frac{4}{x^3}\ dx$

18. $\displaystyle\int_{0}^{1} \frac{dx}{1+x^2}$

19. $\displaystyle\int_{1/4}^{1/2} \frac{dx}{\sqrt{1-x^2}}$

In Exercises 20 to 25 find the average value of the given function over the given interval.

20. x^2; $[3, 5]$

21. x^4; $[1, 2]$

22. $\sin(x)$; $[0, \pi]$

23. $\cos(x)$; $[0, \pi/2]$

24. $(\sec(x))^2$; $[\pi/6, \pi/4]$

25. $\sec(2x)\tan(2x)$; $[\pi/8, \pi/6]$

In Exercises 26 to 33 evaluate the given quantities.

26. The area of the region under the curve $y = 3x^2$ and above $[1, 4]$.

27. The area of the region under the curve $y = 1/x^2$ and above $[2, 3]$.

28. The area of the region under the curve $y = 6x^4$ and above $[-1, 1]$.

29. The area of the region under the curve $y = \sqrt{x}$ and above $[25, 36]$.

30. The distance an object travels from time $t = 1$ second to time $t = 2$ seconds, if its velocity at time t seconds is t^5 feet per second.

31. The distance an object travels from time $t = 1$ second to time $t = 8$ seconds, if its velocity at time t seconds is $7\sqrt[3]{t}$ feet per second.

32. The volume of a solid located between a plane at $x = 1$ and a plane located at $x = 5$ if the cross-sectional area of the intersection of the solid with the plane perpendicular to the x-axis through the point $(x, 0)$ has area $6x^3$ square centimeters. (See Figure 6.4.6.)

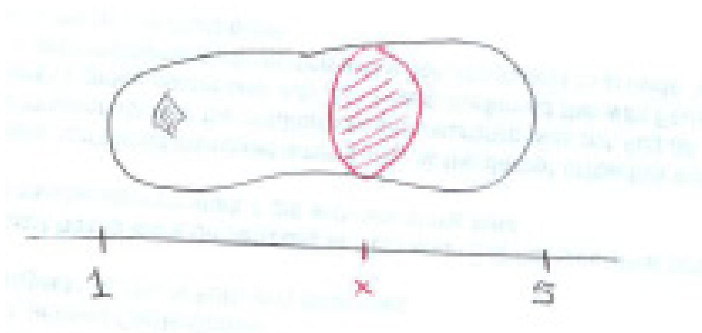


Figure 6.4.6

33. The volume of a solid located between a plane at $x = 1$ and a plane located at $x = 5$ if the cross-sectional area of the intersection of the solid with the plane perpendicular to the x-axis through the point $(x, 0)$ has area $1/x^3$ square centimeters.

34. Let f be a continuous function. Estimate $f(7)$ if $\int_5^7 f(x)dx = 20.4$ and $\int_5^{7.05} f(x)dx = 20.53$.

35. Determine if each of the following expressions is a function or a number.

(a) $\int x^2\ dx$

(b) $\int x^2\ dx\big|_1^3$

(c) $\int_1^3 x^2\ dx$

36.

(a) Which of these two numbers is defined as a limit of sums?

$$\left. \int x^2 \, dx \right|_1^2 \qquad \text{and} \qquad \int_1^2 x^2 \, dx$$

(b) How is the other number defined?

(c) Why are the two numbers in (a) equal?

37. There is no elementary antiderivative of $\sin(x^2)$. Does $\sin(x^2)$ have an antiderivative? Explain.

38. True or false:

(a) Every elementary function has an elementary derivative.

(b) Every elementary function has an elementary antiderivative.

Explain.

39.

(a) Draw the slope field for $\dfrac{dy}{dx} = \dfrac{e^{-x}}{x}$ for $x > 0$.

(b) Use (a) to sketch the graph of an antiderivative of $\frac{e^{-x}}{x}$.

(c) On the slope field drawn in (a), sketch the graph of $f(x) = \displaystyle\int_1^x \frac{e^{-t}}{t} \, dt$. (For which one value of x is $f(x)$ easy to compute?)

Exercises 40 and 41 illustrate why FTC I can be applied using any antiderivative of the integrand.

40. Evaluate the definite integral $\int_a^b x \, dx$ using each of the following antiderivatives of $f(x) = x$.

(a) $F(x) = \frac{1}{2}x^2 + 1$.

(b) $F(x) = \frac{1}{2}x^2 - 3$.

(c) $F(x) = \frac{1}{2}x^2 + C$.

41. Evaluate the definite integral $\int_a^b 2^x \, dx$ using each of the following antiderivatives of $f(x) = 2^x$.

(a) $F(x) = \frac{1}{\ln(2)} 2^x + 11$.

(b) $F(x) = \frac{1}{\ln(2)} 2^x - 7$.

(c) $F(x) = \frac{1}{\ln(2)} 2^x + C$.

42. Let $F(x) = \int_0^x e^{t^2}\, dt$.

(a) Does the graph of $F(x)$ have inflection points? If so, find them.

(b) Make a rough sketch of the graph of $F(x)$.

43. Area was used in Section 6.2 to develop $\int_a^b x\, dx = \frac{b^2}{2} - \frac{a^2}{2}$ when $0 < a < b$. To see that this result is true for all values of a and b (with $b > a$) we will consider these additional cases:

(a) If $a < b < 0$, work with negative area.

(b) If $a < 0 < b$, divide the interval $[a, b]$ into two pieces and work with signed areas.

44. Find $\frac{dy}{dx}$ if

(a) $y = \int \sin(x^2)\, dx$

(b) $y = 3x + \int_{-2}^{3} \sin(x^2)\, dx$

(c) $y = \int_{-2}^{x} \sin(t^2)\, dt$

In Exercises 45 to 48 differentiate the given functions.
45.

(a) $\int_1^x t^4\, dt$

(b) $\int_x^1 t^4\, dt$ (Re-write this integral with x as the upper limit of integration.)

46.

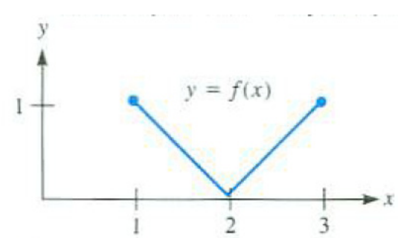
(a) $\int_1^x \sqrt[3]{1 + \sin(t)}\, dt$

(b) $\int_1^{x^2} \sqrt[3]{1+\sin(t)}\ dt$ (Use the Chain Rule.)

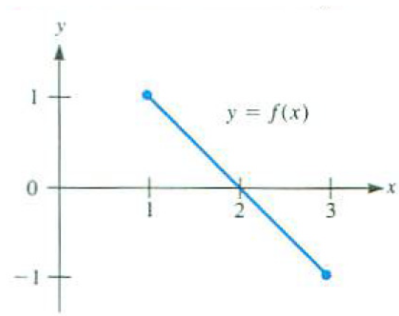
47. $\int_{-1}^{x} 3^{-t}\ dt$

48. $\int_{2x}^{3x} t\tan(t)\ dt$ (Assume x is in the interval $(-\pi/6, \pi/6)$.) (First rewrite the integral as $\int_{2x}^{0} t\tan(t)\ dt + \int_0^{3x} t\tan(t)\ dt$.)

49. Figure 6.4.7(a) shows the graph of a function $f(x)$ for x in $[1,3]$. Let $G(x) = \int_1^x f(t)\ dt$. Graph $y = G(x)$ for x in $[1,3]$ as well as you can. Explain your reasoning.



(a)



(b)

Figure 6.4.7

50. Figure 6.4.7(b) shows the graph of a function $f(x)$ for x in $[1,3]$. Let $G(x) = \int_1^x f(t)\ dt$. Graph $y = G(x)$ for x in $[1,3]$ as well as you can. Explain your reasoning.

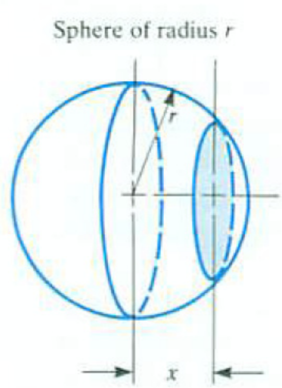


Figure 6.4.8 ARTIST: Change “Sphere” to “Ball”

51. A plane at a distance x from the center of the ball of radius r, $0 \le x \le 4$, meets the ball in a disk. (See Figure 6.4.8.)

(a) Show that the radius of the disk is $\sqrt{r^2 - x^2}$.

(b) Show that the area of the disk is $\pi r^2 - \pi x^2$.

(c) Using the FTC, find the volume of the ball.

52. Let $v(t)$ be the velocity at time t of an object moving on a straight line. The velocity may be positive or negative.

(a) What is the physical meaning of $\int_a^b v(t)\ dt$? Explain.

(b) What is the physical meaning of the slope of the graph of $y = v(t)$? Explain.

(c) What is the physical meaning of $\int_a^b |v(t)|\ dt$? Explain.

53. Give an example of a function f such that $f(4) = 0$ and $f'(x) = \sqrt[3]{1 + x^2}$.

54. Let f be a continuous function. Show that $\frac{d}{dx} \int_x^b f(x)\ dx = -f(x)$

(a) by using the definition of derivative as a limit

(b) by using properties of the definite integral and FTC II.

55. If $f(x) = \int_{-1}^x \sin^3\left(e^{t^2}\right) dt$, find $f'(1)$.

56. If $\int_1^x f(t)dt = \sin^3(5x)$, find $f'(3)$.

57. Figure 6.4.9 shows the graph of a function f. Let $A(x)$ be the area under the graph of f and above the interval $[1, x]$.

(a) Find $A(1)$, $A(2)$, and $A(3)$.

(b) Find $A'(1)$, $A'(2)$, and $A'(3)$.

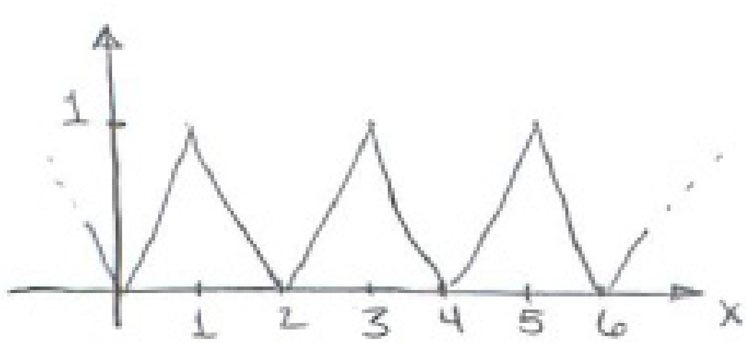


Figure 6.4.9

58.

(a) If $\int_x^{x+4} g(t)dt = 5$ for all x, what can be said about the graph of g?

(b) How would you construct such a function?

59. Find $D\left(\int_{x^2}^{x^3} e^{t^2} dt\right)$.

60. Find $D\left(\int_{x^2}^{5} \sin^{10}(3t)dt\right)$.

61. Find the derivative of $\cos(t^2)\big|_{2x}^{3x}$.

62. How often should a machine be overhauled? This depends on the rate $f(t)$ at which it depreciates and the cost A of overhaul. Denote the time between overhauls by T.

(a) Explain why you would like to minimize $g(T) = \frac{1}{T}(A + \int_0^T f(t)\ dt)$.

(b) Find $\frac{dg}{dT}$.

(c) Show that if $\frac{dg}{dT} = 0$, then $f(T) = g(T)$.

(d) Is this reasonable? Explain.

63. Let $f(x)$ be a continuous function with only positive values. Define $H(x) = \int_x^b f(t)\ dt$ for all $a \le x \le b$. Let Δx be positive.

(a) Interpreting the definite integral as an area of a region, draw the regions whose areas are $H(x)$ and $H(x + \Delta x)$.

(b) Is $H(x + \Delta x) - H(x)$ positive or negative?

(c) Draw the region whose area is related to $H(x + \Delta x) - H(x)$.

(d) When Δx is small, estimate $H(x + \Delta x) - H(x)$ in terms of the integrand f.

(e) Use (d) to evaluate the derivative $H'(x)$:

$$\frac{dH}{dx} = \lim_{\Delta x \to 0} \frac{H(x + \Delta x) - H(x)}{\Delta x}.$$

64. Say that you want to find the area of a certain planar cross-section of a rock. One way to find it is by sawing the rock in two and measuring the area directly. But suppose you do not want to ruin the rock. However, you do have a measuring glass, as shown in Figure 6.4.10, which gives you excellent volume measurements. How could you use the glass to get a good estimate of the cross-sectional area?

Figure 6.4.10

65. Let R be a function with continuous second derivative R''. Assume $R(1) = 2$, $R'(1) = 6$, $R(3) = 5$, and $R'(3) = 8$. Evaluate $\int_1^3 R''(x)\ dx$. Not all of the information provided is needed.

66. Two conscientious calculus students are having an argument:

Jane: $\int_a^b f(x)\ dx$ is a number.

Sam: But if I treat b as a variable, then it is a function.

Jane: How can it be both a number and a function?

Sam: It depends on what "it" means.

Jane: You can't get out of this so easily.

Which student is correct? That is, either give two interpretations of "it" or explain why "it" has only one meaning.

67. The function $\frac{e^x}{x}$ does not have an elementary antiderivative. Show that its reciprocal, $\frac{x}{e^x}$, does have an elementary antiderivative. (Write $\frac{x}{e^x}$ as xe^{-x} and then experiment for a few minutes.)

68. Show that if we knew that every continuous function has an antiderivative, then FTC I would imply FTC II.

69.

(a) Show that for any constant function, $f(x) = c$, the average value of f over $[a, b]$ is the same as the value of the function at the midpoint of the interval $[a, b]$.

(b) Give an example of a non-constant function f such that for any interval $[a, b]$,

$$\frac{\int_a^b f(t)dt}{b-a} = f\left(\frac{a+b}{2}\right).$$

(c) Show that if a continuous function f on $(-\infty, \infty)$ satisfies the equation in (b), it is differentiable.

(d) Find all continuous functions that satisfy the equation in (b).

70. Find all continuous functions f such that their average over $[0, x]$ always equals $f(x)$.

71. Give a geometric explanation of the following properties of definite integrals:

(a) if f is an even function, then $\int_{-a}^{a} f(t)dt = 2\int_0^a f(t)dt$.

(b) if f is an odd function, then $\int_{-a}^{a} f(t)dt = 0$.

(c) if f is a periodic function with period p, then, for any integers m and n, $\int_{mp}^{np} f(t)dt = (n-m)\int_0^p f(t)dt$.

72. Use FTC II to explain why, if u and v are differentiable functions,

(a) $\frac{d}{dx}\int_a^{v(x)} f(t)\ dt = f(v(x))v'(x)$

(b) $\frac{d}{dx}\int_{u(x)}^{b} f(t)\ dt = -f(u(x))u'(x)$

(c) $\frac{d}{dx}\int_{u(x)}^{v(x)}\ dt = f(v(x))v'(x) - f(u(x))u'(x)$

(In (c), break the integral into two convenient integrals.)

73. For which continuous functions f is the average value of f on the interval $[0, b]$ a non-decreasing function of b?

6.5 Estimating a Definite Integral

It is easy to evaluate $\int_0^1 x^2\sqrt{1+x^3}\,dx$ by the Fundamental Theorem of Calculus, for the integrand has an elementary antiderivative, $\frac{2}{9}(1+x^3)^{3/2}$. (Check that $\frac{d}{dx}\frac{2}{9}(1+x^3)^{3/2}$ simplifies to $x^2\sqrt{1+x^3}$.) However, an antiderivative of $\sqrt{1+x^3}$ is not elementary, so $\int_0^1 \sqrt{1+x^3}\,dx$ cannot be evaluated so easily. In this case we have to estimate it. This section describes three ways to do this.

Approximation by Rectangles

The definite integral $\int_a^b f(x)\,dx$ is, by definition, a limit of sums of the form

$$\sum_{i=1}^{n} f(c_i)(x_i - x_{i-1}). \qquad (6.5.1)$$

Any such sum is an estimate of $\int_a^b f(x)\,dx$.

In terms of area, the area of a rectangle gives a local estimate of the area under the graph of $y = f(x)$ above the interval $[x_{i-1}, x_i]$. See Figure 6.5.1. The sum of the areas of individual rectangles is an estimate the area under the curve.

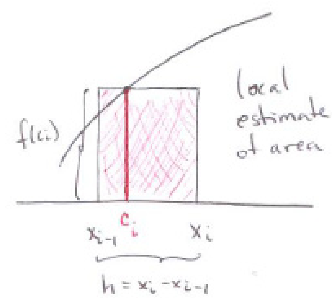


Figure 6.5.1

To use rectangles to estimate $\int_a^b f(x)\,dx$, divide the interval $[a, b]$ into n sections of equal length by the $n+1$ numbers $a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$. (Choosing the sections to have the same length simplifies the arithmetic.) The width of each section is $h = (b-a)/n$. Then choose a sampling number c_i in the i^{th} section, $i = 1, 2, \ldots, n$ and form the Riemann sum $\sum_{i=1}^{n} f(c_i)h$. By the very definition of the definite integral, this sum is an estimate of the definite integral.

Denoting $f(x_i)$ by y_i, and using the left endpoint x_{i-1} of each interval $[x_{i-1}, x_i]$ as the sampling number, we have this **left endpoint rectangular estimate**

$$\int_a^b f(x)\,dx \approx h(y_0 + y_1 + y_2 + \cdots + y_{n-2} + y_{n-1}), \qquad (h = (b-a)/n).$$

If the right endpoints are used, we have the **right endpoint rectangular estimate**:

$$\int_a^b f(x)\,dx \approx h(y_1 + y_2 + \cdots + y_{n-1} + y_n), \qquad (h = (b-a)/n).$$

We will illustrate this and other ways to estimate a definite integral by estimating $\int_0^1 \frac{dx}{1+x^2}$. We chose this integral because it can be easily computed by the FTC:

$$\int_0^1 \frac{dx}{1+x^2} = \arctan(x)|_0^1 = \arctan(1) - \arctan(0) = \frac{\pi}{4} \approx 0.785398.$$

That enables us to judge the accuracy of each method.

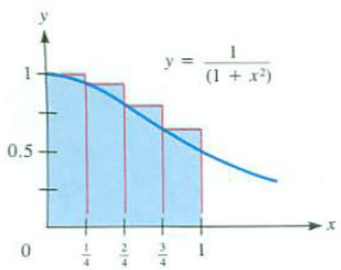


Figure 6.5.2

EXAMPLE 1 Use four rectangles with equal widths to estimate $\int_0^1 \frac{dx}{1+x^2}$. Use the left endpoint of each section as the sampling number to determine the height of each rectangle.
SOLUTION Since the length of $[0,1]$ is 1, each of the four sections of equal length has length $\frac{1}{4}$. See Figure 6.5.2. The sum of the areas of the rectangles is

$$\frac{1}{1+0^2}\cdot\frac{1}{4} + \frac{1}{1+\left(\frac{1}{4}\right)^2}\cdot\frac{1}{4} + \frac{1}{1+\left(\frac{2}{4}\right)^2}\cdot\frac{1}{4} + \frac{1}{1+\left(\frac{3}{4}\right)^2}\cdot\frac{1}{4},$$

which equals $$\frac{1}{4}\left(1 + \frac{16}{17} + \frac{16}{20} + \frac{16}{25}\right).$$

This is approximately

$$\frac{1}{4}(1.0000 + 0.9411 + 0.8000 + 0.6400) \quad = \quad \frac{1}{4}(3.3811) \approx 0.845294.$$

$\diamond$

As Figure 6.5.2 shows, it is an overestimate; it exceeds the definite integral by about 0.06.

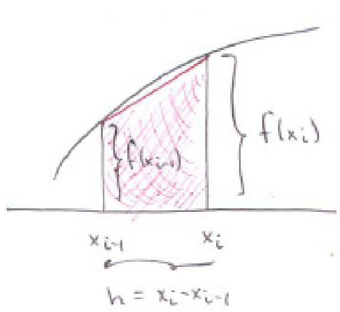


Figure 6.5.3 ARTIST: indicate height is $f(x_i) = y_i$

Approximation by Trapezoids

Trapezoids can also be used to find a local estimate of the area under the graph of $y = f(x)$ above the interval $[x_{i-1}, x_i]$. The basic idea is shown in Figure 6.5.3.

The area, A, of a trapezoid with base width h and side lengths b_1 and b_2 is the product of the base width and the average of the two side lengths: $A = \frac{1}{2}(b_1 + b_2)h$. (See Figure 6.5.4.)

The formula for the trapezoidal estimate of $\int_a^b f(x)\ dx$ follows from an argument like the one for the rectangular estimate.

Let n be a positive integer. Divide the interval $[a, b]$ into n sections of equal length $h = (b - a)/n$ with

$$x_0 = a,\ x_1 = a + h,\ x_2 = a + 2h,\ \ldots,\ x_n = a + nh = b.$$

Denote $f(x_i)$ by y_i. The local estimate of the area under $y = f(x)$ and above $[x_{i-1}, x_i]$ is

$$\frac{1}{2}(y_{i-1} + y_i)h.$$

Summing the n local estimates of area gives the formula for the trapezoidal estimate of $\int_a^b f(x)\ dx$:

$$\frac{y_0 + y_1}{2} \cdot h + \frac{y_1 + y_2}{2} \cdot h + \cdots + \frac{y_{n-1} + y_n}{2} \cdot h$$

Factoring out $h/2$ and collecting like terms gives us the **trapezoidal estimate**:

$$\int_a^b f(x)\ dx \approx \frac{h}{2}\left(y_0 + 2y_1 + 2y_2 + \cdots + 2y_{n-1} + y_n\right). \qquad (6.5.2)$$

There are n sections of width $h = (b - a)/n$, each corresponding to one trapezoid. However, the function is evaluated at $n + 1$ points, including both ends of the interval $[a, b]$.

Note that y_0 and y_n have coefficient 1 while all other y_i's have coefficient 2. This is due to the double counting of the edges common to two trapezoids.

If $f(x)$ is a polynomial of the form $A + Bx$, its graph is a straight line. The top edge of each approximating trapezoid coincides with the graph. The approximation (6.5.2) in this special case gives the exact value of $\int_a^b f(x)\ dx$. There is no error.

Figures 6.5.5 and 6.5.6 illustrate the trapezoidal estimate for the case $n = 4$. Notice that in Figure 6.5.5 the function is concave down and the trapezoidal estimate underestimates $\int_a^b f(x)\ dx$. On the other hand, when the curve is concave up the trapezoids overestimate, as shown in Figure 6.5.6. In both cases the trapezoids appear to give a better approximation of $\int_a^b f(x)\ dx$ than the same number of rectangles. For this reason we expect the trapezoidal

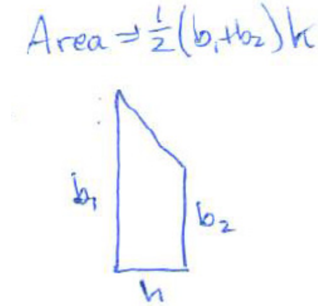


Figure 6.5.4

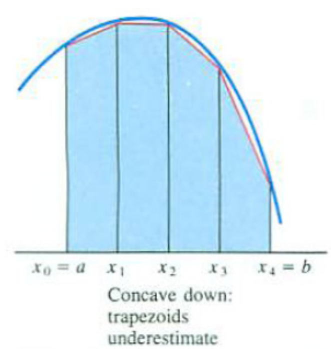


Figure 6.5.5

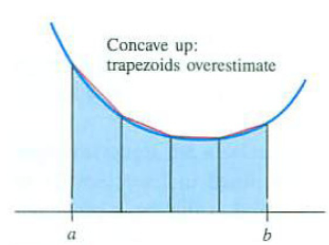


Figure 6.5.6

method to provide better estimates of a definite integral than we obtain by rectangles.

EXAMPLE 2 Use the trapezoidal method with $n = 4$ to estimate $\int_0^1 \frac{dx}{1+x^2}$.
SOLUTION In this case $a = 0$, $b = 1$, and $n = 4$, so $h = (1-0)/4 = \frac{1}{4}$. The four trapezoids are shown in Figure 6.5.7. The trapezoidal estimate is

$$\frac{h}{2}\left(f(0) + 2f\left(\frac{1}{4}\right) + 2f\left(\frac{2}{4}\right) + 2f\left(\frac{3}{4}\right) + f(1)\right).$$

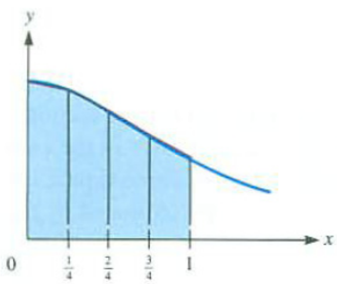


Figure 6.5.7 ARTIST: Try to make top side of trapezoids more visible.

Now, $h/2 = \frac{1}{4}/2 = 1/8$. To compute the sum of the five terms involving values of $f(x) = \frac{1}{1+x^2}$, make a list as shown in Table 6.5.1.

x_i	$f(x_i)$	coefficient	summand	decimal form
0	$\frac{1}{1+0^2}$	1	$1 \cdot \frac{1}{1+0}$	1.0000
$\frac{1}{4}$	$\frac{1}{1+\left(\frac{1}{4}\right)^2}$	2	$2 \cdot \frac{1}{1+\frac{1}{16}}$	1.8823
$\frac{2}{4}$	$\frac{1}{1+\left(\frac{2}{4}\right)^2}$	2	$2 \cdot \frac{1}{1+\frac{4}{16}}$	1.6000
$\frac{3}{4}$	$\frac{1}{1+\left(\frac{3}{4}\right)^2}$	2	$2 \cdot \frac{1}{1+\frac{9}{16}}$	1.2800
$\frac{4}{4}$	$\frac{1}{1+\left(\frac{4}{4}\right)^2}$	1	$1 \cdot \frac{1}{1+\frac{16}{16}}$	0.5000

Table 6.5.1

The trapezoidal sum is therefore, approximately,

$$\frac{1}{8}(1.0000 + 1.8823 + 1.6000 + 1.2800 + 0.5000) \approx \frac{1}{8}(6.2623) \approx 0.7827.$$

Thus

$$\int_0^1 \frac{dx}{1+x^2} \approx 0.782794.$$

This estimate differs from the definite integral by about 0.0026, which is much smaller than the error in the rectangular method, which had an error of 0.06. $\diamond$

Comparison of Rectangular and Trapezoidal Estimates

If we divide out the 2 in the trapezoidal estimate, it takes the form

$$h\left(\frac{y_0}{2} + y_1 + y_2 + \cdots + y_{n-1} + \frac{y_n}{2}\right). \tag{6.5.3}$$

In this form it looks much like the rectangular estimate. It has $n+1$ summands, while the rectangular estimate has only n summands. However, if $f(a)$ happens to equal $f(b)$, that is, $y_0 = y_n$, then (6.5.3) can be written either as $h(y_0 + y_1 + y_2 + \cdots + y_{n-1})$ (the left endpoint rectangular estimate) or as $h(y_1 + y_2 + \cdots + y_{n-1} + y_n)$ (the right endpoint rectangular estimate). In this special case when $f(a) = f(b)$ the three estimates for $\int_a^b f(x)\,dx$ coincide.

Simpson's Estimate: Approximation by Parabolas

In the trapezoidal estimate a curve is approximated by chords. Simpson's estimate for $\int_a^b f(x)\,dx$ approximates the curve by parabolas. Given *three* points on a curve, there is a unique parabola of the form $y = Ax^2 + Bx + C$ that passes through them, as shown in Figure 6.5.8. (See Exercise 28.) The area under the parabola is then used to approximate the area under the curve.

Thomas Simpson, 1710–1761, `http://en.wikipedia.org/wiki/Thomas_Simpson`

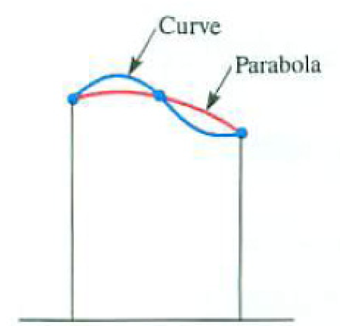
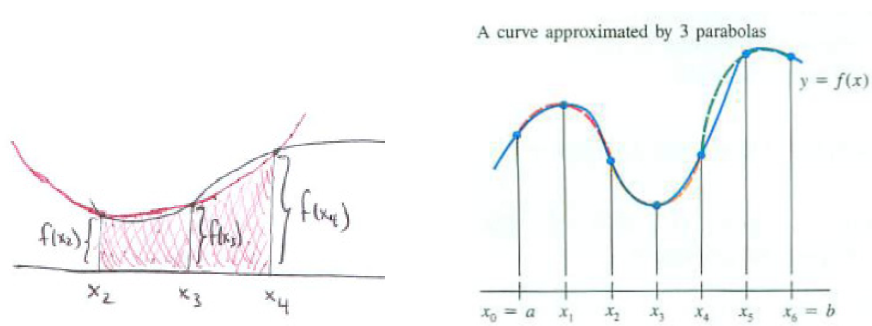


Figure 6.5.8
Curve: $y = f(x)$,
Parabola: $y = Ax^2 + Bx + C$

The computations leading to the formula for the area under the parabola are more involved than those for the area of a trapezoid. (They are outlined in Exercises 28 to 29.) However, the final formula is fairly simple. Let the three points be $(x_1, f(x_1))$, $(x_2, f(x_2))$, $(x_3, f(x_3))$, with $x_1 < x_2 < x_3$, $x_2 - x_1 = h$, and $x_3 - x_2 = h$, as shown in Figure 6.5.9(a). The shaded area under the parabola turns out to be

$$\frac{h}{3}\left(f(x_1) + 4f(x_2) + f(x_3)\right). \tag{6.5.4}$$



(a) (b)

Figure 6.5.9 ARTIST: In (a), x_2, x_3, and x_4 should be labeled as x_1, x_2, and x_3.

To estimate $\int_a^b f(x)\,dx$, we pick an *even* number n and use $n/2$ parabolic arcs, each of width $2h$. As in the trapezoidal method, we start with a partition of $[a, b]$ into n sections of equal width, h: $a = x_0 < x_1 < x_2 < \cdots < x_{n-1} <$

$x_n = b$. Denoting $f(x_i)$ by y_i, form the sum

$$\frac{h}{3}\left((y_0 + 4y_1 + y_2) + (y_2 + 4y_3 + y_4) + \cdots + (y_{n-2} + 4y_{n-1} + y_n)\right).$$

Collecting like terms gives us **Simpson's estimate** for the definite integral $\int_a^b f(x)\,dx$:

$$\boxed{\frac{h}{3}\left(y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \cdots + 2y_{n-2} + 4y_{n-1} + y_n\right) \qquad (6.5.5)}$$

Except for the first and last terms, the coefficients alternate 4, 2, 4, 2, ..., 2, 4. To apply (6.5.5), pick an even number n. Then $h = (b-a)/n$. The estimate uses $n+1$ points, x_0, x_1, ..., x_n, and $n/2$ parabolas. Example 3 illustrates the method, with $n = 4$.

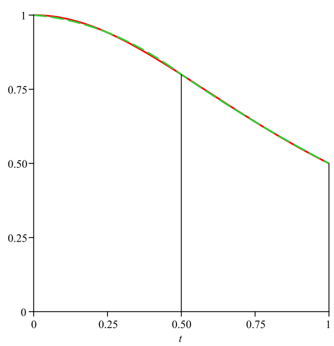


Figure 6.5.10

EXAMPLE 3 Use Simpson's method with $n = 4$ to estimate $\int_0^1 \frac{dx}{1+x^2}$.

SOLUTION In this case, the estimate takes the form

$$\frac{h}{3}\left(y_0 + 4y_1 + 2y_2 + 4y_3 + y_4\right)$$

with $h = (1-0)/4 = 1/4$. There are two parabolas, shown in Figure 6.5.10. Because the parabolas look almost like the curve, we expect Simpson's estimate to be even better than the trapezoidal estimate.

The computations are shown in Table 6.5.2.

x_i	$f(x_i)$	coefficient	summand	decimal form
0	$\frac{1}{1+0^2}$	1	$1 \cdot \frac{1}{1+0}$	1.0000
$\frac{1}{4}$	$\frac{1}{1+\left(\frac{1}{4}\right)^2}$	4	$4 \cdot \frac{1}{1+\frac{1}{16}}$	3.7647
$\frac{2}{4}$	$\frac{1}{1+\left(\frac{2}{4}\right)^2}$	2	$2 \cdot \frac{1}{1+\frac{4}{16}}$	1.6000
$\frac{3}{4}$	$\frac{1}{1+\left(\frac{3}{4}\right)^2}$	4	$4 \cdot \frac{1}{1+\frac{9}{16}}$	2.5600
$\frac{4}{4}$	$\frac{1}{1+\left(\frac{4}{4}\right)^2}$	1	$1 \cdot \frac{1}{1+\frac{16}{16}}$	0.5000

Table 6.5.2

Combining the data in the table with the factor $h/3 = 1/12$ provides the estimate

$$\frac{1}{12}\left(1.0000 + 3.7647 + 1.6000 + 2.5600 + 0.5000\right) = \frac{1}{12}(9.4247) \approx 0.7853.$$

As the decimal form of $\int_0^1 dx/(1+x^2)$ begins 0.78539, this Simpson estimate is accurate to all four decimal places given. ◇

Comparison of the Three Methods

We know the value of $\int_0^1 \frac{dx}{1+x^2}$ is 0.78539816, to eight decimal places. Table 6.5.3 compares the estimates made in the three examples to this value.

Error = |Estimate − Exact|

Method	Estimate	Error
Rectangles	0.845294	0.059896
Trapezoids	0.782794	0.002604
Simpson's (Parabolas)	0.785392	0.000006

Table 6.5.3

Though each method takes about the same amount of work, the table shows that Simpson's method gives the best estimate. The trapezoidal method is next best. The rectangular method has the largest error. These results should not come as a surprise. Parabolas should fit the curve better than chords do, and chords should fit better than horizontal line segments. Note that the trapezoidal and Simpson's methods in Examples 2 and 3 used the same sampling numbers to evaluate the integrand; their only difference is in the "weights" (coefficients) given the outputs of the integrand.

The size of the error is closely connected to the derivatives of the integrand. For a positive number k, let M_k be the largest value of $\left|f^{(k)}(x)\right|$ for x in $[a, b]$. Table 6.5.4 lists the general upper bounds for the error when $\int_a^b f(x)\,dx$ is estimated by sections of length $h = (b-a)/n$. These results are usually developed in a course on numerical analysis. They can also be obtained by a straightforward use of the Growth Theorem of Section 5.4 and the Fundamental Theorem of Calculus. (See Exercises 44 and 45 in this section and Exercise 76 in the Chapter 6 Summary.) They offer a good review of basic ideas.

Recall that $f^{(k)}(x)$ is the k^{th} derivative of f. For instance, $f^{(2)}(x)$ is the second derivative.

Table 6.5.4 expresses the bounds on the size of the error for each method in terms of $h = (b-a)/n$ and n.

Method	Bound on Error in Terms of h	Bound on Error in Terms of n
Rectangles	$M_1(b-a)h$	$M_1(b-a)^2/n$
Trapezoids	$\frac{1}{12}M_2(b-a)h^2$	$\frac{1}{12}M_2(b-a)^3/n^2$
Simpson's (Parabolas)	$\frac{1}{180}M_4(b-a)h^4$	$\frac{1}{180}M_4(b-a)^5/n^4$

Table 6.5.4

The coefficients in the error bounds tell us a great deal. For instance, if $M_4 = 0$, then there is no error in Simpson's method. That is, if $f^{(4)}(x) = 0$ for all x in $[a, b]$, then Simpson's method produces an exact answer. For in this case the error is $M_4(b-a)h^4/180 = 0$. As a consequence, for polynomials of at most degree 3, Simpson's approximation is exact. (See Exercise 79 in Section 6.5.)

We know that the trapezoidal method is exact for polynomials of degree at most one, in other words, for functions whose second derivative is zero. That suggests that the error in this method is controlled by the size of the second derivative; Table 6.5.4 shows that it is.

The power of h that appears in the error bound is even more important. For instance, if you reduce the width h by a factor of 10 (using 10 times as many sections) you expect the error of the rectangular method to shrink by a factor of 10, the error in the trapezoidal method to shrink by a factor of $10^2 = 100$, and the error in Simpson's method by a factor of $10^4 = 10,000$. These observations are recorded in Table 6.5.5.

Method	Reduction Factor of h	Expected Reduction Factor of Error
Rectangles	10	10
Trapezoids	10	100
Simpson's (Parabolas)	10	10,000

Table 6.5.5

Because the error in the rectangular method approaches 0 so slowly as $h \to 0$, we will not refer to it further.

Technology and Definite Integrals

The trapezoidal method and Simpson's method are just two examples of what is called **numerical integration**. Such techniques are studied in detail in courses on numerical analysis. While the Fundamental Theorem of Calculus is useful for evaluating definite integrals, it applies only when an antiderivative is readily available. Numerical integration is an important tool in estimating definite integrals, particularly when the FTC cannot be applied. Numerical integration can always be used to find out something about the value of a definite integral.

The design of an efficient and accurate general-purpose numerical integration algorithm is harder than it might seem. Effective algorithms typically divide the interval into unequal-length sections. The sections can be longer where the function is tame, that is, almost constant. Shorter sections are used where the function is wild, that is, changes very rapidly. Large, even unbounded, intervals can lead to another set of difficulties. Some examples of challenging definite integrals include:

$$\int_0^2 \sqrt{x(4-x)}\, dx \qquad \int_{-1}^{1} \frac{dx}{x^2+10^{-10}} \qquad \int_0^{600\pi} \frac{(\sin(x))^2}{\sqrt{x}+\sqrt{x+\pi}}\, dx$$

The HP-34C was, in 1980, the first handheld calculator to perform numerical integration. Now this is a common feature on most scientific calculators. The algorithms used vary greatly, and the details are often corporate secrets. The techniques are similar to those presented in this section and in Exercise 40.

Reference: Handheld Calculator Evaluates Integrals, William Kahan, Hewlett-Packard Journal, vol. 31, no. 8, Aug. 1980, pp. 23–32, `http://www.cs.berkeley.edu/~wkahan/Math128/INTGTkey.pdf`.

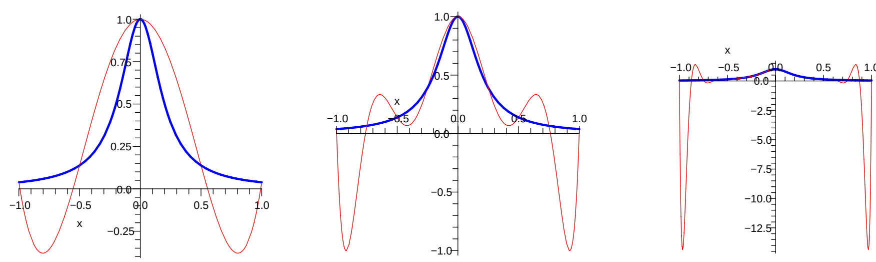
Summary

Three techniques for estimating definite integral are suggested by the areas of rectangles, the areas of trapezoids, and the areas under parabolas. We observed that the error in each method is influenced by a derivative of the integrand and the distance, $h = (b-a)/n$, between the numbers at which we evaluate the integrand. The main difference between the methods is the coefficients used to weight the function values $y_i = f(x_i)$. In the left-hand rectangular estimate the coefficients are 1, 1, 1, ..., 1, 0 (because $y_n = f(b)$ is not used). In the right-hand rectangular estimate the coefficients are 0, 1, 1, ..., 1. In the trapezoidal estimate, they are 1, 2, 2, ..., 2, 1 and in Simpson's estimate they are 1, 4, 2, 4, 2, ..., 2, 4, 1. A course in numerical analysis presents several other ways to estimate a definite integral.

Carle Runge, 1856–1927, http://en.wikipedia.org/wiki/Carle_David_Tolm%C3%A9_Runge

Higher-Order Interpolation Methods and Runge's Counterexample
In the trapezoidal method you pass a line through two points to approximate the curve. That uses a first-degree polynomial, $Ax + B$. In Simpson's method you pass a parabola through three points, using a second-degree polynomial, $Ax^2 + Bx + C$. You would expect that as you pass higher-degree polynomials through more points on the curve you would get even better approximations. This is not always the case.

For the function $f(x) = 1/(1 + 25x^2)$, defined on $[-1, 1]$, known as Runge's Counterexample, the higher-degree polynomials passing through equally-spaced points do not resemble the function. Figure 6.5.11 shows the **interpolating polynomials** of degree 4 (a), 8 (b), and 16 (c). Notice how the approximations improve away from the endpoints and exhibit increasingly large oscillations near the endpoints. These oscillations result in poor estimates of $\int_{-1}^{1} \frac{dx}{1+25x^2}$. A Google search for "Runge's Counterexample" yields more information on this function.



(a) (b) (c)

Figure 6.5.11 In each figure the thick curve is the graph of Runge's Counterexample and the thin curve is the graph of the interpolating polynomials of degree 4 (a), 8 (b), and 12 (c). Notice the very different vertical scales in these three graphs. EDITOR: Please move these figures inside the box.

EXERCISES for Section 6.5

In the Exercises, T_n refers to the trapezoidal estimate with n trapezoids (partition with n sections and $n+1$ points), and S_n refers to Simpson's estimate with $n/2$ parabolas (partition with n sections and $n+1$ points)
In Exercises 1 to 8 approximate the given definite integrals by the trapezoidal estimate with the indicated T_n.

1. $\int_0^2 \frac{dx}{1+x^2}$, T_2

2. $\int_0^2 \frac{dx}{1+x^2}$, T_4

3. $\int_0^2 \sin(\sqrt{x})\ dx$, T_2

4. $\int_0^2 \sin(\sqrt{x})\ dx$, T_3

5. $\int_1^3 \frac{2^x}{x}\ dx$, T_3

6. $\int_1^3 \frac{2^x}{x}\ dx$, T_6

7. $\int_1^3 \cos(x^2)\ dx$, T_2

8. $\int_1^3 \cos(x^2)\ dx$, T_4

In Exercises 9 to 12 use Simpson's estimate to approximate each definite integral with the given S_n.

9. $\int_0^1 \frac{dx}{1+x^3}$, S_2

10. $\int_0^1 \frac{dx}{1+x^3}$, S_4

11. $\int_0^1 \frac{dx}{1+x^4}$, S_2

12. $\displaystyle\int_0^1 \frac{dx}{1+x^4}$, S_4

13. Write out T_6 for $\int_1^4 5^x\, dx$ but do not carry out any of the calculations.

14. Write out S_{10} for $\int_0^1 e^{x^2}\, dx$ but do not carry out any of the calculations.

15. By a direct computation, show that the trapezoidal estimate is not exact for second-order polynomials. (Take the simplest case, $\int_0^1 x^2\, dx$.)

16. By a direct computation, show that the Simpson's estimate is not exact for fourth-order polynomials. (Take the simplest case, $\int_0^1 x^4\, dx$.)

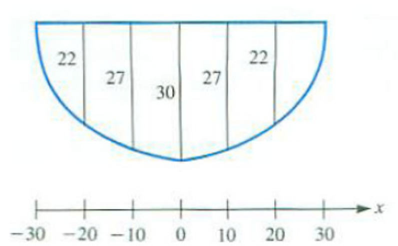
17. In an interval $[a, b]$ in which $f''(x)$ is positive, do trapezoidal estimates of $\int_a^b f(x)\, dx$ underestimate or overestimate the definite integral? Explain.

18. The cross section of a ship's hull is shown in Figure 6.5.12(a). Estimate the area of this cross section by

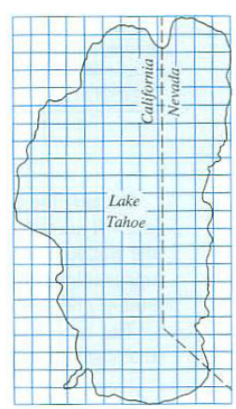
(a) T_6

(b) S_6

Dimensions are in feet. Give your answer to four decimal places.



(a)



(b)

Figure 6.5.12

19. A ship is 120 feet long. The area of the cross section of its hull is given at intervals in the table below:

x	0	20	40	60	80	100	120	feet
area	0	200	400	450	420	300	150	square feet

Estimate the volume of the hull in cubic feet by

(a) the trapezoidal estimate and

(b) Simpson's estimate.

Give your answer to four decimal places. (What is largest n you can use in this problem?)

20. A map of Lake Tahoe is shown in Figure 6.5.12(b). Use Simpson's method and data from the map to estimate the surface area of the lake. Use cross sections parallel to the side of the page. (Each little square represents a square mile.)

Exercises 21 and 22 present cases in which the maximum bound on the error is assumed.

21. Show that the error for the trapezoidal estimate of $\int_0^1 x^2dx$ is exactly $(b-a)M_2h^2/12$ where $a=0$, $b=1$, $h=1$, and M_2 is the maximum value of $\left|D^2(x^2)\right|$ for x in $[0,1]$.

22. Show that the error for the Simpson estimate of $\int_0^1 x^4dx$ is exactly $(b-a)M_4h^4/180$ where $a=0$, $b=1$, $h=1/2$, and M_4 is the maximum value of $\left|D^4(x^4)\right|$ for x in $[0,1]$.

23. Figure 6.5.13(b) shows cross sections of a pond in two directions. Use Simpson's method to estimate the area of the pond using

(a) vertical cross sections, three parabolas and

(b) horizontal cross sections, two parabolas.

24. In the case of trapezoidal estimates, if you double the length of the interval $[a,b]$ and also the number of trapezoids, would you expect the error in the estimates to increase, decrease, or stay about the same? Explain.

25. In the case of Simpson estimates, if you double the length of the interval $[a,b]$ and also the number of parabolas, would you expect the error in the estimates to increase, decrease, or stay about the same? Explain.

26.

(a) Fill in this table concerning $\int_0^6 x^2\ dx$ and its trapezoidal estimates.

	$\int_0^6 x^2\ dx$	T_1	T_2	T_3
Value				
Error		—		

(b) Do the errors in (a) seem to be proportional to h^c for some constant c? (Recall that h is the width of the trapezoids.)

27.

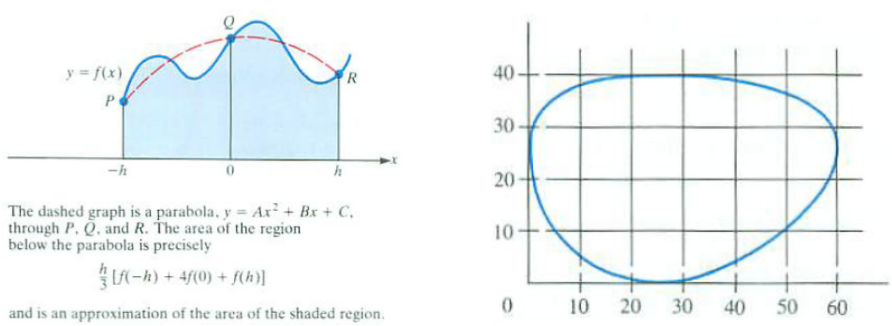
(a) Fill in this table concerning $\int_1^7 dx/(1+x)^2$ and its Simpson estimates.

	$\int_1^7 dx/(1+x)^2$	S_2	S_4	S_6
Value				
Error		—		

(b) Do the errors in (a) seem to be proportional to h^c for some constant c? (Recall that h is the width of the sections.)

Exercises 28 to 30 provide the basis of Simpson estimates. For convenience we place the origin of the x-axis at the midpoint of the interval for which a single parabola will approximate the function. Because the interval has length $2h$, its ends are $-h$ and h.

28. Let $f(x)$ be a function defined on at least $[-h,h]$, with $f(-h)=y_1$, $f(0)=y_2$, and $f(h)=y_3$. Show that there is exactly one parabola $P(x)=Ax^2+Bx+C$ that passes through the three points $(-h,y_1)$, $(0,y_2)$, and (h,y_3). (See Figure 6.5.13(a).)



(a) (b)

Figure 6.5.13

29. Let $p(x)=Ax^2+Bx+C$. Show, by computing both sides of the equation, that

$$\int_{-h}^{h} p(x)\ dx = \frac{h}{3}\left(p(-h)+4p(0)+p(h)\right).$$

This equation, expressed geometrically, was known to the ancient Greeks.

30. Let $f(x) = x^3$. Show that

$$\int_{-h}^{h} f(x)\, dx = \frac{h}{3}\left(f(-h) + 4f(0) + f(h)\right).$$

This information, combined with Exercise 29, implies that Simpson's method is exact for polynomials of degree at most 3.

31. The table lists the values of a function f at the given points.

x	1	2	3	4	5	6	7
$f(x)$	1	2	1.5	1	1.5	3	3

(a) Plot the corresponding seven points on the graph of f.

(b) Sketch six trapezoids that can be used to estimate $\int_1^7 f(x)\, dx$.

(c) Find the trapezoidal estimate of $\int_1^7 f(x)\, dx$.

(d) Sketch, by eye, the three parabolas used in Simpson's method to estimate $\int_1^7 f(x)\, dx$.

(e) Find Simpson's estimate of $\int_1^7 f(x)\, dx$.

32. A function f is defined on $[a, b]$ and $f(x)$, $f'(x)$, and $f''(x)$ are all positive for x in that interval. Arrange the following quantities in order of size, from smallest to largest. (Some may be equal.) Sketches may help.

(a) the area of the trapezoid with base $[a, b]$ and parallel sides of lengths $f(a)$ and $f(b)$

(b) the area of the "midpoint" rectangle with base $[a, b]$ and height $f((a + b)/2)$

(c) the area of the "right-endpoint" rectangle with base $[a, b]$ and height $f(b)$

(d) the area of the "left-endpoint" rectangle with base $[a, b]$ and height $f(a)$

(e) the average of (c) and (d)

(f) the trapezoid whose base is $[a, b]$ and whose top edge lies on the tangent line at $((a + b)/2, f((a + b)/2))$

(g) $\int_a^b f(x)\, dx$.

Exercises 33 to 35 describe the **midpoint estimate**, yet another way to estimate a definite integral.

33. Another way to estimate a definite integral is by a Riemann sum $\sum_{i=1}^{n} f(c_i)h$, where the c_i are the midpoints of the intervals. Call such an estimate with n sections, M_n. Find M_4 for $\int_0^1 dx/(1+x^2)$.

34. With the aid of a diagram, show that the midpoint estimate is exact for functions of the form $f(x) = Ax + B$.

35. Assume that $f''(x)$ is negative for x in $[a, b]$. With the aid of a diagram, show that the midpoint method overestimates $\int_a^b f(x)\ dx$. (Draw a tangent at the point $((a+b)/2, f((a+b)/2))$.)

36. If the Simpson estimate with 4 parabolas estimate a certain definite integral with an error of 0.35, what error would you expect with (a) 8 parabolas? (b) 5 parabolas?

37. The equation in Exercise 28 is called the **prismoidal formula**. Use it to compute the volume of

(a) a sphere of radius a and

(b) a right circular cone of radius a and height h.

The prisomoidal formula was known to the Greeks. Reference: `http://www.mathpages.com/home/kmath189/kmath189.htm`

Exercise 38 provides a review of several basic ideas as it involves the Fundamental Theorem of Calculus (FTC I), the chain rule, l'Hôpital's rule, and the intermediate-value theorem. The midpoint estimate is defined in Exercise 33.

38. Assume that $f''(x)$ is continuous and negative for x in $[0, 2h]$. Then the midpoint estimate, M, for $\int_{-h}^{h} f(x)\ dx$ is too large and the trapezoidal estimate, T, is too small. The error of the first is $M - \int_{-h}^{h} f(x)\ dx$ and of the second is $\int_{-h}^{h} f(x)\ dx - T$. Show that

$$\lim_{h \to 0} \frac{M - \int_{-h}^{h} f(x)\ dx}{\int_{-h}^{h} f(x)\ dx - T} = \frac{1}{2}.$$

This suggests that the error in the midpoint estimate when h is small is about half the error of the trapezoidal estimate. However, the midpoint estimate is seldom used because data at midpoints are usually not available (and because the Simpson estimate provides an even more accurate estimate using the same data as the trapezoidal estimate).

39. Simpson's estimate is not exact for fourth-degree polynomials.

(a) Estimate $\int_0^h x^4\,dx$ by S_2.

(b) What is the ratio between that estimate and $\int_0^h x^4\,dx$?

(c) What does (b) imply about the ratio between Simpson's estimate and $\int_0^h P(x)\,dx$ for any polynomial of degree at most 4?

40. There are many other methods for estimating definite integrals. Some old methods, which had been of only theoretical interest because of their messy arithmetic, have, with the advent of computers, assumed practical importance. This exercise illustrates the simplest of the so-called **Gaussian quadrature** formulas. For convenience, we consider only integrals over $[-1, 1]$.

(a) Show that

$$\int_{-1}^{1} f(x)\,dx = f\left(\frac{-1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right)$$

for $f(x) = 1$, x, x^2, and x^3.

(b) Let a and b be two numbers, $-1 \le a < b \le 1$, such that

$$\int_{-1}^{1} f(x)\,dx = f(a) + f(b)$$

for $f(x) = 1$, x, x^2, and x^3. Show that only $a = \frac{-1}{\sqrt{3}}$ and $b = \frac{1}{\sqrt{3}}$ (or $a = \frac{1}{\sqrt{3}}$ and $b = \frac{-1}{\sqrt{3}}$) satisfy this equation.

(c) Show that the Gaussian approximation

$$\int_{-1}^{1} f(x)\,dx \approx f\left(\frac{-1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right)$$

has no error when f is a polynomial of degree at most 3.

(d) Use the formula in (a) to estimate $\displaystyle\int_{-1}^{1} \frac{dx}{1+x^2}$.

(e) Compare the answer in (d) to the exact value of $\displaystyle\int_{-1}^{1} \frac{dx}{1+x^2}$. How large is the error?

41. Let f be a function such that $\left|f^{(2)}(x)\right| \leq 10$ and $\left|f^{(4)}(x)\right| \leq 50$ for all x in $[1,5]$. If $\int_1^5 f(x)\, dx$ is to be estimated with an error of at most 0.01, how small must h be in

(a) the trapezoidal approximation?

(b) Simpson's approximation?

42.

Sam: I bet I can find a better way than Simpson's estimate to approximate $\int_{-h}^{h} f(x)\, dx$ using the same three arguments ($-h$, 0, and h).

Jane: How so?

Sam: Look at his formula $\frac{h}{3}(f(-h)+4f(0)+f(h))$, which equals $2h\left(\frac{1}{6}f(-h) + \frac{4}{6}f(0) + \frac{1}{6}f(h\right.$ The $2h$ is the width of the interval. I can't change that.

Jane: What would you change?

Sam: The weights $\frac{1}{6}$, $\frac{4}{6}$, and $\frac{1}{6}$. I'll use weights w_1, w_2, and w_3 and demand that the estimates I get be exact when the function $f(x)$ is either constant, x, or x^2.

Jane: Go ahead.

Sam: If $f(x) = c$, a constant, then, because $\int_{-h}^{h} c\, dx = 2hc$, I must have $2hc = 2h(w_1c + w_2c + w_3c)$. That tells me that $w_1 + w_2 + w_3$ must be 1.

Jane: But you need three equations for three unknowns.

Sam: When $f(x) = x$, I get $\int_{-h}^{h} f(x)\, dx = 0$, so $0 = 2h(-w_1h + w_2 0 + w_3h)$. Now I know that w_1 equals w_3.

Jane: And the third equation?

Sam: With $f(x) = x^2$, I find that $\frac{2}{3}h^3 = 2h^3(w_1 + w_3)$.

Jane: So what are your three w's?

Sam: A little high school algebra shows they are $\frac{1}{6}$, $\frac{4}{6}$, and $\frac{1}{6}$. What a disappointment. But at least I avoided all the geometry of parabolas. It's really all about assigning proper weights.

Check the missing details and show that Sam is right.

43. Another way to estimate a definite integral is to use Taylor polynomials (discussed in Section 5.5). If the Maclaurin polynomial $P_2(x)$ for $f(x)$ of degree 2 is used to approximate $f(x)$ for x in $[0,h]$, express the possible error in using $\int_0^h P_2(x)\ dx$ to estimate $\int_0^h f(x)\ dx$.

In Section 5.5 we showed why a higher derivative controls the error in using a Taylor polynomial to approximate a function value. Exercises 44 and 45 show why a higher derivative controls the error in using the trapezoidal or Simpson estimate of a definite integral $\int_a^b f(x)dx$. (See Exercise 76 in the Chapter 6 Summary for the derivation of the corresponding error estimate for the midpoint estimate.) In each case $h = (b-a)/n$ and a function $E(t)$, $0 \le t \le h$, is introduced. The "local error" is $E(h)$, that is, the error in using one trapezoid of width h or one parabola of width $2h$. Once $E(h)$ is controlled by a higher derivative, we multiply by n, where $nh = b-a$, to obtain a measure of the total error in estimating $\int_a^b f(x)\ dx$. The argument involves both FTC I and FTC II and provides a review of basic concepts.

44. *(The error in the trapezoid estimate.)* As usual, let $h = (b-a)/n$. We will estimate the error for a single section of width h and then multiply by n to find the error in estimating $\int_a^b f(x)\ dx$. For convenience, we move the graph so the interval (of length h) is $[0,h]$.

(a) Show that the error when using T_1 is $E(h) = \int_0^h f(x)\ dx - \frac{h}{2}(f(0)+f(h))$.

(b) For t in $[0,h]$ let $E(t) = \int_0^t f(x)\ dx - \frac{t}{2}(f(0)+f(t))$. Show that $E(0)=0$, $E'(0)=0$, and $E''(t) = -\frac{t}{2}f''(t)$.

(c) Let M be the maximum of $f''(x)$ on $[a,b]$ and m be the minimum. Show that $\frac{-mt}{2} \ge E''(t) \ge \frac{-Mt}{2}$.

(d) Using (b) and (c), show that $\frac{-mt^2}{4} \ge E'(t) \ge \frac{-Mt^2}{4}$.

(e) Show that $\frac{-mt^3}{12} \ge E(t) \ge \frac{-Mt^3}{12}$.

(f) Show that $\frac{-mh^3}{12} \ge E(h) \ge \frac{-Mh^3}{12}$.

(g) Show that $\frac{-m(b-a)h^2}{12} \ge \int_a^b f(x)\ dx - T_n \ge \frac{-M(b-a)h^2}{12}$.

(h) Show that $\int_a^b f(x)\ dx - T_n = \frac{-f''(c)(b-a)h^2}{12}$ for some number c in $[a,b]$.

(i) Deduce that $\left|\int_a^b f(x)\ dx - T_n\right| \le \frac{M_2(b-a)h^2}{12}$, where M_2 is the maximum of $|f''(x)|$ for x in $[a,b]$.

45. *(The error in the Simpson estimate.)* Now n is even and $[a, b]$ is divided into n sections of width $h = (b - a)/n$. The Simpson estimate is based on $n/2$ intervals of length $2h$. We will place the origin at the midpoint of an interval, so that its ends are $-h$ and h. In this case we wish to control the size of $E(h) = \int_{-h}^{h} f(x)\ dx - \frac{h}{3}(f(-h)+4f(0)+f(h))$. Introduce the function $E(t)$, for $-h \leq t \leq h$, defined by $E(t) = \int_{-t}^{t} f(x)\ dx - \frac{t}{3}(f(-t) + 4f(0) + f(t))$.

(a) Show that

$$E'(t) = \frac{2}{3}(f(t) + f(-t)) - \frac{4}{3}f(0) - \frac{t}{3}(f'(t) - f'(-t)).$$

(b) Show that $E''(t) = \frac{1}{3}(f'(t) - f'(-t)) - \frac{t}{3}(f''(t) + f''(-t))$.

(c) Show that $E'''(t) = -\frac{t}{3}(f'''(t) - f'''(-t))$.

(d) Show that $E'''(t) = \frac{-2t^2}{3} f^{(4)}(c)$ for some c in $[-h, h]$.

(e) Show that $E(0) = E'(0) = E''(0) = 0$.

(f) Let M_4 be the maximum of $|f^{(4)}(t)$ on $[a, b]$. Show that $|E(t)| \leq \frac{2t^5}{180} M_4$.

(g) Deduce that $\left| \int_a^b f(x)\ dx - S_n \right| \leq \frac{M_4(b-a)h^4}{180}$.

6.S Chapter Summary

Chapter 6 introduced the second major concept in calculus, the definite integral, defined as a limit:

$$\int_a^b f(x)\ dx = \lim_{\max \Delta x_i \to 0} \sum_{i=1}^{n} f(c_i)\Delta x_i$$

For a continuous function this limit always exists and $\int_a^b f(x)\ dx$ can be viewed as the (net) area under the graph of $y = f(x)$ on the interval $[a, b]$. Both the definite integral and an antiderivative of a function f are called "integrals." Context tells which is meant. An antiderivative is also called an "indefinite integral."

The definite integral, in contrast to the derivative, gives global information.

Integrand: $f(x)$	Integral: $\int_a^b f(x)\ dx$
velocity	change in position
speed (|velocity|)	distance traveled
cross-sectional length of plane region	area of a plane region
cross-sectional area of solid	volume of solid
rate bacterial colony grows	total growth

As the first and last of these applications show, if you compute the definite integral of the rate at which some quantity is changing, you get the total change. To put this in mathematical symbols, let $F(x)$ be the quantity present at time x. Then $F'(x)$ is the rate at which the quantity changes. Thus $\int_a^b F'(x)\ dx$ equals the change in $F(x)$ as x goes from a to b, which is $F(b) - F(a)$. In short, $\int_a^b F'(x)\ dx = F(b) - F(a)$. This is another way of stating the Fundamental Theorem of Calculus, because F is an antiderivative of F'.

That equation gives a shortcut for evaluating many common definite integrals. However, finding an antiderivative can be tedious or impossible. For instance, $\exp(x^2)$ does not have an elementary antiderivative. However, continuous functions do have antiderivatives, as slope fields suggest. Indeed $G(x) = \int_a^x f(t)\ dt$ is an antiderivative of the integrand.

One way to estimate a definite integral is to employ one of the sums $\sum_{i=1}^{n} f(c_i)\ \Delta x_i$ that appear in its definition.

A more accurate method, which involves the same amount of arithmetic, uses trapezoids. Then the estimate takes the form

$$\int_a^b f(x)\ dx \approx \frac{h}{2}\left(f(x_0) + 2f(x_1) + 2f(x_2) + \ldots + 2f(x_{n-1}) + f(x_n)\right),$$

where consecutive x_i's are a fixed distance $h = (b-a)/n$ apart. In Simpson's method the graph is approximated by parts of parabolas, n is even, and the estimate is

$$\frac{h}{3}\left(f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \ldots + 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)\right).$$

The remaining chapters are simply elaborations of the derivative and the definite integral or further applications of them. For instance, instead of integrals over intervals, Chapter 17 deals with integrals over sets in the plane or in space. Chapter 15 treats derivatives of functions of several variables. In both cases the definitions involve limits similar to those that appear in the definitions of the derivative and the definite integral. That is one reason not to lose sight of those two definitions in their many applications.

EXERCISES for 6.S

1. State FTC I in words, using no mathematical symbols. (It refers to $F(b)-F(a)$.)

2. State FTC II in words, using no mathematical symbols. (It refers to the derivative of $\int_a^x f(t)\ dt$.)

Evaluate the definite integrals in Exercises 3 to 16.

3. $\int_1^2 (2x^3 + 3x - 5)\ dx$

4. $\int_5^7 \frac{3}{x}\ dx$

5. $\int_1^4 \frac{dx}{\sqrt{x}}$

6. $\int_1^4 \frac{x+2x^3}{\sqrt{x}}\ dx$

7. $\int_0^1 x(3+x)\ dx$

8. $\int_0^2 (2+3x)^2\ dx$

9. $\int_1^2 \frac{(2+3x)^2}{x^2}\ dx$

10. $\int_1^2 e^{2x}\ dx$

11. $\int_0^\pi \sin(3x)\ dx$

12. $\int_0^{\pi/4} \sec^2(x)\ dx$

13. $\int_0^{\sqrt{2}/2} \frac{3\ dx}{\sqrt{1-x^2}}\ dx$

14. $\int_0^{\pi/4} \cos(x)\ dx$

15. $\int_0^{\pi/4} \sec(x)\tan(x)\ dx$

16. $\int_{1/2}^{\sqrt{2}/2} \frac{dx}{x\sqrt{x^2-1}}$

In Exercises 17 to 24 find an antiderivative of the given function by guess and experiment. Check your answer by differentiating it.

17. $(2x+1)^5$

18. $\frac{1}{(2x+1)^5}$

19. $\frac{1}{x+1}$

20. $\frac{1}{2x+1}$

21. $\ln(x)$

22. $x\sin(x)$

23. $\sin(2x)$

24. xe^{x^2}

Use Simpson's estimate with three parabolas ($n = 6$) to approximate the definite integrals in Exercises 25 and 26.

25. $\int_0^{\pi/2} \sin(x^2)\ dx$

26. $\int_1 2\sqrt{1+x^2}\ dx$

27. Use the trapezoidal estimate with $n = 6$ to estimate the integral in Exercise 25.

28. Use the trapezoidal estimate with $n = 6$ to estimate the integral in Exercise 26.

Exercises 29 and 30 provide additional detail for the historical discussion (see page 1082) about Newton's calculation of the area under a hyperbola to more than 50 decimal places. (See also Exercise 29 in Section 6.5.)

29. Let c be a positive constant.

(a) Show that the area under the curve $y = 1/(1+x)$ above the interval $[0, c]$ is $\ln(1+c)$.

(b) Show that the area under the curve $y = 1/(1+x)$ above the interval $[-c, 0]$ is $-\ln(1-c)$.

30.

(a) In his approximation of $\ln(1.1)$ to 53 decimal places Newton used, in effect, $P_{53}(0.1; 0)$ for $f(x) = \ln(1+x)$. What is the bound on the error for this approximation?

(b) Could Newton have used fewer terms to obtain an equally accurate answer? Explain your answer.

31.

(a) What is the area under $y = 1/x$ and above $[1, b]$, $b > 1$?

(b) Is the area under $y = 1/x$ and above $[1, \infty)$ finite or infinite?

(c) The region under $y = 1/x$ and above $[1, b]$ is rotated around the x-axis. What is the volume of the solid produced?

32. The basis for this chapter is that if f is continuous and $x > a$, then $\frac{d}{dx} \int_a^x f(t)\, dt = f(x)$.

(a) Review how this equation was obtained.

(b) Use a similar method to show that, if $x < b$, then $\frac{d}{dx} \int_x^b f(t)\, dt = -f(x)$.

33. Let $f(x)$ and $g(x)$ be differentiable functions with $f(x) \geq g(x)$ for all x in $[a, b]$, $a < b$.

(a) Is $f'(x) \geq g'(x)$ for all x in $[a, b]$? Explain.

(b) Is $\int_a^b f(x)dx \geq \int_a^b g(x)dx$? Explain.

34. Find $D\left(\int_{x^2}^{x^3} e^{-t^2} dt\right)$.

35.

Jane: I'm not happy. The text says that a definite integral measures area. But they never defined "area under a curve." I know what the area of a rectangle is: width times length. But what is meant by the area under a curve? If they say, "Well, its the definite integral of the cross sections," that won't do. What if I integrate cross sections that are parallel to the x-axis instead of the y-axis? How do I know I'll get the same answer? Once again, the authors are hoping no one will notice a big gap in their logic.

Is Jane right? Have the authors tried to slip something past the reader?

36. Let T_n be the trapezoidal estimate of $\int_a^b f(x)\,dx$ with n trapezoids and M_n be the midpoint estimate with n sections. Show that $\frac{1}{3}T_n + \frac{2}{3}M_n$ equals the Simpson estimate S_{2n} with n parabolas. (Consider a typical interval of length h.)

37. A river flows at the (varying) rate of $r(t)$ cubic feet per second.

(a) Approximate how many cubic feet passes during the short time interval from time t to time $t + \Delta t$ seconds.

(b) How much passes from time t_1 to time t_2 seconds?

38. Let $f(x) = xe^{-x}$ for $x \geq 0$. For which interval of length 1 is the area below the graph of f and above that interval a maximum?

39. Let $f(x) = x/(x+1)^2$ for $x \geq 0$.

(a) Graph f, showing any extrema.

(b) Looking at your graph, estimate for which interval of length one, the area below the graph of f and above the interval is a maximum.

(c) Using calculus, find the interval in (b) that yields the maximum area.

40.

(a) Estimate $\int_0^1 \frac{\sin(x)}{x} dx$ by approximating $\sin(x)$ by the Taylor polynomial $P_6(x;0)$.

(b) Use the Lagrange bound on the error to bound the error in (a).

41.

(a) Estimate $\int_1^3 \frac{e^x}{x} dx$ by using the Taylor polynomial $P_3(x;2)$ to approximate e^x. (To avoid computing e^2, approximate e by 2.71828.)

(b) Use the Lagrange bound on the error to bound the error in (a).

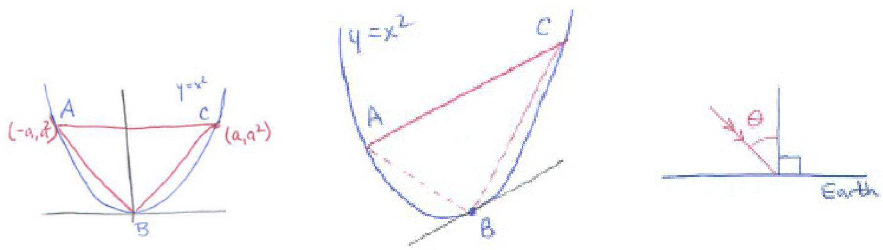
42. Assume $f(2) = 0$ and $f'(2) = 0$ and $f''(x) \leq 5$ for all x in $[0,7]$. Show that $\int_2^3 f(x)dx \leq 5/6$.

43. Find $\lim_{t\to 0} \dfrac{\int_0^t \left(e^{x^2}-1\right)\,dx}{\int_0^t \sin(2x^2)\,dx}$.

44. Let $G(t) = \int_0^t \cos^5(\theta)\,d\theta$ for t in $[0, 2\pi]$.

(a) Sketch a rough graph of $y = G'(t)$.

(b) Sketch a rough graph of $y = G(t)$.



(a) (b) (c)

Figure 6.S.1

45. Figure 6.S.1(a) shows a triangle ABC inscribed in the parabola $y = x^2$ $A = (-a, a^2)$, $B = (0, 0)$, and $C = (a, a^2)$. Let $T(a)$ be its area and $P(a)$ the area bounded by AC and the parabola above the interval $[-a, a]$. Find $\lim_{a\to 0} \frac{T(a)}{P(a)}$. Archimedes established a much more general result. In Figure 6.S.1(b) the tangent line at B is parallel to AC. He determined for any chord AC the ratio between the area of triangle ABC and the area of the parabolic section .

Usually we use a sum to estimate a definite integral. We can also use a definite integral to estimate a sum. In Exercises 46 and 47, rewrite each sum so that it becomes the sum estimating a definite integral. Then use the definite integral to estimate the sum.

46. $\frac{1}{100}\sum_{i=1}^{100} \frac{1}{i^2}$

47. $\sum_{n=51}^{100} \frac{1}{n}$

48.

(a) Show that the average value of $\cos(\theta)$ for θ in $[0, \pi/2]$ is about 0.637.

(b) The average in (a) is fairly large, being much more than half of the maximum value of $\cos(\theta)$. Why is that good news for a farmer or solar engineer on Earth who depends on heat from the sun? (See Figure 6.S.1(c).)

49. Assume f' is continuous on $[0, t]$.

(a) Find the derivative of $F(t) = 2\int_0^t f(x)f'(x)\ dx - f(t)^2$.

(b) Give a shorter formula for $F(t)$.

50. Find a simple expression for the function $F(t) = \int_1^t \cos(x^2)\ dx - \int_1^{t^2} \frac{\cos(u)}{2\sqrt{u}}\ du$.

51. A tent has a square base of side b and a pole of length $b/2$ above the center of the base.

(a) Set up a definite integral for the volume of the tent.

(b) Evaluate the integral in (a) by the Fundamental Theorem of Calculus.

(c) Find the volume of the tent by showing that six copies of it fill up a cube of side b.

52.

Sam: I can get the first FTC, the one about $F(b) - F(a)$, without all that stuff in the second FTC.

Jane: That would be nice.

Sam: As usual, I assume F' is continuous and $\int_a^b F'(x)dx$ exists. Now, $F(b) - F(a)$ is the total change in F. Well, bust up $[a, b]$ by t_0, t_1, ..., t_n in the usual way. Then the total change is just the sum of the changes in F over each of the n intervals, $[t_{i-1}, t_i]$, $i = 1, 2, \ldots, n$.

Jane: That's a no-brainer, but then what?

Sam: The change in F over the typical interval is $F(t_i) - F(t_{i-1})$. By the Mean Value Theorem for F, that equals $F'(t_i^*)(t_i - t_{i-1})$ for some t_i^* in the i^{th} interval. The rest is automatic.

Jane: I see. You let all the intervals get shorter and shorter and the sums of the $F'(t_i^*)(t_i - t_{i-1})$ approach $\int_a^b F'(x)\ dx$. But they are all already equal to $F(b) - F(a)$.

Sam: Pretty neat, yes?

Jane: Something must be wrong.

Is anything wrong?

53.

Sam: There are two authors and they are both wrong.

Jane: How so?

Sam: Light can be both a wave and a particle, right?

Jane: Yes.

Sam: Well the definite integral is both a number and a function.

Jane: Did you get enough sleep?

Sam: This is serious. Take $\int_0^b x^2\ dx$. That equals $b^3/3$. Right?

Jane: So far, right.

Sam: Well, as b varies, so does $b^3/3$. So it's a function.

Jane: ...

What is Jane's reply?

54.

(a) Graph $y = e^x$ for x in $[0, 1]$.

(b) Let c be the number such that the area under the graph of $y = e^x$ above $[0, c]$ equals the area under the graph above $[c, 1]$. Looking at the graph in (a), decide whether c is bigger or smaller than $1/2$.

(c) Find c.

55. Find $\lim_{\Delta x \to 0} \left(\frac{1}{\Delta x} \int_5^{7+\Delta x} e^{x^3}\ dx - \frac{1}{\Delta x} \int_5^7 e^{x^3}\ dx \right)$.

56. Find $\lim_{\Delta x \to 0} \left(\frac{1}{\Delta x} \int_{5+\Delta x}^{7} e^{x^3}\ dx - \frac{1}{\Delta x} \int_5^7 e^{x^3}\ dx \right)$.

57. A company is founded with capital investment A. It plans to have its rate of investment proportional to its total investment at any time. Let $f(t)$ denote the rate of investment at time t.

(a) Show that there is a constant k such that $f(t) = k(A + \int_0^t f(x)dx)$ for any $t \geq 0$.

(b) Find a formula for f.

There are two definite integrals in each of Exercises 58 to 61. One can be evaluated by the FTC, the other not. Evaluate the one that can be evaluated by the FTC and approximate the other by Simpson's estimate with $n = 4$ (2 parabolas).

58. $\int_0^1 (e^x)^2 \ dx$; $\int_0^1 e^{x^2} \ dx$.

59. $\int_0^{\pi/4} \sec(x^2) \ dx$; $\int_0^{\pi/4} (\sec(x))^2 \ dx$.

60. $\int_1^3 e^{x^2} x \ dx$; $\int_1^3 \frac{e^{x^2}}{x} \ dx$.

61. $\int_{0.2}^{0.4} \frac{dx}{\sqrt{1-x^2}}$; $\int_{0.2}^{0.4} \frac{dx}{\sqrt{1-x^3}}$.

62. If $F'(x) = f(x)$, find an antiderivative for (a) $g(x) = x + f(x)$, (b) $g(x) = 2f(x)$, and (c) $g(x) = f(2x)$.

63. John M. Robson in *The Physics of Fly Casting*, American J. Physics 58(1990), pp. 234–240, lets the reader fill in the calculus steps. For instance, he has the equation

$$\mu(4z + h)\dot{z}^2 = 2 \int_0^t c r h \rho \dot{z}^3 dt + T(0)$$

where z is a function of time t, $\dot{z} = dz/dt$, and $\ddot{z} = d^z/dt^2$. He then states, "differentiating this gives

$$(2\mu - c r h \rho)\dot{z}^2 + (4z + h)\mu\ddot{z} = 0."$$

Check that he is correct.

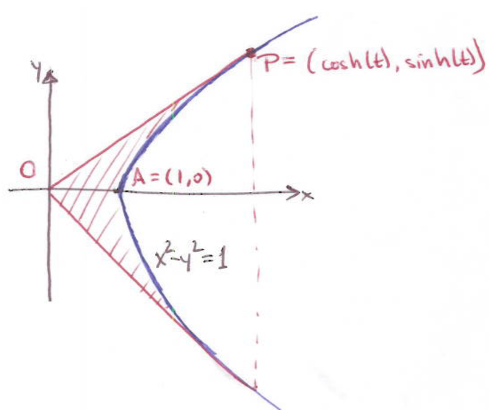


Figure 6.S.2

64. This exercise verifies the claims made in the last paragraph of Section 5.8.

(a) Explain why, for each angle θ in $[0, \pi]$, a sector of the unit circle with angle 2θ has area θ.

(b) In Figure 6.S.2, the area of the shaded region is twice the area of region OAP. The area of OAP is the area of a triangle less the area under the hyperbola. Express this area in terms of the parameter t. (This will include a definite integral with integrand $\sqrt{x^2-1}$.)

(c) Verify that $\frac{1}{2}\left(x\sqrt{x^2-1} - \ln(x+\sqrt{x^2-1})\right)$ is an antiderivative of $\sqrt{x^2-1}$ for $x > 1$.

(d) Show that the area of the shaded region in Figure 6.S.2 is t.

Alternate ways to compute the area of the shaded region are found in Exercises 77 on page 778 and 8 on page 1291.

65. Jane is running from a to b, on the x-axis. When she is at x, her speed is $v(x)$. How long does it take her to go from a to b?

66.

(a) Find all continuous functions $f(t)$, $t \geq 0$, such that $\int_0^{x^2} f(t)\,dt = 3x^3$. $x \geq 0$.

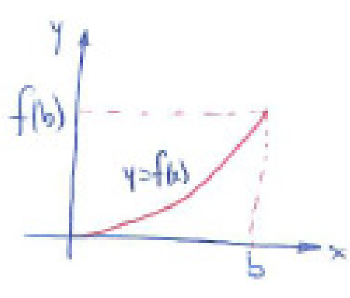
(b) Check that they satisfy the equation in (a).

67. Let $f(x)$ be defined for x in $[0, b]$, $b > 0$. Assume that $f(0) = 0$ and $f'(x)$ is positive.

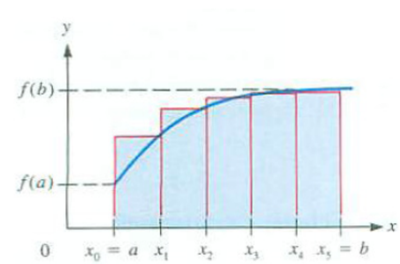
(a) Use Figure 6.S.3(a) to show that $\int_0^b f(x)\,dx + \int_0^{f(b)} (\mathsf{inv}f)(x)\,dx = bf(b)$.

(b) As a check on the equation in (a), differentiate both sides of it with respect to b. You should get a valid equation.

(c) Use (a) to evaluate $\int_0^1 \arcsin(x)\,dx$.



(a)



(b)

Figure 6.S.3

68.

(a) Verify, without using the FTC, that $\int_0^2 \sqrt{x(4-x)}\, dx = \pi$. (What region has an area give by that integral?)

(b) Approximate the definite integral in (a) by the trapezoidal estimate with 4 trapezoids and also with 8 trapezoids.

(c) Compute the error in each case.

(d) By trial-and-error, estimate how many trapezoids are needed to have an approximation that is accurate to three decimal places?

(e) Why is the error bound for the trapezoidal estimate of no use in (d)?

69.

(a) Approximate the definite integral in Exercise 68 by Simpson's estimate with 2 parabolas and again with 4 parabolas. (These use the same number of arguments as in Exercise 68.)

(b) Compute the error in each case.

(c) By trial-and-error, estimate how many parabolas are needed to have an estimate accurate to 3 decimal places. (Use your calculator or computer to automate the calculations.)

(d) Why is the error bound for the Simpson's estimate of no use in (c)?

70. In his *Principia*, published in 1607, Newton examined the error in approximating an area by rectangles. He considered an increasing, differentiable function f defined on the interval $[a, b]$ and drew a figure similar to Figure 6.S.3(b). All rectangles have the same width h. Let R equal the sum of the areas of the rectangles using right endpoints and let L equal the sum of the areas of the rectangles using left endpoints. Let A be the area under the curve $y = f(x)$ and above $[a, b]$.

(a) Why is $R - L = (f(b) - f(a))h$?

(b) Show that any approximating sum for A, formed with rectangles of equal width h and any sampling points, differs from A by at most $(f(b) - f(a))h$.

(c) Let M_1 be the maximum value of $|f'(x)|$ for x in $[a, b]$. Show that any approximating sum for A formed with equal widths h differs from A by at most $M_1(b - a)h$.

(d) Newton also considered the case where the rectangles do not necessarily have the same widths. Let h be the largest of their widths. What can be said about the error in this case?

71. Let f be a continuous function such that $f(x) > 0$ for $x > 0$ and $\int_0^x f(t)\ dt = (f(x))^2$ for $x \geq 0$.

(a) Find $f(0)$.

(b) Find $f(x)$ for $x > 0$.

72. A particle moves on a line in such a way that its average velocity over any interval of time $[a, b]$ is the same as its velocity at $(a + b)/2$. Prove that the velocity $v(t)$ must be of the form $ct + d$ for some constants c and d. (Differentiate the relationship $\int_a^b v(t)\ dt = v\left(\frac{a+b}{2}\right)(b-a)$ with respect to b and with respect to a.)

73. A particle moves on a line in such a way that the average velocity over any interval of the form $[a, b]$ is equal to the average of the velocities at the beginning and the end of the interval of time. Prove that the velocity $v(t)$ must be of the form $ct + d$ for some constants c and d.

Exercises 74 and 75 present Archimedes' derivations for the area of a disk and the volume of a ball. He viewed these explanations as informal, and also presented rigorous proofs for them.

74. Archimedes pictured a disk as made up of "almost" isosceles triangles, with one vertex of each triangle at the center of the disk and the base of the triangle part of the boundary of the disk. On the basis of this he conjectured that the area of a disk is one-half the product of the radius and its circumference. Explain why Archimedes' reasoning is plausible.

75. Archimedes pictured a ball as made up of "almost" pyramids, with the vertex of each pyramid at the center of the ball and the base of the pyramid as part of the surface of the ball. On the basis of this he conjectured that the volume of a ball is one-third the product of the radius and its surface area. Explain why Archimedes' reasoning is plausible.

76. (The midpoint estimates for a definite integral is described in Exercises 33 to 35 in Section 6.5.) Let M_n be the midpoint estimate of $\int_a^b f(x)\ dx$ based on n sections of width $h = (b - a)/n$. This exercise shows that the bound on the error, $\left|\int_a^b f(x)\ dx - M_n\right|$ is half of the bound on the trapezoidal estimate. The argument is like that in Exercises 44 and 45 of Section 6.5, a direct application of the Growth Theorem of Section 5.4.
Let $E(t) = \int_{-t/2}^{t/2} f(x)\ dx - f(0)t$.

(a) Show that $E(0) = E'(0) = 0$, and that $E''(t) = \frac{1}{4}\left(f'\left(\frac{t}{2}\right) - f'\left(\frac{-t}{2}\right)\right)$.

(b) Show that $\left|\int_a^b f(x)\, dx - M_n\right| \leq \frac{1}{24} M(b-a)h^2$, where M is the maximum of $|f''(x)|$ for x in $[a, b]$.

77. Let $y = f(x)$ be a function such that $f(x) \geq 0$, $f'(x) \geq 0$, and $f''(x) \geq 0$ for all x in $[1, 4]$. An estimate of the area under $y = f(x)$ is made by dividing the interval into sections and forming rectangles. The height of each rectangle is the value of $f(x)$ at the midpoint of the corresponding section.

(a) Show that the estimate is less than or equal to the area under the curve. (Draw a tangent to the curve at each of the midpoints.)

(b) How does the estimate compare to the area under the curve if, instead, $f''(x) \leq 0$ for all x in $[1, 4]$?

78. The definite integral $\int_0^1 \sqrt{x}\, dx$ gives numerical analysts a pain. The integrand is not differentiable at 0. What is worse, the derivatives (first, second, etc.) of $\sqrt{x}$ become arbitrarily large for x near 0. It is instructive, therefore, to see how the error in Simpson's estimate behaves as h is made small.

(a) Use the FTC to show that $\int_0^1 \sqrt{x}\, dx = \frac{2}{3}$.

(b) Fill in the table. (Keep at least 7 decimal places in each answer.)

h	Simpson's Estimate	Error
$\frac{1}{2}$		
$\frac{1}{4}$		
$\frac{1}{8}$		
$\frac{1}{16}$		
$\frac{1}{32}$		
$\frac{1}{64}$		

(c) In the typical application of Simpson's method, when you cut h by a factor of 2, you find that the error is cut by a factor of $2^4 = 16$. (That is, the ratio of the two errors would be $\frac{1}{16} = 0.0625$.) Examine the five ratios of consecutive errors in the table.

(d) Let $E(h)$ be the error in using Simpson's method to estimate $\int_0^1 \sqrt{x}\, dx$ with sections of length h. Assume that $E(h) = Ah^k$ for some constants k and A. Estimate k and A.

79. Since Simpson's method was designed to be exact when $f(x) = Ax^2+Bx+C$, one would expect the error associated with it to involve $f^{(3)}(x)$. By a quirk of good fortune, Simpson's method happens to be exact even when $f(x)$ is a cubic, $Ax^3 + Bx^2 + Cx + D$. This suggests that the error involves $f^{(4)}(x)$, not $f^{(3)}(x)$. Confirm that this is the case. Exercise 45 in Section 6.5 does this using the Growth Theorem.

(a) Show that $\displaystyle\int_c^d x^3\, dx = \frac{d-c}{6}\left(f(c) + 4f\left(\frac{c+d}{2}\right) + f(d)\right).$

(b) Why is Simpson's estimate exact for cubic polynomials?

80. A producer of wine can choose to store it and sell it at a higher price after it has aged. However, he also must consider storage costs, which should not exceed the revenue.

Assume the revenue he would receive when selling the wine at time t is $V(t)$. If the interest rate on bank balances is r, which we will assume is constant, the present value of that sale is $V(t)e^{-rt}$.

The cost of storing the wine varies with time. Assume $c(t)$ represents that cost, that is, the cost of storing the wine during the short interval $[t, t+\Delta t]$ is approximately $c(t)\Delta t$.

(a) What is the present value of storing the wine for the period $[0, x]$?

(b) What is the present value, $P(x)$, of the profit (or loss) selling all the wine at time x? That is, the present value of the revenue minus the present value of the storage cost if sold at time x?

(c) Show that $P'(x) = V'(x)e^{-rx} - rV(x)e^{-rx} - c(x)e^{-rx}$.

(d) Show that if $V'(x)e^{-rx} > rV(x)e^{-rx} + c(x)e^{-rx}$, then $P'(x)$ is positive, and he should continue to store the wine.

(e) What is the meaning of each of the three terms in the inequality in (d)? Why does that inequality make economic sense?

81. The average of a function that we have defined is called the arithmetic average. In some applications the geometric average is more appropriate and useful.

The **geometric average** of n positive numbers is defined as the n^{th} root of their product.

(a) If the positive numbers are $p_1,\ p_2,\ \ldots,\ p_n$, their geometric average G is $(p_1 p_2 \cdots p_n)^{1/n}$. Show that $\ln(G)$ is the arithmetic average of the n numbers $\ln(p_1),\ \ln(p_2),\ \ldots,\ \ln(p_n)$.

(b) Now let f be a continuous positive function on $[a, b]$. How would you define its "geometric average of f on $[a, b]$"?

(c) Check that your definition in (b) is between the minimum and maximum of f on $[a, b]$.

(d) How would you define the geometric average of a continuous positive function defined on $(0, \infty)$?

Skill Drill: Derivatives

Exercises 82 to 87 offer an opportunity to practice differentiation skills. In each case, verify that the derivative of the first function is the second function.

82. $\ln\left(\frac{e^x}{1+e^x}\right)$; $\frac{1}{1+e^x}$ (To simplify, first take logs.)

83. $\frac{1}{m}\arctan\left(e^{mx}\right)$; $\frac{1}{e^{mx}+e^{-mx}}$ (m is a constant).

84. $\ln(\tan(x))$; $\frac{1}{\sin(x)\cos(x)}$

85. $\tan\left(\frac{x}{2}\right)$; $\frac{1}{1+\cos(x)}$

86. $\frac{1}{2}\ln\left(\frac{1+\sin(x)}{1-\sin(x)}\right)$; $\sec(x) = \frac{1}{\cos(x)}$

87. $\arcsin(x) - \sqrt{1-x^2}$; $\sqrt{\frac{1+x}{1-x}}$

In Exercises 88 to 90 differentiate the given functions.

88. $\frac{\sin(2x)\tan(3x)}{x^3}$

89. $2^{x^2} x^3 \cos(4x)$

90. $\frac{x^2 e^{3x}}{\sqrt{1+x^2}}$

Calculus is Everywhere # 8
Peak Oil Production

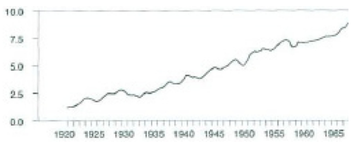


Figure C.8.1

The United States in 1956 produced most of the oil it consumed, and the rate of production was increasing. Even so, M. King Hubbert, a geologist at Shell Oil, predicted that production would peak near 1970 and then gradually decline. His prediction did not convince geologists, who were reassured by the rising curve in Figure C.8.1.

Hubbert was right and the moment of maximum production is known today as Hubbert's Peak.

We present below Hubbert's reasoning in his own words, drawn from "Nuclear Energy and the Fossil Fuels," available at `http://www.hubbertpeak.com/hubbert/1956/1956.pdf`. In it he uses an integral over the entire positive x-axis, a concept we will define in Section 7.8. However, since a finite resource is exhausted in a finite time, his integral is an ordinary definite integral, whose upper bound is not known.

First he stated two principles when analyzing curves that describe the rate of exploitation of a finite resource:

1. For any production curve of a finite resource of fixed amount, two points on the curve are known at the outset, namely that at $t = 0$ and again at $t = \infty$. The production rate will be zero when the reference time is zero, and the rate will again be zero when the resource is exhausted; that is to say, in the production of any resource of fixed magnitude, the production rate must begin at zero, and then after passing through one or several maxima, it must decline again to zero.

2. The second consideration arises from the fundamental theorem of integral calculus; namely, if there exists a single-valued function $y = f(x)$, then

$$\int_0^{x_1} y \, dx = A, \tag{C.8.1}$$

where A is the area between the curve $y = f(x)$ and the x-axis from the origin out to the distance x_1.

In the case of the production curve plotted against time on an arithmetical scale, we have as the ordinate

$$P = \frac{dQ}{dt}, \tag{C.8.2}$$

where dQ is the quantity of the resource produced in time dt. Likewise, from equation (C.8.1) the area under the curve up to any time t is given by

$$A = \int_0^t P\,dt = \int_0^t \left(\frac{dQ}{dt}\right)\,dt = Q, \qquad \text{(C.8.3)}$$

where Q is the cumulative production up to the time t. Likewise, the ultimate production will be given by

$$Q_{max} = \int_0^\infty P\,dt, \qquad \text{(C.8.4)}$$

and will be represented on the graph of production-versus-time as the total area beneath the curve.

These basic relationships are indicated in Figure C.8.2. The only a priori information concerning the magnitude of the ultimate cumulative production of which we may be certain is that it will be less than, or at most equal to, the quantity of the resource initially present. Consequently, if we knew the production curves, all of which would exhibit the common property of beginning and ending at zero, and encompassing an area equal to or less than the initial quantity.

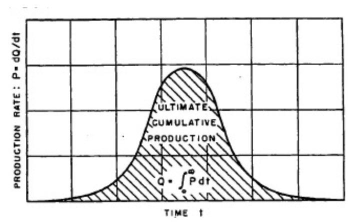


Figure C.8.2

That the production of exhaustible resources does behave this way can be seen by examining the production curves of some of the older producing areas.

He then examines those curves for Ohio and Illinois. They resembled the curves below, which describe more recent data on production in Alaska, the United States, the North Sea, and Mexico.

Hubbert did not use a particular formula. Instead he employed the key idea in calculus, expressed in terms of production of oil, "The definite integral of the rate of production equals the total production."

He looked at the data up to 1956 and extrapolated the curve by eye, and by logic. This is his reasoning:

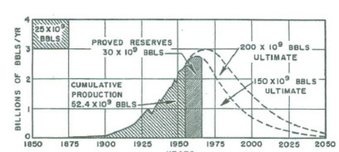
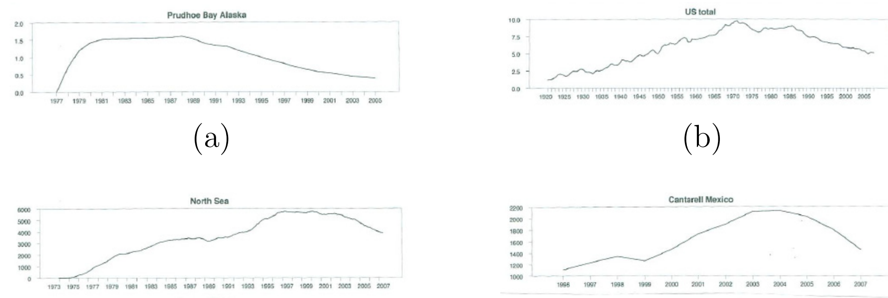


Figure C.8.4 Ultimate United States crude-oil production based on assumed initial reserves of 150 and 200 billion barrels.

> Figure C.8.4 shows "a graph of the production up to the present, and two extrapolations into the future. The unit rectangle in this case represents 25 billion barrels so that if the ultimate potential production is 150 billion barrels, then the graph can encompass but six rectangles before returning to zero. Since the cumulative production is already a little more than 50 billion barrels, then only four more rectangles are available for future production. Also, since the production rate is still increasing, the ultimate production peak must be greater than the present rate of production and must occur sometime in the future. At the same time it is possible to delay



(a) (b)

(c) (d)

Figure C.8.3 Annual production of oil in millions of barrels per day for (a) Annual oil production for Prudhoe Bay in Alaska, 1977–2005 [Alaska Department of Revenue], (b) moving average of preceding 12 months of monthly oil production for the United States, 1920–2008 [EIA, "Crude Oil Production"], (c) moving average of preceding 12 months of sum of U.K. and Norway crude oil production, 1973–2007 [EIA, Table 11.1b], and (d) annual production from Cantarell complex in Mexico, 1996–2007 [Pemex 2007 Statistical Yearbook and Green Car Congress (`http://www.greencarcongress.com/2008/01/mexicos-cantare.html`).

> the peak for more than a few years and still allow time for the unavoidable prolonged period of decline due to the slowing rates of extraction from depleting reservoirs.
>
> With due regard for these considerations, it is almost impossible to draw the production curve based upon an assumed ultimate production of 150 billion barrels in any manner differing significantly from that shown in Figure C.8.4, according to which the curve must culminate in about 1965 and then must decline at a rate comparable to its earlier rate of growth.
>
> If we suppose the figure of 150 billion barrels to be 50 billion barrels too low — an amount equal to eight East Texas oil fields — then the ultimate potential reserve would be 200 billion barrels. The second of the two extrapolations shown in Figure C.8.4 is based upon this assumption; but it is interesting to note that even then the date of culmination is retarded only until about 1970."

Geologists are now trying to predict when world production of oil will peak. (Hubbert predicted the peak to occur in the year 2000.) In 2009 oil was being extracted at the rate of 85 million barrels per day. Some say the peak occurred as early as 2005, but others believe it may not occur until after 2020.

To see some of the latest estimates, do a web search for "Hubbert peak oil estimate".

What is just as alarming is that the world is burning oil faster than we are discovering new deposits.

In the CIE on Hubbert's Peak in Chapter 10 (see page 915) we present a later work of Hubbert, in which he uses a specific formula to analyze oil use and depletion.

Chapter 7

Applications of the Definite Integral

This chapter develops four applications of the definite integral. Sections 7.1 and 7.4 describe two geometric applications: finding total area from the length of each cross section and finding total volume from the area of each cross section. Section 7.5 gives an alternate way to compute volumes. Sections 7.6 and 7.7 present two applications in physics, the first, on water pressure, the second on the work accomplished by a force.

Section 7.8 generalizes the definite integral to cases when either the integrand becomes infinite or the interval of integration is infinite. The case where you "integrate from zero to infinity," for instance, is surprisingly common in physics and statistics.

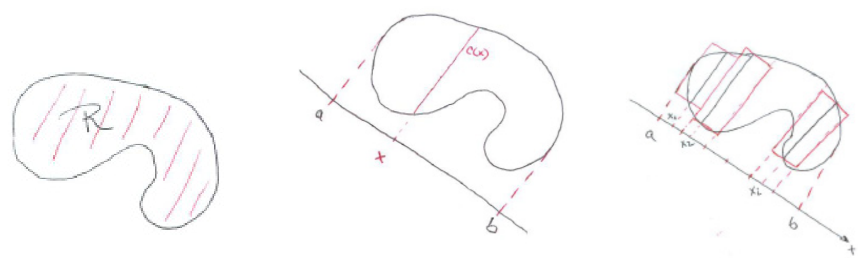
Advice on drawing and setting up definite integrals, two very useful but often overlooked skills, is found in Sections 7.2 and 7.3.

7.1 Computing Area by Parallel Cross-Sections

In Section 6.1 we computed the area under $y = x^2$ and above the interval $[a, b]$, and later saw that it equals the definite integral $\int_a^b x^2\,dx$. Now we generalize the idea behind this example.

Area as a Definite Integral of Cross Sections

How can we express the area of the region R shown in Figure 7.1.1(a) as a definite integral?



(a) (b) (c)

Figure 7.1.1

First, we introduce an "x-axis", as in Figure 7.1.1(b).

Assume that lines perpendicular to the axis for x in $[a, b]$, intersect the region R in an interval of length $c(x)$. The interval is called the **cross section** of R at x.

We approximate R by a collection of rectangles, just as we estimated the area of the region under $y = x^2$.

Pick an integer n, and divide the interval $[a, b]$ on the x-axis into n congruent sections. The total length of the interval $[a, b]$ is $b - a$; each section has width $\Delta x = \frac{b-a}{n}$. Then, in the i^{th} section, $i = 1, 2, \ldots, n$, we pick a "sampling number" x_i. For each of the n sections we form a rectangle of width Δx and height $c(x_i)$. These are indicated in Figure 7.1.1(c).

Since we use c for the "cross-sectional" length, we cannot use use c_i to name the sampling point. Instead, x_i is used to denote the sampling point. This does not cause any confusion since we are not using x_i to describe the endpoints of a partition.

Since the i^{th} rectangle has area $c(x_i)\Delta x$, the total area of the n rectangles is $\sum_{i=1}^{n} c(x_i)\Delta x$. As n increases, the collection of rectangles provides a better approximation to the area of R. This suggests that:

$$\lim_{n\to\infty} \sum_{i=1}^{n} c(x_i)\Delta x = \text{area of region } R$$

But, by the definition of a definite integral,

$$\lim_{n\to\infty} \sum_{i=1}^{n} c(x_i)\Delta x = \int_a^b c(x)\ dx.$$

Thus,

$$\boxed{\text{area of } R = \int_a^b c(x)\ dx.}$$

Or, informally,

Area of a region equals the integral of its cross-sectional lengths.

Note that x need not refer to the x-axis of the xy-plane; it may refer to any conveniently chosen line in the plane. It may even refer to the y-axis; in this case the cross-sectional length would be denoted by $c(y)$.

To compute an area:

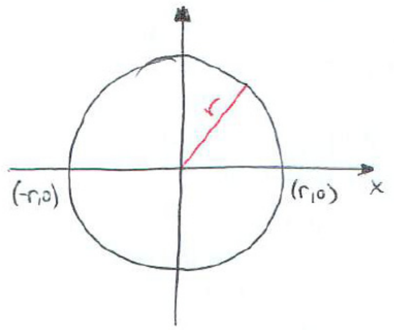
1. Find the endpoints a and b, and the cross-sectional length $c(x)$.

2. Evaluate $\int_a^b c(x)\ dx$ by the Fundamental Theorem of Calculus, if the antiderivative of $c(x)$ is elementary.

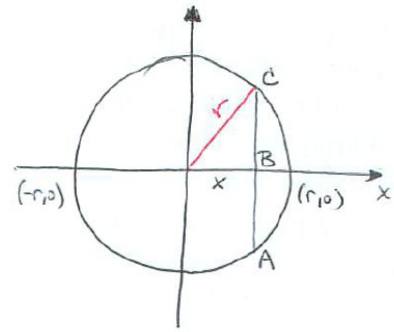
Chapter 6 showed how to accomplish Step 2. FTC I is used when the antiderivative is an elementary function, and other cases can be approximated numerically. The present section is concerned primarily with Step 1, how to find the cross-sectional length $c(x)$ and set up the definite integral.

If the region R happens to be the region under the graph of $f(x)$ and above the interval $[a, b]$, then the cross-sectional length is simply $f(x)$. We have already met this special case in Sections 6.2–6.4 with $f(x) = x^2$ and $f(x) = 2^x$.

EXAMPLE 1 Find the area of a disk of radius r.
SOLUTION Introduce an xy-coordinate system with its origin at the center of the disk, as in Figure 7.1.2(a).



(a)



(b)

Figure 7.1.2

The typical cross section perpendicular to the x-axis is shown in Figure 7.1.2(b). The length of the cross section, $\overline{AC}$, is twice $\overline{BC}$. By the Pythagorean Theorem,

$$x^2 + \overline{BC}^2 = r^2.$$

Then

$$\overline{BC}^2 = r^2 - x^2$$

and, because $\overline{BC}$, a length, is positive

$$\overline{BC} = \sqrt{r^2 - x^2}.$$

Because x is in $[-r, r]$,

$$\text{area of disk of radius } r = \int_{-r}^{r} 2\sqrt{r^2 - x^2}\, dx. \tag{7.1.1}$$

Equation (7.1.2) is preferable because it reduces the chance of making an error when working with the subtraction of negative numbers.

By symmetry, we can also say that the total area is four times the area of a quadrant:

$$\text{area of disk of radius } r = 4\int_{0}^{r} \sqrt{r^2 - x^2}\, dx. \tag{7.1.2}$$

This completes the setup of the integral for the area of the region.

The next chapter presents a technique for finding an antiderivative of $\sqrt{r^2 - x^2}$. In the mean time, we use the table of integrals on the inside cover. According to formula 32,

$$\int \sqrt{r^2 - x^2}\, dx = \frac{r^2}{2}\left(\arcsin\left(\frac{x}{r}\right) + \frac{x}{r^2}\sqrt{r^2 - x^2}\right).$$

By FTC I,

See Exercise 43

$$\begin{aligned}
\int_0^r \sqrt{r^2-x^2}\,dx &= \frac{r^2}{2}\left(\arcsin\left(\frac{x}{r}\right)+\frac{x}{r^2}\sqrt{r^2-x^2}\right)\Bigg|_0^r \\
&= \frac{r^2}{2}\left(\arcsin\left(\frac{r}{r}\right)+\frac{r}{r^2}\sqrt{r^2-r^2}\right)-\frac{r^2}{2}\left(\arcsin\left(\frac{0}{r}\right)+\frac{0}{r^2}\sqrt{r^2-0^2}\right) \\
&= \frac{r^2}{2}\left(\frac{\pi}{2}\right)=\frac{\pi r^2}{4}.
\end{aligned}$$

Thus one quarter of the disk has area $\frac{\pi r^2}{4}$ and the whole disk has area πr^2. $\diamond$

Archimedes found the area in the next example, expressing it in terms of the area of a certain triangle (see Exercise 41). He used geometric properties of a parabola, since calculus was not invented until some 1900 years later.

Reference: S. Stein: *Archimedes: What did he do besides cry Eureka?*, MAA, 1999.

EXAMPLE 2 Set up a definite integral for the area of a region above the parabola $y = x^2$ and below the line through $(2, 0)$ and $(0, 1)$ shown in Figure 7.1.3.

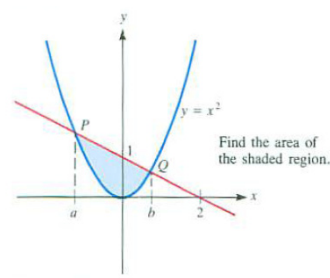


Figure 7.1.3

SOLUTION Since the x-intercept of the line is 2 and the y-intercept is 1, an equation for the line is

$$\frac{x}{2}+\frac{y}{1}=1.$$

Hence $y = 1 - x/2$. The length $c(x)$ of a cross section of the region taken parallel to the y-axis is, therefore

$$c(x)=\left(1-\frac{x}{2}\right)-x^2=1-\frac{x}{2}-x^2.$$

To find the interval $[a, b]$ of integration, we must find the x-coordinates of the points P and Q in Figure 7.1.2(b) where the line meets the parabola. For these values of x,

$$x^2=1-\frac{x}{2},$$

so

$$2x^2+x-2=0. \qquad (7.1.3)$$

The solutions to (7.1.3) are

$$x=\frac{-1\pm\sqrt{17}}{4}.$$

Hence

$$\text{area}=\int_{(-1-\sqrt{17})/4}^{(-1+\sqrt{17})/4}\left(1-\frac{x}{2}-x^2\right)\,dx.$$

The value of this definite integral is found in Exercise 33. ◇

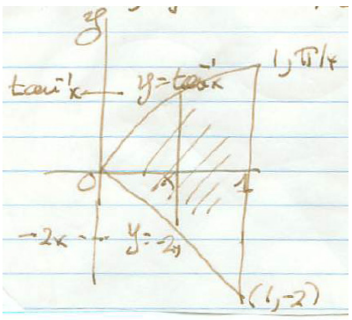


Figure 7.1.4

EXAMPLE 3 Find the area of the region in Figure 7.1.4, bounded by $y = \arctan(x)$, $y = -2x$, and $x = 1$.

SOLUTION We will find the area two ways, first (a) with cross sections parallel to the y-axis, then (b) with cross sections parallel to the x-axis.

(a) The typical cross section has length $\arctan(x) - (-2x) = \arctan(x) + 2x$. Thus the area is

$$\int_0^1 (\arctan(x) + 2x)\ dx.$$

Formula 72 in the cover of this book tells us that $\int \arctan(x)\ dx$ is $x\arctan(x) - \frac{1}{2}\ln(1+x^2)$. Use differentiation to check that this is correct.

It's easy to find $\int 2x\ dx$; it's just x^2. By the FTC,

$$\begin{aligned}
\int_0^1 (\arctan(x) + 2x)\ dx &= \left(x\arctan(x) - \frac{1}{2}\ln(1+x^2) + x^2 \right)\Big|_0^1 \\
&= \left(1\arctan(1) - \frac{1}{2}\ln(1+1^2) + 1^2 \right) \\
&\quad - \left(0\arctan(0) - \frac{1}{2}\ln(1+0^2) + 0^2 \right) \\
&= \left(\frac{\pi}{4} - \frac{1}{2}\ln(2) + 1 \right) - 0 \\
&= \frac{\pi}{4} + 1 - \frac{1}{2}\ln(2). \approx 1.4388 \qquad (7.1.4)
\end{aligned}$$

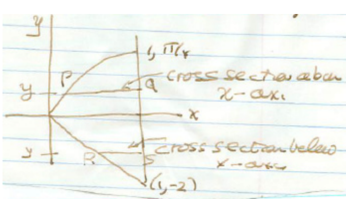


Figure 7.1.5

(b) Now we use cross sections parallel to the x-axis, as indicated in Figure 7.1.5.

Cross-sections above the x-axis involve the curved part of the boundary, while those below the x-axis involve the slanted line.

We must find the cross-sectional length as a function of y. That means we should first find the x-coordinates of P and Q, the ends of the typical cross section above the x-axis. The x-coordinate of Q is 1. Let the x-coordinate of P be x, then $y = \arctan(x)$, so $x = \tan(y)$. Hence

$$c(y) = 1 - \tan(y), \qquad \text{for } y \geq 0.$$

The length of RS, a typical cross section below the x-axis, is $1-$(x-coordinate of Since R is on the line $y = -2x$, we have $x = -y/2$. Thus

$$c(y) = 1 - (-y/2) = 1 + y/2, \qquad \text{for } -2 \leq y \leq 0.$$

Note that the interval of integration is $[-2, \pi/4]$. Hence

$$\text{area of } R = \int_{-2}^{\pi/4} c(y)\ dy.$$

We have to break this integral into two separate ones:

$$\int_{-2}^{0} \left(1 + \frac{y}{2}\right)\ dy \text{ and } \int_{0}^{\pi/4} (1 - \tan(y))\ dy \tag{7.1.5}$$

It will be shown Example 3 in Section 8.5 that

Differentiate $\ln(\sec(y))$ to check this antiderivative. Because $\sec(y)$ is positive for $-\pi/2 < y < \pi/2$ it is not necessary to write the antiderivative as $\ln|\sec(y)|$; see Exercise 31.

$$\int \tan(y)\ dy = \ln(\sec(y)).$$

First,

$$\begin{aligned} \int_{-2}^{0} \left(1 + \frac{y}{2}\right)\ dy &= \left(y + \frac{y^2}{4}\right)\Big|_{-2}^{0} \\ &= \left(0 + \frac{0^2}{4}\right) - \left((-2) + \frac{(-2)^2}{4}\right) \\ &= 1 \end{aligned} \tag{7.1.6}$$

Second,

$$\begin{aligned} \int_{0}^{\pi/4} (1 - \tan(y))\ dy &= (y - \ln\sec(y))|_0^{\pi/4} \\ &= \left(\frac{\pi}{4} - \ln(\sec(\frac{\pi}{4}))\right) - (0 - \ln(\sec(0))) \\ &= \frac{\pi}{4} - \ln(\sqrt{2}) \end{aligned} \tag{7.1.7}$$

Adding (7.1.6) and (7.1.7) gives

$$\text{area of } R = \frac{\pi}{4} - \ln(\sqrt{2}) + 1 \tag{7.1.8}$$

The two answers (7.1.4) and (7.1.8) may look different but they agree, as you may show in Exercise 32. ◇

See Exercise 32.

In this example we could have simplified the solution by observing that the area below the x-axis is a triangle of area 1. But the purpose of Example 3 is to illustrate a general approach.

Summary

The key idea in this section, "area of a region equals integral of cross-sectional length," was already anticipated in Chapter 6. There we met the special case where the region is bounded by the graph of a function, the x-axis, and two lines perpendicular to the axis. In this section the concept was extended to more general regions.

EXERCISES for Section 7.1

In each of Exercises 1 to 6 (a) draw the region, (b) compute the lengths of vertical cross sections ($c(x)$), and (c) compute the lengths of horizontal cross sections ($c(y)$).

1. The finite region bounded by $y = \sqrt{x}$ and $y = x^2$.

2. The finite region bounded by $y = x^2$ and $y = x^3$.

3. The finite region bounded by $y = 2x$, $y = 3x$, and $x = 1$.

4. The finite region bounded by $y = x^2$, $y = 2x$, and $x = 1$.

5. The triangle with vertices $(0, 0)$, $(3, 0)$, and $(0, 4)$.

6. The triangle with vertices $(1, 0)$, $(3, 0)$, and $(2, 1)$.

In Exercises 7 to 12 find the indicated areas. Use the table of integrals provided inside the cover of this textbook to find antiderivatives, if necessary.

7. Under $y = \sqrt{x}$ and above $[1, 2]$

8. Under $y = \sin(2x)$ and above $[\pi/6, \pi/3]$

9. Under $y = e^{2x}$ and above $[0, 1]$

10. Under $y = 1/\sqrt{1 - x^2}$ and above $[0, 1/2]$.

11. Under $y = \ln(x)$ and above $[1, e]$

12. Under $y = \cos(x)$ and above $[-\pi/2, \pi/2]$

In Exercises 13 to 20 find the indicated areas using cross sections parallel to the x-axis.

13. Between $y = x^2$ and $y = x^3$.

14. Between $y = 2^x$ and $y = 2x$.

15. Between $y = \arcsin(x)$ and $y = 2x/\pi$ (to the right of the y-axis).

16. Between $y = 2^x$ and $y = 3^x$ (to the right of the y-axis).

17. Between $y = \sin(x)$ and $y = \cos(x)$ (above $0, \pi/2]$.

18. Between $y = x^3$ and $y = -x$ for x in $[1, 2]$.

19. Between $y = x^3$ and $y = \sqrt[3]{2x - 1}$ for x in $[1, 2]$.

20. Between $y = 1 + x$ and $y = \ln(x)$ for x in $[1, e]$.

In Exercises 21 to 27 set up a definite integral for the area of the given region. *Do not attempt to evaluate the integral.* These integrals will be evaluated in Exercises 36 to 42 in the Chapter 8 Summary.

21. The region under the curve $y = \arctan(2x)$ and above the interval $[1/2, 1/\sqrt{3}]$.

22. The region in the first quadrant below $y = -7x + 29$ and above the portion of $y = 8/(x^2 - 8)$ that lies in the first quadrant.

23. The region below $y = 10^x$ and above $y = \log_{10}(x)$ for x in $[1, 10]$.

24. The region under the curve $y = x/(x^2+5x+6)$ and above the interval $[1, 2]$.

25. The region below $y = (2x + 1)/(x^2 + x)$ and above the interval $[2, 3]$.

26. The region bounded by $y = \tan(x)$, $y = 0$, $x = 0$, and $x = \pi/2$ by (a) vertical cross sections and (b) horizontal cross sections.

27. The region bounded by $y = \sin(x)$, $y = 0$, and $x = \pi/4$ (consider only $x \geq 0$) by (a) vertical cross sections and (b) horizontal cross sections.

28.

(a) Draw the region inside the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

(b) Find a definite integral for the area of the ellipse in (a) with horizontal cross sections.

(c) Find a definite integral for the area of the ellipse in (a) with vertical cross sections.

See Exercise 43 in Chapter 8 Summary.

29. Cross-sections in different directions lead to different definite integrals for the same area. While both integrals must give the same area, one of the two integrals can be easier to evaluate.

(a) Identify and evaluate the easier definite integral found in Exercise 26.

(b) Identify and evaluate the easier definite integral found in Exercise 27.

30. Set up the definite integral for the area $A(b)$ of the region in the first quadrant under the curve $y = e^{-x}(\cos(x))^2$ and above the interval $[0, b]$.

31. In Example 3 it is asserted that $\int \tan(y)\ dy = \ln(\sec(y))$. Verify this result, by differentiating.

32. In Example 3 the area of the region bounded by $y = \arctan(x)$, $y = 2x$, and $x = 1$ is found to be both

$$\frac{\pi}{4} + 1 - \frac{1}{2}\ln(2) \qquad \text{and} \qquad \frac{\pi}{4} - \ln(\sqrt{2}) + 1.$$

Explain why these two answers are equal.

33. In Example 2 the area of the region above the parabola $y = x^2$ and below the line through $(2, 0)$ and $(0, 1)$ is found to be

$$\text{area} = \int_{(-1-\sqrt{17})/4}^{(-1+\sqrt{17})/4} \left(1 - \frac{x}{2} - x^2\right)\, dx.$$

Find the value of this definite integral.

34. Let R be the region bounded by $y = x^3$, $y = x + 2$, and the x-axis.

(a) Find a definite integral for the area of R. (You may have to solve an equation to find an endpoint of the interval of integration.)

(b) Use a graph or other method to approximate the endpoints.

(c) Use the estimates in (b) to obtain an estimate of the area of R.

35. Let R be the region to the right of the y-axis bounded by $y = 3$ and $y = e^x/x$.

(a) Graph the region R.

(b) Find a definite integral for the area of R. (You will encounter an equation that cannot be solved exactly. Identify the endpoints on the graph found in (a).)

(c) Find approximate values for the endpoints of the definite integral for the area in (b).

(d) Because the antiderivative of e^x/x is not elementary, it is still not easy to estimate the area of R. What methods do we have for estimating this definite integral? Use one of these definite integrals to find an approximate value for the area of R.

36. Let a be a positive number. What fraction of the rectangle whose vertices are $(0, 0)$, $(a, 0)$, (a, a^4), and $(0, a^4)$,is occupied by the region under the curve $y = x^4$ and above $[0, a]$?

37. Let $A(t)$ be the area of the region in the first quadrant between $y = x^2$ and $y = 2x^2$ and inside the rectangle bounded by $x = t$, $y = t^2$, and the coordinate axes. (See the shaded region in Figure 7.1.6.) If $R(t)$ is the area of the rectangle, find

(a) $\displaystyle\lim_{t\to 0}\frac{A(t)}{R(t)}$

(b) $\displaystyle\lim_{t\to\infty}\frac{A(t)}{R(t)}$

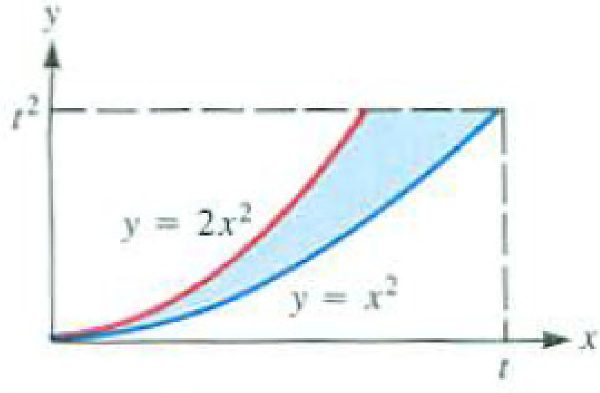


Figure 7.1.6

38. Figure 7.1.7 shows the graph of an increasing function $y = f(x)$ with $f(0) = 0$. Assume that $f'(x)$ is continuous and $f'(0) > 0$. Do not assume that $f''(x)$ exists. Our objective is to investigate

$$\frac{\text{area of shaded region under the curve}}{\text{area of triangle } ABC} \quad \text{as } t \text{ decreases toward } 0. \tag{7.1.9}$$

(a) Experiment with various functions, including some trigonometric functions and polynomials. Make sure that $f'(0) > 0$.

(b) Make a conjecture about (7.1.9) and explain why it is true.

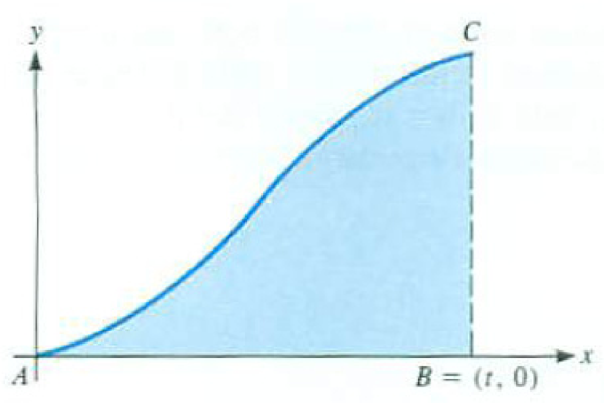
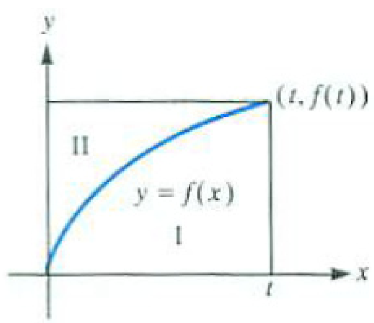


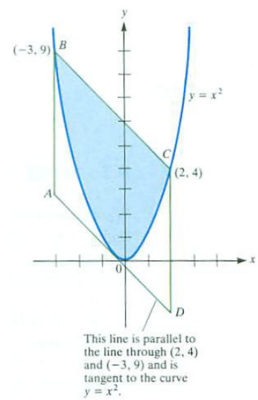
Figure 7.1.7

39. Repeat Exercise 38, but now assume that $f'(0) = 0$, f'' is continuous, and $f''(0) \neq 0$.

40. Let f be an increasing function with $f(0) = 0$, and assume that it has an elementary antiderivative. Then f^{-1} is an increasing function, and $f^{-1}(0) = 0$. Prove that if f^{-1} is elementary, then it also has an elementary antiderivative. (See Figure 7.1.8(a).)



(a)



(b)

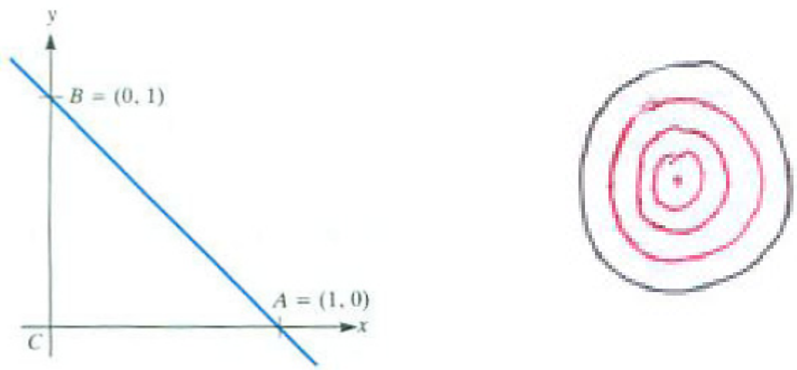
Figure 7.1.8

41. Show that the area of the shaded region in Figure 7.1.8(b) is two-thirds the area of the parallelogram $ABCD$. This is an illustration of a theorem of Archimedes concerning sectors of parabolas. He showed that the shaded area is 4/3 the area of triangle BOC. See also Example 2.

42. Figure 7.1.9(a) shows a right triangle ABC.

(a) Find equations for the lines parallel to each edge, AC, BC, and AB, that cut the triangle into two pieces of equal area.

(b) Are the three lines in (a) concurrent; that is, do they meet at a single point?



(a) (b)

Figure 7.1.9

43. Find the area of a disk of radius r by using concentric rings as suggested in Figure 7.1.9(b). The advantage of this approach is that it leads to an integral with a much simpler antiderivative than in Example 1. (Approximate the area of each ring as the product of a circumference and the width of the ring.)

7.2 Some Pointers on Drawing

None of us were born knowing how to draw solids. As we grew up, we lived in flatland: the surface of the Earth. Few high school math classes cover solid geometry, so calculus is often the first place where you have to think and sketch in terms of three dimensions. That is why we pause for a few words of advice on how to draw. Too often you cannot work a problem simply because your diagrams confuse even yourself. The following guidelines are not based on any profound artistic principles. Instead, they derive from years attempting to sketch diagrams that do more good than harm.

A Few Words of Advice

1. *Draw large.* Many students tend to draw diagrams that are so small that there is no room to place labels or to sketch cross sections.

2. *Draw neatly.* Use a straightedge to make straight lines that are actually straight. Use a compass to make circles that look like circles. Draw each line or curve slowly.

3. *Avoid clutter.* If you end up with too many labels or the cross section doesn't show up well, add separate diagrams for important parts of the figure.

4. *Practice.*

A jar lid or soda can works just fine for drawing circles and circular arcs. Credit cards and ID badges make good straightedges.

EXAMPLE 1 Draw a diagram of a ball of radius a that shows the circular cross section made by a plane at a distance x from the center of the ball. Use the diagram to help find the radius of the cross section as a function of x.

This example is continued in Example 1 in Section 7.4.

TERRIBLE SOLUTION Is Figure 7.2.1 a potato or a ball? What segment has length r? What's x? What does the cross section look like?

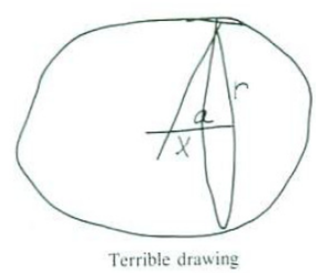


Figure 7.2.1

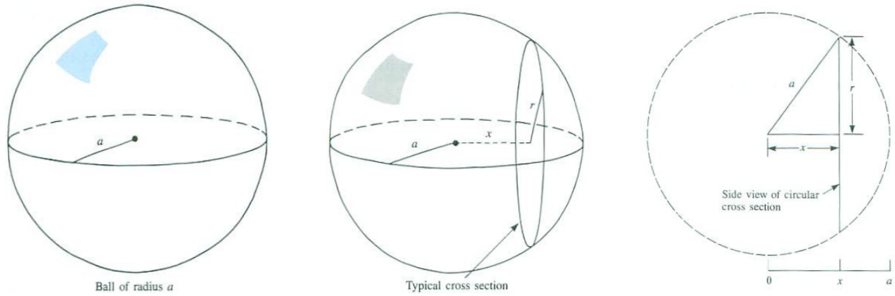
REASONABLE SOLUTION First, draw the ball carefully, as in Figure 7.2.2(a). The equator is drawn to give it perspective. Add a little shading.

Next show a typical cross section at a distance x from the center, as in Figure 7.2.2(b). Shading the cross section helps, too.

To find r, the radius of the cross section, in terms of x, sketch a companion diagram. The radius we want is part of a right triangle. In order to avoid clutter, draw only the part of interest in a convenient side view, as in Figure 7.2.4(c).

Inspection of the right triangle in this figure shows that

$$r^2 + x^2 = a^2, \quad \text{hence that} \quad r = \sqrt{a^2 - x^2}.$$



(a) (b) (c)

Figure 7.2.2 NOTE: Add shading to cross section in (b).

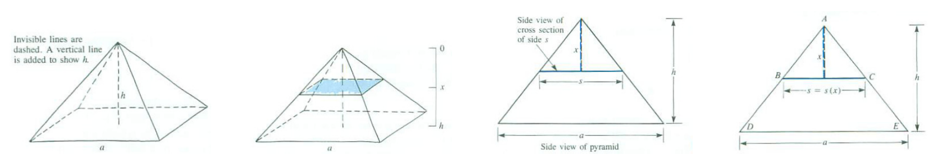
$\diamond$

This example is continued in Example 2 in Section 7.4.

EXAMPLE 2 A pyramid has a square base with a side of length a. The top of the pyramid is above the center of the base at a height h. Draw the pyramid and its cross sections by planes parallel to the base. Then find the area of the cross sections in terms of their distance x from the top.

Figure 7.2.3 Terrible drawing

TERRIBLE SOLUTION Figure 7.2.3 is too small; there's no room for the symbols. While it's pretty clear which side has length a, to what are the x and h attached? Also, without the hidden edges of the pyramid the shape of the base is not clear.



(a) (b) (c) (d)

Figure 7.2.4

REASONABLE SOLUTION First draw a large pyramid with a square base, as in Figure 7.2.4(a). Note that the opposite edges of the base are drawn as parallel lines. While artists draw parallel lines as meeting in a point to enhance the sense of perspective, for our purposes it is more useful to use parallel lines to depict parallel lines. Then show a typical cross section in perspective and side views, as in Figures 7.2.4(b) and (c). Note the x-axis, which is drawn separate from the pyramid.

As x increases, so does s, the width of the square cross section. Thus s is a function of x, which we could call $s(x)$ (or $f(x)$, if you prefer). A glance

The use of s is recommended because it suggests its meaning - side.

at Figure 7.2.4(b) shows that $s(0) = 0$ and $s(h) = 1$. To find $s(x)$ for all x in $[0, h]$, use the similar triangles ABC and ADE, shown in Figure 7.2.4(c). These triangles show that

$$\frac{x}{s} = \frac{h}{a}; \qquad \text{hence} \qquad s = \frac{ax}{h}. \tag{7.2.1}$$

Notice that $s = \frac{ax}{h}$ expresses s is a linear function of x. As a check on (7.2.1), replace x by 0 and by h; we get 0 and a for the respective values of s, as expected. Finally, the area A of the cross sections is given by

$$A = s^2 = \left(\frac{ax}{h}\right)^2.$$

$\diamond$

EXAMPLE 3 A cylindrical drinking glass of height h and radius a is full of water. It is tilted until the remaining water covers exactly half the base.

This example is continued in Example 2 and Exercise 18, both in Section Section 7.4.

A. Draw a diagram of the glass and water.

B. Show a cross section of the water that is a triangle.

C. Find the area of the triangle in terms of the distance x of the cross section from the axis of the glass.

TERRIBLE SOLUTION The diagram in Figure 7.2.5 is too small. It is not clear what has length a. The cross section is unclear. What does x refer to?

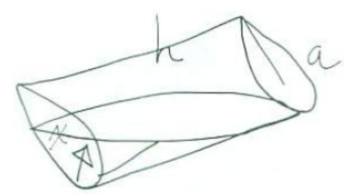
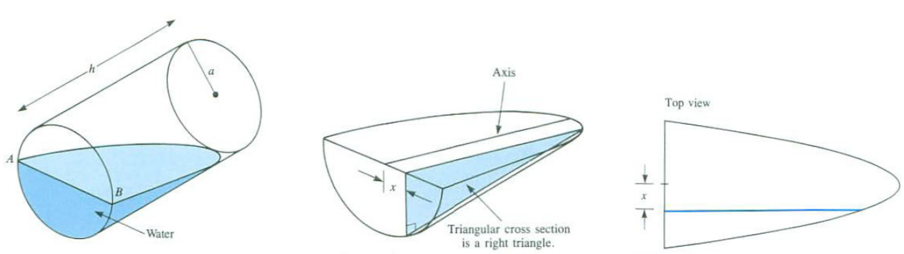


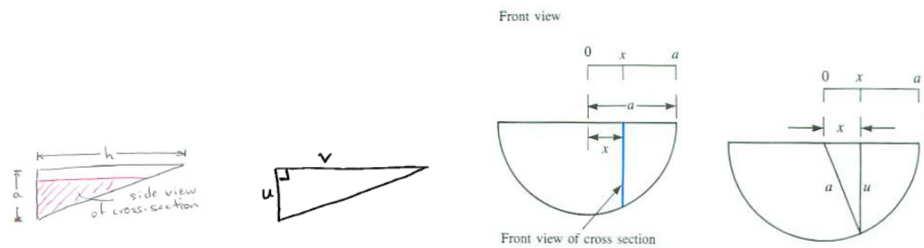
Figure 7.2.5 ARTIST: Make this sketch smaller (it's supposed to be "too small").



(a) (b) (c)

Figure 7.2.6

REASONABLE SOLUTION First, draw a *neat, large* diagram of a slanted cylinder, as in Figure 7.2.6. Don't put in too much detail at first. When showing the cross section, draw only the water. Figures 7.2.6 and 7.2.7 show



(a) (b) (c) (d)

Figure 7.2.7

various views. Let u and v be the lengths of the two legs of the cross section, as shown in Figure 7.2.7(d).

Comparing Figures 7.2.7(a) and (b), we have, by similar triangles, the relation

$$\frac{u}{a} = \frac{v}{h} \qquad \text{hence} \qquad v = \frac{h}{a}u.$$

Let $A(x)$ be the area of the cross section at a distance x from the center of the base, as shown in Figure 7.2.6(b). If we can find u and v as functions of x, we will be able to write a formula for $A(x) = \frac{1}{2}uv$ in terms of x.

Figure 7.2.7(b) suggests how to find u. Copy it and draw in the necessary radius, as in Figure 7.2.7(d). By the Pythagorean Theorem,

$$u = \sqrt{a^2 - x^2}.$$

All told,

$$A(x) = \frac{1}{2}uv = \frac{1}{2}u\left(\frac{h}{a}u\right) = \frac{h}{2a}u^2 = \frac{h}{2a}(a^2 - x^2). \tag{7.2.2}$$

As a check, note that

$$A(a) = \frac{h}{2a}(a^2 - a^2) = 0,$$

which makes sense. Also the formula (7.2.2) gives

$$A(0) = \frac{h}{2a}(a^2 - 0^2) = \frac{1}{2}ah,$$

again agreeing with the geometry of, say, Figure 7.2.6(b). ◇

Summary

When you look back at these three examples, you will see that most of the work is spent on making clear diagrams. If you can't draw a straight line free hand, use a straightedge. If you can't draw a circle, use a compass.

EXERCISES for Section 7.2

1. Cross-sections of the pyramid in Example 2 are made by using planes perpendicular to the base and parallel to the edge of the base. What is the area of the cross section made by a plane that is a distance x from the top of the pyramid?

(a) Draw a large perspective view of the pyramid.

(b) Copy the diagram in (a) and show the typical cross section shaded.

(c) Draw a side view that shows the shape of the cross section.

2. Cross-sections of the water in Example 3 are made by using planes parallel to the plane that passes through the horizontal diameter of the base and the axis of the glass. What is the area of the cross section made by a plane that is a distance x from the center of the base?

(a) Draw a large perspective view of the water and glass.

(b) Copy the diagram in (a) and show the typical cross section shaded.

(c) Draw a side view that clearly shows the shape of the cross section.

(d) Draw a different side view.

(e) Put necessary labels, such as x, a, and h, on the diagrams, where appropriate. (You will need to introduce more labels.)

(f) Find the area of the cross section, $A(x)$, as a function of x.

3. Cross-sections of the water in Example 3 are made by using planes perpendicular to the axis of the glass. Make clear diagrams, including perspective and side views, that show the typical cross sections. Do not find its area.

4. A lumberjack saws a wedge out of a cylindrical tree of radius a. His first cut is parallel to the ground and stops at the axis of the tree. His second cut makes an angle θ with the first cut and meets it along a diameter.

(a) Draw a typical cross section that is a triangle.

(b) Find the area of the triangle as a function of x, the distance of the plane from the axis of the tree.

(c) Draw a typical cross section that is a rectangle.

(d) Find the area of the rectangle as a function of x, the distance of the plane from the axis of the tree.

5. A cylindrical glass is full of water. The glass is tilted until the remaining water just covers the base of the glass. (Try it.) The radius of the glass is a and its height is h. Consider parallel planes such that cross sections of the water are rectangles.

(a) Make clear diagrams that show the situation. (Include a top view to show the cross sections.)

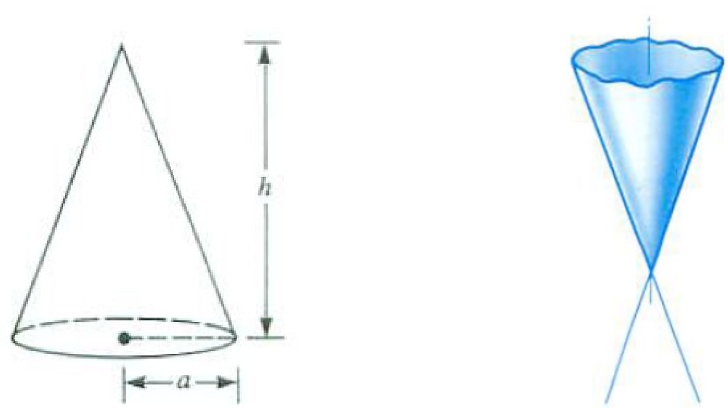
(b) Obtain a formula for the area of the cross sections. *Advice:* The two planes at the same distance x from the axis of the glass cut out cross sections of different areas. So introduce an x-axis with 0 at the center of the base and extending from $-a$ to a in a convenient direction.

6. Repeat Exercise 5, but this time consider parallel planes such that the cross sections are trapezoids.

7. A right circular cone has a radius a and height h as shown in Figure 7.2.8(a). Consider cross sections made by planes parallel to the base of the cone.

(a) Draw perspective and side views of the cone and typical cross sections.

(b) Drawing as many diagrams as necessary, find the area of the cross section made by a plane at a distance x from the vertex of the cone.



(a) (b)

Figure 7.2.8

8. Draw the typical cross section made by a plane parallel to the axis of the cone. Draw perspective and side views, but do not find a formula for the area of the cross section. See Exercise 7

9. Figure 7.2.8(b) indicates an unbounded, solid right circular cone. Draw a cross section that is bounded by (a) a circle, (b) an ellipse (but not a circle), (c) a parabola, and (d) a hyperbola.

10. Draw a cross section of a right circular cylinder that is (a) a circle, (b) an ellipse that is not a circle, and (c) a rectangle.

11. Draw a cross section of a solid cube that is (a) a square, (b) an equilateral triangle, (c) a five-sided polygon, and (d) a regular hexagon.

12. The plane region between the curves $y = x$ and $y = x^2$ is spun around the x-axis to produce a solid resembling the bell of a trumpet.

(a) Draw the plane region.

(b) Draw the solid region produced by spinning this region around the x-axis.

(c) Draw the typical cross section made by a plane perpendicular to the x-axis. Show this in both perspective and side views.

(d) Find the area of the cross section in terms of the distance x of the plane from the origin to the x-axis.

13. Obtain a circular stick such as a broom handle or a dowel. Saw off a piece, making one cut perpendicular to the axis and the second cut at an angle to the axis. Mark on the surface of the piece you cut out the borders of cross sections that are (a) rectangles and (b) trapezoids.

7.3 Setting Up a Definite Integral

This section presents an informal shortcut for setting up a definite integral to evaluate some quantity. First, the formal and informal approaches are contrasted in the case of setting up the definite integral for area. Then the informal approach will be illustrated as commonly applied in a variety of fields.

The Complete Approach

Recall how the formula $A = \int_a^b f(x)\,dx$ was obtained (in Section 7.1). The interval $[a, b]$ was partitioned by the numbers $x_0 < x_1 < x_2 < \cdots < x_n$ with $x_0 = a$ and $x_n = b$. A sampling number c_i was chosen in each section $[x_{i-1}, x_i]$. For convenience, all the sections were of equal length, $\Delta x = (b-a)/n$. (See Figure 7.3.1.) We then formed the sum

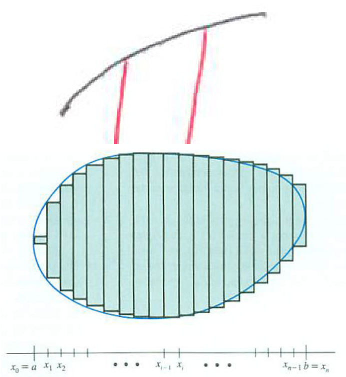


Figure 7.3.1

Figure 7.3.2 NOTE: Revise figure so not left-hand sum.

$$\sum_{i=1}^{n} f(c_i)\Delta x \tag{7.3.1}$$

It equaled the total area of the rectangular approximation in Figure 7.3.2.

As Δx approaches 0, the sum (7.3.1) approaches the area of the region under consideration. But, by the definition of the definite integral, the sum (7.3.1) approaches

$$\int_a^b f(x)\,dx.$$

Thus

$$\text{Area} = \int_a^b f(x)\,dx. \tag{7.3.2}$$

That is the complete or "formal" approach to obtain formula (7.3.2). Now consider the "informal" approach, which is just a shorthand for the complete approach.

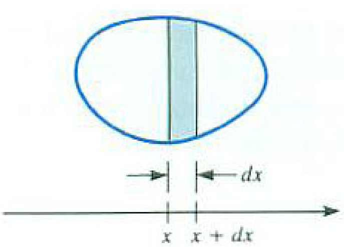


Figure 7.3.3

The Shorthand Approach

The heart of the complete approach is the *local estimate* $f(c_i)\Delta x$, the area of a rectangle of height $f(c_i)$ and width Δx, which is shown in Figure 7.3.4.

In the shorthand approach to setting up a definite integral attention is focused on the *local approximation*. No mention is made of the partition or the sampling numbers. We illustrate this shorthand approach by obtaining formula (7.3.2) informally. This is *not* a new method of integration, but just a way to save time when setting up an integral - finding out the integrand and the interval of integration.

For example, consider a small positive number dx. What would be a good estimate of the area of the region corresponding to the short interval $[x, x+dx]$ of width dx shown in Figure 7.3.3? The area of the rectangle of width dx and height $f(x)$ shown in Figure 7.3.4 would seem to be a plausible estimate. The area of this thin rectangle is

$$f(x)\ dx. \tag{7.3.3}$$

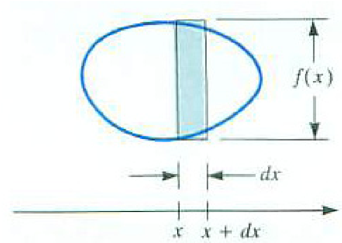


Figure 7.3.4

Without further ado, we then write

$$\text{Area} = \int_a^b f(x)\ dx, \tag{7.3.4}$$

which is formula (7.3.2). The leap from the local approximation (7.3.3) to the definite integral (7.3.4) omits many steps of the complete approach. This informal approach is the shorthand commonly used in applications of calculus. It is the way engineers, physicists, biologists, economists, and mathematicians set up integrals.

It should be emphasized that it is only an abbreviation of the formal approach, which deals with approximating sums.

The Volume of a Ball

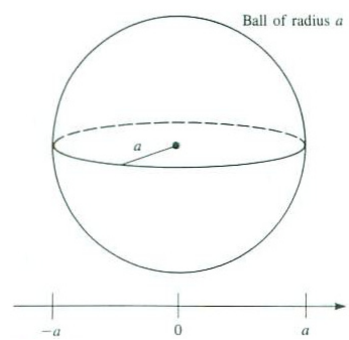


Figure 7.3.5

EXAMPLE 1 Find the volume of a ball of radius a. First use the complete approach. Then use the shorthand approach.

SOLUTION Both approaches require good diagrams. In the complete approach we show an x-axis, a partition into sections of equal lengths, sampling numbers c_i, and the approximating disks. See Figures 7.3.5 and 7.3.6(a). The thickness of a disk is Δx, as shown in the side view of Figure 7.3.6(b), while its radius is labeled r_i, as shown in the end view of Figure 7.3.6(c). The volume of this typical disk is

$$\pi r_i^2(\Delta x). \tag{7.3.5}$$

All that remains is to determine r_i. Figure 7.3.6(d) helps us do that. By the Pythagorean Theorem,

$$r_i^2 = a^2 - c_i^2. \tag{7.3.6}$$

Combining (7.3.1), (7.3.5), and (7.3.6) gives the typical estimate of the volume of a sphere of radius a:

$$\sum_{i=1}^{n} \pi(a^2 - c_i^2)\Delta x. \tag{7.3.7}$$

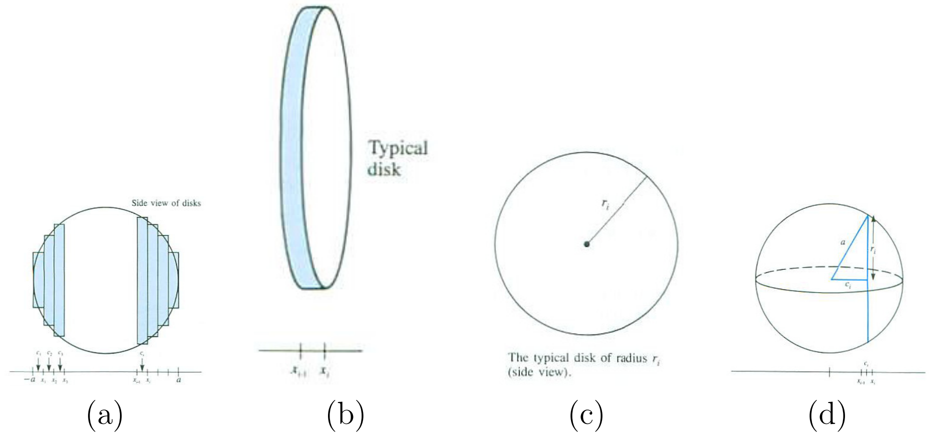


Figure 7.3.6

By the definition of the definite integral,

$$\lim_{\Delta x \to 0} \Sigma_{i=1}^{n} \pi(a^2 - c_i^2)\Delta x = \int_{-a}^{a} \pi(a^2 - x^2)\ dx.$$

Hence

$$\text{Volume of ball of radius } a = \int_{-a}^{a} \pi(a^2 - x^2)\ dx.$$

By the Fundamental Theorem of Calculus, the integral equals $4\pi a^3/3$.

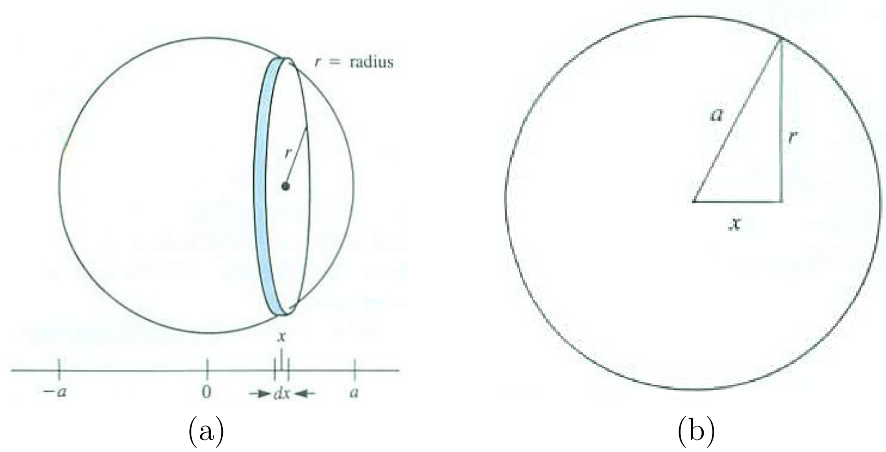


Figure 7.3.7

Now for the shorthand approach. We draw only a short section of an x-axis and label its length dx. Then we draw an approximating disk, whose radius we label r, as in Figure 7.3.7(a). Since the disk has a base of area πr^2 and thickness dx, its volume is $\pi r^2\ dx$. Moreover, as Figure 7.3.7(b) shows, $r^2 = a^2 - x^2$. Hence the local approximation is

$$\pi(a^2 - x^2)\ dx. \tag{7.3.8}$$

Then, without further ado, without choosing any c_i or showing any approximating sum, we have

$$\text{Volume of ball of radius } a = \int_{-a}^{a} \pi(a^2 - x^2)\ dx.$$

The key to this bookkeeping is the local approximation (7.3.8) in differential form, which gives the necessary integrand. The limits of integration are determined separately. ◇

Volcanic Ash

EXAMPLE 2 After the explosion of a volcano, ash gradually settles from the atmosphere and falls on the ground. The depth diminishes with distance from the volcano. Assume that the depth of the ash at a distance x feet from the volcano is Ae^{-kx} feet, where A and k are positive constants. Set up a definite integral for the total volume of ash that falls within a distance b of the volcano.

SOLUTION First estimate the volume of ash that falls on a very narrow ring of width dx and inner radius x centered at the volcano. (See Figure 7.3.8(a).) This estimate can be made since the depth of the ash depends only on the distance from the volcano. On this ring the depth is almost constant.

The area of this ring is approximately that of a rectangle of length $2\pi x$ and width dx. (See Figure 7.3.8(b)) So the area of the ring is approximately

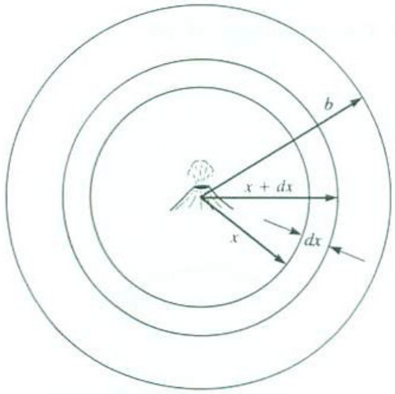
$$2\pi x\ dx.$$

Although the depth of the ash on this narrow ring is not constant, it does not vary much. A good estimate of the depth throughout the ring is Ae^{-kx}. Thus the volume of the ash that falls on the typical ring of inner radius x and outer radius $x + dx$ is approximately

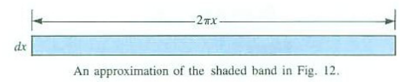
Exercise 4 shows that its area is $2\pi x\ dx + \pi(dx)^2$.

$$Ae^{-kx}(2\pi x)\ dx \text{ cubic feet.} \tag{7.3.9}$$

Once we have the key local estimate (7.3.9), we immediately write down the definite integral for the total volume of ash that falls within a distance b of the volcano:



(a)



(b)

Figure 7.3.8

$$\text{Total volume} = \int_0^b Ae^{-kx} 2\pi x \, dx.$$

The limits of integration must be determined just as in the formal approach.

This completes the shorthand setting up the definite integral. (To evaluate this integral, use a formula from the inside front cover of this book or a technique in Chapter 8.) $\diamond$

Kinetic Energy

The next example of the informal approach to setting up definite integrals concerns kinetic energy. The kinetic energy associated with an object of mass m kilograms and velocity v meters per second is defined as

$$\text{Kinetic energy} = \frac{mv^2}{2} \text{joules}.$$

If the various parts of the objects are not all moving at the same speed, an integral is needed to express the total kinetic energy. We develop this integral in the next example.

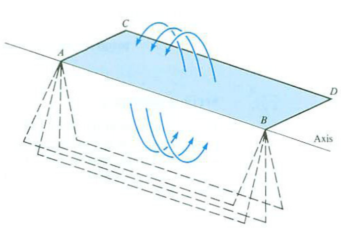


Figure 7.3.9

EXAMPLE 3 A thin rectangular piece of sheet metal is spinning around one of its longer edges 3 times per second, as shown in Figure 7.3.9. The length of its shorter edge is 6 meters and the length of its longer edge is 10 meters. The density of the sheet metal is 4 kilograms per square meter. Find the kinetic energy of the spinning rectangle.

SOLUTION The farther a mass is from the axis, the faster it moves, and therefore the larger its kinetic energy. To find the total kinetic energy of the rotating piece of sheet metal, imagine it divided into narrow rectangles of length 10 meters and width dx meters parallel to the edge $\overline{AB}$; a typical one is shown in Figure 7.3.10. (Introduce an x-axis parallel to edge $\overline{AC}$ with the origin corresponding to A.) Since all points of the typical narrow rectangle move at roughly the same speed, we will be able to estimate its kinetic energy. That estimate will provide the key local approximation in the informal approach to setting up a definite integral.

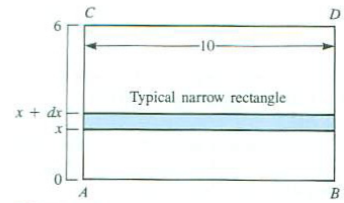


Figure 7.3.10

First of all, the mass of the typical rectangle is

$$4 \cdot 10\ dx \text{ kilograms,}$$

since its area is $10\ dx$ square meters and the density is 4 kilograms per square meter.

Second, we must estimate its velocity. The narrow rectangle is spun 3 times per second around a circle of radius x. In 1 second each point in it covers a distance of about

$$3 \cdot 2\pi x = 6\pi x \text{ meters.}$$

Consequently, the velocity of the typical rectangle is

$$6\pi x \text{ meters per second.}$$

The local estimate of the kinetic energy associated with the typical rectangle is therefore

$$\frac{1}{2} \underbrace{40\ dx}_{\text{mass}} \underbrace{(6\pi x)^2}_{\text{velocity squared}} \quad \text{joules}$$

or simply

The local approximation

$$720\pi^2 x^2\ dx \text{ joules.} \qquad (7.3.10)$$

Having obtained the local estimate (7.3.10), we jump directly to the definite integral and conclude that

$$\text{Total energy of spinning rectangle} = \int_0^6 720\pi^2 x^2\ dx \text{ joules.}$$

◇

Summary

This section presented a shorthand approach to setting up a definite integral for a quantity Q. In this method we estimate how much of the quantity Q corresponds to a very short section $[x, x+dx]$ of the x-axis, say $f(x)\ dx$. Then $Q = \int_a^b f(x)\ dx$, where a and b are determined by the particular situation.

EXERCISES for Section 7.3

1. In Section 6.4 we showed that if $f(t)$ is the velocity at time t of an object moving along the x-axis, then $\int_a^b f(t)\ dt$ is the change in position during the time interval $[a, b]$. Develop this fact in the informal style of this section. Keep in mind that $f(t)$ may be positive or negative.

2. The depth of rain at a distance r feet from the center of a storm is $g(r)$ feet.

(a) Estimate the total volume of rain that falls between a distance r feet and a distance $r + dr$ feet from the center of the storm. (Assume that dr is a small positive number.)

(b) Using (a), set up a definite integral for the total volume of rain that falls between $1,000$ and $2,000$ feet from the center of the storm.

3. Consider a disk of radius a with the home base of production at the center. Let $G(r)$ denote the density of foodstuffs (in calories per square meter) at radius r meters from the home base. Then the total number of calories produced in the range is given by what definite integral?
This analysis of primitive agriculture is taken from *Is There an Optimum Level of Population?*, edited by S. Fred Singer, McGraw-Hill, New York, 1971.

4. In Example 2 the area of the ring with inner radius x and outer radius $x + dx$ was estimated to be about $2\pi x\ dx$.

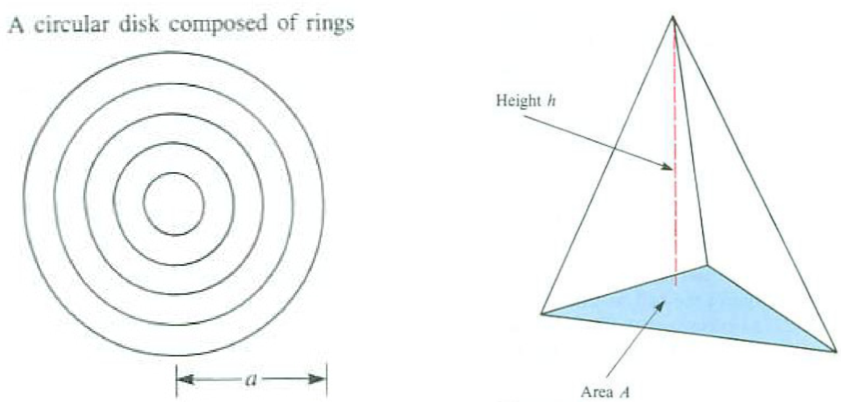
(a) Using the formula for the area of a disk, show that the area of the ring is $2\pi x\ dx + \pi(\ dx)^2$.

(b) Show that the ring has the same area as a trapezoid of height dx and bases of lengths $2\pi x$ and $2\pi(x +\ dx)$.

5. Think of a circular disk of radius a as being composed of concentric circular rings, as in Figure 7.3.11(a).

(a) Using the shorthand approach, set up a definite integral for the area of the disk. (Draw a good picture of the local approximation.)

(b) Evaluate the integral in (a).



(a) (b)

Figure 7.3.11

Exercises 6 to 8 concern the volumes of solids. In each case (a) draw a good picture of the local approximation of width dx, (b) set up the appropriate definite integral, and (c) evaluate the integral.

6. A right circular cone of radius a and height h.

7. A pyramid with a square base of side a and of height h. Its top vertex is above one corner of the base. (Use square cross sections.)

8. A pyramid with a triangular base of area A and of height h. (The triangle can be any shape. See Figure 7.3.11(b).)

9. At the time t hours, $0 \le t \le 24$, a firm uses electricity at the rate of $e(t)$ joules per hour. The rate schedule indicates that the cost per joule at time t is $c(t)$ dollars. Assume that both e and c are continuous functions.

(a) Estimate the cost of electricity consumed between times t and $t + dt$, where dt is a small positive number.

(b) Using (a), set up a definite integral for the total cost of electricity for the 24-hour period.

10. The **present value** of a promise to pay one dollar t years from now is $g(t)$ dollars.

(a) What is $g(0)$?

(b) Why is it reasonable to assume that $g(t) \le 1$ and that g is a decreasing function of t?

(c) What is the present value of a promise to pay q dollars t years from now?

(d) Assume that an investment made now will result in an income flow at the rate of $f(t)$ dollars per year t years from now. (Assume that f is a continuous function.) Estimate informally the present value of the income to be earned between time t and time $t + dt$, where dt is a small positive number.

(e) On the basis of the local estimate made in (d), set up a definite integral for the present value of all the income to be earned during the next b years.

11. Let the number of females in a certain population in the age range from x years to $x + dx$ years, where dx is a small positive number, be approximately $f(x)\,dx$. Assume that, on average, women of age x produce $m(x)$ offspring during the year before they reach age $x + 1$. Assume that both f and m are continuous functions.

(a) What definite integral represents the number of women between ages a and b years?

(b) What definite integral represents the total number of offspring during the calendar year produced by women whose ages at the beginning of the calendar year were between a and b years?

Exercises 12 to 17 concern **kinetic energy**. They are all based on the concept that a particle of mass m moving with velocity v has the kinetic energy $mv^2/2$. (See Example 3.) An object whose density is the same at all its points is called **homogeneous**. If the object is planar, such as a square or disk, and has mass m kilograms and area A square meters, its density is m/A kilograms per square meter.

12. The piece of sheet metal in Example 3 is rotated around the line midway between the edges AB and CD at the rate of 5 revolutions per second.

(a) Using the informal approach, obtain a local approximation for the kinetic energy of a narrow strip of the metal.

(b) Using (a), set up a definite integral for the kinetic energy of the piece of sheet metal.

(c) Evaluate the integral in (b).

13. A circular piece of metal of radius 7 meters has a density of 3 kilograms per square meter. It rotates 5 times per second around an axis perpendicular to the circle and passing through the center of the circle.

(a) Devise a local approximation for the kinetic energy of a narrow ring in the circle.

(b) With the aid of (a), set up a definite integral for the kinetic energy of the rotating metal.

(c) Evaluate the integral in (b).

14. The density of a rod x centimeters from its left end is $g(x)$ grams per centimeter. The rod has a length of b centimeters. The rod is spun around its left end 7 times per second.

(a) Estimate the mass of the rod in the section that is between x and $x + dx$ centimeters from the left end. (Assume that dx is small.)

(b) Estimate the kinetic energy of the mass in (a).

(c) Set up a definite integral for the kinetic energy of the rotating rod.

15. A homogeneous square of mass m kilograms and side a meters rotates around an edge 5 times per second.

(a) Obtain a local estimate of the kinetic energy. What part of the square would you use? Why? Draw it.

(b) What is the local estimate?

(c) What definite integral represents the total kinetic energy of the square?

(d) Evaluate it.

16. Repeat Exercise 15 for a square spun around a line through its center and parallel to an edge.

17. Repeat Exercise 15 for a disk of radius a and mass M spinning around a line through its center and perpendicular to it. It is spinning at the rate of ω radians per second. (See Figure 7.3.12.)

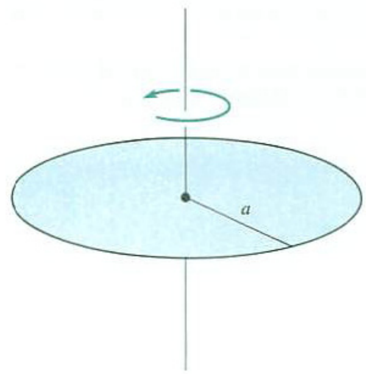


Figure 7.3.12

In Exercises 18 and 19 you will meet definite integrals that cannot be evaluated by the Fundamental Theorem of Calculus (since the desired antiderivative is not elementary). Use (a) the trapezoidal and (b) Simpson's method with six sections to estimate the definite integrals.

18. A homogeneous object of mass M occupies the region under $y = e^{x^2}$ and above $[0, 1]$. It is spun at the rate of ω radians per second around the y-axis. Estimate its kinetic energy.

19. A homogeneous object of mass M occupies the region under $y = \sin(x)/x$ and above $[\pi/2, \pi]$. It is spun around the line $x = 1$ at the rate of ω radians per second. Estimate its kinetic energy.

In each of Exercises 20 to 23, find the kinetic energy of a planar homogeneous object that occupies the given region, has mass M, and is spun around the y-axis ω radians per second.

20. The region under $y = e^x$ and above the interval $[1, 2]$.

21. The region under $y = \arctan(x)$ and above the interval $[0, 1]$.

22. The region under $y = 1/(1 + x)$ and above $[2, 4]$.

23. The region under $y = \sqrt{1 + x^2}$ and above $[0, 2]$.

24. A solid homogeneous right circular cylinder of radius a, height h, and mass M is spun at the rate of ω radians per second around its axis. Find its kinetic energy. (Include a good picture on which your local approximation is based.)

25. A solid homogeneous ball of radius a and mass M is spun at the rate of ω radians per second around a diameter. Find its kinetic energy. (Include a good picture on which your local approximation is based.)

26. Find the surface area of a sphere of radius a. (Begin by estimating the area of the narrow band shown in Figure 7.3.13.)

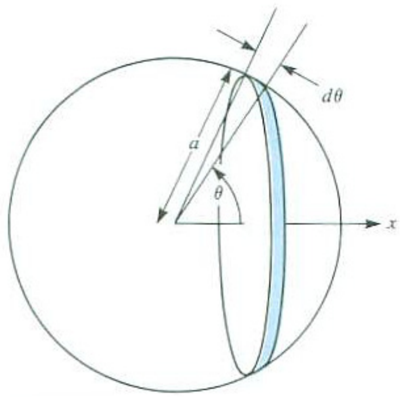


Figure 7.3.13

27. *[Actuarial tables]* Let $F(t)$ be the fraction of people born in 1900 who are alive t years later, $0 \leq F(t) \leq 1$.

(a) What is $F(150)$, probably?

(b) What is $F(0)$?

(c) Sketch the general shape of the graph of $y = F(t)$.

(d) Let $f(t) = F'(t)$. (Assume F is differentiable.) Is $f(t)$ positive or negative?

(e) What fraction of the people born in 1900 die during the time interval $[t, t+dt]$? (Express your answer in terms of F.)

(f) Answer (e), but express your answer in terms of f.

(g) Evaluate $\int_0^{150} f(t)\ dt$.

(h) What integral would you propose to call "the average life span of the people born in 1900"? Why?

28. Let $F(t)$ be the fraction of ball bearings that wear out during the first t hours of use. Thus $F(0) = 0$ and $F(t) \leq 1$.

(a) As t increases, what would you think happens to $F(t)$?

(b) Show that during the short interval of time $[t, t+\ dt]$, the fraction of ball bearings that wear out is approximately $F'(t)\ dt$. (Assume F is differentiable.)

(c) Assume all wear out in at most $1,000$ hours. What is $F(1,000)$?

(d) Using the assumption in (b) and (c) devise a definite integral for the average life of the ball bearings.

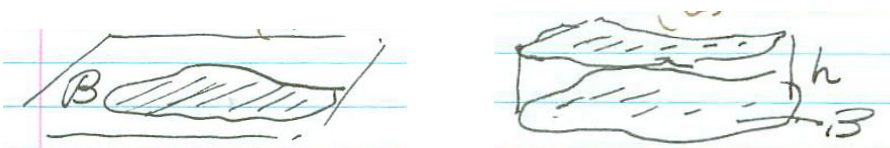
29. The density of the earth at a distance of r miles from its center is $g(r)$ pounds per cubic mile. Set up a definite integral for the total mass of the earth. (Take the radius of the earth to be $4,000$ miles.)

7.4 Computing Volumes by Parallel Cross-Sections

In Section 6.1 we computed areas by integrating lengths of cross sections made by parallel lines. In this section we will use a similar approach, finding volumes by integrating areas of cross sections made by parallel planes. We already saw an example of this method when we represented the volume of a tent as a definite integral.

See Problem 3 in Section 6.1.

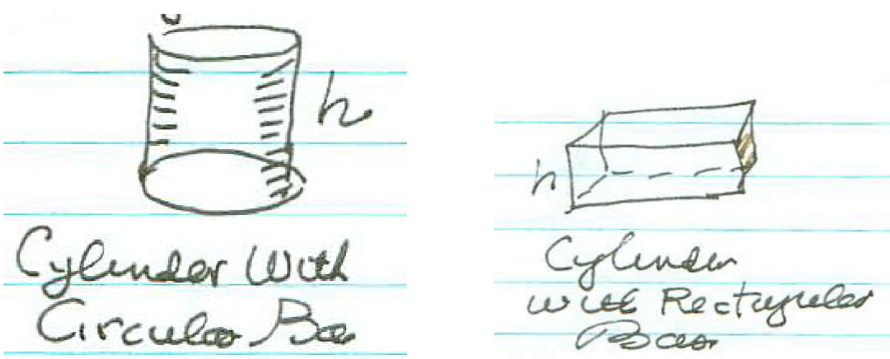
Cylinders



(a) (b)

Figure 7.4.1

Let $\mathcal{B}$ be a region in the plane (see Figure 7.4.1(a) and h a positive number. The **cylinder with base** $\mathcal{B}$ **and height** h consists of all line segments of length h perpendicular to $\mathcal{B}$, one end of which is in $\mathcal{B}$ and the other end is on a fixed side (above or below) of $\mathcal{B}$. This typical cylinder is shown in Figure 7.4.1(b). The top of the cylinder is congruent to $\mathcal{B}$. If $\mathcal{B}$ is a disk, the



(a) (b)

Figure 7.4.2 ARTIST: Final word in each caption is "Base"

cylinder is the customary circular cylinder of daily life (see Figure 7.4.2(a)). If $\mathcal{B}$ is a rectangle, the cylinder is a rectangular box (see Figure 7.4.2(b)).

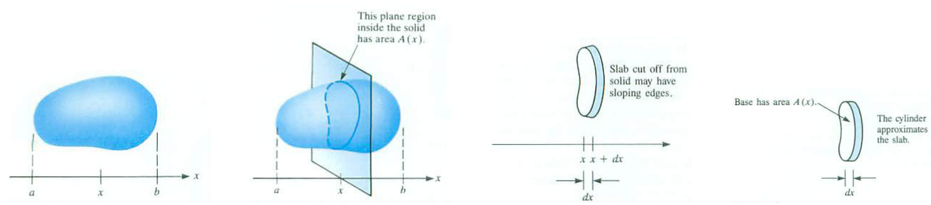
We will make use of the formula for the volume of a cylinder:

> The volume of a cylinder with base $\mathcal{B}$ and height h is
>
> $$V = \text{Area of Base} \times \text{Height} = (\text{Area of } \mathcal{B}) \times h.$$

Volume as the Definite Integral of Cross-Sectional Area

Let's use the informal approach for setting up a definite integral to see how to use integration to calculate volumes of solids.

Consider the solid region $\mathcal{R}$ shown in Figure 7.4.3(a), which lies between the planes perpendicular to the x-axis at $x = a$ and at $x = b$. We use a cylinder to estimate the volume of the part of $\mathcal{R}$ that lies between two parallel planes a "small distance" dx apart, shown in perspective in Figure 7.4.3(b). This thin



(a) (b) (c) (d)

Figure 7.4.3

slab is not usually a cylinder (Figure 7.4.3(c)). However, we can approximate it by a cylinder. To do this, let x be, say, the left endpoint of an interval of width dx. The plane perpendicular to the x-axis at x intersects $\mathcal{R}$ in a plane cross section of area $A(x)$. The cylinder whose base is that cross section and whose height is dx is a good approximation of the part of $\mathcal{R}$ between the planes corresponding to x and $x + dx$. It is the slab shown in Figure 7.4.3(d).

We therefore have

$$\text{Local Approximation to Volume} = A(x)dx.$$

Then

> $$\text{Volume of Solid} = \int\limits_a^b A(x)\ dx.$$

In short, "volume equals the integral of cross-sectional area." To apply this idea, we compute $A(x)$. That is a where good drawings come in handy.

Given a particular solid, one just has to find a, b and the cross-sectional area $A(x)$ in order to construct a definite integral for its volume. These are the steps for finding the volume of a solid:

1. Choose a line to serve as an x-axis.

See Figure 7.4.3(a).

2. For each plane perpendicular to that axis, find the area of the cross section of the solid made by the plane. Call this area $A(x)$.

See Figure 7.4.3(b).

3. Determine the limits of integration, a and b, for the region.

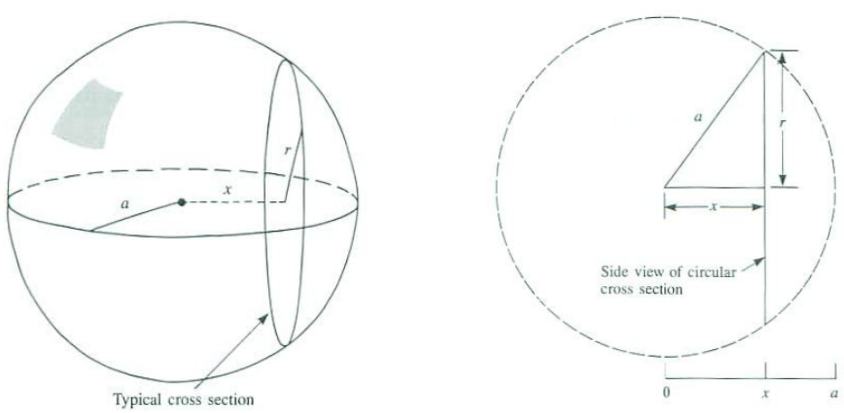
4. Evaluate the definite integral $\int_a^b A(x)\,dx$.

Most of the effort is usually spent in finding the integrand $A(x)$.

In addition to the Pythagorean Theorem and properties of similar triangles, formulas for the areas of familiar plane figures may be needed. Also keep in mind that if corresponding dimensions of similar figures have a ratio k, then their areas have the ratio k^2; that is, area is proportional to the square of the ratios of the lengths of corresponding line segments.

Formulas for the area of familiar plane regions are on the inside back cover.

EXAMPLE 1 Find the volume of a ball of radius a.



(a) (b)

Figure 7.4.4 Cross-section (a) viewed in perspective and (b) from the side.

Archimedes was the first person to find the volume of a ball. He did not express the volume as a number. Rather, in the style of mathematics of the 3rd century BC, he expressed the volume in terms of the volume of a simpler object: the volume of a ball is two-thirds the volume of the smallest cylinder that contains it. That he considered this one of his greatest accomplishments is evidenced by his request that his tomb be topped with a carving of a ball within a cylinder.

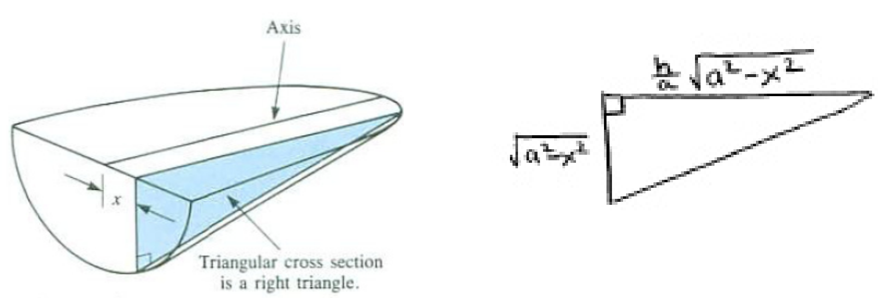
SOLUTION We sketch the typical cross section in perspective and in side view (see Figure 7.4.4). The cross section is a disk of radius r, which depends on x. The area of the cross section is πr^2. To express this area in terms of x, use the Pythagorean Theorem, which tells us that $a^2 = x^2 + r^2$, hence $r^2 = a^2 - x^2$. So we have

$$\begin{aligned} \text{Volume} &= \int_{-a}^{a} \pi(a^2 - x^2)\, dx = \pi \left(a^2x - \frac{x^3}{3} \right)\Big|_{-a}^{a} && \text{by FTC I} \\ &= \pi \left(\left(a^3 - \frac{a^3}{3} \right) - \left((-a)^3 - \frac{(-a)^3}{3} \right) \right) = \frac{4\pi}{3} a^3. \end{aligned}$$

$\diamond$

The next example concerns the solid region discussed in Example 3 of Section 7.2.

EXAMPLE 2 A cylindrical glass of height h and radius a is full of water. It is tilted until the remaining water covers exactly half the base. Find the volume of the remaining water.
SOLUTION We use the triangular cross section shown in Figure 7.2.6.



(a) (b)

Figure 7.4.5

Introduce the x-axis as in Figure 7.4.5. It was shown that the area of the cross section at x is $\frac{1}{2}\frac{h}{a}(a^2 - x^2)$. Thus,

$$\begin{aligned} \text{Volume} &= \int_{-a}^{a} \frac{h}{2a}(a^2 - x^2)\, dx = \frac{h}{2a} \left(a^2x - \frac{x^3}{3} \right)\Bigg|_{-a}^{a} && \text{by FTC I} \\ &= \frac{h}{2a} \left(\left(a^3 - \frac{a^3}{3} \right) - \left(-a^3 + \frac{a^3}{3} \right) \right) = \frac{h}{2a} \left(\frac{4}{3} a^3 \right) = \frac{2}{3} h a^2. \end{aligned}$$

That's about 21% of the volume of the glass.

This calculation of the integral could be simplified by noting that the integrand is an even function (the volume to the right of 0 equals the volume to the left of 0). In this method we have

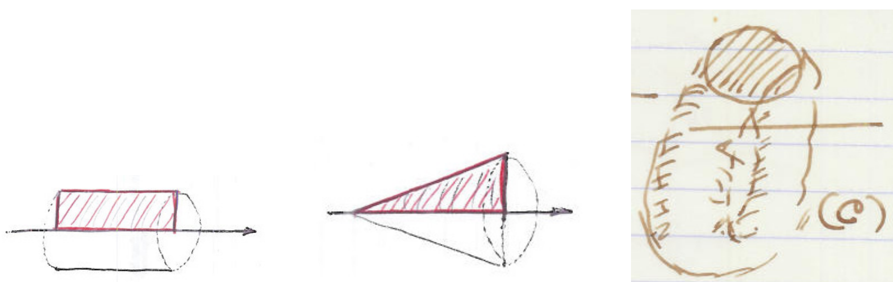
There is a much less chance for arithmetical error in this calculation.

$$\begin{aligned}\text{Volume} &= 2\int_0^a \frac{h}{2a}\left(a^2 - x^2\right)\,dx = \frac{h}{a}\left(a^2x - \frac{x^3}{3}\right)\Big|_0^a \\ &= \tfrac{h}{a}\left(\left(a^3 - \tfrac{a^3}{3}\right) - (0-0)\right) = \tfrac{2}{3}ha^2\end{aligned}$$

The two solutions yield the same result. The second way avoids a lot of arithmetic with negative numbers, thus reducing the chance of making a mistake. ◇

Solids of Revolution

The solid formed by revolving a region $\mathcal{R}$ in the plane about a line in that plane that does not intersect the interior of $\mathcal{R}$ is called a **solid of revolution**.



(a) (b) (c)

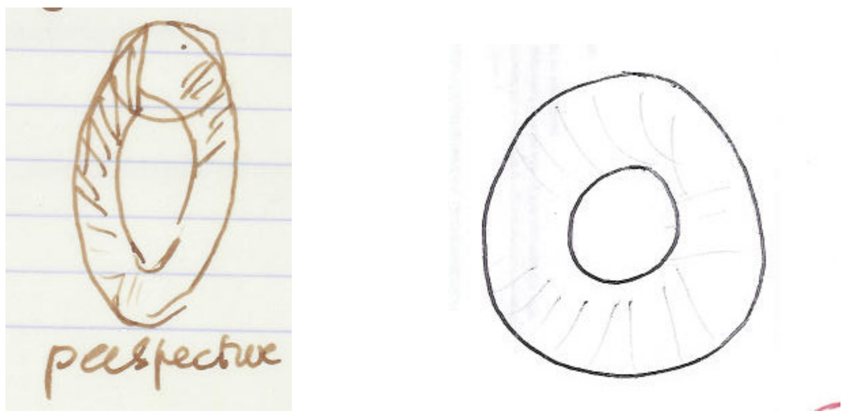
Figure 7.4.6

Figure 7.4.6 shows three examples: (a) a circular cylinder obtained by revolving a rectangle about one of its edges, (b) a cone obtained by revolving a right triangle about one of its two legs, and (c) a torus ("doughnut" or "ring") formed by revolving a disk about a line outside the disk.

The cross sections by planes perpendicular to the line around which the figure is revolved is either a disk or a "washer". The latter is a disk with a round hole. The cross sections in Figure 7.4.6(a) and (b) are disks. In Figure 7.4.6(c) the cross sections are washers. Figure 7.4.7 shows that the typical cross section is a washer.

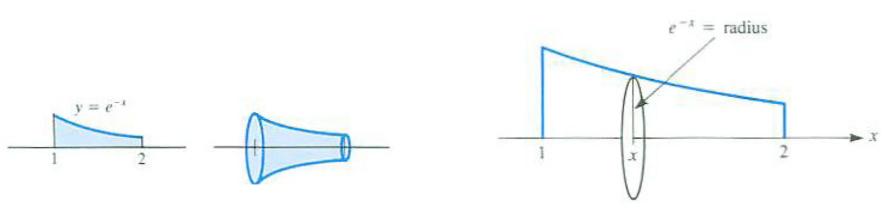
EXAMPLE 3 The region under $y = e^{-x}$ and above $[1, 2]$ is revolved about the x-axis. Find the volume of the resulting solid of revolution. (See Figure 7.4.8(a).)

SOLUTION The typical cross section by a plane perpendicular to the x-axis



(a) (b)

Figure 7.4.7 (a) perspective (b) side view



(a) (b)

Figure 7.4.8

is a disk of radius e^{-x}, as shown in Figure 7.4.8(b). The cross-sectional area is

$$\pi\left(e^{-x}\right)^2 = \pi e^{-2x}.$$

The volume of the solid is therefore

$$\int_1^2 \pi e^{-2x}\, dx.$$

Recall that $\frac{d}{dx}\left(e^{ax}\right) = ae^{ax}$, so that an antiderivative of e^{ax} is $\frac{1}{a}e^{ax}$. Hence,

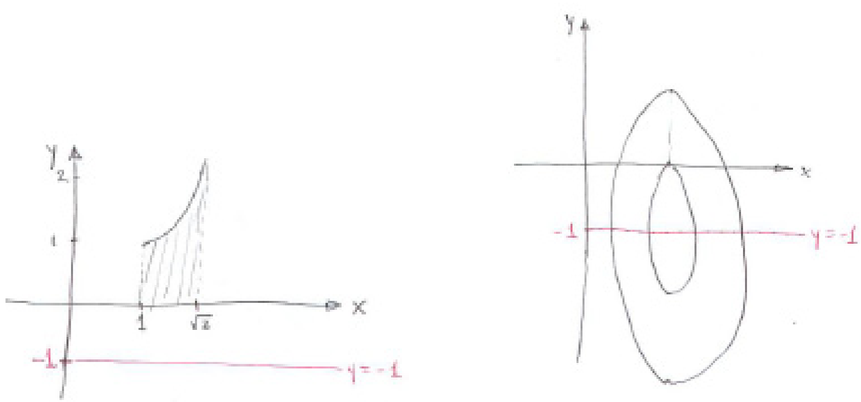
$$\int_1^2 \pi e^{-2x}\, dx \quad = \quad \frac{\pi}{-2}e^{-2x}\Big|_1^2 = \frac{\pi}{-2}\left(e^{-4} - e^{-2}\right) = \frac{\pi}{2}\left(e^{-2} - e^{-4}\right).$$

◇

The final two examples illustrate two themes: draw a good picture of the cross section and integrate the cross-sectional area.

EXAMPLE 4 The region bounded by $y = x^2$, the lines $x = 1$ and $x = \sqrt{2}$, and the x-axis ($y = 0$). is revolved around the line $y = -1$. Find the volume of the resulting region $\mathcal{R}$.

SOLUTION Figure 7.4.9(a) shows the region being revolved and the line around which it is revolved. Figure 7.4.9(b) shows a perspective view of the typical cross section.



(a) (b)

Figure 7.4.9

The typical cross section is a washer, with inner radius 1 and outer radius $1 + x^2$. Its area is therefore $\pi(1 + x^2)^2 - \pi(1)^2$.

Consequently, since "volume equals integral of cross-sectional area,"

$$\begin{aligned}
\text{Volume} &= \int_1^{\sqrt{2}} \left(\pi(1+x^2)^2 - \pi(1)^2\right)\ dx \\
&= \pi \int_1^{\sqrt{2}} (1+2x^2+x^4-1)\ dx && \text{algebra} \\
&= \pi \int_1^{\sqrt{2}} (2x^2+x^4)\ dx \\
&= \pi \left(\frac{2x^3}{3}+\frac{x^5}{5}\right)\Big|_1^{\sqrt{2}} && \text{FTC I} \\
&= \pi \left(\frac{32\sqrt{2}}{15}-\frac{13}{15}\right) && \text{arithmetic.}
\end{aligned}$$

◇

EXAMPLE 5 Find the volume of the solid formed by revolving the region in Figure 7.4.9(a) around the y-axis ($x = 0$).

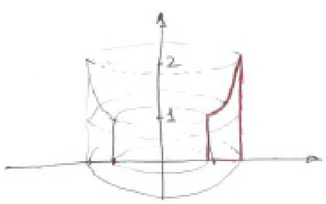


Figure 7.4.10

SOLUTION The cross sections by planes perpendicular to the y-axis are again washers (not disks). But something new enters the scene. For $0 \le y \le 1$ the cross sections are between the vertical lines $x = 1$ and $x = \sqrt{2}$. For $1 \le y \le 2$ they are determined by the curve and the line $x = \sqrt{2}$. (See Figure 7.4.10.)

The cross sections for $0 \le y \le 1$, when rotated about the y-axis, fill out a cylinder whose height is 1 and whose base is a washer of area $\pi(\sqrt{2})^2 - \pi(1)^2 = \pi$. Thus, its volume (height times area of base) is $\pi(1) = \pi$. We did not need an integral for this.

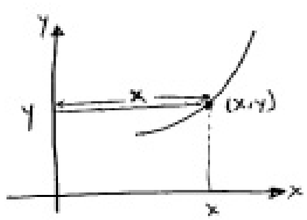


Figure 7.4.11

The cross sections for $1 \le y \le \sqrt{2}$ are washers whose outer radius is $\sqrt{2}$ and inner radius is determined by the curve $y = x^2$, as shown in Figure 7.4.11. Since $y = x^2$, the inner radius is $x = \sqrt{y}$. The area of these typical cross sections is

$$\pi(\sqrt{2})^2 - \pi(\sqrt{y})^2.$$

Thus the typical local estimate of volume is

$$\left(\pi(\sqrt{2})^2 - \pi(\sqrt{y})^2\right)\ dy = (2\pi - \pi y)\ dy.$$

Therefore the volume swept out by these cross sections is

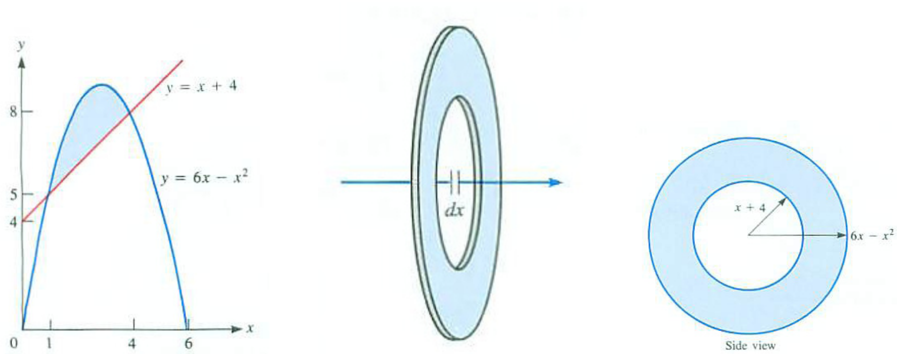
$$\begin{aligned}
\int_1^{\sqrt{2}} (2\pi - \pi y)\ dy &= \left(2\pi y - \pi\frac{y^2}{2}\right)\Big|_1^{\sqrt{2}} && \text{FTC I} \\
&= \left(2\pi\sqrt{2} - \pi\right) - \left(2\pi - \frac{\pi}{2}\right) \\
&= 2\pi\sqrt{2} - \frac{5}{2}\pi.
\end{aligned}$$

Adding this to the volume obtained for the cylinder gives

$$\begin{aligned}
\text{total volume} &= \left(2\pi\sqrt{2} - \frac{5}{2}\pi\right) + \pi \\
&= 2\pi\sqrt{2} - \frac{3}{2}\pi \approx 4.1734.
\end{aligned}$$

$\diamond$

EXAMPLE 6 The region bounded by the graphs of $y = x + 4$ and $y = 6x - x^2$, shown in Figure 7.4.12(a), is revolved about the x-axis to form a solid of revolution. Express the volume as a definite integral.



(a) (b) (c)

Figure 7.4.12

SOLUTION We first draw a local approximation to a thin slice of the solid (see Figure 7.4.12(b)). The side view in Figure 7.4.12(c) shows the area of the typical cross section is

$$\pi\left(6x - x^2\right)^2 - \pi\left(x + 4\right)^2.$$

This is the integrand. Next we find the interval of integration. The ends of the interval are determined by where the curves cross: when $x + 4 = 6x - x^2$. Moving all terms to the left-hand side yields: $x^2 - 5x + 4 = 0$, or $(x-1)(x-4) = 0$. So the endpoints of the interval are $x = 1$ and $x = 4$. The volume of the solid is given by the definite integral

$$\int_1^4 \left(\pi\left(6x - x^2\right)^2 - \pi\left(x + 4\right)^2\right)\,dx.$$

$\diamond$

Summary

The key idea in this section is that "volume is the definite integral of cross-sectional area". To implement this idea we have to find that varying area and also the interval of integration. A solid of revolution, where the cross section may be a disk or a washer, is just a special case.

EXERCISES for Section 7.4

In Exercises 1 to 8, (a) draw the solid, (b) draw the typical cross section in perspective and side view, (c) find the area of the typical cross section, (d) set up the definite integral for the volume, and (e) evaluate the definite integral (if possible).

1. Find the volume of a cone of radius a and height h.

2. The base of a solid is a disk of radius 3. Each plane perpendicular to a given diameter meets the solid in a square, one side of which is in the base of the solid. (See Figure 7.4.13(a).) Find its volume.

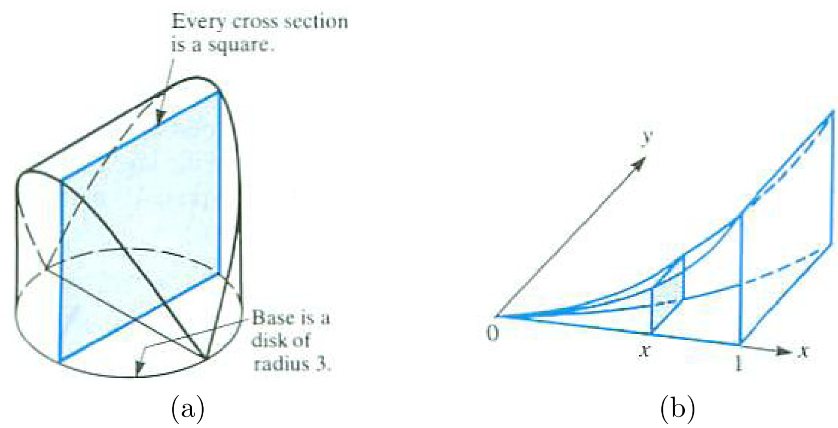


Figure 7.4.13

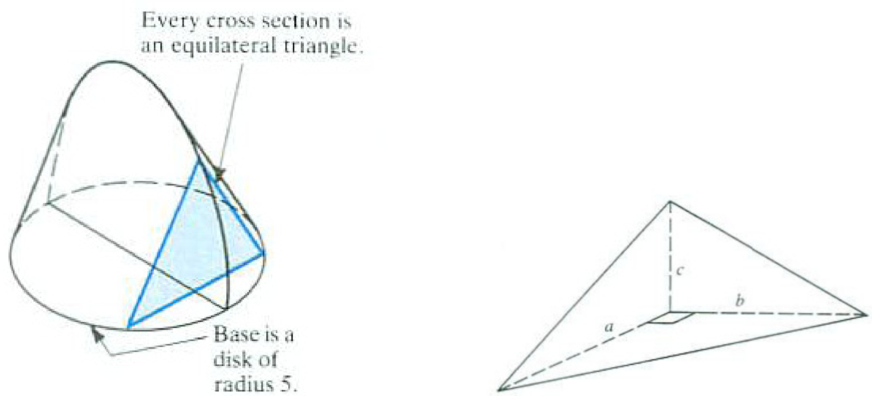
3. The base of a solid is the region bounded by $y = x^2$, the line $x = 1$, and the x- and y-axes. Each cross section perpendicular to the x-axis is a square. (See Figure 7.4.13(b).) Find the volume of the solid.

4. Repeat Exercise 3 except that the cross sections perpendicular to the base are equilateral triangles.

5. Find the volume of a pyramid with a square base of side a and height h, using square cross sections perpendicular to the base. The top of the pyramid is above the center of the base.

6. Repeat Exercise 5, but using trapezoidal cross sections perpendicular to the base.

7. Find the volume of the solid whose base is the disk of radius 5 and whose cross sections perpendicular to a diameter are equilateral triangles. (See Figure 7.4.14(a).)



(a) (b)

Figure 7.4.14

8. Find the volume of the pyramid shown in Figure 7.4.14(b) by using cross sections perpendicular to the edge of length c.

In Exercises 9 to 14 set up a definite integral for the volume of the solid formed by revolving the given region R about the given axis.

9. R is bounded by $y = \sqrt{x}$, $x = 1$, $x = 2$, and the x-axis, about the x-axis.

10. R is bounded by $y = \frac{1}{\sqrt{1+x^2}}$, $x = 0$, $x = 1$, and the x-axis, about the x-axis.

11. R is bounded by $y = x^{-1/2}$, $y = x^{-1}$, $x = 1$, and $x = 2$, about the x-axis.

12. R is bounded by $y = x^2$ and $y = x^3$, about the y-axis.

13. R is bounded by $y = \tan(x)$, $y = \sin(x)$, $x = 0$, and $x = \pi/4$, about the x-axis.

14. R is bounded by $y = \sec(x)$, $y = \cos(x)$, $x = \pi/6$, and $x = \pi/3$, about the x-axis.

15. A cylindrical drinking glass of height h and radius a, full of water, is tilted until the water just covers the base. Set up a definite integral that represents the amount of water left in the glass. Use rectangular cross sections. Refer to Figure 7.4.15 and follow the directions preceding Exercise 1.

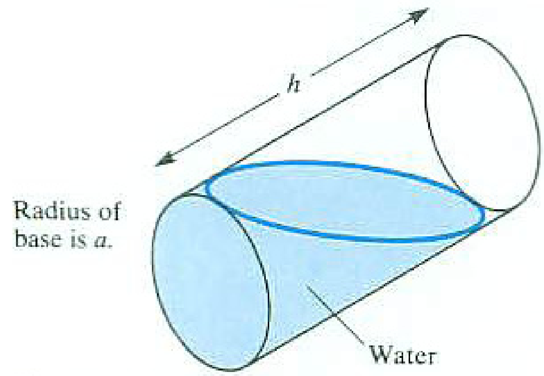


Figure 7.4.15

16. Repeat Exercise 15, but use trapezoidal cross sections.

17. Repeat Exercise 15 using only common sense. Don't use any calculus.

18. A cylindrical drinking glass of height h and radius a, full of water, is tilted until the water remaining covers half the base.

(a) Set up a definite integral for the volume of water in the glass, using cross sections that are parts of disks.

(b) Compare yours answer in (a) with the definite integral found in Example 2. Which definite integral looks easier to evaluate?

19. Repeat Exercise 18, but use rectangular cross sections.

20. A solid is formed in the following manner. A plane region R and a point P not in the plane are given. The solid consists of all line segments joining P to points in R. If R has area A and P is a distance h from the plane R, show that the volume of the solid is $Ah/3$. (See Figure 7.4.16.)

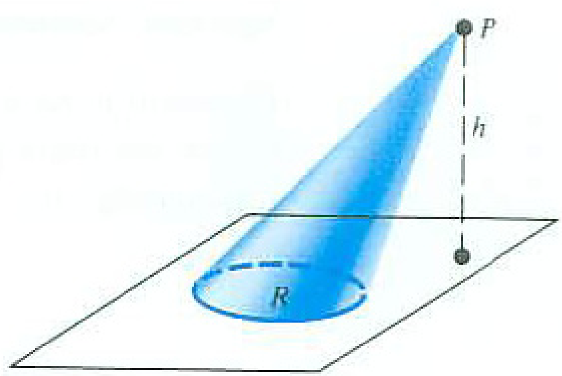


Figure 7.4.16

21. A drill of radius 4 inches bores a hole through a wooden sphere of radius 5 inches, passing symmetrically through the center of the sphere.

(a) Draw the part of the sphere removed by the drill.

(b) Find $A(x)$, the area of a cross section of the region in (a) made by a plane perpendicular to the axis of the drill and at a distance x from the center of the sphere.

(c) Set up the definite integral for the volume of wood removed.

22. What fraction of the volume of a sphere is contained between parallel planes that trisect the diameter to which they are perpendicular? (Leave your answer in terms of a definite integral.)

23. The disk bounded by the circle $(x-b)^2+y^2=a^2$, where $0<a<b$, is revolved around the y-axis. Set up a definite integral for the volume of the doughnut (torus) produced.

In Exercises 24 to 27 set up definite integrals for (a) the area of R, (b) the volume formed when R is revolved around the x-axis, and (c) the volume formed when R is revolved around the y-axis.

24. R is the region under $y=\tan(x)$ and above the interval $[0,\pi/4]$.

25. R is the region under $y=e^x$ and above the interval $[-1,1]$.

26. R is the region under $y=1/\sqrt{1-x^2}$ and above the interval $[0,1/2]$.

27. R is the region under $y=\sin(x)$ and above the interval $[0,\pi]$.

28. Set up a definite integral for the volume of one octant of the region common to two right circular cylinders of radius 1 whose axes intersect at right angles, as shown in Figure 7.4.17. Contributed by Archimedes.

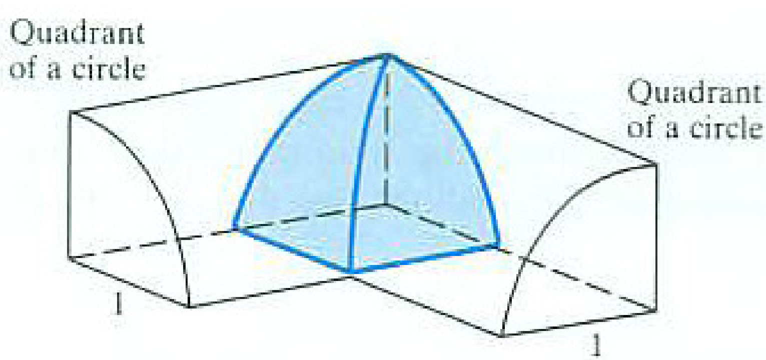


Figure 7.4.17

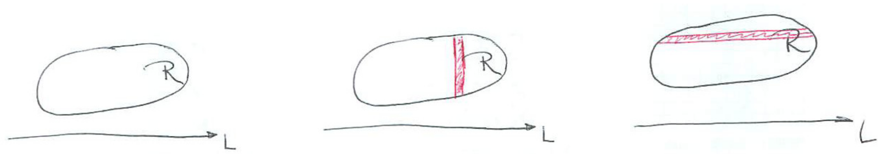
29. When a convex region R of area A situated to the right of the y-axis is revolved around the y-axis, the resulting solid of revolution has volume V. When R is revolved around the line $x=-k$, the volume of the resulting solid is V^*. Express V^* in terms of k, A, and V. The definition of convex can be found on page 131 in Section 2.5.

30. Archimedes viewed a ball as a cone whose height is the radius of the ball and whose base is the surface of the ball. On that basis he computed that the volume of the ball is one third the product of the radius and the surface area. He then gave a rigorous proof of his conjecture.

Clever Sam, inspired by this, said "I'm going to get the volume of a circular cylinder in a new way. Say its radius is r and height is h. Then I'll view it as a cylinder made up of "r by h" rectangles, all of which have the axis as an edge. Then I pile them up to make a box whose base is an r by h rectangle and whose height is $2\pi r$ (the circumference of the cylinder's base). So the volume would be $2\pi r$ times rh, or $2\pi r^2 h$. That's twice the usual volume, so the standard formula is wrong." Is Sam right? (Explain.)

7.5 Computing Volumes by Shells

Imagine revolving the planar region $\mathcal{R}$ about the line L, as in Figure 7.5.1(a). We may think of $\mathcal{R}$ as being formed from narrow strips perpendicular to L, as in Figure 7.5.1(b). Revolving such a strip around L produces a washer (or disk). This is the approach used in the preceding section.



(a) (b) (c)

Figure 7.5.1

However, we can also think of $\mathcal{R}$ as being formed from narrow strips *parallel* to L, as in Figure 7.5.1(c). Revolving such a strip around L produces a solid shaped like a bracelet or part of a drinking straw, as shown, in perspective, in Figure 7.5.2. We will call such a solid a **shell**. (Perhaps "tube" or "pipe" might be a better choice, but "shell" is standard in the world of calculus.)

This section describes how to find the volume of a solid of revolution using shells (instead of disks). Sometimes this approach provides an easier calculation.

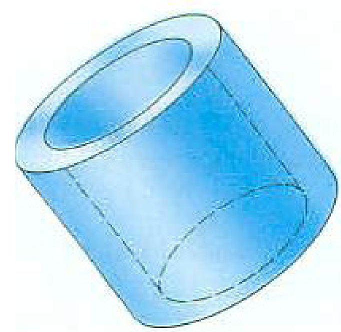
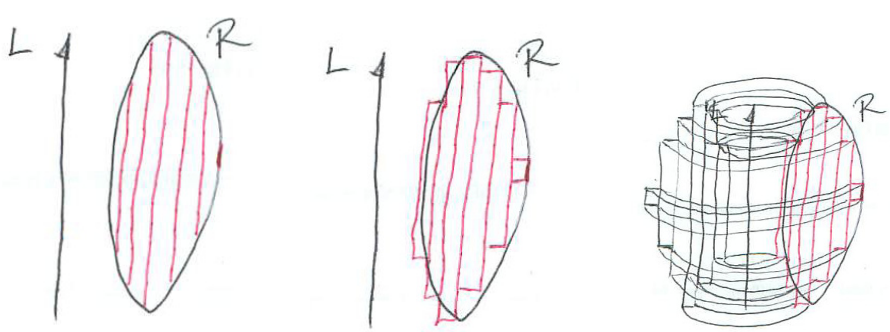


Figure 7.5.2

The Shell Technique

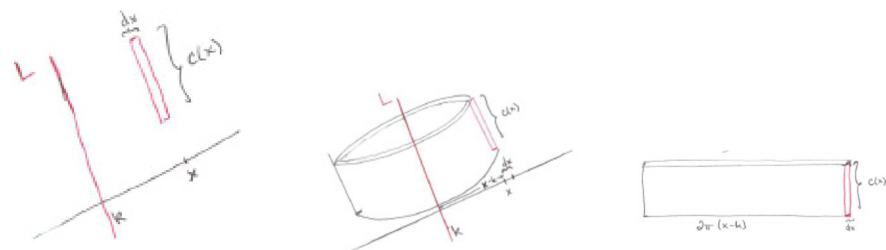


(a) (b) (c)

Figure 7.5.3

To apply the shell technique we first imagine cutting the plane region $\mathcal{R}$ in Figure 7.5.3(a) into a finite number of narrow strips by lines parallel to L. Each strip is then approximated by a rectangle of width dx as in Figure 7.5.3(b). Then we approximate the solid of revolution by a collection of tubes (like the parts of a collapsible telescope), as in Figure 7.5.3(c).

The key to this method is estimating the volume of each shell. Figure 7.5.4(a)



(a) (b) (c)

Figure 7.5.4

shows the typical local approximation. Its height, $c(x)$, is the length of the cross section of $\mathcal{R}$ corresponding to the value x on a line that we will call the x-axis. The radius of the shell, shown in Figure 7.5.4(b), is $x - k$, where k is the x-coordinate of the equation of the axis of rotation. Imagine cutting the shell along a direction parallel to L, unrolling it, and then laying it flat like a carpet. When laid flat, the shell resembles a thin slab of thickness dx, width $c(x)$, and length $2\pi(x - k)$, as shown in Figure 7.5.4(c). The volume of this shell, therefore, is about

The exact volume of the shell is found in Exercise 23.

$$\text{Local Approximation to Volume of a Shell} = 2\pi(x-k)c(x)\ dx \qquad (7.5.1)$$

With the aid of the local approximation (7.5.1), we conclude that

$$\text{Volume of Solid of Revolution} = \int_a^b 2\pi(x-k)c(x)\ dx. \qquad (7.5.2)$$

If $x - k$ is denoted $R(x)$, the "radius of the shell," as in Figure 7.5.5, then

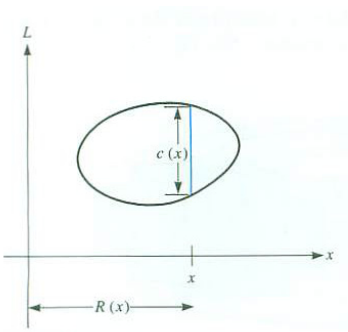
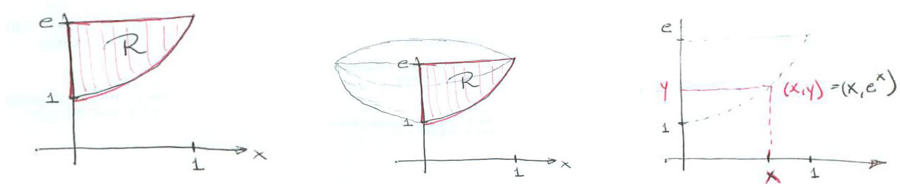


Figure 7.5.5 ARTIST: Add k as label on x-axis (at origin)

$$\text{Volume of Solid of Revolution} = \int_a^b 2\pi R(x)c(x)\, dx.$$

EXAMPLE 1 The region $\mathcal{R}$ below the line $y = e$, above $y = e^x$, and to the right of the y-axis is revolved around the y-axis to produce a solid $\mathcal{S}$. Set up the definite integrals for the volume of $\mathcal{S}$ using (a) disks and (b) shells.



(a) (b) (c)

Figure 7.5.6

SOLUTION Figure 7.5.6(a) shows the region $\mathcal{R}$ and Figure 7.5.6(b) shows the solid $\mathcal{S}$.

(a) If we use cross sections perpendicular to the y-axis, as in the preceding section, we find that

$$\text{Volume} = \int_1^e \pi(\ln(y))^2\, dy.$$

This integrand has an elementary antiderivative, and we will learn how to find one in Chapter 8. Formula 66 (with $a = 1$) in the table on the inside cover of this book has $\int (\ln(x))^2\, dx = x((\ln(x))^2 - 2\ln(x) + 2)$, which you may check by differentiation. Thus

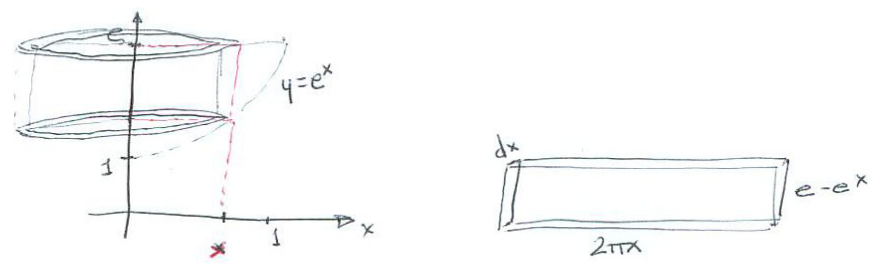
See Exercise 21

$$\text{Volume} = \pi(e - 2) \approx 2.2565.$$

(b) If we use cross sections parallel to the x-axis, we meet a much simpler integration. The typical shell has radius x, height $e - e^x$, and thickness dx as shown in Figure 7.5.7(a).

The local approximation to the total volume of the shell is

$$\underbrace{2\pi x}_{\text{circumference}}\ \underbrace{(e - e^x)}_{\text{height}}\ \underbrace{dx}_{\text{thickness}},$$



(a) (b)

Figure 7.5.7

so the volume of $\mathcal{S}$ is

$$\int_0^1 2\pi x\left(e-e^x\right)\ dx. = 2\pi\int_0^1 ex - xe^x\ dx.$$

Now one needs an antiderivatives of ex and xe^x. The first part is trivial, $\int ex\ dx = \frac{e}{2}x^2$. In Chapter 8 we will learn how to find an antiderivative of xe^x, and we will find it is much easier to find than $\int (\ln(y))^2\ dy$. and then formula 59 on the inside cover gives $\int xe^x\ dx = xe^x - e^x$. As expected, once again the volume is $\pi(e-2)$.

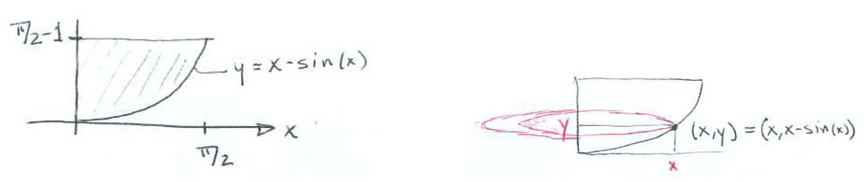
See Exercise 22

◇

It is not unusual to find one formulation much easier than the other. In Example 1 both methods were feasible. In the next, the shell technique is clearly preferable.

EXAMPLE 2 The region $\mathcal{R}$ bounded by the line $y=\frac{\pi}{2}-1$, the y-axis, and the curve $y=x-\sin(x)$ is revolved around the y-axis. Try to set up definite integrals for the volume of this solid using (a) disks and (b) shells.

$y=x-\sin(x)$ is Kepler's equation, with $\mathrm{e}=1$. See Exercise 28 on page 58.



(a) (b)

Figure 7.5.8

SOLUTION The region $\mathcal{R}$ is displayed in Figure 7.5.8(a).

For instance, when $y=0$, then $x=0$. When $y=\frac{\pi}{2}-1$, then $x=\frac{\pi}{2}$.

(a) To use the method of parallel cross sections you would have to find the radius of the typical disk shown in Figure 7.5.8(b). The radius for each value of y is the value of x for which $x - \sin(x) = y$. In other words, we have to express x as a function of y. This inverse function is not elementary, ending our hopes of using the FTC.

(b) On the other hand, the shell technique goes through smoothly. The typical shell, shown in Figure 7.5.9, has radius x and height $\frac{\pi}{2} - 1 - (x - \sin(x))$. The volume of the local approximation is

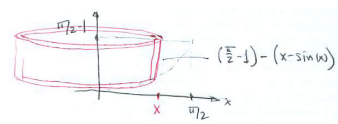


Figure 7.5.9

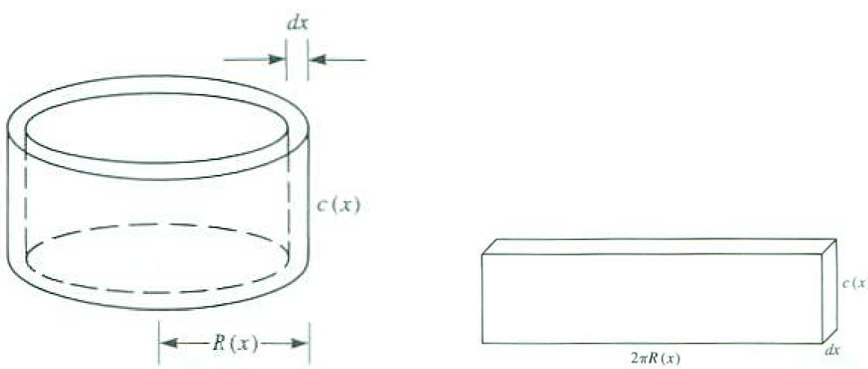
$$\underbrace{2\pi x}_{\text{circumference}} \underbrace{\left(\frac{\pi}{2} - 1 - (x - \sin(x))\right)}_{\text{height}} \underbrace{dx}_{\text{thickness}} .$$

The total volume of the bowl is then

$$\int_0^{\pi/2} 2\pi x \left(\frac{\pi}{2} - 1 - (x - \sin(x))\right)\, dx.$$

The value of this definite integral is found in Exercise 50 on page 773. ◇

Summary



(a) (b)

Figure 7.5.10

The volume of a solid of revolution may be found by approximating the solid by concentric thin shells. The volume of such a shell is approximately $2\pi R(x)\ c(x)\ dx$. (See Figure 7.5.10.) The shell technique is often useful even when integration by cross sections is difficult or impossible.

EXERCISES for Section 7.5

In Exercises 1 to 4 draw a typical approximating cylindrical shell for the solid described, and set up a definite integral for the volume of the given solid. When evaluating your definite integral, feel free to use the tables of antiderivatives in the inside covers of the text.

1. The trapezoid bounded by $y = x$, $x = 1$, $x = 2$, and the x-axis is revolved around the x-axis.

2. The trapezoid in Exercise 1 is revolved about the line $y = -3$.

(a) Repeat this problem when the trapezoid is revolved around the y-axis.

(b) Repeat this problem when the trapezoid is revolved around the line $x = -3$.

3. The triangle with vertices $(0, 0)$, $(1, 0)$, and $(0, 2)$ is revolved around the y-axis.

4. The triangle in Exercise 3 is revolved about the x-axis.

5. Find a definite integral for the volume of the solid produced by revolving about the y-axis the finite region bounded by $y = x^2$ and $y = x^3$.

6. Repeat Exercise 5, except the region is revolved around the x-axis.

7. Set up a definite integral for the volume of the solid produced by revolving about the x-axis the finite region bounded by $y = \sqrt{x}$ and $y = \sqrt[3]{x}$.

8. Repeat Exercise 7, except the region is revolved about the y-axis.

9. Find a definite integral for the volume of the right circular cone of radius a and height h by the shell method.

10. Let R be the region bounded by $y = x + x^3$, $x = 1$, $x = 2$, and the x-axis. Set up a definite integral for the volume of the solid produced by revolving R about (a) the x-axis and (b) the line $x = 3$.

11. Set up a definite integral for the volume of the solid produced by revolving the region R in Exercise 10 about (a) the x-axis and (b) the line $y = -2$.

12. Set up a definite integral for the volume of the solid of revolution formed by revolving the region bounded by $y = 2 + \cos(x)$, $x = \pi$, $x = 10\pi$, and the x-axis around (a) the y-axis and (b) the x-axis.

13. The region below $y = \cos(x)$, above the x-axis, and between $x = 0$ and $x = \frac{\pi}{2}$ is revolved around the x-axis. Find a definite integral for the volume of the resulting solid of revolution by (a) parallel cross sections and (b) concentric shells.

14. Let R be the region below $y = 1/\left(1 + x^2\right)^2$ and above $[0, 1]$. Set up a definite integral for the volume of the solid produced by revolving R about the y-axis.

15. The region between $y = e^{x^2}$, the x-axis, $x = 0$, and $x = 1$ is revolved about the y-axis.

(a) Set up a definite integral for the area of this region.

(b) Set up a definite integral for the volume of the solid produced.

The FTC is of no use in evaluating the *area* of this region, but it easily evaluates the volume of the solid.

16. Set up a definite integral for the volume of the doughnut (torus) produced by revolving the disk of radius a about a line L at a distance $b > a$ from its center. (See Figure 7.5.11.)

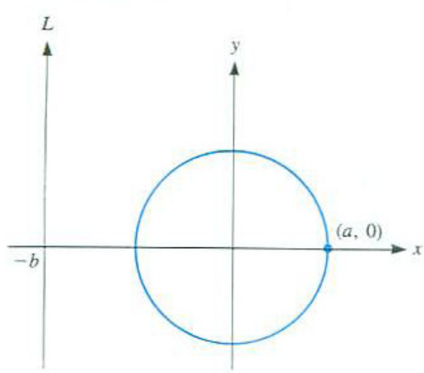


Figure 7.5.11

17. The region R below $y = e^x\left(1 + \sin(x)\right)/x$ and above $[0, 10\pi]$ is revolved about the y-axis to produce a solid of revolution. (a) Find a definite integral for the volume of the solid by parallel cross sections. (b) Find a definite integral for the volume of the solid by concentric shells. (c) Which definite integral do you think is easier to evaluate? Why?

18. Let R be the region below $y = \ln(x)$ and above $[1, e]$. Find a definite integral for the volume of the solid produced by revolving R about (a) the x-axis and (b) the y-axis.

19. Let R be the region below $y = 1/\left(x^2 + 4x + 1\right)$ and above $[0, 1]$. Find a definite integral for the volume of the solid produced by revolving R about the line $x = -2$.

20. Let R be the region below $y = 1/\sqrt{2 + x^2}$ and above $[\sqrt{3}, \sqrt{8}]$. Set up a definite integral for the volume of the solid produced by revolving R about the (a) the x-axis and (b) the y-axis.

Exercises 21 and 22 complete Exercise 1. In that Example the region below $y = e$, above $y = e^x$, and to the right of the y-axis is revolved around the y-axis to form a solid $\mathcal{S}$.

21. The volume of $\mathcal{S}$ using cross sections perpendicular to the y-axis was found to be $\int_1^e \pi\left(\ln(y)\right)^2 \ dy$.

(a) Verify that $x\left((\ln(x))^2 - 2\ln(x) + 2\right)$ is an antiderivative of $(\ln(x))^2$.

(b) Find the volume of $\mathcal{S}$. (Use FTC I.)

22. The volume of $\mathcal{S}$ using cross sections parallel to the y-axis was found to be $\int_0^1 2\pi x\left(e - e^x\right) \ dx$.

(a) Verify that $xe^x - e^x$ is an antiderivative of xe^x.

(b) Find the volume of $\mathcal{S}$. (Use FTC I.)

23. When we unrolled the shell as a carpet we pictured it as a rectangular solid whose faces meet at right angles. However, since the inner radius is x and the outer radius is $x + dx$ the circumference of the inside of the shell is less than the outer circumference.

(a) By viewing the shell as the difference between two circular cylinders, compute its exact volume.

(b) Show that this volume is $2\pi\left(x + \frac{dx}{2}\right)c(x)$.

This means that if we used $x + \frac{dx}{2}$ as our sampling number in the interval $[x, x + dx]$ instead of x, our local approximation to the volume of the shell would be exact.

The **kinetic energy** of a particle of mass m grams moving at a velocity of v centimeters per second is $mv^2/2$ ergs. Exercises 24 and 25 ask for the kinetic energy of rotating objects.

24. A solid cylinder of radius r and height h centimeters has a uniform density of g grams per cubic centimeter. It is rotating at the rate of two revolutions per second around its axis.

(a) Find the speed of a particle at a distance x from the axis.

(b) Find a definite integral for the kinetic energy of the rotating cylinder.

25. A solid ball of radius r centimeters has a uniform density of g grams per cubic centimeter. It is rotating around a diameter at the rate of three revolutions per second around its axis.

(a) Find the speed of a particle at a distance x from the diameter.

(b) Find a definite integral for the kinetic energy of the rotating ball.

26. When a region $\mathcal{R}$ in the first quadrant is revolved around the y-axis, a solid of volume 24 is produced. When $\mathcal{R}$ is revolved around the line $x = -3$, a solid of volume 82 is produced. What is the area of $\mathcal{R}$?

27. Let $\mathcal{R}$ be a region in the first quadrant. When it is revolved around the x-axis, a solid of revolution is produced. When it is revolved around the y-axis, another solid of revolution is produced. Give an example of such a region $\mathcal{R}$ with the property that the volume of the first solid *cannot* be evaluated by the FTC, but the volume of the second solid can be evaluated by the FTC.

7.6 Water Pressure Against a Flat Surface

This section shows how to use integration to compute the force of water against a submerged flat surface.

Introduction

Imagine the portion of the Earth's atmosphere directly above one square inch at sea level. That air forms a column some hundred miles high which weighs about 14.7 pounds. It exerts a pressure of 14.7 pounds per square inch (14.7 psi).

This pressure does not crush us because the cells in our body are at the same pressure. If we were to go into a vacuum, we would explode.

This is why astronauts wear pressurized suits.

The pressure inside a flat tire is 14.7 psi. When you pump up a bicycle tire so that the gauge reads 60 psi, the pressure is actually $60 + 14.7 = 74.7$ psi. The tire must be strong enough to avoid bursting.

Next imagine diving into a lake and descending 33 feet (10 meters). Extending that 100-mile-high column 33 feet into the water adds $(33)(12)(0.036227) = 14.7$ pounds of water. The pressure is now twice 14.7 psi. The pressure is now twice 14.7, or 29.4 psi. You cannot escape that pressure by turning your body, since at a given depth the pressure is the same in all directions.

One cubic foot of water weighs 62.6 pounds, so one cubic inch weighs $\frac{62.6}{1728} = 0.036227$ pounds and the density is 0.036227 pounds per cubic inch.

Pressure and force are closely related. If the force is the same throughout a region, then the pressure is simply "total force divided by area":

$$\text{pressure} = \frac{\text{force}}{\text{area}}.$$

Equivalently,

$$\text{force} = \text{pressure} \times \text{area}.$$

Thus, when the pressure is constant in a plane region it is easy to find the total force against it: multiply the pressure and the area of the region.

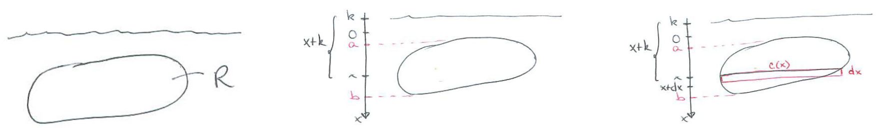
If the pressure varies in the region, we must make use of integration.

Using an Integral to Find the Force of Water

We will see how to find the total force on a flat submerged object due to the water. We will disregard the pressure due to the atmosphere. (See Figure 7.6.1(a).)

At a depth of h inches, water exerts a pressure of about $0.036h$ psi. Therefore the water exerts a force on a flat *horizontal* object of area $\mathcal{A}$ square inches, at a depth of h inches equal to $0.036h\mathcal{A}$ pounds.

To deal with, say, a vertical submerged surface takes more calculation, since the pressure is not constant over that surface. Imagine the surface $\mathcal{R}$, shown



(a) (b) (c)

Figure 7.6.1

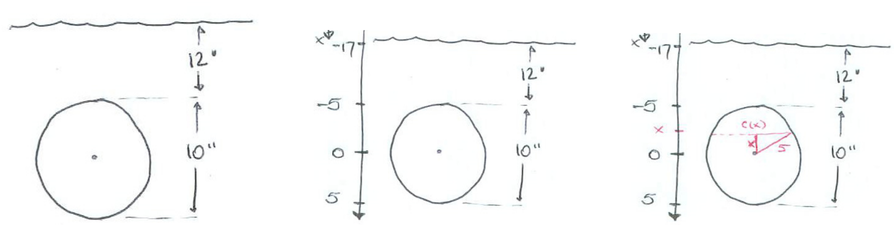
in Figure 7.6.1(b). Introduce a vertical x-axis, pointed down, with its origin $\mathcal{O}$, a distance k below the water's surface. $\mathcal{R}$ lies between lines corresponding to $x = a$ and $x = b$. The depth of the water corresponding to x is not x but $x + k$. (If the origin is at the water's surface, then $k = 0$.)

As usual, we will find the local approximation of the force by considering a narrow horizontal strip corresponding to the interval $[x, x + dx]$ of the x-axis, as in Figure 7.6.1(c). Letting $c(x)$ denote the cross-sectional length, we see that the force of the water on this strip is approximately

$$\underbrace{(0.036)}_{\text{density of } H_2O} \quad \underbrace{(x+k)}_{\text{depth}} \quad \underbrace{c(x)\ dx}_{\text{area of strip}} \quad \text{pounds.}$$

Therefore

$$\boxed{\text{Force against } \mathcal{R} \text{ is } 0.036 \int_a^b (x+k)c(x)\ dx \text{ pounds.}}$$



(a) (b) (c)

Figure 7.6.2

EXAMPLE 1 A circular tank is submerged in water. An end is a disk 10 inches in diameter. The top of the tank is 12 inches below the surface of the water. Find the force against one end.

SOLUTION The end of the tank is shown in Figure 7.6.2(a). Introduce

This placement of $\mathcal{O}$ will make it easier to compute the cross section lengths.

a vertical x-axis with its origin $\mathcal{O}$ level with the center of the disk. (See Figure 7.6.2(b).) To find the cross section $c(x)$ we use Figure 7.6.2(c).

By the Pythagorean Theorem applied to the right triangle in Figure 7.6.2(c) we have

For any number x, $|x|^2 = x^2$.

$$\left(\frac{c(x)}{2}\right)^2 + |x|^2 = 5^2.$$

Thus

$$(c(x))^2 + 4x^2 = 100.$$

So

$$c(x) = \sqrt{100-4x^2}.$$

Having found the cross section as a function of x, we still must find the depth as a function of x. To do this, inspect Figure 7.6.3.

As a check, let $x = 0$, when the depth is clearly 17 inches.

The depth $\overline{AC}$ equals $\overline{AB} + \overline{BC} = 12 + (x - (-5)) = 17 + x$.

We have

$$\text{Local Estimate of Force} = \underbrace{(0.036)(x+17)}_{\text{pressure}}\underbrace{\sqrt{100-4x^2}\,dx}_{\text{area}}.$$

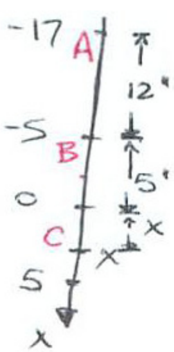


Figure 7.6.3

From this we obtain

$$\begin{aligned}\text{Total Force} &= \int_{-5}^{5} (0.036)(x+17)\sqrt{100-4x^2}\,dx \text{ pounds}\\ &= 0.036\int_{-5}^{5} x\sqrt{100-4x^2}\,dx + 0.036\int_{-5}^{5} 17\sqrt{100-4x^2}\,dx \text{ pounds.}\end{aligned}$$

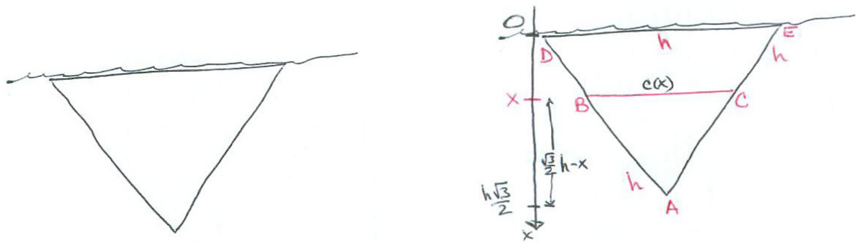
The first integral is 0 because the integrand, $x\sqrt{100-4x^2}$, is an odd function and the interval of integration is symmetric about $x = 0$. The integrand in the second integral is even, so, after factoring out the 17, we have

$$\begin{aligned}\int_{-5}^{5}\sqrt{100-4x^2}\,dx &= 2\int_{0}^{5}\sqrt{100-4x^2}\,dx = 4\int_{0}^{5}\sqrt{25-x^2}\,dx\\ &= 4\,(\text{Area of one quarter of disk of radius 5}) = 4\left(\frac{1}{4}\pi 5^2\right) = 5^2\pi\end{aligned}$$

Thus,

$$\text{Total Force} = (0.036)(17)(25\pi) \text{ pounds} \approx 48 \text{ pounds.}$$

$\diamond$



(a) (b)

Figure 7.6.4

EXAMPLE 2 Figure 7.6.4(a) shows a submerged equilateral triangle of side h. Find the force of water against it.

SOLUTION In this case we place the origin of the vertical axis at the surface of the water (see Figure 7.6.4(b)). To set up an integral we must compute $c(x)$. Note $\frac{\sqrt{3}h}{2}$ is marked on the x-axis; it is the length of an altitude in the triangle.

The similar triangles ABC and ADE give us

$$\frac{c(x)}{h} = \frac{\frac{\sqrt{3}}{2}h - x}{\frac{\sqrt{3}}{2}h}.$$

Thus,

$$c(x) = h - \frac{2x}{\sqrt{3}}.$$

Observe that $c(0) = h$ and $c(\frac{\sqrt{3}}{2}h) = 0$ and c is linear, which agree with Figure 7.6.4(b).

The local estimate of force is therefore

$$\underbrace{0.036x}_{\text{pressure}} \underbrace{\left(h - \frac{2x}{\sqrt{3}}\right) dx}_{\text{area}}.$$

Hence

$$\begin{aligned}\text{Total Force} &= \int_0^{\frac{\sqrt{3}}{2}h} 0.036x\left(h - \frac{2x}{\sqrt{3}}\right) dx = 0.036\int_0^{\frac{\sqrt{3}}{2}h}\left(hx - \frac{2x^2}{\sqrt{3}}\right) dx \\ &= 0.036\left(h\frac{x^2}{2} - \frac{2}{\sqrt{3}}\frac{x^3}{3}\right)\Bigg|_0^{\frac{\sqrt{3}}{2}h} = 0.036\frac{h^3}{8} \text{ pounds.}\end{aligned}$$

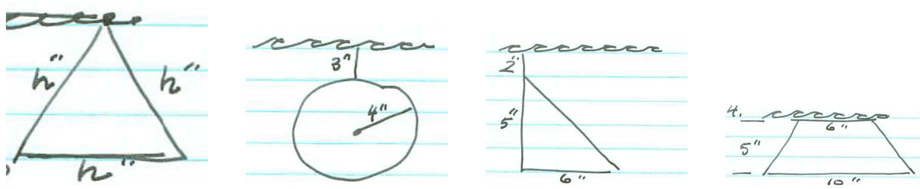
◇

Summary

We introduced the notion of water pressure defined as "force divided by area" or "force per unit area." If the pressure is constant over a flat region of area $\mathcal{A}$, the force is the product: pressure times area. When $p(x)$ is the pressure and $c(x)$ is the length of the typical cross section, then $p(x)c(x)\ dx$ is a local approximation to the force. The water pressure $p(x)$ is 0.036 times the depth. The dimensions are in inches and the force is in pounds.

EXERCISES for Section 7.6

A cubic inch of water weighs about 0.036 pounds. (All dimensions are in inches.)



(a) (b) (c) (d)

Figure 7.6.5

In Exercises 1 to 4 find a definite integral for the force of water on the indicated surface.

1. The triangular surface in Figure 7.6.5(a).

2. The circular surface in Figure 7.6.5(b).

3. The triangular surface in Figure 7.6.5(c).

4. The trapezoidal surface in Figure 7.6.5(d).

In Exercises 5 and 6 the surfaces are tilted like the bottoms of many swimming pools. Find the force of the water against them.

5. The surface is an a by b rectangle inclined at an angle of $30°$ ($\pi/6$ radians) to the horizontal. The top of the surface is at a depth k. (See Figure 7.6.6.)

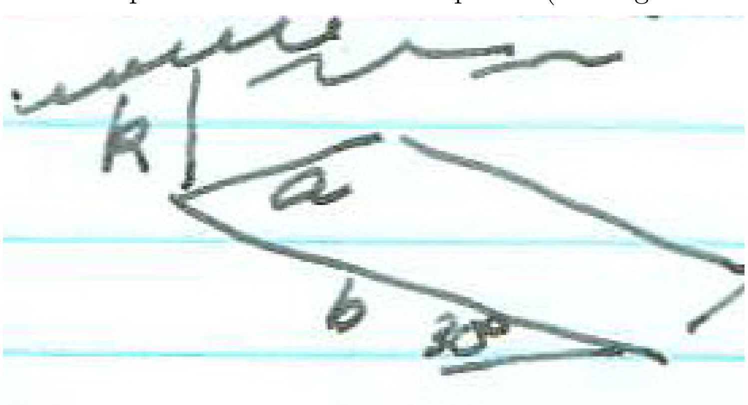


Figure 7.6.6

6. The surface is a disk of radius r tilted at an angle of $45°$ ($\pi/4$ radians) to the horizontal. Its top is at the surface of the water.

7. A vertical disk is totally submerged. Show that the force of the water against it is the same as the product of its area and the pressure at its center.

8. If the region in Exercise 7 is not vertical, is the same conclusion true?

9. Let $\mathcal{R}$ be a convex planar region. $\mathcal{R}$ is called **centrally symmetric** if it contains a point P such that P is the midpoint of every chord of $\mathcal{R}$ that passes through P. For instance, a parallelogram is centrally symmetric. No triangle is. Now, assume that a centrally symmetric region is placed vertically in water and is completely submerged. Show that the force against it equals the product of its area and the pressure at P.

10. Why is finding volume by shells essentially the same as finding the force against a submerged object?

7.7 Work

In this section we treat the work accomplished by a force operating along a line, for example the work done when you stretch a spring. If the force has the *constant* value F and it operates over a distance s in the *direction* of the force, then the work W accomplished is simply

$$\text{Work} = \text{Force} \cdot \text{Distance} \qquad \text{or} \qquad W = F \cdot s.$$

If force is measured in newtons and distance in meters, work is measured in newton-meters or joules. For example, the force needed to lift a mass of m kilograms at the surface of the earth is about $9.8m$ newtons.

A weightlifter who raises 100 kilograms a distance of 0.5 meter accomplishes $9.8(100)(0.5) = 490$ joules of work. On the other hand, the weightlifter who just carries the barbell from one place to another in the weightlifting room, without raising or lowering it, accomplishes no work because the barbell was moved a distance zero in the direction of the force.

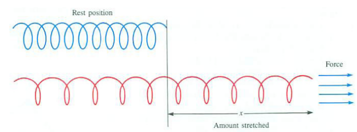


Figure 7.7.1

The Stretched Spring

As you stretch a spring (or rubber band) from its rest position, the further you stretch it the harder you have to pull. According to Hooke's law, the force you must exert is proportional to the distance that the spring is stretched, as shown in Figure 7.7.1. In symbols, $F = kx$, where F is the force and x is the distance from the rest position.

Hooke's law says a spring's force is proportional to the distance it is stretched.

Because the force is *not* constant, we cannot compute the work accomplished just by multiplying force times distance. As usual, we need an integral, as the next example illustrates.

EXAMPLE 1 A spring is stretched 0.5 meter longer than its rest length. The force required to keep it at that length is 3 newtons. Find the total work accomplished in stretching the spring 0.5 meter from its rest position.

SOLUTION Let us estimate the work involved in stretching the spring from x to $x + dx$. (See Figure 7.7.2.)

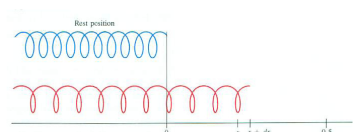


Figure 7.7.2

The distance dx is small. As the end of the spring is stretched from x to $x + dx$, the force is almost constant. Since the force is proportional to x, it is of the form kx for some constant k. We know that the force, F, is 3 when $x = 0.5$, so

$$F = kx \qquad \text{gives} \qquad 3 = k(0.5) \qquad \text{which implies} \qquad k = 6.$$

The work accomplished in stretching the spring from x to $x + dx$ is then approximately

$$\underbrace{kx}_{\text{force}} \cdot \underbrace{dx}_{\text{distance}} \text{ joule.}$$

Hence the total work is

$$\int_a^b kx\,dx = \int_0^{0.5} 6x\,dx = 3x^2\Big|_0^{0.5} = 0.75 \text{ joule.}$$

◇

Work in Launching a Rocket

The earth's surface is about 4,000 miles from its center.

The force of gravity that the earth exerts on an object diminishes as the object gets further away from the earth. The work required to lift an object 1 foot at sea level is greater than the work required to lift the same object the same distance at the top of Mt. Everest. However, the difference in altitudes is so small in comparison to the radius of the earth that the difference in work is negligible. On the other hand, when an object is rocketed into space, that the force of gravity diminishes with distance from the center of the earth is critical.

According to Newton, the force of gravity on a given mass is proportional to the reciprocal of the square of the distance of that mass from the center of the earth. That is, there is a constant k such that the gravitational force at distance r from the center of the earth, $F(r)$, is given by

$$F(r) = \frac{k}{r^2}.$$

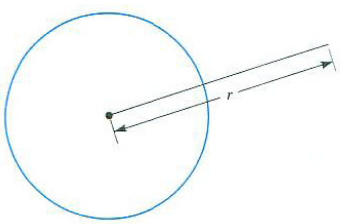


Figure 7.7.3

(See Figure 7.7.3.)

WARNING It is important to remember that r is "distance to the center of the earth," not "distance to the surface."

EXAMPLE 2 How much work is required to lift a 1 pound payload from the surface of the earth to the moon, which is about 240,000 miles away?

The unit for work is **joule**. 1 joule = 1 newton meter = 1 watt second = 0.7376 foot pound.

SOLUTION The work necessary to lift an object a distance x against a constant vertical force F is the product of force times distance:

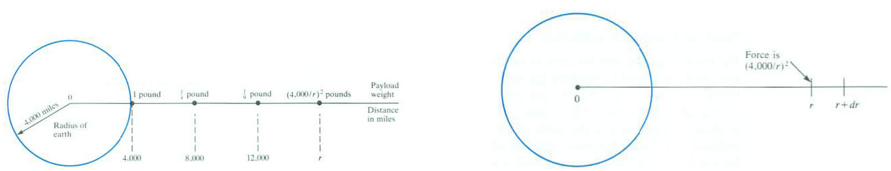
$$\text{Work} = F \cdot x.$$

Since the gravitational pull of the earth on the payload *changes* with distance from the center of the earth, an integral will be needed to express the total work.

The payload weighs 1 pound at the surface of the earth. The farther it is from the center of the earth, the less it weighs, for the force of the earth on the mass is inversely proportional to the square of the distance of the mass from the center of the earth. Thus the force on the payload is given by k/r^2 pounds, where k is a constant, which will be determined in a moment, and r is the distance in miles form the payload to the center of the earth. When $r = 4,000$ (miles), the force is 1 pound; thus

$$1 \text{ pound} = \frac{k}{(4,000 \text{ miles})^2}.$$

From this it follows that $k = 4,000^2$, and therefore the gravitational force on a 1-pound mass is, in general, $(4,000/r)^2$ pounds. As the payload recedes from the earth, it loses weight (but not mass), as recorded in Figure 7.7.4(a). The



(a) (b)

Figure 7.7.4

work done in lifting the payload from a distance r to a distance $r + dr$ from the center of the earth is approximately

$$\underbrace{\left(\frac{4,000}{r}\right)^2}_{\text{force}} \underbrace{(dr)}_{\text{distance}} \text{ miles-pounds.}$$

(See Figure 7.7.4(b).)

Hence the work required to move the 1 pound mass from the surface of the earth to the moon is given by the integral

$$\begin{aligned} \int_{4,000}^{240,000} \left(\frac{4,000}{r}\right)^2 dr &= -\frac{4,000^2}{r}\Bigg|_{4,000}^{240,000} = -4,000^2\left(\frac{1}{240,000} - \frac{1}{4,000}\right) \\ &= -\frac{4,000}{60} + 4,000 \approx 3,933 \text{ miles-pounds} \end{aligned}$$

$$= 2.8154 \times 10^7 \text{ joules.}$$

The work is just a little less than if the payload were lifted 4,000 miles against a constant gravitational force equal to that at the surface of the earth. ◇

Summary

The work accomplished by a constant force F that moves an object a distance x in the direction of the force is the product Fx, "force times distance." The work by a variable force, $F(x)$, moving an object over the interval $[a, b]$ is measured by an integral $\int_a^b F(x)\, dx$.

EXERCISES for Section 7.7

1. A spring is stretched 0.20 meters from its rest length. The force required to keep it at that length is 5 newtons. Assuming that the force of the spring is proportional to the distance it is stretched, find the work accomplished in stretching the spring

(a) 0.20 meters from its rest length;

(b) 0.30 meters from its rest length.

2. A spring is stretched 3 meters from its rest length. The force required to keep it at that length is 24 newtons. Assume that the force of the spring is proportional to the distance it is stretched. Find the work accomplished in stretching the spring

(a) 3 meters from its rest length;

(b) 4 meters from its rest length.

3. Suppose a spring does not obey Hooke's law. Instead, the force it exerts when stretched x meters from its rest length is $F(x) = 3x^2$ Newtons. Find the work done in stretching the spring 0.80 meter from its rest length.

4. Suppose a spring does not obey Hooke's law. Instead, the force it exerts when stretched x meters from its rest length is $F(x) = 2\sqrt{x}$ Newtons. Find the work done in stretching the spring 0.50 meter from its rest length.

5. How much work is done in lifting the 1 pound payload the first $4,000$ miles of its journey to the moon? See Example 2.

6. If a mass that weighs 1 pound at the surface of the earth were launched from a position $20,000$ miles from the center of the earth, how much work would be required to send it to the moon ($240,000$ miles from the center of the earth)?

7. Assume that the force of gravity obeys an inverse cube law, so that the force on a 1 pound payload a distance r miles from the center of the earth ($r \geq 4,000$) is $(4,000/r)^3$ pounds. How much work would be required to lift a 1 pound payload from the surface of the earth to the moon?

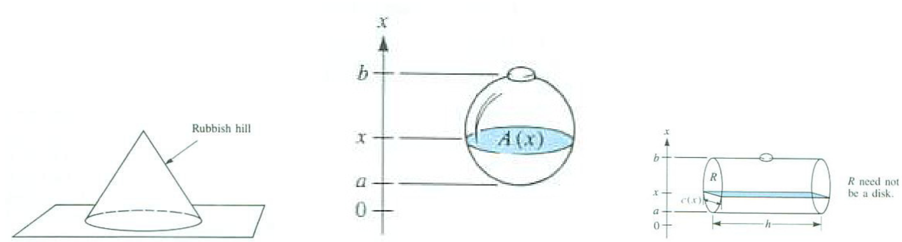
8. Geologists, when considering the origin of mountain ranges, estimate the energy required to lift a mountain up from sea level. Assume that two mountains are composed of the same type of matter, which weighs k pounds per cubic foot. Both are right circular cones in which the height is equal to the radius. One mountain is twice as high as the other. The base of each is at sea level. If the work required to lift the matter in the smaller mountain above sea level is W, what is the corresponding work for the larger mountain?

9. Assume that Mt. Everest has a shape of a right circular cone of height 30,000 feet and radius 150,000 feet, with unifrom density of 200 pounds per cubic foot.

(a) How much work was required to lift the material in Mt. Everest if it was initially all at sea level?

(b) How does this work compare with the energy of a 1 megaton hydrogen bomb? (One megaton is the energy in a million tons of TNT: about 3×10^{14} foot-pounds.)

10. A town in a flat valley made a conical hill out of its rubbish, as shown in Figure 7.7.5(a). The work requireed to lift all the rubbish was W. Happy with the result, the town decided to make another hill with twice the volume, but of the same shape. How much work will be required to build this hill? Explain.



(a) (b) (c)

Figure 7.7.5

11. A container is full of water which weighs 64.2 pounds per cubic foot. All the water is pumped out of an opening at the top of the container. Develop a definite integral for the work accomplished. (The integral involves only a, b, and $A(x)$, the cross-sectional area shown in Figure 7.7.5(b).)

12. A horizontal tank in the form of a cylinder with base R is full of water. The cylinder has height h feet. (See Figure 7.7.5(c).) Develop a definite integral for the total work accomplished when all the water is pumped out an opening at the top. (Express the integral in terms of a, b, $c(x)$, and h.)

Skill Drill: Derivatives

In Exercises 13 to 17 a and b are constants. In each case verify that the derivative of the first function is the second function.

13. $\ln\left(x+\sqrt{a^2+x^2}\right)$; $1/\sqrt{a^2+x^2}$

14. $\frac{1}{2ab}\ln\left|\frac{a+bx}{a-bx}\right|$; $1/(a^2-b^2x^2)$

15. $\frac{x^4}{8}-\left(\frac{x^3}{4}-\frac{3x}{8}\right)\sin(2x)$; $x^3\sin^2(x)$

16. $x-\ln(1+e^x)$; $1/(1+e^x)$

17. $\frac{e^{ax}}{a^2+1}\left(a\sin(x)-\cos(x)\right)$; $e^{ax}\sin(x)$

7.8 Improper Integrals

This section develops the analog of a definite integral when the interval of integration is infinite or the integrand becomes arbitrarily large in the interval of integration. The definition of a definite integral does not cover these cases.

Improper Integrals: Interval Unbounded

A question about areas will introduce the notion of an "improper integral." Figure 7.8.1 shows the region under $y = 1/x$ and above the interval $[1, \infty)$. Figure 7.8.2 shows the region under $y = 1/x^2$ and above the same interval.

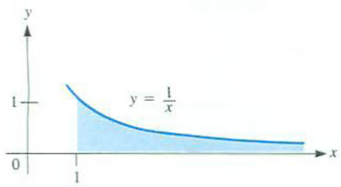


Figure 7.8.1

Let us compute the areas of the two regions. We might be tempted to say that the area in Figure 7.8.1 is $\int_1^\infty 1/x\ dx$. Unfortunately, the symbol $\int_1^\infty f(x)\ dx$ has not been given any meaning so far in this book. The definition of the definite integral $\int_a^b f(x)\ dx$ involves a limit of sums of the form

$$\sum_{i=1}^{n} f(c_i)(x_i - x_{x-1}),$$

where each $x_i - x_{i-1}$ is the length of an interval $[x_{i-1}, x_i]$. If you cut the interval $[1, \infty)$ into a finite number of intervals, then at least one section has infinite length, and such a sum is meaningless.

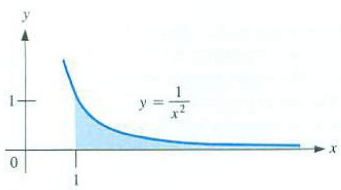


Figure 7.8.2

It does make sense, however, to find the area of that part of the region in Figure 7.8.1 from $x = 1$ to $x = b$, where $b > 1$, and find what happens to that area as $b \to \infty$. To do this, first calculate $\int_1^b (1/x)\ dx$:

$$\int_1^b \frac{dx}{x} = \ln(x)\Big|_1^b = \ln(b) - \ln(1) = \ln(b).$$

Then

$$\lim_{b\to\infty} \int_1^b \frac{dx}{x} = \lim_{b\to\infty} \ln(b) = \infty.$$

So the area of the region in Figure 7.8.1 is infinite.

Next, examine the area of the region in Figure 7.8.2. We first find

$$\int_1^b \frac{dx}{x^2} = -\frac{1}{x}\Big|_1^b = -\frac{1}{b} - \left(-\frac{1}{1}\right) = 1 - \frac{1}{b}$$

Thus,

$$\lim_{b\to\infty} \int_1^b \frac{dx}{x^2} = \lim_{b\to\infty} \left(1 - \frac{1}{b}\right) = 1.$$

In this case the area is finite. Though the regions in Figures 7.8.1 and 7.8.2 look alike, one has an infinite area, and the other, a finite area. This contrast suggests the following definitions.

> **DEFINITION** *(Convergent improper integral $\int_a^\infty f(x)\ dx$.)* Let f be continuous for $x \geq a$. If $\lim_{b\to\infty} \int_a^b f(x)\ dx$ exists, the function f is said to have a **convergent improper integral** from a to ∞. The value of the limit is denoted by $\int_a^\infty f(x)\ dx$:
>
> $$\int_a^\infty f(x)\ dx = \lim_{b\to\infty} \int_a^b f(x)\ dx.$$

We saw that $\int_1^\infty\ dx/x^2$ is a convergent improper integral with value 1.

> **DEFINITION** *(Divergent improper integral $\int_a^\infty f(x)\ dx$.)* Let f be a continuous function for $x \geq a$. If $\lim_{b\to\infty} \int_a^b f(x)\ dx$ does not exist, the function f is said to have a **divergent improper integral** from a to ∞.

As we saw, $\int_1^\infty\ dx/x$ is a divergent improper integral.

The improper integral $\int_1^\infty\ dx/x$ is divergent because $\int_1^b\ dx/x \to \infty$ as $b \to \infty$. But an improper integral $\int_a^\infty f(x)\ dx$ can be divergent without being infinite. Consider, for instance, $\int_0^\infty \cos(x)\ dx$. We have

$$\int_0^b \cos(x)\ dx = \sin(x)|_0^b = \sin(b).$$

As $b \to \infty$, $\sin(b)$ does not approach a limit, nor does it become arbitrarily large. As $b \to \infty$, $\sin(b)$ just keeps going up and down in the range -1 to 1 infinitely often. Thus $\int_0^\infty \cos(x)\ dx$ is divergent.

The improper integral $\int_{-\infty}^b f(x)\ dx$ is defined similarly:

The improper integral $\int_{-\infty}^b f(x)\ dx$.

$$\int_{-\infty}^b f(x)\ dx = \lim_{a\to-\infty} \int_a^b f(x)\ dx.$$

If the limit exists, $\int_{-\infty}^b f(x)\ dx$ is a *convergent* improper integral. If the limit does not exist, it is a *divergent* improper integral.

To deal with improper integrals over the entire x-axis, define

The improper integral $\int_{-\infty}^\infty f(x)\ dx$

$$\int_{-\infty}^\infty f(x)\ dx$$

to be the sum

$$\int_{-\infty}^{0} f(x)\ dx + \int_{0}^{\infty} f(x)\ dx,$$

which will be called *convergent* if both

$$\int_{-\infty}^{0} f(x)\ dx \qquad \text{and} \qquad \int_{0}^{\infty} f(x)\ dx$$

are convergent. If at least one of the two is divergent, $\int_{-\infty}^{\infty} f(x)\ dx$ will be called divergent.

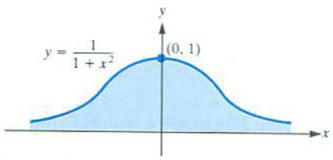


Figure 7.8.3

EXAMPLE 1 Is the area of the region bounded by the curve $y = 1/(1+x^2)$ and the x-axis finite or infinite (see Figure 7.8.3).
SOLUTION The area in question equals $\int_{-\infty}^{\infty} dx/(1+x^2)$. Now,

$$\int_{0}^{\infty} \frac{dx}{1+x^2} = \lim_{b\to\infty} \int_{0}^{b} \frac{dx}{1+x^2} = \lim_{b\to\infty} (\tan^{-1}(b) - \tan^{-1}(0)) = \frac{\pi}{2} - 0 = \frac{\pi}{2}.$$

Because $1/(1+x^2)$ is an even function, we deduce immediately that

$$\int_{-\infty}^{0} \frac{dx}{1+x^2} = \frac{\pi}{2}.$$

Hence,

$$\int_{-\infty}^{\infty} \frac{dx}{1+x^2} = \frac{\pi}{2} + \frac{\pi}{2} = \pi,$$

the integral is convergent and the area in question is π. ◇

Shorthand Notation for $\int_a^{\infty} f(x)\ dx$

OBSERVATION *(Shorthand Notation for $\int_a^{\infty} f(x)\ dx$)* If $\int_a^{\infty} f(x)\ dx$ is convergent and $F(x)$ is an antiderivative of $f(x)$, then $\int_a^{\infty} f(x)\ dx = \lim_{b\to\infty} F(b) - F(a)$. In these situation we could write

$$\int_{a}^{\infty} f(x)\ dx = F(x)|_a^{\infty}$$

where it is understood that $F(\infty)$ is short for $\lim_{b\to\infty} F(b)$.

Comparison Test for Convergence of $\int_a^\infty f(x)\ dx$, $f(x) \geq 0$

The integral $\int_0^\infty e^{-x^2}\ dx$ is important in statistics. Is it convergent or divergent? We cannot evaluate $\int_0^b e^{-x^2}\ dx$ by the Fundamental Theorem since e^{-x^2} does not have an elementary antiderivative. Even so, there is a way of showing that $\int_0^\infty e^{-x^2}\ dx$ is convergent without finding its exact value. The method is described in Theorem 1.

Theorem 7.8.1. *Comparison test for convergence of improper integrals. Let $f(x)$ and $g(x)$ be continuous functions for $x \geq a$. Assume that $0 \leq f(x) \leq g(x)$ and that $\int_a^\infty g(x)\ dx$ is convergent. Then $\int_a^\infty f(x)\ dx$ is convergent and*

$$\int_a^\infty f(x)\ dx \leq \int_a^\infty g(x)\ dx.$$

The geometric interpretation of Theorem 7.8.1is that if the area under $y = g(x)$ is finite, so is the area under $y = f(x)$. (See Figure 7.8.4.) While we do not provide a proof of Theorem 7.8.1, or Theorem 7.8.2, a proof of the less intuitive, Theorem 7.8.3.

A similar convergence test holds for $g(x) \leq f(x) \leq 0$. If $\int_a^\infty g(x)\ dx$ converges, so does $\int_a^\infty f(x)\ dx$.

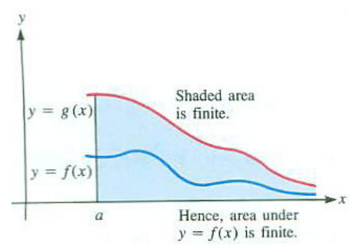


Figure 7.8.4

EXAMPLE 2 Show that $\int_0^\infty e^{-x^2}\ dx$ is convergent and put a bound on its value.

SOLUTION Since e^{-x^2} does not have an elementary antiderivative, we compare $\int_0^\infty e^{-x^2}\ dx$ to an improper integral that we know converges.

For $x \geq 1, x^2 \geq x$; hence $e^{-x^2} \leq e^{-x}$. (See Figure 7.8.5.) Now,

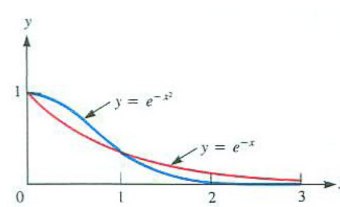


Figure 7.8.5

$$\int_1^b e^{-x}\ dx = -e^{-x}\Big|_1^b = e^{-1} - e^{-b}.$$

Thus

$$\lim_{b\to\infty} \int_1^b e^{-x}\ dx = \frac{1}{e}$$

and the improper integral $\int_1^\infty e^{-x}\ dx$ is convergent.

The comparison test for convergence tells us that $\int_1^\infty e^{-x^2}\ dx$ is also convergent. Furthermore,

$$\int_1^\infty e^{-x^2}\ dx \leq \int_1^\infty e^{-x}\ dx = \frac{1}{e}.$$

In Exercise 32 of Section 17.3 we show that $\int_0^\infty e^{-x^2}\,dx$ equals $\sqrt{\pi/2} \approx 1.25331$.

Thus

$$\int_0^\infty e^{-x^2}\,dx = \int_0^1 e^{-x^2}\,dx + \int_1^\infty e^{-x^2}\,dx \le \int_0^1 e^{-x^2}\,dx + \frac{1}{e}.$$

Since $e^{-x^2} \le 1$ for $0 < x \le 1$, we conclude that

$$\int_0^\infty e^{-x^2}\,dx \le 1 + \frac{1}{e}.$$

$\diamond$

Comparison Test for Divergence of $\int_a^\infty f(x)\,dx$.

Theorem 7.8.2. *Comparison test for divergence of improper integrals. Let $f(x)$ and $g(x)$ be continuous functions for $x \ge a$. Assume that $0 \le g(x) \le f(x)$ and that $\int_a^\infty g(x)\,dx$ is divergent. Then $\int_a^\infty f(x)\,dx$ is also divergent.*

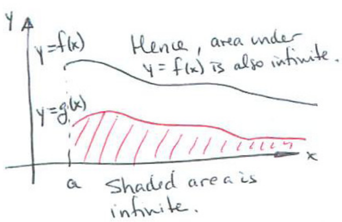


Figure 7.8.6

A glance at Figure 7.8.6 suggests why this theorem is true. The area under $f(x)$ is larger than the area under $g(x)$. When the area under $g(x)$ is infinite, the area under f must also be infinite.

EXAMPLE 3 Show that $\int_1^\infty (x^2+1)/x^3\,dx$ is divergent.

SOLUTION For $x > 0$,

$$\frac{x^2+1}{x^3} > \frac{x^2}{x^3} = \frac{1}{x}.$$

Since $\int_1^\infty \frac{dx}{x} = \infty$, it follows that $\int_1^\infty (x^2+1)/x^3\,dx = \infty$. $\diamond$

Convergence of $\int_a^\infty f(x)\,dx$ When $\int_a^\infty |f(x)|\,dx$ Converges

Is $\int_0^\infty e^{-x}\sin(x)\,dx$ convergent or divergent? Because $\sin(x)$ takes on both positive and negative values, the integrand is not always positive, nor is it always negative. So we can't just compare it with $\int_0^\infty e^{-x}\,dx$.

The next theorem provides a way to establish the convergence of $\int_a^\infty f(x)\,dx$ when $f(x)$ is a function that takes on both positive and negative values. It says that if $\int_a^\infty |f(x)|\,dx$ converges, so does $\int_a^\infty f(x)\,dx$. The argument for this depends on showing that the "negative and positive parts of the function" both have convergent integrals.

Theorem 7.8.3. *Absolute-convergence test for improper integrals. If $f(x)$ is continuous for $x \ge a$ and $\int_a^\infty |f(x)|\,dx$ converges to the number L, then $\int_a^\infty f(x)\,dx$ is convergent and converges to a number between L and $-L$.*

Proof

We will introduce two function, $g(x)$ which is non-negative, and $h(x)$ which is non-positive. That they are both continuous is shown in Exercise 42. That will enable us to use our comparison tests. Figure 7.8.7 shows the graphs of $y = f(x)$ and four functions closely related to $f(x)$.

$$g(x) = \begin{cases} f(x) & \text{if } f(x) \text{ is positive} \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad h(x) = \begin{cases} f(x) & \text{if } f(x) \text{ is negative} \\ 0 & \text{otherwise} \end{cases}$$

Note that $f(x) = g(x) + h(x)$. We will show that $\int_a^\infty g(x)\,dx$ and $\int_a^\infty h(x)\,dx$ both converge.

First, since $\int_a^\infty |f(x)|\,dx$ converges, has value L, and $0 \le g(x) \le |f(x)|$, we conclude that $\int_a^\infty g(x)\,dx$ converges, and the value of the integral is a nonnegative number A between 0 and L:

$$0 \le A \le \int_a^\infty |f(x)|\,dx = L.$$

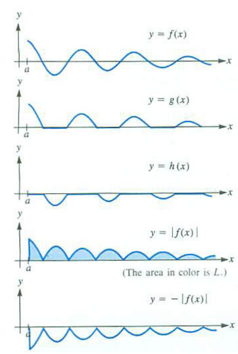


Figure 7.8.7

Second, since $\int_a^\infty -|f(x)|\,dx$ converges, has value $-L$, and $0 \ge h(x) \ge -|f(x)|$, it follows that $\int_a^\infty h(x)\,dx$ converges to a nonpositive number B between $-L$ and 0:

$$0 \ge B \ge \int_a^\infty |f(x)|\,dx = -L.$$

Thus $\int_a^\infty f(x)\,dx = \int_a^\infty (g(x)+h(x))\,dx$ converges to $A+B$, which is a number somewhere in the interval $[-L, L]$. •

EXAMPLE 4 Show that $\int_0^\infty e^{-x}\sin(x)\,dx$ is convergent.

SOLUTION Since $|\sin(x)| \le 1$, we have $|e^{-x}\sin(x)| \le e^{-x}$. Now, $\int_0^\infty e^{-x}\,dx$ is convergent, as we saw in Example 2. Thus $\int_0^\infty e^{-x}\sin(x)\,dx$ is convergent. ◇

See Exercise 29.

Improper Integrals: Integrand Unbounded

The second type of improper integral concerns functions which become infinite in an interval $[a, b]$. For any partition of $[a, b]$, the approximating sum $\sum_{i=1}^n f(c_i)(x_i - x_{i-1})$ can be made arbitrarily large when c_i is chosen so that $f(c_i)$ is very large. The next example shows how to get around this difficulty.

See Exercise 38.

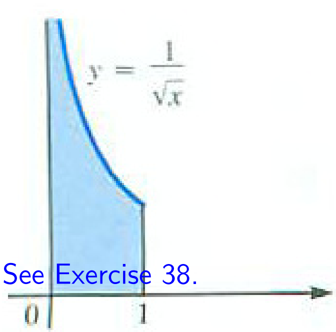


Figure 7.8.8

EXAMPLE 5 Determine the area of the region bounded by $y = 1/\sqrt{x}$, $x = 1$, and the coordinate axes shown in Figure 7.8.8.

SOLUTION Resist for the moment the temptation to write "Area $= \int_0^1 1/\sqrt{x}\, dx$". The integral $\int_0^1 1/\sqrt{x}\, dx$ is not defined since its integrand is unbounded in $[0,1]$. Instead, consider the behavior of $\int_t^1 1/\sqrt{x}\, dx$ as t approaches 0 from the right. Since

$$\int_t^1 \frac{1}{\sqrt{x}}\, dx = 2\sqrt{x}\Big|_t^1 = 2\sqrt{1} - 2\sqrt{t} = 2(1-\sqrt{t}),$$

it follows that

$$\lim_{t\to 0^+} \int_t^1 \frac{dx}{\sqrt{x}} = 2.$$

The area in question is 2.

In Exercise 30 the same value for the area is obtained by taking horizontal cross sections and evaluating an improper integral from 0 to ∞. ◇

The reasoning in Example 5 motivates the definition of the second type of improper integral, in which the integrand rather than the interval is unbounded.

Convergent and Divergent Improper Integrals $\int_a^b f(x)\, dx$.

DEFINITION *(Convergent and Divergent Improper Integrals $\int_a^b f(x)\, dx$.)* Let f be continuous at every number in $[a,b]$ except at a. If $\lim_{t\to a^+} \int_t^b f(x)dx$ exists, the function f is said to have a **convergent improper integral** from a to b. The value of the limit is denoted $\int_a^b f(x)\, dx$.

If $\lim_{t\to a^+} \int_t^b f(x)dx$ does not exist, the function f is said to have a **divergent improper integral** from a to b; in brief, $\int_a^b f(x)\, dx$ does not exist.

In a similar manner, if f is not defined at b, define $\int_a^b f(x)\, dx$ as $\lim_{t\to b^-} \int_a^t f(x)dx$, if this limit exists.

As Example 5 showed, the improper integral $\int_0^1 1/\sqrt{x}\, dx$ is convergent and has the value 2.

More generally, if a function $f(x)$ is not defined at certain isolated numbers, break the domain of $f(x)$ into intervals $[a,b]$ for which $\int_a^b f(x)\, dx$ is either improper or "proper"– that is, an ordinary definite integral.

For instance, the improper integral $\int_{-\infty}^{\infty} 1/x^2\, dx$ is troublesome for four reasons: $\lim_{x\to 0^-} 1/x^2 = \infty$, $\lim_{x\to 0^+} 1/x^2 = \infty$, and the range extends infinitely to the left and also to the right. (See Figure 7.8.9.) To treat the integral, write it as the sum of four improper integrals of the two basic types:

$$\int_{-\infty}^{\infty} \frac{1}{x^2}\,dx = \int_{-\infty}^{-1} \frac{1}{x^2}\,dx + \int_{-1}^{0} \frac{1}{x^2}\,dx + \int_{0}^{1} \frac{1}{x^2}\,dx + \int_{1}^{\infty} \frac{1}{x^2}\,dx.$$

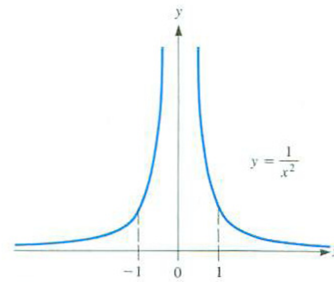


Figure 7.8.9

Each of the four integrals on the right must be convergent in order for $\int_{-\infty}^{\infty} 1/x^2\,dx$ to be convergent. Only the first and last are, so $\int_{-\infty}^{\infty} 1/x^2\,dx$ is divergent.

Summary

We introduced two types of integrals that are not definite integrals, but are defined as limits of definite integrals. The "improper integral" $\int_a^{\infty} f(x)\,dx$ is defined as $\lim_{b\to\infty} \int_a^b f(x)\,dx$. If $f(x)$ is continuous in $[a, b]$ except at a, then $\int_a^b f(x)\,dx$ is defined as $\lim_{t\to a^+} \int_t^b f(x)\,dx$. The first type is far more common in applications. We also developed two comparison tests for convergence or divergence of $\int_a^{\infty} f(x)\,dx$, where the integrand keeps a constant sign. In the case where the integrand $f(x)$ may have both positive and negative values, we showed that if $\int_a^{\infty} |f(x)|\,dx$ converges, so does $\int_a^{\infty} f(x)\,dx$.

EXERCISES for Section 7.8

In Exercises 1 to 9 determine whether the improper integral is convergent or divergent. Evaluate the convergent ones if possible. Some exercises may require using the integral table in the back of the book.

1. $\int_1^\infty \frac{dx}{x^3}$

2. $\int_1^\infty \frac{dx}{\sqrt[3]{x}}$

3. $\int_0^\infty e^{-x}\ dx$

4. $\int_0^\infty \frac{dx}{x+100}$

5. $\int_0^\infty \frac{x^3\ dx}{x^4+1}$

6. $\int_1^\infty x^{-1.01}\ dx$

7. $\int_0^\infty \frac{dx}{(x+2)^3}$

8. $\int_0^\infty \sin(2x)\ dx$

9. $\int_1^\infty x^{-0.99}\ dx$

10. $\int_0^\infty \frac{e^{-x}\sin(x^2)}{x+1}\ dx$

11. $\int_0^\infty \frac{dx}{x^2+4}$

12. $\int_0^\infty \frac{x^2\ dx}{2x^3+5}$

13. $\int_0^\infty \frac{dx}{(x+1)(x+2)(x+3)}$

14. $\int_0^\infty \frac{\sin(x)}{x^2}\ dx$

15. $\int_1^\infty \frac{\ln x\ dx}{x}$

16. $\int_0^\infty e^{-2x}\sin(3x)\ dx$

In Exercises 17 to 21 determine whether the improper integral is convergent or divergent. Evaluate the convergent ones if possible. Some exercises may require using the integral table in the back of the book.

17. $\int_0^1 \frac{dx}{\sqrt[3]{x}}$

18. $\int_0^1 \frac{dx}{\sqrt[3]{x}}$

19. $\int_0^1 \frac{dx}{(x-1)^2}$

20. $\int_0^\infty \frac{e^{-x}}{\sqrt{x}}\ dx$

21. $\int_0^1 \frac{dx}{\sqrt{x}\sqrt{1-x}}$ This integrand is undefined at both endpoints, $x=0$ and $x=1$.

22.

(a) For which values of k is $\int_0^1 x^k\ dx$ improper.

(b) For which values of k is $\int_0^1 x^k$ a convergent improper integral?

(c) For which values of k is $\int_0^1 x^k$ a divergent improper integral?

23.

(a) For which values of k is $\int_1^\infty x^k\, dx$ convergent?

(b) For which values of k is $\int_1^\infty x^k\, dx$ divergent?

24.

(a) For which positive constants p is $\int_0^1\, dx/x^p$ convergent? divergent?

(b) For which positive constants p is $\int_1^\infty\, dx/x^p$ convergent? divergent?

(c) For which positive constants p is $\int_0^\infty\, dx/x^p$ convergent? divergent?

25. Let R be the region between the curves $y = 1/x$ and $y = 1/(x+1)$ to the right of the line $x = 1$. Is the area of R finite or infinite? If it is finite, evaluate it.

26. Let R be the region between the curves $y = 1/x$ and $y = 1/x^2$ to the right of $x = 1$. Is the area of R finite or infinite? If it is finite, evaluate it.

27. Describe how you would go about estimating $\int_0^\infty e^{-x^2}\, dx$ with an error less than 0.02. (Do not do the arithmetic.)

28. Describe how you would go about estimating $\int_0^\infty \frac{dx}{\sqrt{1+x^4}}$ with an error less than 0.01. (Do not do the arithmetic.)

29. Example 4 showed that $\int_0^\infty e^{-x}\sin(x)\, dx$ is convergent. Find its value. (First find constants A and B such that $Ae^{-x}\sin(x) + Be^{-x}\cos(x)$ is an antiderivative of $e^{-x}\sin(x)$.)

30. In Example 5 the area of the region bounded by $y = 1/\sqrt{x}$, $x = 1$, and the coordinate axes was found to have area 2. Confirm this result by using horizontal cross sections and evaluating an improper integral from 0 to ∞.

31. The function $f(x) = \frac{\sin(x)}{x}$ for $x \neq 0$ and $f(0) = 1$ occurs in communication theory. Show that the energy E of the signal represented by f is finite, where

$$E = \int_{-\infty}^{\infty} (f(x))^2 \, dx.$$

32. Let $f(x)$ be a positive function and let R be the region under $y = f(x)$ and above $[1, \infty]$. Assume that the area of R is infinite. Does it follow that the volume of the solid of revolution formed by revolving R about the x-axis is infinite?

33. Let $f(x)$ be a positive function and let R be the region under $y = f(x)$ and above $[1, \infty]$. Assume that the area of R is finite. Does it follow that the volume of the solid of revolution formed by revolving R about the x-axis is infinite?

34.

(a) Sketch the graph of $y = 1/x$, for $x > 0$.

(b) Is the part below the graph and above $(0, 1]$ congruent to the part below the graph and above $[1, \infty)$?

(c) What does this say about the convergence or divergence of $\int_0^1 \frac{dx}{x}$ and $\int_1^\infty \frac{dx}{x}$?

35.

(a) Sketch the graph of $y = 1/x^2$ for $x > 0$.

(b) Is the part below the graph and above $(0, 1]$ congruent to the part below the graph and above $[1, \infty)$?

(c) What does this say about the convergence or divergence of $\int_0^1 \frac{dx}{x^2}$ and $\int_1^\infty \frac{dx}{x^2}$?

(d) What does this say about the convergence or divergence of $\int_0^1 \frac{dx}{\sqrt{x}}$ and $\int_1^\infty \frac{dx}{\sqrt{x}}$?

36. In the study of the harmonic oscillator one meets the integral

$$\int_{-\infty}^{\infty} \frac{dx}{(1 + kx^2)^3},$$

where k is a positive constant. Show this improper integral is convergent.

37. If $\int_0^\infty e^{-x^2}\ dx = \sqrt{\pi}/2$, show that $\int_0^\infty 2^{-x^2}\ dx = \sqrt{\pi}/\ln(4)$.

38. Consider the improper integral $\int_0^1 \frac{dx}{x^2}$. Suppose the interval $[0, 1]$ is partitioned into n equal-width pieces. That is $x_i = i/n$ for all $i = 0, 1, \ldots, n$.

(a) Show that the approximating sum $S_n = \sum_{i=1}^n \frac{1}{c_i^2}\Delta x_i = \sum_{i=1}^n \frac{n}{i^2}$.

(b) Show that $\lim_{n\to\infty} S_n$ does not exist. (Show that $S_n \geq n$ for all positive integers n.)

39. Plankton are small football-shaped organisms. The resistance they meet when falling through water is proportional to the integral

$$\int_0^\infty \frac{dx}{\sqrt{(a^2+x)(b^2+x)(c^2+x)}},$$

where a, b, and c describe the dimensions of the plankton. Is this improper integral convergent or divergent? (Explain.)

40. In R. P. Feynman, *Lectures on Physics*, Addison-Wesley, Reading, MA, 1963, appears this remark: "...the expression becomes

$$\frac{U}{V} = \frac{(kT)^4}{\hbar^3\pi^2c^3}\int_0^\infty \frac{x^3\ dx}{e^x - 1}.$$

This integral is just some number that we can get, approximately, by drawing a curve and taking the area by counting squares. It is roughly 6.5. The mathematicians among us can show that the integral is exactly $\pi^4/15$." Show at least that the integral is convergent.

41.

(a) Assume that $f(x)$ is continuous and nonnegative and that $\int_1^\infty f(x)\ dx$ is convergent. Show by sketching a graph that $\lim_{x\to\infty} f(x)$ may not exist.

(b) Show that if we add the condition that f is a decreasing function, then $\lim_{x\to\infty} f(x) = 0$.

42. Here is the standard proof of the absolute convergence test. Assume that $\int_0^\infty |f(x)|\ dx$ converges. Let $g(x) = f(x) + |f(x)|$. Note that $0 \leq g(x) \leq 2|f(x)|$. Thus $\int_0^\infty g(x)\ dx$ converges, that is, $\int_0^\infty (f(x) + |f(x)|)\ dx$ converges. It follows, since $f(x) = (f(x) + |f(x)|) - |f(x)|$, that $\int_0^\infty f(x)\ dx$ converges.

(a) Study this proof.

(b) State the advantages and disadvantages of each proof, the standard one and the proof in the text.

*****43.** In the proof of the Absolute Convergence Test for Improper Integrals (Theorem 7.8.3), we assumed that the functions g and h are continuous. They are, as the following steps show:

(a) Show that $|f(x)| = \sqrt{(f(x))^2)}$.

(b) Show that if $f(x)$ is continuous, so is $|f(x)|$.

(c) Show that $g(x) = \frac{1}{2}(f(x) + |f(x)|)$.

(d) Deduce that g is continuous.

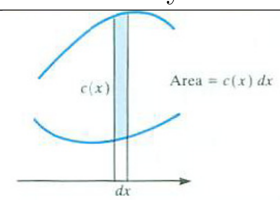
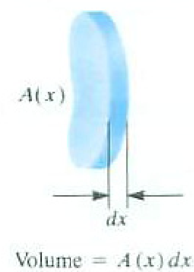
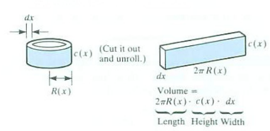
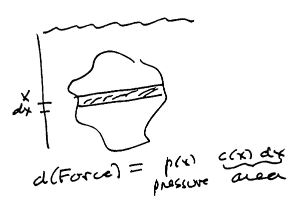
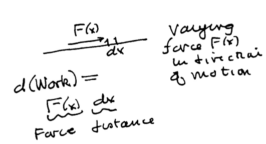
(e) Deduce that h is continuous.

44. If A is in $[0, L]$ and B is in $[-L, 0]$, why is $A + B$ in $[-L, L]$?

7.S Chapter Summary

There are two ideas in this chapter. One is "make a large, clear drawing when setting up a definite integral." The other is "make a local estimate of the total quantity" — whether that quantity is area, volume, force of water, work, or something altogether different. If the local estimate is $f(x)\ dx$, the total quantity is represented by a definite integral $\int_a^b f(x)\ dx$ (or an improper integral).

The following table summarizes some of the applications of the definite integral.

<table>
<tr><th>Section</th><th>Concept</th><th>Memory Aid</th></tr>
<tr><td>7.1</td><td>Area $= \int_a^b c(x)\ dx$</td><td>
</td></tr>
<tr><td>7.4</td><td>Volume $= \int_a^b A(x)\ dx$
Special Case: Solid of revolution (perpendicular cross sections)</td><td>
</td></tr>
<tr><td>7.5</td><td>Volume $= \int_a^b 2\pi R(x)c(x)\ dx$
Special Case: Solid of revolution (parallel cross sections)</td><td>
</td></tr>
<tr><td>7.6</td><td>Force of water</td><td>
</td></tr>
<tr><td>7.7</td><td>Work</td><td>
</td></tr>
</table>

The final section, on improper integrals, shows how to deal with integrals over infinite intervals, which are surprisingly common, and integrands that become infinite (much less common).

EXERCISES for 7.S

1. Consider the parabola $y = x^2$ and two points on it, $P = (a, a^2)$ and $Q = (b, b^2)$.

(a) Show that the tangent to the parabola at the midpoint between P and Q, $R = \left(\frac{a+b}{2}, \left(\frac{a+b}{2}\right)^2\right)$ is parallel to the chord PQ.

(b) Show that the area of the parabola below the chord is $(b-a)^3/6$.

(c) Show that the area of triangle PQR is $(b-a)^3/4$.

Archimedes proved that the area of the parabolic section under PQ is $4/3$ the area of triangle PQR. See S. Stein, *Archimedes: What did he do besides cry Eureka?*, MAA, Washington, DC, 1999 (pp. 51–60).

2.

(a) The exponential function is an increasing function for all x. Use this fact to show that $e^x > 1$ for all $x > 0$.

(b) Suppose $f(t) > g(t)$ for all $t > a$. Explain why $\int_a^x f(t)\,dt > \int_a^x g(t)\,dt$ for all $x > a$.

(c) Use (b) to show that $e^x > 1 + x$ for all $x > 0$.

(d) Use (b) and (c) to show that $e^x > 1 + x + \frac{x^2}{2}$ for all $x > 0$.

3. Extend the argument in Exercise 2 to show that $e^x > \sum_{i=0}^{n+1} \frac{x^i}{i!}$. Use this fact to show that for any fixed integer n, $\lim_{x\to\infty} \frac{x^n}{e^x} = 0$.

4. The average distance of an electron from the nucleus of a hydrogen atom involves the integral

$$\int_0^\infty e^{-x} x^5\,dx.$$

Show that it is convergent. (Its value is $5! = 120$).

5. If $\int_0^\infty f(x)\,dx$ is convergent, does it follow that

(a) $\lim_{x\to\infty} f(x) = 0$?

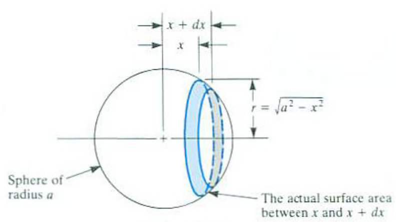
(b) $\lim_{x\to\infty} \int_x^{x+0.1} f(t)\ dt = 0$?

(c) $\lim_{x\to\infty} \int_x^{2x} f(t)\ dt = 0$?

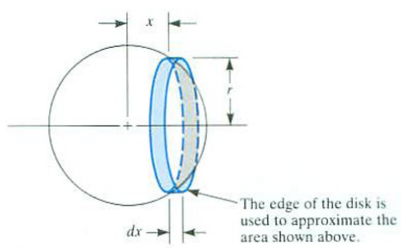
(d) $\lim_{x\to\infty} \int_x^{\infty} f(t)\ dt = 0$?

Compare with Exercise 18 in Chapter 11.

6. Consider the following argument: "Approximate the surface area of the sphere of radius a shown in Figure 7.S.1(a) as follows. To approximate the surface area between x and $x+dx$, let us try using the area of the narrow curved part of the cylinder used to approximate the volume between x and $x+dx$. (This part is shaded in Figure 7.S.1(b).) This local approximation can be pictured (when unrolled and laid flat) as a rectangle of width dx and length $2\pi r$. The surface area of a sphere is $\int_{-a}^{a} 2\pi r\ dx = 4\pi \int_0^a \sqrt{a^2-x^2}\ dx$. But $\int_0^a \sqrt{a^2-x^2}\ dx = \pi a^2/4$, since it equals the area of a quadrant of a disk. Hence the area of the sphere is then $\pi^2 a^2$." This does not agree with the correct value, $4\pi a^2$, which was discovered by Archimedes in the third century B.C. What is wrong with this argument?



(a)



(b)

Figure 7.S.1

7. Determine if the following improper integral converges or diverges: $\int_0^{\infty} \frac{x\ dx}{\sqrt{1+x^4}}$

8. The probability that ball bearing A survives at least until time t will be denoted as $F(t)$. For ball bearing B let $G(t)$ be the probability that it survives at least until time t.

(a) Show that the probability that A lasts at least as long as B is $-\int_0^{\infty} F(t)G'(t)\ dt$.

(b) Similarly, the probability that B lasts at least as long as A is $-\int_0^{\infty} G(t)F'(t)\ dt$. Assume that the probability that A and B last exactly the same time is 0. Why should $-\int_0^{\infty} F(t)G'(t)\ dt - \int_0^{\infty} G(t)F'(t)\ dt = 1$? Show that it does equal 1.

In Exercise 9 assume $\int_{-\infty}^{\infty} e^{-x^2}\ dx = \sqrt{\pi}$, which will be established in Section 17.3 (see Exercise 32 on 1478).
Let μ and σ be constants. The normal distribution, also called the Gaussian distribution and the bell curve, is given by the density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right).$$

9.

(a) Show that the graph of f is symmetric with respect to the line $x = \mu$.

(b) Show that $\int_{-\infty}^{\infty} f(x)\ dx = 1$.

(c) Show that $\int_{-\infty}^{\infty} xf(x)\ dx = \mu$. μ is the average value of x, and is called the **mean of the distribution**.

(d) Show that $\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\ dx = \sigma^2$. σ^2, called the **variance**, measures the deviation of x from the mean. The number σ is called the **standard deviation** of the distribution. Both measure the tendency of the data to spread out away from the mean.

(e) Show that $f(x)$ has two inflection points, which occur when $x = \mu + \sigma$ or $x = \mu - \sigma$.

(f) Sketch the graph of a typical $f(x)$.

The normal distribution, first introduced in Exercises 100 to 104 in Section 5.8, is defined for a variable that can take on both positive and negative values. However, such variables as incomes, life spans, amounts of rainfall, scores on examinations, and ages of first marriages, do not assume negative values. In these cases it may be more appropriate to use a **log-normal distribution**, which is defined only for $(0, \infty)$. (See, for instance, *The Lognormal Distribution*, by economists J. Atchison and J. A. C. Brown, 1957.)
Let $f(x)$ be the density in a normal distribution. The density, $g(x)$, of the log-normal distribution is defined, for $a > 0$, by the equation

$$\int_0^a g(x)\ dx = \int_{-\infty}^{\ln(a)} f(x)\ dx.$$

This says, "the probability that x is at most a is the probability that $\ln(x)$ is at most $\ln(a)$, as given by the normal distribution."
10. In this problem $f(x)$ is the density of a normal distribution with mean μ and variance σ^2 and $g(x)$ is the density of the corresponding log-normal distribution.

(a) Show that $g(x) = \frac{1}{x} f(\ln(x))$ for $x > 0$.

(b) Show that $\int_0^\infty g(x)\ dx = 1$.

(c) Show that the mean value of the log-normal distribution, $\int_0^\infty xg(x)\ dx$, equals $e^{\mu+\frac{\sigma^2}{2}}$.

(d) Show that $\lim_{x\to\infty} g(x) = 0$.

(e) Show that $\lim_{x\to 0^+} g(x) = 0$.

(f) Show that the maximum of $g(x)$ occurs when x is $e^{\mu-\sigma^2}$.

(g) What is the maximum of $g(x)$?

(h) Show that $\int_0^{e^\mu} g(x)\ dx = \int_{e^\mu}^\infty g(x)\ dx$. Thus, half the area under the curve $y = g(x)$ lies to the left of e^μ.

(i) Sketch the general shape of the log-normal distribution. Remember that $g(x)$ is defined only for x in $(0, \infty)$.

11.

(a) Draw the curve $y = e^x/x$ for $x > 0$, showing any asymptotes or critical points.

(b) Find the number t such that the area below $y = e^x/x$ and above the interval $[t, t+1]$ is a minimum.

(Write $A(t) = \int_t^{t+1} f(x)\ dx = \int_0^{t+1} f(x)\ dx - \int_0^t f(x)\ dx$, then use FTC II.)

Skill Drill: Derivatives

In Exercises 12 to 14 a, b, c, m, and p are constants. In each case verify that the derivative of the first function is the second function.

12. $\frac{x}{a} - \frac{1}{ap}\ln\left(a + be^{px}\right)$; $\frac{1}{a+be^{px}}$.

13. $\frac{1}{\sqrt{-c}}\arcsin\left(\frac{-cx-b}{\sqrt{b^2-4ac}}\right)$; $\frac{1}{\sqrt{a+bx+cx^2}}$, for any negative number c.

14. $\frac{1}{c}\ln\left(\sqrt{z+bx+cx^2} + x\sqrt{c} + \frac{b}{2\sqrt{c}}\right)$; $\frac{1}{\sqrt{a+bx+cx^2}}$, for any positive number c.

Calculus is Everywhere # 9

Escape Velocity

In Example 2 in Section 7.7 we saw that the total work required to lift a 1-pound payload from the surface of the earth to the moon is 3,933 mile-pounds. Since the radius of the earth is about 4,000 miles, the work required to launch a payload on an endless journey is given by the improper integral

$$\int_{4{,}000}^{\infty} \left(\frac{4,000}{r}\right)^2 dr = 4,000 \text{ mile-pounds.}$$

Because the integral is convergent, only a finite amount of energy is needed to send a payload on an endless journey — as the Voyager spacecraft has demonstrated. It takes only a little more energy than is required to lift the payload to the moon.

That the work required for the endless journey is finite raises the question "With what initial velocity must we launch the payload so that it never falls back, but continues to rise forever away from the earth?" If the initial velocity is too small, the payload will rise for a while, then fall back, as anyone who has thrown a ball straight up knows quite well.

The energy we supply the payload is kinetic energy. The force of gravity slows the payload and reduces its kinetic energy. We do not want the kinetic energy to shrink to zero. It it were ever zero, then the velocity of the payload would be zero. At that point the payload would start to fall back to earth.

As we will show, the kinetic energy of the payload is reduced by *exactly* the amount of work done on the payload by gravity. If v_{esc} is the minimal velocity needed for the payload to "escape" and not fall back, then

$$\frac{1}{2}mv_{\text{esc}}^2 = 4,000 \text{ mile-pounds,} \tag{C.9.1}$$

where m is the mass of the payload. Equation (C.9.1) can be solved for v_{esc}, the **escape velocity**.

In order to solve (C.9.1) for v_{esc}, we must calculate the mass of a payload that weighs 1 pound at the surface of the earth. The weight of 1 pound is the gravitational force of the earth pulling on the payload. Newton's equation

$$\text{Force} = \text{Mass} \times \text{Acceleration}, \tag{C.9.2}$$

known as his "second law of motion," provides the relationship among force, mass, and the acceleration of that mass that is needed.

The acceleration of an object at the surface of the earth is 32 feet per second per second, or 0.0061 miles per second per second. Then (C.9.2), for the 1-pound payload, becomes

$$1 = m(0.0061). \tag{C.9.3}$$

Combining (C.9.1) and (C.9.3) gives

$$\begin{aligned} \frac{1}{2}\frac{1}{0.0061}\left(v_{\text{esc}}\right)^2 &= 4,000 \\ \text{or} \qquad \left(v_{\text{esc}}\right)^2 = (8,000)(0.0061) &= 48.8. \end{aligned}$$

Hence $v_{\text{esc}} \approx 7$ miles per second, which is about 25,000 miles per hour, a speed first attained by human beings when Apollo 8 traveled to the moon in December 1968. All that remains is to justify the claim that the change in kinetic energy equals the work done by the force.

Let $v(r)$ be the velocity of the payload when it is r miles from the center of the earth. Let $F(r)$ be the force on the payload when it is r miles from the center of the earth. Since the force is in the opposite direction from the motion, we will define $F(r)$ to be negative.

Let a and b be numbers, $4,000 \leq a < b$. (See Figure C.9.1.) We wish to show that

$$\underbrace{\frac{1}{2}m(v(b))^2 - \frac{1}{2}m(v(a))^2}_{\text{change in kinetic energy}} = \underbrace{\int_a^b F(r)dr}_{\text{work done by gravity}} \quad . \tag{C.9.4}$$

Figure C.9.1

In this equation m is the payload mass. Note that both sides of (C.9.4) are negative.

Equation (C.9.4) resembles the Fundamental Theorem of Calculus. If we could show that $\frac{1}{2}m(v(r))^2$ is an antiderivative of $F(r)$, then (C.9.4) would follow immediately. Let us find the derivative of $\frac{1}{2}m(v(r))^2$ with respect to r and show that it equals $F(r)$:

$$\begin{aligned} \frac{d}{dr}\left(\frac{1}{2}m(v(r))^2\right) &= mv(r)\frac{dv}{dr} = mv(r)\frac{dv/dt}{dr/dt} && \text{(Chain Rule; } t \text{ is time)} \\ &= mv(r)\frac{d^2r/dt^2}{v(r)} = m\frac{d^2r}{dt^2} && \left(v(r) = \frac{dr}{dt}\right) \\ &= \text{mass} \times \text{acceleration} \\ &= F(r) && \text{(Newton's 2}^{\text{nd}} \text{ Law of Motion.} \end{aligned}$$

Hence (C.9.4) is valid and we have justified our calculation of escape velocity.

Incidentally, the escape velocity is $\sqrt{2}$ times the velocity required for a satellite to orbit the earth (and not fall into the atmosphere and burn up).

EXERCISES 1. The earth is not a perfect sphere. The"mean radius" of the earth is about 3,959 miles. A more accurate value for the force of gravity is 32.174 feet per second per second. Repeat the derivation of the escape velocity using these values. References: `http://en.wikipedia.org/wiki/Earth_radius` and `http://en.wikipedia.org/wiki/Standard_gravity`.

2. Repeat the derivation of the escape velocity using CGS units. That is, assume the radius of the earth is 6,371 kilometers and the force of gravity is 9.80665 meters per second per second.

3. Determine the escape velocity from the moon. Find the information you need to complete this calculation?

4. Determine the escape velocity from the sun.

Calculus is Everywhere # 10
Average Speed and Class Size

There are two ways to define your average speed when jogging or driving a car. You could jot down your speed at regular intervals of time, say, every second. Then you would just average those speeds. That average is called an average *with respect to time*. Or, you could jot down your velocity at regular intervals of distance, say, every hundred feet. The average of those velocities is called an average *with respect to distance*.

How do you think they would compare? If you kept a constant speed, c, the averages would both be c. Are they always equal, even if your speed varies? Would one of the averages always tend to be larger? Try to answer the question before we analyze it mathematically, with the aid of the **Cauchy-Schwarz inequality**.

pronounced: "ko-shee' shwartz"

There are several versions of the Cauchy-Schwarz inequality. The version we need here concerns two continuous functions, f and g, defined on an interval $[a, b]$. If $\int_a^b f(x)^2\ dx$ and $\int_a^b g(x)^2\ dx$ are small, then the absolute value of $\int_a^b f(x)g(x)\ dx$ ought to be small too. It is, as the following Cauchy-Schwarz inequality implies:

$$\left(\int_a^b f(x)g(x)\ dx\right)^2 \leq \int_a^b f(x)^2\ dx \int_a^b g(x)^2\ dx. \tag{C.10.1}$$

After showing some of its applications, we will use the quadratic formula to show that it is true.

First we use the inequality (C.10.1) to answer the question, "Which average of speed is larger, the one with respect to time or the one with respect to distance?"

Let the speed at time t be $v(t)$ and let $s(t)$ be the distance traveled up to time t. During the time interval from time a to time b the average of velocity with respect to time is

$$\frac{\int_a^b v(t)\ dt}{b-a} = \frac{s(b)-s(a)}{b-a}.$$

On the other hand, the average of velocity with respect to distance is defined as

$$\frac{\int_{s(a)}^{s(b)} v(s)\ ds}{s(b)-s(a)}, \tag{C.10.2}$$

where $v(s)$ denotes the velocity when the distance covered is s. Changing the independent variable in the numerator of (C.10.2) from s to t by the relation $ds = v(t)\ dt$, we obtain

$$\frac{\int_{s(a)}^{s(b)} v(s)\ ds}{s(b)-s(a)} = \frac{\int_a^b v(t)v(t)\ dt}{s(b)-s(a)}.$$

Noting that $s(b) - s(a) = \int_a^b v(t)\ dt$ and $b - a = \int_a^b 1\ dt$, we will show that the average with respect to time is less than or equal to the average with respect to distance, that is,

$$\frac{\int_a^b v(t)\ dt}{\int_a^b 1\ dt} \leq \frac{\int_a^b v(t)^2\ dt}{\int_a^b v(t)\ dt}.$$

Or, equivalently,

$$\left(\int_a^b v(t)\ dt\right)^2 \leq \int_a^b 1\ dt \int_a^b v(t)^2\ dt. \tag{C.10.3}$$

But, (C.10.3) is a special case of (C.10.1), with $f(t) = 1$ and $g(t) = v(t)$.

This implies that the average with respect to time is always less than or equal to the average with respect to distance. Exercise 1 shows a bit more: if the speed is not constant, then the average with respect to time is less than the average with respect to distance.

The way to show that inequality (C.10.1) holds is indirect but short. Introduce a new function, $h(t)$, defined by

$$h(t) = \int_a^b (f(x) - tg(x))^2\ dx = \int_a^b f(x)^2\ dx - 2t\int_a^b f(x)g(x)\ dx + t^2\int_a^b g(x)^2\ dx. \tag{C.10.4}$$

Because the first integrand in (C.10.4) is never negative, $h(t) \geq 0$. Now, $h(t) = pt^2 + qt + r$, where

$$p = \int_a^b g(x)^2\ dx, \qquad q = -2\int_a^b f(x)g(x)\ dx, \quad \text{and} \quad r = \int_a^b f(x)^2\ dx.$$

The parabola $y = h(t)$ never drops below the t-axis, and touches the t-axis at at most one point. Otherwise, if it touches the t-axis at two points, it would dip below that axis, forcing $h(t)$ to take on some negative values.

Because the equation $h(t) = 0$ has at most one solution, the discriminant $q^2 - 4pr$ must not be positive. Thus, $q^2 - 4pr \leq 0$, from which the Cauchy-Schwarz inequality follows.

EXERCISES

1. Show that the only case when equality holds in (C.10.1) is when $g(x)$ is a constant times $f(x)$.

2. The "discrete" form of the Cauchy-Schwarz inequality asserts that if a_1, a_2, a_3, $\ldots$, a_n and b_1, b_2, b_3, $\ldots$, b_n are real numbers, then

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leq \sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} a_i^2.$$

(a) Prove this inequality.

(b) When does equality hold?

3. Use the inequality in Exercise 2 to show that the average class size at a university as viewed by the registrar is usually smaller than the average class size as viewed by the students.

It is also the case that the average time between buses as viewed by the dispatcher is usually shorter than the average time between buses as viewed by passengers arriving randomly at a bus stop.
Reference: S. K. Stein, *An Inequality Between Two Average Speeds*, Transportation Research 22B (1988), pp. 469–471.

4. A region R is bounded by the x-axis, the lines $x = 2$ and $x = 5$, and the curve $y = f(x)$, where f is a positive function. The area of R is A. When revolved around the x-axis it produces a solid of volume V.

(a) How large can V be?

(b) How small can V be?

(In one of these two cases the Cauchy-Schwarz inequality on 679 may help.)

5. If the region R in the preceding exercise is revolved around the y-axis, what can be said about the maximum and minimum values for the volume of the resulting solid? Explain.

Chapter 8

Computing Antiderivatives

In Chapter 7 we saw several uses for definite integrals in geometry and physics. Similar applications of integration can be found in many other fields, including economics, engineering, biology, and statistics. Definite integrals are usually either evaluated using the Fundamental Theorem of Calculus or estimated numerically, as discussed in Section 6.5.

To evaluate $\int_a^b f(x)\ dx$ by the Fundamental Theorem of Calculus (FTC I) we need to find an antiderivative $F(x)$ of the integrand $f(x)$; then $\int_a^b f(x)\ dx$ is simply $F(b) - F(a)$. This chapter describes techniques for finding an antiderivative.

The problem of finding an antiderivative differs from that of finding a derivative in two important ways. First, the antiderivatives of some elementary functions, such as e^{x^2}, are not elementary. On the other hand, as we saw in Chapter 3, the derivatives of all elementary functions are elementary.

Second, a slight change in the form of a function can cause a great change in the form of its antiderivative. For instance,

$$\int \frac{dx}{x^2+1} = \arctan(x) + C \qquad \text{while} \qquad \int \frac{x\ dx}{x^2+1} = \frac{1}{2}\ln(x^2+1) + C,$$

as you may check by differentiating $\arctan(x)$ and $\frac{1}{2}\ln(x^2+1)$. On the other hand, a slight change in the form of an elementary function produces only a slight change in the form of its derivative.

There are three ways to find an antiderivative:

- By hand, using techniques described in this chapter
- By an integral table
- By computer, calculator, or other automated integrator.

Section 8.1 illustrates a few shortcuts, describes how to use integral tables, and discusses the strengths and weaknesses of computer-based evaluation of integrals.

Section 8.2 presents "substitution," the most important technique for finding an antiderivative.

Section 8.3 describes"integration by parts," a technique that has many uses, such as in solving differential equations, besides finding antiderivatives.

Section 8.4 discusses the integration of rational functions.

Section 8.5 describes how to integrate some special integrands.

Section 8.6 offers an opportunity to practice the techniques when there is no clue as to which is the best to use.

8.1 Shortcuts, Tables, and Technology

In this section we list antiderivatives of some common functions and some shortcuts. Then we describe integral tables and the computation of antiderivatives by computers.

Some Common Integrands

Every formula for a derivative provides a corresponding formula for an antiderivative. For instance, since $(x^3/3)' = x^2$, it follows that

$$\int x^2\ dx = \frac{x^3}{3} + C.$$

The following miniature integral table lists a few formulas that should be memorized. Each can be checked by differentiating the right-hand side of the equation.

$\int x^a\ dx$	$=$	$\frac{x^{a+1}}{a+1} + C$	for $a \neq -1$
$\int \frac{1}{x}\ dx$	$=$	$\ln\lvert x\rvert + C$	This is $\int x^a\ dx$ for $a = -1$.
$\int \frac{f'(x)}{f(x)}\ dx$	$=$	$\ln\lvert f(x)\rvert + C$	if $f(x) > 0$, the absolute value can be omitted.
$\int (f(x))^n\ f'(x)\ dx$	$=$	$\frac{(f(x))^{n+1}}{n+1} + C$	for $n \neq -1$
$\int e^{ax}\ dx$	$=$	$\frac{e^{ax}}{a} + C$	
$\int \sin(ax)\ dx$	$=$	$\frac{-1}{a}\cos(ax) + C$	remember the negative sign
$\int \cos(ax)\ dx$	$=$	$\frac{1}{a}\sin(ax) + C$	
$\int \frac{1}{\sqrt{a^2-x^2}}\ dx$	$=$	$\arcsin\left(\frac{x}{a}\right) + C$	
$\int \frac{1}{a^2+x^2}\ dx$	$=$	$\frac{1}{a}\arctan\left(\frac{x}{a}\right) + C$	
$\int \frac{1}{\lvert x\rvert\sqrt{x^2-a^2}}\ dx$	$=$	$\frac{1}{a}\,\mathrm{arcsec}\left(\frac{x}{a}\right) + C$	

Antiderivative of a polynomial

EXAMPLE 1 Find $\int (2x^4 - 3x + 2)\ dx$.
SOLUTION

$$\begin{aligned}\int (2x^4 - 3x + 2)\ dx &= \int 2x^4\ dx - \int 3x dx + \int 2\ dx \\ &= 2\int x^4\ dx - 3\int x\ dx + 2\int 1\ dx \\ &= 2\frac{x^5}{5} - 3\frac{x^2}{2} + 2x + C\end{aligned}$$

One constant of integration is enough

$\diamond$

EXAMPLE 2 Find $\int \frac{4x^3}{x^4+1}\ dx$
SOLUTION The numerator is precisely the derivative of the denominator. Hence

Antiderivative of f'/f

$$\int \frac{4x^3}{x^4+1}\ dx = \ln|x^4+1| + C$$

Since $x^4 + 1$ is always positive, the absolute-value sign is not needed, and $\int \frac{4x^3}{x^4+1}\ dx = \ln(x^4+1) + C$. $\diamond$

Antiderivative of x^a

EXAMPLE 3 Find $\int \sqrt{x}\ dx$.
SOLUTION

$$\int \sqrt{x}\ dx \quad = \quad \int x^{1/2}\ dx = \frac{x^{1/2+1}}{\frac{1}{2}+1} + C = \frac{2}{3}x^{3/2} + C$$

$\diamond$

EXAMPLE 4 Find $\int \frac{1}{x^3}\ dx$.
SOLUTION

$$\int \frac{1}{x^3}\ dx \quad = \quad \int x^{-3}\ dx = \frac{x^{-3+1}}{-3+1} + C = -\frac{1}{2}x^{-2} + C = -\frac{1}{2x^2} + C.$$

$\diamond$

EXAMPLE 5 Find $\int (3\cos(x) - 4\sin(2x) + \frac{1}{x^2})\ dx$.
SOLUTION

$$\begin{aligned}\int (3\cos(x) - 4\sin(2x) + \frac{1}{x^2})\ dx &= 3\int \cos(x)\ dx - 4\int \sin(2x)\ dx + \int \frac{1}{x^2} dx \\ &= 3\sin(x) + 2\cos(2x) - \frac{1}{x} + C.\end{aligned}$$

$\diamond$

EXAMPLE 6 Find $\int \frac{x}{1+x^2}\, dx$.
SOLUTION If the numerator were exactly $2x$, it would be the derivative of the denominator and we would have the case $\int (f'(x)/f(x))\, dx$: the antiderivative would be $\ln(1+x^2)$. But the numerator can be multiplied by 2 if we simultaneously divide by 2:

Multiplying the integrand by a constant

$$\int \frac{x}{1+x^2}\, dx = \frac{1}{2}\int \frac{2x}{1+x^2}\, dx.$$

This step depends on the fact that a constant can be moved past the integral sign:

$$\frac{1}{2}\int \frac{2x}{1+x^2}\, dx = \frac{1}{2}\cdot 2\int \frac{x}{1+x^2}\, dx = \int \frac{x}{1+x^2}\, dx.$$

Thus

Since $1+x^2>0$, the absolute value is not needed in $\ln(1+x^2)$.

$$\int \frac{x}{1+x^2}\, dx \;=\; \frac{1}{2}\int \frac{2x}{1+x^2}\, dx = \frac{1}{2}\ln(1+x^2)+C.$$

◇

Special Shortcuts

We present three shortcuts for evaluating some special but fairly common definite integrals. These shortcuts can save a lot of work. **Shortcut 1** If f is an odd function, then

$$\int_{-a}^{a} f(x)\, dx = 0. \tag{8.1.1}$$

Explanation. Recall that for an odd function $f(-x)=-f(x)$. Figure 8.1.1 suggests why (8.1.1) holds. The shaded area to the left of the y-axis equals the shaded area to the right. As integrals, however, these two areas represent quantities of opposite sign: $\int_{-a}^{0} f(x)\, dx = -\int_{0}^{a} f(x)\, dx$.

Therefore, the definite integral over the entire interval is 0.

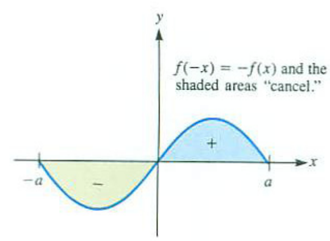


Figure 8.1.1

EXAMPLE 7 Find $\int_{-2}^{2} x^3\sqrt{4-x^2}\, dx$.
SOLUTION The function $f(x) = x^3\sqrt{4-x^2}$ is odd. (Check it.) By the shortcut,

$$\int_{-2}^{2} x^3\sqrt{4-x^2} = 0.$$

◇

Shortcut 2 $\int_0^a \sqrt{a^2-x^2}\, dx = \frac{1}{4}\pi a^2$.

Note that this shortcut applies to a particular function over a particular interval.

Explanation The graph of $y = \sqrt{a^2 - x^2}$ is part of a circle of radius a. The definite integral $\int_0^a \sqrt{a^2 - x^2}\ dx$ is a quarter of the area of that circle. (See Figure 8.1.2.)

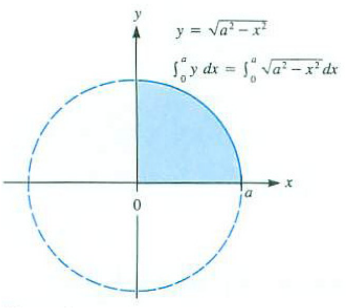


Figure 8.1.2

EXAMPLE 8 Find $\int_0^1 \sqrt{1 - x^2}\ dx$
SOLUTION Use Shortcut 2, with $a = 1$, to get

$$\int_0^1 \sqrt{1 - x^2}\ dx = \frac{\pi}{4}.$$

◇

Shortcut 3 If f is an even function,

$$\int_{-a}^{a} f(x)\ dx = 2\int_0^a f(x)\ dx.$$

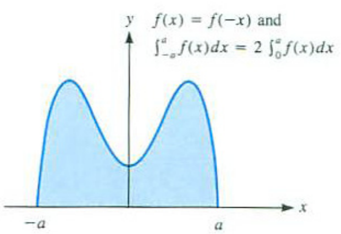


Figure 8.1.3

Explanation A glance at Figure 8.1.3 suggests why this shortcut is valid.

EXAMPLE 9 Find $\int_{-1}^1 \sqrt{1 - x^2}\ dx$.
SOLUTION Since $\sqrt{1 - x^2}$ is an even function, by Shortcut 3:

$$\int_{-1}^{1} \sqrt{1 - x^2}\ dx = 2\int_0^1 \sqrt{1 - x^2}\ dx.$$

So, by Example 8, with $a = 1$,

$$\int_{-1}^{1} \sqrt{1 - x^2} dx = 2 \cdot \frac{\pi}{4} = \frac{\pi}{2}.$$

◇

Using an Integral Table

An integral table lists antiderivatives. You will find a short integral table on the inside covers of this book. *Burington's Handbook of Mathematical Tables and Formulas*, 5th edition, McGraw-Hill, 1973, lists over 300 integrals in 33 pages. *CRC Standard Math Tables*, 30th edition, CRC Press, 1996, lists more than 700 integrals in almost 60 pages. Two Wikipedia topics devoted to tables of integration are `http://en.wikipedia.org/wiki/List_of_integrals` and `http://en.wikipedia.org/wiki/Table_of_integrals`.

Often integral tables use "log" to denote "ln"; it is understood that e is the base. Most integral tables omit the constant of integration ($+C$).

The best way to use an integral table is to browse through one (buy one, check one out from the library, or navigate to an online table). Notice how the formulas are grouped. First might come the forms that everyone uses most frequently. Then may come "forms containing $ax+b$," then "forms containing $a^2 \pm x^2$," then "forms containing ax^2+bx+c," and so on, running through many different algebraic forms. There are separate sections with trigonometric forms, logarithmic, and exponential functions. The integral table on the inside front cover is similarly grouped.

EXAMPLE 10 Use the integral table inside the cover to integrate

$$\int \frac{dx}{x\sqrt{3x+2}}.$$

SOLUTION Search until you find Formula 23,

$$\int \frac{dx}{x\sqrt{ax+b}} = \frac{1}{\sqrt{b}} \ln \left| \frac{\sqrt{ax+b}-\sqrt{b}}{\sqrt{ax+b}+\sqrt{b}} \right| \qquad b > 0,$$

and replace $ax+b$ by $3x+2$ and b by 2. Thus

$$\int \frac{dx}{x\sqrt{3x+2}} = \frac{1}{\sqrt{2}} \ln \left| \frac{\sqrt{3x+2}-\sqrt{2}}{\sqrt{3x+2}+\sqrt{2}} \right| + C.$$

◇

EXAMPLE 11 Use the integral table to integrate

$$\int \frac{dx}{x\sqrt{3x-2}}, \qquad x > 2/3.$$

SOLUTION This time we need Formula 24 with $b=-2$,

$$\int \frac{dx}{x\sqrt{ax+b}} = \frac{2}{\sqrt{-b}} \arctan\left(\sqrt{\frac{ax+b}{-b}}\right) \qquad b < 0.$$

Thus,

$$\int \frac{dx}{x\sqrt{3x-2}} = \frac{2}{\sqrt{2}} \arctan\left(\sqrt{\frac{3x-2}{2}}\right) + C$$

◇

Though the integrands in Examples 10 and 11 are similar, their antiderivatives are not.

There is no need to make a big fuss about integral tables. Be cautious and keep a cool head. Just match the patterns carefully, including any conditions on the variables and their coefficients. Note that some formulas are expressed in terms of an integral of a different integrand. In these cases you will have to search through the table more than once. (Exercises 35 and 36 illustrate this.)

Computers, Calculators, and Other Automated Integrators

Using an integral table is an exercise in "pattern matching", where you hunt for the formula that fits a particular integral. Computers are good at pattern matching, so it is not surprising that for many years computers have been used to find antiderivatives. MACSYMA is one of the earliest computer-based programs that perform the basic operations of calculus: limits, derivatives, integrals. Today, the most widely used computer algebra systems are Maple and Mathematica.

This technology is slowly creeping to handheld calculators. With such wide-ranging aids at our fingertips, calculus users do not need to rely as much on formal integration techniques or tables of integrals. What is essential is that they understand what an integral is, what it can represent, and how to utilize information obtained from an integral.

In addition to matching problems with formulas from large tables of integrals, these programs utilize various substitutions and computations to transform integrals into forms that can be evaluated.

In spite of the availability of integral tables, and computer programs, it is often simpler to use one of the techniques described later in this chapter.

EXERCISES for Section 8.1

In Exercises 1 to 14 find the integrals. Use the short list at the beginning of the section.

1. $\int 5x^3 \; dx$

2. $\int (8 + 11x) \; dx$

3. $\int x^{1/3} \; dx$

4. $\int \sqrt[3]{x^2} \; dx$

5. $\int \frac{6 \; dx}{x^2}$

6. $\int \frac{dx}{x^3}$

7. $\int 5e^{-2x} \; dx$

8. $\int \frac{5 \; dx}{1 + x^2}$

9. $\int \frac{6 \; dx}{|x|\sqrt{x^2 - 1}}$

10. $\int \frac{5 \; dx}{\sqrt{1 - x^2}}$

11. $\int \frac{4x^3 \; dx}{1 + x^4}$

12. $\int \frac{e^x \; dx}{1 + e^x}$

13. $\int \frac{\sin(x) \; dx}{1 + \cos(x)}$

14. $\int \frac{dx}{1 + 3x}$

In Exercises 15 to 20, change the integrand into an easier one by algebra and find the antiderivative.

15. $\int \frac{1 + 2x}{x^2} \; dx$ ($\frac{a+b}{c} = \frac{a}{c} + \frac{b}{c}$)

16. $\int \frac{1 + 2x}{1 + x^2} \; dx$

17. $\int (x^2 + 3)^2 \; dx$ (First multiply out the integrand.)

18. $\int (1 + e^x)^2 \; dx$

19. $\int (1+3x)x^2 \, dx$

20. $\int \frac{1+\sqrt{x}}{x} \, dx$

21. A shortcut for $\int_0^{\pi/2} \sin^2(\theta) \, d\theta$.)

(a) Why would you expect $\int_0^{\pi/2} \cos^2(\theta) \, d\theta$ to equal $\int_0^{\pi/2} \sin^2(\theta) \, d\theta$?

(b) Why is $\int_0^{\pi/2} \sin^2(\theta) \, d\theta + \int_0^{\pi/2} \cos^2(\theta) \, d\theta = \pi/2$.

(c) Conclude that $\int_0^{\pi/2} \sin^2(\theta) \, d\theta = \pi/4$.

The integrals in Exercises 22 to 28 can be evaluated using one of the shortcuts. (Is the integrand even or odd? Can you relate the integral to a known area? Recall the result of Exercise 21: $\int_0^{\pi/2} \cos^2(x) \, dx = \frac{\pi}{4} = \int_0^{\pi/2} \sin^2(x) \, dx$.)

22. $\int_{-1}^{1} x^5 \sqrt{1+x^2} \, dx$

23. $\int_{-\pi/2}^{\pi/2} \sin(3x)\cos(5x) \, dx$

24. $\int_{-1}^{1} x^5 \sqrt[4]{1-x^2} \, dx$

25. $\int_{-\pi}^{\pi} \sin^3(x) \, dx$

26. $\int_0^5 \sqrt{25-x^2} \, dx$

27. $\int_{-3}^{3} \sqrt{9-x^2} \, dx$

28. $\displaystyle\int_{-3}^{3} (x^3\sqrt{9-x^2} + 10\sqrt{9-x^2})\ dx$

In Exercises 29 to 34 find the antiderivative with the aid of a table of integrals, such as the one inside the front cover.

29.

(a) $\displaystyle\int \frac{dx}{(3x+2)^2}$

(b) $\displaystyle\int \frac{dx}{x(3x+2)}$

30.

(a) $\displaystyle\int \frac{dx}{x\sqrt{3x+4}}$

(b) $\displaystyle\int \frac{dx}{x^2\sqrt{3x+4}}$

31.

(a) $\displaystyle\int \frac{dx}{x\sqrt{3x-4}}$

(b) $\displaystyle\int \frac{dx}{x^2\sqrt{3x-4}}$

32.

(a) $\displaystyle\int \frac{dx}{4x^2+9}$

(b) $\displaystyle\int \frac{dx}{4x^2-9}$

33.

(a) $\displaystyle\int \frac{dx}{x^2+8x+7}$

(b) $\displaystyle\int \frac{dx}{x^2+2x+5}$

34.

(a) $\displaystyle\int \frac{dx}{\sqrt{11-x^2}}$

(b) $\displaystyle\int \frac{dx}{\sqrt{11+x^2}}$

35. Using the integral table on the inside front cover, find $\displaystyle\int \frac{x\ dx}{\sqrt{2x^2+x+5}}$. (Use Formula 39 first, followed by Formula 38.)

36. Using the integral table in the inside front cover, find

(a) $\displaystyle\int \frac{dx}{\sqrt{3x^2+x+2}}$

(b) $\displaystyle\int \frac{dx}{\sqrt{-3x^2+x+2}}$

8.2 The Substitution Method

This section describes the substitution method, which changes an integrand, preferably to one that we can integrate more easily. Several examples will illustrate the technique, which is the chain rule in disguise. Sometimes we can use a substitution to transform an integral not listed in an integral table to one that is listed. After the examples, the basis of the substitution method is provided.

The Substitution Method

EXAMPLE 1 Find $\int \sin(x^2)\ 2x\ dx$.
SOLUTION Note that $2x$ is the derivative of x^2. Make the substitution $u = x^2$. The differential of u is $du = \frac{d}{dx}(x^2)\ dx = 2x\ dx$ and so

$$\int (\sin(x^2))2x\ dx = \int \sin(u)\ du.$$

It is easy to find $\int \sin(u)\ du$:

$$\int \sin(u)\ du = -\cos(u) + C.$$

Replacing u by x^2 in $-\cos(u)$ yields $-\cos(x^2)$. Thus

Check the answer using the chain rule

$$\int \sin(x^2)2x\ dx = -\cos(x^2) + C.$$

◇

Contrast Example 1 with $\int \sin(x^2)\ dx$, which is not elementary. The presence of $2x$, the derivative of x^2, made it easy to find $\int (\sin(x^2))2x\ dx$.

Description of the Substitution Method

In Example 1, the integrand $f(x)$ could be written in the form

$$\boxed{f(x) = \underbrace{g(h(x))}_{\text{function of } h(x)} \times \underbrace{h'(x)}_{\text{derivative of } h(x),} \qquad (8.2.1)}$$

for some function $h(x)$. To put it another way, the expression $f(x)\ dx$ could be written as

$$f(x)\ dx = \underbrace{g(h(x))}_{\text{function of } h(x)} \times \underbrace{h'(x)}_{\text{derivative of } h(x),} dx \qquad (8.2.2)$$

Whenever this is the case, the substitution of u for $h(x)$ and du for $h'(x)\ dx$ transforms $\int f(x)\ dx$ to another integral, one involving u instead of x, $\int g(u)\ du$.

If you can find an antiderivative $G(u)$ of $g(u)$, replace u by $h(x)$. The resulting function, $G(h(x))$, is an antiderivative of $f(x)$. (This claim will be justified at the end of the section.)

The process of using substitution to evaluate an indefinite integral can be summarized as follows:

$$\int f(x)\ dx = \int g(h(x))\ h'(x)\ dx = \int g(u)\ du = G(u) + C = G(h(x)) + C.$$

EXAMPLE 2 Find $\int (1+x^3)^5\ x^2\ dx$.
SOLUTION The derivative of $1+x^3$ is $3x^2$, which differs from the x^2 in the integrand only by the constant factor 3. So let $u = 1+x^3$. Hence

$$du = 3x^2\ dx \qquad \text{and} \qquad \frac{du}{3} = x^2\ dx. \qquad (8.2.3)$$

Then

$$\int (1+x^3)^5 x^2\ dx \ = \ \int u^5 \frac{du}{3} = \frac{1}{3}\int u^5\ du = \frac{1}{3}\frac{u^6}{6} + C = \frac{(1+x^3)^6}{18} + C.$$

◇

If the factor x^2 were not present in the integrand in Example 2, you could still compute $\int (1+x^3)^5\ dx$. In this case you would have to multiply out $(1+x^3)^5$, which would be a polynomial of degree 15.

As Example 2 shows, you don't need exactly "derivative of $h(x)$" as a factor. Just "a constant times the derivative of $h(x)$" will do.

Similarly, $\int \frac{x^2}{\sqrt{1+x^3}}\ dx$ is easy (use $u = 1+x^3$), but $\int \frac{dx}{\sqrt{1+x^3}}$ is not elementary. The presence of x^2 makes a great difference.

Substitution in a Definite Integral

The substitution technique, or "change of variables," extends to definite integrals, $\int_a^b f(x)\ dx$, with one important proviso:

When making the substitution from x to u, be sure to replace the interval $[a,b]$ by the interval whose endpoints are $u(a)$ and $u(b)$.

An example will illustrate the necessary change in the limits of integration. The technique is justified in Theorem 8.2.2.

EXAMPLE 3 Evaluate $\int_1^2 3(1+x^3)^5\, x^2\, dx$.
SOLUTION Let $u = 1+x^3$. Then $du = 3x^2\, dx$. Furthermore, as x goes from 1 to 2, $u = 1+x^3$ goes from $1+1^3 = 2$ to $1+2^3 = 9$. Thus

This is the last you see of x.

$$\int_1^2 3(1+x^3)^5 x^2\, dx = \int_2^9 u^5\, du = \left.\frac{u^6}{6}\right|_2^9 = \frac{9^6-2^6}{6}.$$

Once you make the substitution in the integrand and the limits of integration, you work only with expressions involving u. There is no need to bring back x. ◇

The remaining examples present integrals needed in Section 8.4. They also show how some formulas in integral tables are obtained.

EXAMPLE 4 Integral tables include a formula for (a) $\int dx/(ax+b)$ and (b) $\int dx/(ax+b)^n$, $n \neq 1$. Obtain the formulas by using the substitution $u = ax+b$.
SOLUTION (a) Let $u = ax+b$. Hence $du = a\, dx$ and therefore $dx = du/a$. Thus

This is Formula 12 from the integral table.

$$\int \frac{dx}{ax+b} = \int \frac{du/a}{u} = \frac{1}{a}\int \frac{du}{u} = \frac{1}{a}\ln|u| + C = \frac{1}{a}\ln|ax+b| + C.$$

(b) The same substitution $u = ax+b$ gives

$$\begin{aligned} \int \frac{dx}{(ax+b)^n} &= \int \frac{du/a}{u^n} = \frac{1}{a}\int u^{-n}\, du = \frac{1}{a}\frac{u^{-n+1}}{(-n+1)} + C \\ &= \frac{(ax+b)^{-n+1}}{a(-n+1)} + C = \frac{1}{a(-n+1)(ax+b)^{n-1}} + C. \end{aligned}$$

◇

In the next Example we use u instead of x, to simplify Example 6.

EXAMPLE 5 Find $\displaystyle\int \frac{du}{4u^2+9}$.
SOLUTION $\displaystyle\int \frac{du}{4u^2+9}$ resembles $\displaystyle\int \frac{du}{u^2+1}$. This suggests rewriting $4u^2$ as $9t^2$, so we could then factor the 9 out of $9t^2+9$, getting $9(t^2+1)$. Here are the details.

Introduce t so $4u^2 = 9t^2$. To do this let $2u = 3t$, so $u = (3/2)\ t$. Then $du = (3/2)\ dt$. Also, $t = (2/3)u$. With this substitution we have

$$\begin{aligned}\int \frac{du}{4u^2+9} &= \int \frac{(3/2)\ dt}{9t^2+9} = \frac{3}{2}\cdot\frac{1}{9}\int\frac{dt}{t^2+1}\\ &= \frac{1}{6}\arctan(t) + C = \frac{1}{6}\arctan\left(\frac{2u}{3}\right) + C.\end{aligned}$$

◇

The next example uses a substitution together with "completing the square." To complete the square in the quadratic expression $x^2 + bx + c$ means adding and subtracting $(b/2)^2$ so that we get the simpler form "$v^2 + k$" where k is a constant:

$$x^2 + bx + \left(\frac{b}{2}\right)^2 + c - \left(\frac{b}{2}\right)^2 = \left(x + \frac{b}{2}\right)^2 + c - \frac{b^2}{4}.$$

One squares half the coefficient of b: $(b/2)^2$. To complete the square in $ax^2 + bx + c$, where a is not 1, factor a out first:

$$ax^2 + bx + c = a\left(x^2 + \frac{b}{a}x + \frac{c}{a}\right).$$

Then complete the square in $x^2 + (b/a)x + c/a$.

EXAMPLE 6 Find $\int \frac{dx}{4x^2+8x+13}$.

Note the subtraction of $4(1^2)$, not 1^2.

SOLUTION First complete the square in the denominator:

$$\begin{aligned}4x^2 + 8x + 13 &= 4(x^2 + 2x + \boxed{}) + 13 - 4\boxed{}\\ &= 4(x^2 + 2x + 1^2) + 13 - 4(1^2)\\ &= 4(x+1)^2 + 9.\end{aligned}$$

We now can rewrite the integral as

$$\int \frac{dx}{4(x+1)^2 + 9}$$

Let $u = x + 1$, hence $du = dx$ and we have

$$\int \frac{dx}{4(x+1)^2+9} = \int \frac{du}{4u^2+9}.$$

By a piece of good luck, we found in Example 5 that

$$\int \frac{du}{4u^2+9} = \frac{1}{6}\arctan\left(\frac{2u}{3}\right) + C$$

Putting all this together:

Check this by differentiating.

$$\begin{aligned}\int \frac{dx}{4x^2+8x+9} &= \int \frac{dx}{4(x+1)^2+9} = \int \frac{du}{4u^2+9} \\ &= \frac{1}{6}\tan^{-1}\left(\frac{2u}{3}\right) + C = \frac{1}{6}\tan^{-1}\left(\frac{2(x+1)}{3}\right) + C.\end{aligned}$$

◇

The integral

$$\int \frac{2ax+b}{ax^2+bx+c}\,dx \tag{8.2.4}$$

is easy since it has the form $\int \frac{f'}{f}\,dx$. The integral is $\ln|az^2+b+c|+C$. This observation is the key to treating the integral in the next example.

EXAMPLE 7 Find $\displaystyle\int \frac{x}{4x^2+8x+13}\,dx$.
SOLUTION No substitution comes to mind. However, if $8x+8$, were in the numerator, we would have an easy integral, for $8x+8$ is the derivative of the denominator. So we will do a little algebra on x to get $8x+8$ into the numerator. We can write $x = \frac{1}{8}(8x+8) - \frac{8}{8} = \frac{1}{8}(8x+8) - 1$. Then we have

$$\begin{aligned}\int \frac{x\,dx}{4x^2+8x+13} &= \int \frac{\frac{1}{8}(8x+8)-1}{4x^2+8x+13}\,dx \\ &= \frac{1}{8}\int \frac{8x+8}{4x^2+8x+13} - \int \frac{dx}{4x^2+8x+13} \\ &= \frac{1}{8}\ln\left|4x^2+8x+13\right| - \frac{1}{6}\arctan\left(\frac{2(x+1)}{3}\right).\end{aligned}$$

Exercise 6 is used here.

◇

The techniques of completing the square, substitution, and rewriting x in the numerator, illustrated in Examples 6 and 7, show how to integrate any integrand of the form $\dfrac{1}{ax^2+bx+c}$ or $\dfrac{x}{ax^2+b+c}$.

Why Substitution Works

Theorem 8.2.1. *(Substitution in an indefinite integral) Assume that f and g are continuous functions and $u = h(x)$ is differentiable. Suppose that $f(x)$ can be written as $g(u)\frac{du}{dx}$ and that G is an antiderivative of g. Then $G(u(x))$ is an antiderivative of $f(x)$.*

Proof

We differentiate $G(u(x))$ and check that the result is $f(x)$:

$$\begin{array}{rcll} \frac{d}{dx}G(u(x)) & = & \frac{dG}{du}\frac{du}{dx} & \text{(Chain Rule)} \\ & = & g(u)\frac{du}{dx} & \text{(by definition of } G) \\ & = & f(x). & \text{(by assumption)} \end{array}$$

•

Theorem 8.2.2. *(Substitution in a definite integral) In addition to the assumptions in Theorem 8.2.1, assume $u(x)$ is monotonic*

$$\int_a^b f(x)\ dx = \int_{u(a)}^{u(b)} g(u)\ du. \tag{8.2.5}$$

Warning: If x goes from a to b, $u(x)$ goes from $u(a)$ to $u(b)$. Be sure to change the limits of integration

Proof

Let $F(x) = G(u(x))$, where G is defined in the previous proof.

$$\begin{array}{rcll} \int_a^b f(x)\ dx & = & F(b) - F(a) & \text{(FTC I)} \\ & = & G(u(b)) - G(u(a)) & \text{(definition of } F) \\ & = & \int_{u(a)}^{u(b)} g(u)\ du & \text{(FTC, again)} \end{array}$$

•

Summary

This section introduced the most commonly used integration technique, "substitution:" If $f(x)\ dx$ can be written as $g(u(x))\ d(u(x))$ for a function $u(x)$ then $\int f(x)dx = \int g(u)\ du$ and $\int_a^b f(x)\ dx = \int_{u(a)}^{u(b)} g(u)\ du$.

It is to be hoped that finding $\int g(u)\ du$ is easier than finding $\int f(x)\ dx$. If it is not, try another substitution or a method presented in the rest of the chapter. There is no simple routine method for antidifferentiation of elementary functions. Practice in integration pays off in spotting which technique is most promising and also being able to transform an integral into one listed in an integral table.

EXERCISES for Section 8.2

In Exercises 1 to 14 use the given substitution to find the antiderivatives or definte integrals.

1. $\int (1+3x)^5 3\ dx; \qquad u = 1+3x$

2. $\int e^{\sin(\theta)} \cos(\theta)\ d\theta; \qquad u = \sin\theta$

3. $\int_0^1 \frac{x}{\sqrt{1+x^2}}\ dx; \qquad u = 1+x^2$

4. $\int_{\sqrt{8}}^{\sqrt{15}} x\sqrt{1+x^2}\ dx; \qquad u = 1+x^2$

5. $\int \sin(2x)\ dx; \qquad u = 2x$

6. $\int \frac{e^{2x}}{(1+e^{2x})^2}\ dx; \qquad u = 1+e^{2x}$

7. $\int_{-1}^{2} e^{3x}\ dx; \qquad u = 3x$

8. $\int_2^3 \frac{e^{1/x}}{x^2}\ dx; \qquad u = \frac{1}{x}$

9. $\int \frac{1}{\sqrt{1-9x^2}}\ dx; \qquad u = 3x$

10. $\int \frac{t\ dt}{\sqrt{2-5t^2}}; \qquad u = 2-5t^2$

11. $\int_{\pi/6}^{\pi/4} \tan(\theta)\sec^2(\theta)\ d\theta; \qquad u = \tan\theta$

12. $\int_{\pi^2/16}^{\pi^2/4} \frac{\sin(\sqrt{x})}{\sqrt{x}}\ dx; \qquad u = \sqrt{x}$

13. $\int \frac{(\ln x)^4}{x}\ dx; \qquad u = \ln x$

14. $\int \frac{\sin(\ln x)}{x}\ dx; \qquad u = \ln x$

Every antiderivative can be verified by checking that its derivative is the integrand. That is, if $\int f(x)\ dx = F(x)$, then $F'(x) = f(x)$. Exercises 15 to 21 ask you to verify an antiderivative found in one of the examples in this section.

15. $\int (\sin(x^2))2x\ dx = -\cos(x^2) + C$ (Example 1)

16. $\int (1+x^3)^5 x^2\ dx = \dfrac{(1+x^3)^6}{18} + C$ (Example 2)

17. $\int \dfrac{dx}{ax+b} = \dfrac{1}{a}\ln|ax+b| + C$ (Example 4(a))

18. $\int \dfrac{dx}{(ax+b)^n} = \dfrac{1}{a(-n+1)(ax+b)^{n-1}} + C$ (Example 4(b))

19. $\int \dfrac{dx}{4x^2+9} = \dfrac{1}{6}\arctan\left(\dfrac{2x}{3}\right) + C$ (Example 5)

20. $\int \dfrac{dx}{4x^2+8x+9} = \dfrac{1}{6}\tan^{-1}\left(\dfrac{2(x+1)}{3}\right) + C$ (Example 6)

21. $\int \dfrac{x\ dx}{4x^2+8x+13} = \dfrac{1}{8}\ln\left|4x^2+8x+13\right| - \dfrac{1}{6}\arctan\left(\dfrac{2(x+1)}{3}\right)$ (Example 7)

In Exercises 22 to 47 use appropriate substitutions to find the antiderivatives.

22. $\int (1-x^2)^5 x\ dx$

23. $\int \dfrac{x\ dx}{(x^2+1)^3}$

24. $\int x\sqrt[3]{1+x^2}\ dx$

25. $\int \dfrac{\sin(\theta)}{\cos^2(\theta)}\ d\theta$

26. $\int \dfrac{e^{\sqrt{t}}}{\sqrt{t}}\ dt$

27. $\int e^x \sin(e^x)\ dx$

28. $\int \sin(3\theta)\ d\theta$

29. $\int \dfrac{dx}{\sqrt{2x+5}}$

30. $\int (x-3)^{5/2}\ dx$

31. $\int \dfrac{dx}{(4x+3)^3}$

32. $\int \dfrac{2x+3}{x^2+3x+2}\ dx$

33. $\int \dfrac{2x+3}{(x^2+3x+5)^4}\ dx$

34. $\displaystyle\int \frac{x^3}{\sqrt{1-x^8}}\ dx$

35. $\displaystyle\int \frac{dx}{\sqrt{x}(1+\sqrt{x})^3}$

36. $\displaystyle\int x^4 \sin(x^5)\ dx$

37. $\displaystyle\int \frac{\cos(\ln(x))\ dx}{x}$

38. $\displaystyle\int \frac{x}{1+x^4}\ dx$

39. $\displaystyle\int \frac{x^3}{1+x^4}\ dx$

40. $\displaystyle\int \frac{x\ dx}{(1+x)^3}$

41. $\displaystyle\int \frac{x^2\ dx}{(1+x)^3}$

42. $\displaystyle\int \frac{\ln(3x)\ dx}{x}$

43. $\displaystyle\int \frac{\ln(x^2)\ dx}{x}$

44. $\displaystyle\int \frac{(\arcsin(x))^2}{\sqrt{1-x^2}}\ dx$

45. $\displaystyle\int \frac{dx}{\arctan(2x)(1+4x^2)}$

46. $\displaystyle\int \frac{dx}{9x^2+1}$

47. $\displaystyle\int \frac{dx}{9x^2+25}$

In Exercises 48 and 49 complete the square in each expression.

48.

(a) $x^2+6x+10$

(b) $4x^2+6x+11$

49.

(a) $x^2+\frac{5}{3}x+4$

(b) $3x^2+5x+12$

50. Evaluate $\displaystyle\int \frac{dx}{x^2+2x+5}$

51. Evaluate $\displaystyle\int \frac{dx}{2x^2+2x+5}$

52. Evaluate $\displaystyle\int \frac{x}{x^2+2x+5}\,dx$

53. Evaluate $\displaystyle\int \frac{x}{2x^2+2x+5}\,dx$

In Exercises 54 to 59 find the area of the region under the graph of the given function and above the given interval.

54. $f(x) = x^2 e^{x^3}$; $[1, 2]$
55. $f(x) = \sin^3(\theta)\cos(\theta)$; $[0, \pi/2]$
56. $f(x) = \frac{x^2+3}{(x+1)^4}$; $[0, 1]$ (Let $u = x + 1$.)
57. $f(x) = \frac{x^2-x}{(3x+1)^2}$; $[1, 2]$
58. $f(x) = \frac{(\ln(x))^3}{x}$; $[1, e]$
59. $f(x) = \tan^5(\theta)\sec^2(\theta)$; $[0, \frac{\pi}{3}]$

In Exercises 60 to 63 use substitution to evaluate the integral.

60. $\displaystyle\int \frac{x^2}{ax+b}\,dx; \qquad a \neq 0$

61. $\displaystyle\int \frac{x}{(ax+b)^2}\,dx; \qquad a \neq 0$

62. $\displaystyle\int \frac{x^2}{(ax+b)^2}\,dx; \qquad a \neq 0$

63. $\displaystyle\int x(ax+b)^n\,dx$; for (a) $n = -1$, (b) $n = -2$

64. Use a substitution to show that if f is an odd function then $\int_{-a}^{a} f(x)\,dx = 0$. (First show that $\int_{-a}^{0} f(x)\,dx = -\int_0^a f(x)\,dx$ by using the substitution $u = -x$.) (Do not refer to "areas".)

65. Use a substitution to show that if f is an even function, then $\int_{-a}^{a} f(x)\,dx = 2\int_0^a f(x)\,dx$. (First show that $\int_{-a}^{0} f(x)\,dx = \int_0^a f(x)\,dx$ by using the substitution $u = -x$.) (Do not refer to "areas".)

66.

(a) Graph $y = \ln(x)/x$.

(b) Find the area under the curve in (a) and above the interval $[e, e^2]$

67.

Sam: Jane, what did you find for the antiderivative of $\int 2\cos(\theta)\sin(\theta)\ d\theta$?

Jane: I found $\int 2\cos(\theta)\sin(\theta)\ d\theta = \sin^2(\theta) + C$.

Sam: That's too bad.

Jane: Why?

Sam: Because I found $\int 2\cos(\theta)\sin(\theta)\ d\theta = -\cos^2(\theta) + C$.

Jane: I'm pretty sure of my answer. I used the substitution $u = \sin(\theta)$.

Sam: Well, that's the problem. I used $u = \cos(\theta)$.

Who is right? Explain.

68.

Jane: $\int_0^\pi \cos^2(\theta)\ d\theta$ is obviously positive.

Sam: No, it's zero. Just make the substitution $u = \sin(\theta)$; hence $du = \cos(\theta)\ d\theta$. Then I get

$$\int_0^\pi \cos^2(\theta)\ d\theta = \int_0^\pi \cos(\theta)\cos(\theta)\ d\theta = \int_0^0 \sqrt{1-u^2}\ du = 0.$$

Simple.

(a) Who is right? What is the mistake?

(b) Use the identity $\cos^2(\theta) = (1 + \cos(2\theta))/2$ to evaluate the integral without substitution or the shortcut in Section 8.1.

69.

Jane: $\int_{-2}^1 2x^2\ dx$ is obviously positive.

Sam: Why are you so sure of this?

Jane: After all, the integrand is never negative and $-2 < 1$. It equals the area under $y = 2x^2$ and above $[-2, 1]$".

Sam: You're wrong again. It's negative. Here are my computations. Let $u = x^2$; hence $du = 2x\ dx$. Then

$$\int_{-2}^{1} 2x^2\ dx = \int_{-2}^{1} x \cdot 2x\ dx = \int_{4}^{1} \sqrt{u}\ du = -\int_{1}^{4} \sqrt{u}\ du,$$

which is obviously negative.

Who is right? Explain.

8.3 Integration by Parts

Integration by substitution, described in the previous section, is based on the chain rule. The technique called "integration by parts," is based on the product rule for derivatives.

The Basis for "Integration by Parts"

If u and v are differentiable functions then

It is a tradition to use u and v instead of the expected f and g.

$$(uv)' = u'\ v + u\ v'.$$

This tells us that uv is an antiderivative of $u'\ v + u\ v'$:

$$uv = \int (u'\ v + u\ v')\ dx,$$

Then

$$uv = \int u'\ v\ dx + \int u\ v'\ dx,$$

which can be rearranged as

$$\boxed{\int u\ v'\ dx = u\ v - \int u'\ v\ dx} \qquad (8.3.1)$$

This equation tells us, "if you can integrate $u'\,v$, then you can integrate $u\,v'$." Now, $u'\,v$ may look quite different from $u\,v'$. Maybe $\int u'\,v\ dx$ is easier to find than $\int u\,v'\ dx$. The technique based on (8.3.1) is called "Integration by Parts".

Using the differentials $du = u'\ dx$ and $dv = v'\ dx$, we can replace (8.3.1) by the shorter version

$$\boxed{\int u\ dv = uv - \int v\ du} \qquad (8.3.2)$$

Typical Examples

EXAMPLE 1 Find $\int xe^{3x}\,dx$.

SOLUTION Let's see what happens if we let $u = x$. Because $u\,dv$ must equal $xe^{3x}\,dx$, we must choose $dv = e^{3x}\,dx$. That is, $v' = e^{3x}$. Then, differentiating u gives $du = dx$ and integrating v' gives $v = \int e^{3x}\,dx = e^{3x}/3$. The integration by parts formula, (8.3.2), tells us that:

$$\int \underbrace{x}_{u}\underbrace{e^{3x}\,dx}_{dv} = \underbrace{x}_{u}\underbrace{\frac{e^{3x}}{3}}_{v} - \int \underbrace{\frac{e^{3x}}{3}}_{v}\underbrace{dx}_{du} = \frac{xe^{3x}}{3} - \frac{e^{3x}}{9} = e^{3x}\left(\frac{x}{3} - \frac{1}{9}\right) + C$$

To check, differentiate $e^{3x}(\frac{x}{3} - \frac{1}{9}) + C$ and see that it's xe^{3x}. ⋄

Look closely at Example 1 to see why it worked. The key is that the derivative of $u = x$ is simpler than u and also we could integrate $v' = e^{3x}$ to find v.

EXAMPLE 2 Find $\int x\ln(x)\,dx$.

SOLUTION Setting $dv = \ln(x)\,dx$ is not a wise move, since $v = \int \ln(x)\,dx$ is not immediately apparent. But setting $u = \ln(x)$ is promising because $du = d(\ln(x)) = \frac{1}{x}dx$ is much easier to handle than $\ln(x)$. This forces dv to be $x\,dx$. This second approach goes through smoothly:

$$\begin{array}{rclcrcl} u & = & \ln(x) & \quad & dv & = & x\,dx \\ du & = & \frac{dx}{x} & & v & = & \frac{x^2}{2}. \end{array}$$

This antiderivative can be checked by differentiation.

(Note that we needed to find $v = \int x\,dx$.) Thus

$$\begin{aligned} \int x\ln(x)\,dx &= \int \underbrace{\ln(x)}_{u}\underbrace{x\,dx}_{dv} = \underbrace{\ln(x)}_{u}\underbrace{\frac{x^2}{2}}_{v} - \int \underbrace{\frac{x^2}{2}}_{v}\underbrace{\frac{dx}{x}}_{du} \\ &= \frac{x^2\ln(x)}{2} - \int \frac{x\,dx}{2} = \frac{x^2\ln(x)}{2} - \frac{x^2}{4} + C. \end{aligned}$$

⋄

General Guidelines for Applying Integration by Parts

General Guidelines for Applying Integration by Parts

The key to applying integration by parts is the selection of u and dv. The following three conditions should be met:

1. v can be found by integrating and should not be too messy.
2. du should not be messier than u.
3. $\int v\,du$ should be easier than the original $\int u\,dv$

The next example shows the general approach that can be used to integrate any inverse trigonometric function.

EXAMPLE 3 Find $\int \arctan(x)\ dx$.

SOLUTION Recall that the derivative of $\arctan(x)$ is $1/(1+x^2)$, a much simpler function than $\arctan(x)$. This suggests the following integration by parts:

$$\begin{array}{rclcrcl} u & = & \arctan(x) & \qquad & dv & = & dx \\ du & = & \frac{dx}{1+x^2} & & v & = & x \end{array}$$

Integrating an inverse trigonometric function by parts

$$\int \underbrace{\arctan(x)}_{u}\,\underbrace{dx}_{dv} = \underbrace{\arctan(x)}_{u}\,\underbrace{x}_{v} - \int \underbrace{x}_{v}\,\underbrace{\frac{dx}{1+x^2}}_{du}$$
$$= x\arctan(x) - \int \frac{x}{1+x^2}\ dx.$$

It is easy to compute $\int \frac{x\ dx}{1+x^2}$, since the numerator is a constant times the derivative of the denominator:

$$\int \frac{x}{1+x^2}\ dx = \frac{1}{2}\int \frac{2x}{1+x^2}\ dx = \frac{1}{2}\ln(1+x^2).$$

Hence

$$\int \arctan(x)\ dx = x\arctan(x) - \frac{1}{2}\ln(1+x^2) + C.$$

You can check this by differentiation. ◇

Compare your answer with Formula 68 in the integral table in the front cover of the book.

To check that you understand the idea in Example 3, find $\int \arcsin(x)\ dx$ by the same method.

EXAMPLE 4 Find $\int x\sin(x)\ dx$.

SOLUTION There are two approaches. We could choose $u = \sin(x)$ and $dv = x\ dx$ or we could choose $u = x$ and $dv = \sin(x)\ dx$.

Approach 1: $u = \sin(x)$ and $dv = x\ dx$

$$\int x\sin(x)\ dx = \int \underbrace{\sin(x)}_{u}\,\underbrace{(x\ dx)}_{dv}.$$

Then $du = \cos(x)\ dx$, which is not any worse than $u = \sin(x)$. And, since $dv = x\ dx$, $v = x^2/2$. Thus,

$$\int \underbrace{\sin(x)}_{u}\,\underbrace{(x\ dx)}_{dv} = \underbrace{\sin(x)}_{u}\,\underbrace{\frac{x^2}{2}}_{v} - \int \underbrace{\frac{x^2}{2}}_{v}\,\underbrace{\cos(x)\ dx}_{du}.$$

$\int u\ dv$ is more difficult than $\int v\ du$; Guideline 3 is not satisfied.

We have replaced the problem of finding $\int x \sin(x)\ dx$ with the harder problem of finding $1/2 \int x^2 \cos(x)\ dx$. That is *not* progress: we have *raised* the exponent of x in the integrand from 1 to 2.

Approach 2: $u = x$ and $dv = \sin(x)\ dx$
With these choices for u and dv,

$$\begin{aligned} u &= x & dv &= \sin(x)\ dx \\ du &= dx & v &= -\cos(x). \end{aligned}$$

This time integration by parts goes through smoothly:

$$\begin{aligned} \int \underbrace{\sin(x)}_{u} \underbrace{(x\ dx)}_{dv} &= \underbrace{x}_{u} \underbrace{(-\cos(x))}_{v} - \int \underbrace{-\cos(x)}_{v} \underbrace{dx}_{du} \\ &= -x\cos(x) + \int \cos(x) dx = -x\cos(x) + \sin(x) + C. \end{aligned}$$

All 3 Guidelines are satisfied by this choice of u and dv.

◇

EXAMPLE 5 Find $\int x^2 e^{3x}\ dx$.
SOLUTION If we let $u = x^2$, then $du = 2x\ dx$. This is good, for it *lowers* the exponent of x. Hence, try $u = x^2$ and therefore $dv = e^{3x}\ dx$:

$$\begin{aligned} u &= x^2 & dv &= e^{3x}\ dx \\ du &= 2x\ dx & v &= \tfrac{1}{3}e^{3x}. \end{aligned}$$

Thus

$$\begin{aligned} \int \underbrace{x^2}_{u} \underbrace{e^{3x}\ dx}_{dv} &= \underbrace{x^2}_{u} \underbrace{\frac{1}{3}e^{3x}}_{v} - \int \underbrace{\frac{1}{3}e^{3x}}_{v} \underbrace{2x\ dx}_{du} \\ &= \frac{x^2}{3}e^{3x} - \frac{2}{3}\int x e^{3x}\ dx \\ &= \frac{x^2}{3}e^{3x} - \frac{2}{3}\left(e^{3x}\left(\frac{x}{3} - \frac{1}{9}\right) + C\right) \qquad \text{by Example 1} \\ &= e^{3x}\left(\frac{x^2}{3} - \frac{2}{3}\left(\frac{x}{3} - \frac{1}{9}\right)\right) - \frac{2}{3}C \\ &= e^{3x}\left(\frac{x^2}{3} - \frac{2x}{9} + \frac{2}{27}\right) - \frac{2C}{3}. \end{aligned}$$

We may rename $-\frac{2C}{3}$, the arbitrary constant, as K, obtaining

$$\int x^2 e^{3x}\ dx = e^{3x}\left(\frac{x^2}{3} - \frac{2x}{9} + \frac{2}{27}\right) + K.$$

◇

The idea behind Example 5 applies to integrals of the form $\int P(x)g(x)\,dx$, where $P(x)$ is a polynomial and $g(x)$ is a function – such as $\sin(x)$, $\cos(x)$, or e^x – that can be repeatedly integrated. Let $u = P(x)$ and $dv = g(x)\,dx$. Then $du = P'(u)\,dx$ and $\int v\,du = \int P'(x)g(x)\,dx$ where $P'(x)$ has a lower degree that $P(x)$.

Example 5 generalizes.

Definite Integrals and Integration by Parts

Integration by parts of a definite integral $\int_a^b f(x)\,dx$, where $f(x) = u(x)v'(x)$, takes the form

$$\begin{aligned}\int_a^b f(x)\,dx &= \int_a^b u(x)v'(x)\,dx = u(x)v(x)|_a^b - \int_a^b v(x)u'(x)\,dx\\ &= u(v)v(b) - u(a)v(a) - \int_a^b v(x)u'(x)\,dx.\end{aligned}$$

EXAMPLE 6 Find the area under the curve $y = \arctan(x)$ and above [0,1]. (See Figure 8.3.1.)

SOLUTION The area is $\int_0^1 \arctan(x)\,dx$. By Example 3,

$$\int \arctan(x)\,dx = x\arctan(x) - \frac{1}{2}\ln(1+x^2) + C.$$

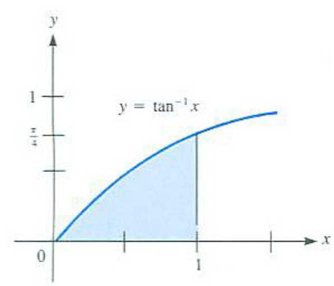


Figure 8.3.1

Since only one antiderivative is needed in order to apply the Fundamental Theorem of Calculus, we may choose $C = 0$. Then

$$\begin{aligned}\int_0^1 \arctan x\,dx &= x\arctan(x)|_0^1 - \frac{1}{2}\ln(1+x^2)\Big|_0^1\\ &= 1\arctan(1) - 0\arctan(0) - \frac{1}{2}\ln(1+1^2) + \frac{1}{2}\ln(1+0^2)\\ &= \frac{\pi}{4} - \frac{1}{2}\ln(2) \approx 0.438824.\end{aligned}$$

◇

Reduction Formulas

Formulas 36, 43, and 46 in the table of integrals on the inside cover of this book express the integral of a function that involves the n^{th} power of some expression in terms of the integral of a function that involves a lower power of the same expression. These are **reduction formulas** or **recursion formulas**. Usually they are obtained by integration by parts.

See Exercise 25 or Formula 43 in the table of integrals.

An example of a reduction formula is

$$\int \sin^n(x)\,dx = -\frac{\sin^{n-1}(x)\cos(x)}{n} + \frac{n-1}{n}\int \sin^{n-2}(x)\,dx \qquad \text{for integer values of } n \geq \tag{8.3.3}$$

EXAMPLE 7 Use the reduction formula (8.3.3) to evaluate $\int \sin^5(x)\,dx$.
SOLUTION In this case $n = 5$. By (8.3.3),

$$\int \sin^5(x)\,dx = -\frac{\sin^4(x)\cos(x)}{5} + \frac{4}{5}\int \sin^3(x)\,dx. \tag{8.3.4}$$

Use (8.3.3) again to dispose of $\int \sin^3(x)\,dx$. In this case $n = 3$:

$$\begin{aligned}\int \sin^3(x)\,dx &= -\frac{\sin^2(x)\cos(x)}{3} + \frac{2}{3}\int \sin(x)\,dx \\ &= -\frac{\sin^2(x)\cos(x)}{3} - \frac{2}{3}\cos(x)\end{aligned} \tag{8.3.5}$$

Combining (8.3.4) and (8.3.5) gives

$$\int \sin^5(x)\,dx = -\frac{\sin^4(x)\cos(x)}{5} + \frac{4}{5}\left(\frac{-\sin^2(x)\cos(x)}{3} - \frac{2}{3}\cos(x)\right) + C.$$

Every time (8.3.3) is used, the exponent of $\sin(x)$ decreases by 2. If you keep applying (8.3.3), you eventually run into the exponent 1 (as we did, because n is odd) or, if n is even, into the exponent 0. ◇

The next example shows how (8.3.3) can be obtained by integration by parts.

See Formula 43 in the inside cover of the text.

EXAMPLE 8 Obtain the reduction formula (8.3.3).
SOLUTION First write $\int \sin^n(x)\,dx$ as $\int \sin^{n-1}(x)\sin(x)\,dx$. Then let $u = \sin^{n-1}(x)$ and $dv = \sin(x)\,dx$. Thus

$$\begin{aligned} u &= \sin^{n-1}(x) & dv &= \sin(x)\,dx \\ du &= (n-1)\sin^{n-2}(x)\cos(x)\,dx & v &= -\cos(x). \end{aligned}$$

Integration by parts yields

$$\begin{aligned}&\int \underbrace{\sin^{n-1}(x)}_{u}\underbrace{\sin(x)\,dx}_{dv} \\ &= \underbrace{(\sin^{n-1}(x))}_{u}\underbrace{(-\cos(x))}_{v} - \int \underbrace{(-\cos(x))}_{v}\underbrace{(n-1)\sin^{n-2}(x)\cos(x)\,dx}_{du}.\end{aligned}$$

The integral on the right of this equation is

$$\begin{aligned} & -\int (n-1)\cos^2(x)\sin^{n-2}(x)\ dx \\ = & -(n-1)\int (1-\sin^2(x))\sin^{n-2}(x)\ dx \\ = & -(n-1)\int \sin^{n-2}(x)\ dx + (n-1)\int \sin^n(x)\ dx. \end{aligned}$$

Thus

$$\begin{aligned} & \int \sin^n(x)\ dx \\ = & -\sin^{n-1}(x)\cos(x) - \left(-(n-1)\int \sin^{n-2}(x)\ dx + (n-1)\int \sin^n(x)\ dx\right) \\ = & -\sin^{n-1}(x)\cos(x) + (n-1)\int \sin^{n-2}(x)\ dx - (n-1)\int \sin^n(x)\ dx. \end{aligned}$$

Rather than being dismayed by the reappearance of $\int \sin^n(x)\ dx$, move that term to the left side to obtain:

$$n\int \sin^n(x)\ dx = -\sin^{n-1}(x)\cos(x) + (n-1)\int \sin^{n-2}(x)\ dx,$$

from which (8.3.3) follows. ◇

The reduction formula for $\int \cos^n x\ dx$ is obtained similarly.

See Formula 46, with $a = 1$, in the table on the front cover.

EXAMPLE 9 Obtain the reduction formula for $\int \frac{dx}{(x^2+c)^n}$ where n is a positive integer.

SOLUTION The only choice that comes to mind for integration by parts is

$$\begin{aligned} u &= \tfrac{1}{(x^2+c)^n} & dv &= dx \\ du &= \tfrac{-2nx}{(x^2+c)^{n+1}} & v &= x. \end{aligned}$$

Integration by parts gives

$$\int \frac{dx}{(x^2+c)^n} = \frac{x}{(x^2+c)^{n+1}} + 2n\int \frac{x^2}{(x^2+c)^{n+1}}\ dx.$$

It looks as though we have just created a more compicated integrand. However, in the numerator of the integrand on the right-hand side, write x^2 as x^2+c-c. We then have

See Formula 28, with $a = 1$, in the table on the front cover.

$$\int \frac{dx}{(x^2+c)^n} = \frac{x}{(x^2+c)^{n+1}} + 2n\int \frac{x^2+c}{(x^2+c)^{n+1}}\ dx - 2nc\int \frac{dx}{(x^2+c)^{n+1}} \quad (8.3.6)$$

Canceling out x^2+c in the first integrand on the right gives us an equation which could be rewritten to express $\int \frac{dx}{(x^2+c)^{n+1}}$ in terms of $\int \frac{dx}{(x^2+c)^n}$. ◇

See also Exercises 46 and 62 in this section.

An Unusual Example

In the next example one integration by parts appears at first to be useless, but two in succession lead to the successful evaluation of the integral.

EXAMPLE 10 Find $\int e^x \cos(x)\ dx$
SOLUTION There are two reasonable choices for applying integration by parts: $u = e^x$, $dv = \cos(x)\ dx$ or $u = \cos(x)$, $dv = e^x\ dx$. In neither case is du "simpler", but watch what happens when integration by parts is applied twice.

Following the first choice:

$$\begin{aligned} u &= e^x & dv &= \cos(x)\ dx \\ du &= e^x\ dx & v &= \sin(x) \end{aligned}$$

The second choice is explored in Exercise 57.

Then integration by parts proceeds as follows:

$$\int \underbrace{e^x}_{u} \underbrace{\cos(x)\ dx}_{dv} = \underbrace{e^x}_{u} \underbrace{\sin(x)}_{v} - \int \underbrace{\sin(x)}_{v} \underbrace{e^x\ dx}_{du}. \tag{8.3.7}$$

Repeated integration by parts

It may seem that nothing useful has been accomplished; $\cos(x)$ is replaced by $\sin(x)$. But watch closely as the new integral is also treated by an integration by parts. Capital letters U and V, instead of u and v, are used to distinguish this computation from the preceeding one.

$$\begin{aligned} U &= e^x & dV &= \sin(x)\ dx \\ dU &= e^x\ dx & V &= -\cos(x). \end{aligned}$$

So

$$\begin{aligned} \int \underbrace{e^x}_{U} \underbrace{\sin(x)\ dx}_{dV} &= \underbrace{e^x}_{U} \underbrace{(-\cos(x))}_{V} - \int \underbrace{(-\cos(x))}_{V} \underbrace{e^x\ dx}_{dU} \\ &= -e^x \cos(x) + \int e^x \cos(x)\ dx. \end{aligned} \tag{8.3.8}$$

Combining (8.3.7) and (8.3.8) yields

$$\begin{aligned} \int e^x \cos(x)\ dx &= e^x \sin(x) - \left(-e^x \cos(x) + \int e^x \cos(x)\ dx\right) \\ &= e^x(\sin(x) + \cos(x)) - \int e^x \cos(x)\ dx. \end{aligned}$$

Bringing $-\int e^x \cos x\ dx$ to the left side of the equation gives

$$2\int e^x \cos(x)\ dx = e^x(\sin(x) + \cos(x)),$$

and we conclude that

See Formula 63, with $a = 1$ and $b = 1$.

$$\int e^x \cos(x)\ dx = \frac{1}{2}e^x(\sin(x) + \cos(x)) + C.$$

◇

See Exercise 60.

Summary

Integration by parts is described by the formula

$$\int u\ dv = uv - \int v\ du.$$

When you break up the original integral into the parts u and dv, try to choose them so that

1. You can find v and it is not too messy.
2. The derivative of u is nicer than u.
3. You can integrate $\int v\ du$.

Sometimes you have to apply integration by parts more than once, for instance, in finding $\int e^x \cos(x)\ dx$.

EXERCISES for Section 8.3

Use integration by parts to evaluate each of the integrals in Exercises 1 to 20.

1. $\int xe^{2x}\ dx$

2. $\int (x+3)e^{-x}\ dx$

3. $\int x\sin(2x)\ dx$

4. $\int (x+3)\cos(2x)\ dx$

5. $\int x\ln(3x)\ dx$

6. $\int (2x+1)\ln(x)\ dx$

7. $\int_1^2 x^2e^{-x}\ dx$

8. $\int_0^1 x^2e^{2x}\ dx$

9. $\int_0^1 \sin^{-1}(x)\ dx$

10. $\int_0^{1/2} \tan^{-1}(2x)\ dx$

11. $\int x^2\ln(x)\ dx$

12. $\int x^3\ln(x)dx$

13. $\int_2^3 (\ln(x))^2\ dx$

14. $\int_2^3 (\ln(x))^3\ dx$

15. $\int_1^e \frac{\ln(x)\ dx}{x^2}$

16. $\displaystyle\int_e^{e^2} \frac{\ln(x)\ dx}{x^3}$

17. $\displaystyle\int e^{3x} \cos(2x)\ dx$

18. $\displaystyle\int e^{-2x} \sin(3x)\ dx$

19. $\displaystyle\int \frac{\ln(1+x^2)\ dx}{x^2}$

20. $\displaystyle\int x \ln(x^2)\ dx$

In Exercises 21 to 24 find the integrals two ways: (a) by substitution, (b) by integration by parts.

21. $\displaystyle\int x\sqrt{3x+7}\ dx$

22. $\displaystyle\int \frac{x\ dx}{\sqrt{2x+7}}$

23. $\displaystyle\int x(ax+b)^3\ dx$

24. $\displaystyle\int \frac{x\ dx}{\sqrt[3]{ax+b}}, \qquad a \neq 0$

25. Use differentiation to verify (8.3.3).

26. Use the recursion in Example 8 to find

(a) $\displaystyle\int \sin^2 x\ dx$

(b) $\displaystyle\int \sin^4 x\ dx$

(c) $\displaystyle\int \sin^6 x\ dx$

27. Use the recursion in Example 8 to find

(a) $\displaystyle\int \sin^3 x\ dx$

(b) $\displaystyle\int \sin^5 x\ dx$

28.

(a) Graph $y = e^x \sin x$ for x in $[0, \pi]$, showing extrema and inflection points.

(b) Find the area of the region below the graph and above the interval $[0, \pi]$.

29.

(a) Graph $y = e^{-x} \sin x$ for x in $[0, \pi]$, showing extrema and inflection points.

(b) Find the area of the region below the graph and above the interval $[0, \pi]$.

30. Figure 8.3.2(a) shows a shaded region whose cross sections by planes perpendicular to the x-axis are squares. Find its volume.

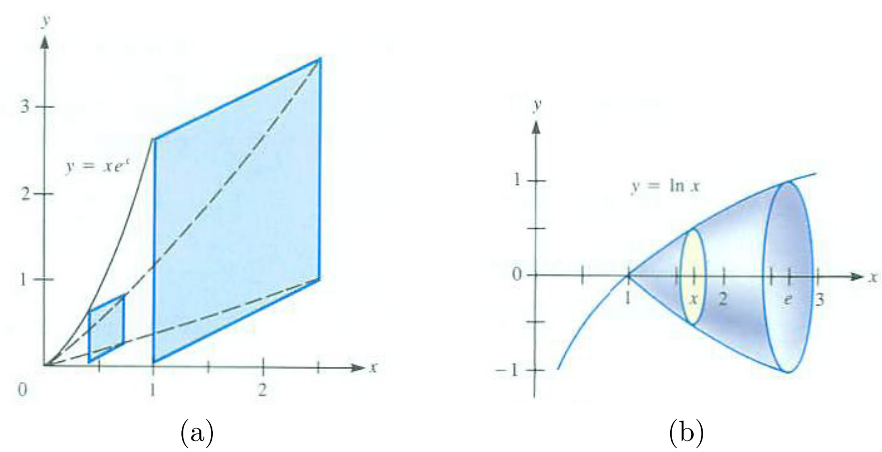


Figure 8.3.2

31. Figure 8.3.2(b) shows a solid whose cross sections by planes perpendicular to the x-axis are disks. The solid meets the x-axis in the interval $[y.e]$. Find its volume.

In Exercises 32 to 37 find the integrals. In each case a substitution is required before integration by parts can be used. In Exercises 36 and 37 the notation $\exp(u)$ is used for e^u. This notation is often used for clarity.

32. $\displaystyle\int \sin(\sqrt{x})\ dx$

33. In Exercise 67 in Section 6.4 it is claimed that $\frac{e^x}{x}$ does not have an elementary antiderivative. From this fact we can show other functions also lack elementary antiderivatives.

(a) Show that $\int \frac{e^x}{x}\,dx$ equals $\ln(x)e^x - \int \ln(x)e^x\,dx$ and also equals $\frac{e^x}{x} + \int \frac{e^x}{x^2}\,dx$ and $\int \frac{du}{\ln(u)}$ (where $u = e^x$). (Each expression can be obtained from the first by an appropriate use of integration by parts or substitution.)

(b) Deduce that $\displaystyle\int \ln(x)e^x\,dx$, $\int (e^x/x^2)\,dx$, and $\displaystyle\int 1/\ln(x)\,dx$ do not have elementary antiderivatives. If one of these integrals has an elementary antiderivative, then they all do.

34. Explain how you would go about finding

$$\int x^{10}(\ln x)^{18}\,dx$$

(Don't just say, "I'd use integral tables or a computer.") Explain why your approach would work, but include only enough calculation to convince a reader that it would succeed.

35. Find $\displaystyle\int \sin(\sqrt[3]{x})\,dx$.

36. Find $\displaystyle\int \exp(\sqrt{x})\,dx$. Recall that $\exp(x) = e^x$.

37. Find $\displaystyle\int \exp(\sqrt[3]{x})\,dx$

38. Given that $\displaystyle\int \frac{\sin(x)}{x}\,dx$ is not elementary, deduce that $\displaystyle\int \cos(x)\ \ln(x)\,dx$ is not elementary.

39. Given that $\displaystyle\int x\tan(x)\,dx$ is not elementary, deduce that $\displaystyle\int (x/\cos(x))^2\,dx$ is not elementary.

40. Let I_n denote $\displaystyle\int_0^{\pi/2} \sin^n(\theta)\,d\theta$, where n is a nonnegative integer.

(a) Evaluate I_0 and I_1.

(b) Using the recursion in Example 8, show that

$$I_n = \frac{n-1}{n} I_{n-2}, \qquad \text{for } n \geq 2.$$

(c) Use (b) to evaluate I_2 and I_3.

(d) Use (c) to evaluate I_4 and I_5.

(e) Explain why $I_n = \frac{2\cdot 4\cdot 6\cdots(n-1)}{3\cdot 5\cdot 7\cdots n}$ when n is odd.

(f) Explain why $I_n = \frac{1\cdot 3\cdot 5\cdots(n-1)}{2\cdot 4\cdot 6\cdots n}\cdot\frac{\pi}{2}$ when n is even.

(g) Explain why $\int\limits_0^{\pi/2} \sin^n(\theta)\ d\theta = \int\limits_0^{\pi/2} \cos^n(\theta)\ d\theta$. (Use the substitution $u = \pi/2 - \theta$.)

41. Find $\int \ln(x+1)\ dx$ using

(a) $u = \ln(x+1)\ dx$, $dv = dx$, $v = x$

(b) $u = \ln(x+1)\ dx$, $dv = dx$, $v = x+1$

(c) Which is easier?

42. Let n be a positive integer and a is a constants. Obtain a formula that expresses $\int x^n e^{-ax}\ dx$ in terms of $\int x^{n-1}e^{-ax}$.

43. Find $\int x\sin(ax)\ dx$

44. Let a be a constant and n a positive integer.

(a) Express $\int x^n \sin(ax)\ dx$ in terms of $\int x^{n-1}\cos(ax)\ dx$.

(b) Express $\int x^n \cos(ax)\ dx$ in terms of $\int x^{n-1}\sin(ax)\ dx$.

(c) Why do (a) and (b) enable us to find $\int x^n \sin(ax)\ dx$?

45.

(a) Express $\int (\ln(x))^n\ dx$ in terms of $\int (\ln(x))^{n-1}\ dx$.

(b) Use (a) to find $\int (\ln(x))^3\ dx$

46.

(a) Show how the integral $\int \frac{dx}{(ax^2+bx+c)^{n+1}}$ can be reduced to an integral of the form $\int \frac{du}{(u^2+p)^{n+1}}$.

(b) Use (a) and the recursion formula obtained in Exercise 62 to find a recursion formula for $\int \frac{dx}{(x^2+bx+c)^n}$. (How does your answer compare with Formula 35 in the integral table on the front cover of the text?)

In Exercises 47 to 50 obtain recursion formulas for the integrals.

47. $\int x^n e^{ax}\, dx$, n a positive integer, a a nonzero constant

48. $\int (\ln(x))^n\, dx$, n a positive integer

49. $\int x^n \sin(x)\, dx$, n a positive integer

50. $\int \cos^n(ax)\, dx$, n a positive integer.

Laplace Transform Let $f(t)$ be a continuous function defined for $t \geq 0$. Assume that, for certain fixed positive numbers r, $\int_0^\infty e^{-rt} f(t)\, dt$ converges and that $e^{-rt} f(t) \to 0$ as $t \to \infty$. Define $P(r)$ to be $\int_0^\infty e^{-rt} f(t)\, dt$. The function P is called the **Laplace transform** of the function f. It is an important tool for solving differential equations, and appears in the CIE on present value of future income (see page 785). In Exercises 51 to 55 find the Laplace transform of the given functions.

51. $f(t) = t$

52. $f(t) = t^2$

53. $f(t) = e^t$ (assume $r > 1$)

54. $f(t) = \sin(t)$

55. $f(t) = \cos(t)$

56. Let $P(x)$ be a polynomial.

(a) Check by differentiation that $(P(x) - P'(x) + P''(x) - \cdots)e^x$ is an antiderivative of $P(x)e^x$. (Note that the signs alternate and that the derivatives are taken to successively higher orders until they are constant, with value 0.)

(b) Use (a) to find $\int (3x^3 - 2x - 2)e^x\, dx$.

(c) Apply integration by parts to $\int P(x)e^x\, dx$ to show how the formula in (a) could be obtained.

57. In Example 10, $\int e^x \cos(x)\, dx$ was evaluated by applying integration by parts twice, each time differentiating an exponential and antidifferentiating a trigonometric function. What happens when integration by parts is applied (twice, if necessary) when a trigonometric function is differentiated and an exponential is antidifferentiated. That is, to get started, apply integration by parts with $u = \cos(x)$ and

$dv = e^x\ dx$.

58. Find $\int_{-1}^{1} x^3\sqrt{1+x^{20}}\ dx$.

59. Find $\int_{-\pi/4}^{\pi/4} \tan(x)(1+\cos(x))^{3/2}\ dx$

60. According to the reasoning in Example 10, it appears that $\int e^x \cos(x)\ dx$ must equal $\frac{1}{2}e^x(\sin(x)+\cos(x))$. This would contradict the fact that for any constant C, $\frac{1}{2}e^x(\sin(x)+\cos(x))+C$ is also an antiderivative of $e^x\sin(x)$. Resolve the paradox.

61.

(a) What does the graph of $y = \cos(ax)$ look like when $a = 1$? when $a = 2$? when $a = 3$? when a is a very large constant? Include graphs and a written description in your answers.

(b) Let $f(x)$ be a function with a continuous derivative. Assume that $f(x)$ is positive. What does the graph of $y = f(x)\cos(ax)$ look like when a is large? Express your response in terms of the graph of $y = f(x)$. Include a sketch of $y = f(x)\cos(ax)$ to give an idea of its shape.

(c) On the basis of (b), what do you think happens to

$$\int_0^1 f(x)\cos(ax)\ dx$$

as $a \to \infty$? Give an intuitive explanation.

(d) Use integration by parts to justify your answer in (c).

62. Solve (8.3.6) in Example 9 to obtain the reduction formula for $\int \frac{dx}{(ax^2+c)^n}$. To check your answer, compare it to Formula 28 in the integral table in the inside cover of this book with $a = 1$.

63. If we have a recursion for $\int \frac{dx}{(ax^2+bx+c)^n}$ why don't we need one for $\int \frac{x\ dx}{(ax^2+bx+c)^n}$?

8.4 Integrating Rational Functions: The Algebra

Every rational function, no matter how complicated, has an elementary integral which is the sum of some or all of these types of functions:

Recall that a rational function is a polynomial or the quotient of two polynomials.

- rational functions (including polynomials),
- logarithms of linear or quadratic polynomials: $\ln(ax+b)$ or $\ln(ax^2+bx+c)$, and
- arctangents of linear or quadratic polynomials: $\arctan(ax+b)$ or $\arctan(ax^2+bx+c)$.

The reason is mainly algebraic. In an advanced algebra course it is proved that every rational function is the sum of much simpler rational functions, namely those of the forms:

$$\text{polynomials}, \frac{k}{(ax+b)^n}, \frac{d}{(ax^2+bx+c)^n}, \text{ and } \frac{ex}{(ax^2+bx+c)^n} \tag{8.4.1}$$

where a, b, c, d, e, k are constants and n is a positive integer. In Sections 8.2 and 8.3 we saw how to integrate each of these integrands. (See Examples 4 to 7 in Section 8.2 and Formulas 13, 14, 15, 35, 36, and 37.)

As this section is completely algebraic, our objective is to see how to express a rational function $f(x)$ as a sum of the functions in (8.4.1), that is, to find the **partial fraction decomposition** of $f(x)$. For instance, we will see how to find the decomposition

$$\frac{1}{2x^2+7x+3} = \frac{2/5}{2x+1} - \frac{1/5}{x+3}.$$

Reducible and Irreducible Polynomials

A polynomial $a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$, where a_n is not 0 is said to have degree n. The polynomials of degree one are called linear; those of degree two, quadratic. A polynomial of degree zero is a constant. If all the coefficients a_i are zero, we have the zero polynomial, which is not assigned a degree.

A polynomial of degree at least one is **reducible** if it is a product of nonconstant polynomials of lower degree. Otherwise, it is **irreducible**.

Every polynomial of degree one, $ax+b$, is clearly irreducible. A polynomial of degree two, ax^2+bx+c, is irreducible if its discriminant b^2-4ac is negative. (See Exercises 37 and 38.) However,

Recall: $a \neq 0$.

FACT 1: Every polynomial of degree three or higher is reducible.

This is far from obvious. For instance, x^4+1 looks like it cannot be factored, but you can check that

$$x^4+1=(x^2+\sqrt{2}x+1)(x^2-\sqrt{2}x+1).$$

On the other hand,

$$x^4-1=(x^2+1)(x^2-1)=(x^2+1)(x+1)(x-1).$$

The next non-obvious fact is that

FACT 2: Every polynomial of degree at least one is either irreducible or the product of irreducible polynomials.

The factoring of x^4+1 and x^4-1, given above, illustrate both Facts 1 and 2.

Proper and Improper Rational Functions

In arithmetic, the rational number m/n is called proper if $|m|$ is less than $|n|$.

Let $A(x)$ and $B(x)$ be polynomials. The rational function $A(x)/B(x)$ is **proper** if the degree of $A(x)$ is less than the degree of $B(x)$. Otherwise, it is **improper**.

Every improper rational function is the sum of a polynomial and a proper rational function. The next example illustrates why this is true. It depends on long division.

EXAMPLE 1 Express $\frac{2x^3+1}{2x^2-x+1}$ as a polynomial plus a proper rational function.

Keep dividing until the degree of the remainder is less than the degree of the divisor, or the remainder is 0.

SOLUTION We carry out long division

$$\begin{array}{rrrrrl}
 & & & x & +1/2 & \leftarrow \text{quotient} \\ \cline{2-5}
2x^2-x+1 \big) & 2x^3 & +0x^2 & +0x & +1 & \\
 & 2x^3 & -x^2 & +x & & \\ \cline{2-4}
 & & x^2 & -x & +1 & \\
 & & x^2 & -x/2 & +1/2 & \\ \cline{3-5}
 & & & -x/2 & +1/2 & \leftarrow \text{remainder}
\end{array}$$

Thus

$$2x^3+1=\left(2x^2-x+1)(x+\frac{1}{2}\right)+\left(-\frac{x}{2}+\frac{1}{2}\right).$$

Division by $2x^2-x+1$ gives us the representation

$$\underbrace{\frac{2x^3+1}{2x^2-x+1}}_{\text{improper}}=\underbrace{x+\frac{1}{2}}_{\text{polynomial}}+\underbrace{\frac{\left(\frac{-x}{2}+\frac{1}{2}\right)}{2x^2-x+1}}_{\text{proper}}.$$

To check this equation, just rewrite the right-hand side as a single fraction. ⋄

To integrate a rational function we first check that it is proper. If it is improper, we carry out long division, and represent the function as the sum of a polynomial and a proper rational function. Since we already know how to integrate a polynomial we consider in the rest of this section only proper rational functions.

Partial Fractions

As mentioned in the introduction, every rational function is the sum of particularly simple rational functions, ones we know how to integrate. Here is a recipe for finding that representation for a proper rational function $A(x)/B(x)$.

1. Write $B(x)$ as a product of first-degree polynomials and irreducible second-degree polynomials.

2. If $px + q$ appears exactly n times in the factorizaiton of $B(x)$, form

   Step 2: List summands of the form $\frac{k_i}{(px+q)^i}$.

$$\frac{k_1}{px+q} + \frac{k_2}{(px+q)^2} + \cdots + \frac{k_n}{(px+q)^n},$$

   where the constants k_1, k_2, ..., k_n are to be determined later.

3. If ax^2+bx+c appears exactly m times in the factorization of $B(x)$, then form the sum

   Step2: List summands of the form $\frac{r_jx+s_j}{(ax^2+bx+c)^j}$.

$$\frac{r_1x+s_1}{ax^2+bx+c} + \frac{r_2x+s_2}{(ax^2+bx+c)^2} + \cdots + \frac{r_mx+s_m}{(ax^2+bx+c)^m},$$

   where the constants r_1, r_2, ..., r_m and s_1, s_2, ..., s_m are to be determined later.

4. Find all the constants (k_i's, r_j's, and s_j's) mentioned in Steps 2 and 3 so that the sum of the rational functions in Steps 2 and 3 equals $A(x)/B(x)$.

The rational functions in Steps 2 and 3 are called the **partial fractions** of $A(x)/B(x)$. This process deserves some comments.

Regarding Step 1

In practice the denominator $B(x)$ often already appears in factored form. If it does not, finding the factorization can be quite a challenge. To find first-degree factors, look for a root of $B(x) = 0$. If r is a root of $B(x)$, then $x-r$ is a factor. Divide $x-r$ into $B(x)$, getting a quotient $Q(x)$; so $B(x) = (x-r)Q(x)$. Repeat the process on $Q(x)$, continuing as long as you can find roots. Already you can see problems. Suppose you find a root numerically to several decimal places. Consequently your results of integration will be approximations. If you want $\int_a^b A(x)/B(x)\ dx$ it might be simpler just to approximate the definite integral.

After finding all the linear factors "what's left" has to be the product of second-degree polynomials. If the degree of "what's left" is just two, then you are happy: you have found the complete factorization. But, if that degree is 4 or 6 or higher, you face a task best to be avoided — or attacked with the assistance of a computer.

Regarding Steps 2 and 3

These steps refer to the number of times a factor occurs in the denominator. If you factor $2x^2+4x+2$, you may obtain $(x+1)(2x+2)$. Note that $2x+2$ is a constant times $x+1$. The factorization may be written as $2(x+1)^2$, where $x+1$ is a repeated factor. We say that "$x+1$ appears exactly two times in the factorization of $2x^2+4x+2$. Always collect factors that are constants times each other.

Regarding Step 4

Finding the unknown constants may take a lot of work. If there are only linear factors without repetition, the method illustrated in Example 3 is quick. Clearing denominators and comparing the corresponding coefficients of the polynomials on both sides of the resulting equation always works. The number of unknown constants always equals the degree of the denominator $B(x)$. If $B(x)$ has repeated linear or second-degree factors and the degree of $B(x)$ is "large," consider using a computing tool to find approximations to the coefficients.

EXAMPLE 2 What is the form of the partial fraction representation of

$$\frac{x^{10}+x+3}{(x+1)^2(2x+2)^3(x-1)^2(x^2+x+3)^2}? \tag{8.4.2}$$

SOLUTION The degree of the denominator is 11 and the degree of the numerator is 10. Thus (8.4.2) is proper. There is no need to divide the numerator by the denominator.

The factor $2x+2$ is $2(x+1)$. So $(x+1)^2(2x+2)^3$ should be written as $8(x+1)^5$. The discriminant of x^2+x+3 is $(1)^2-4(1)(3)=-11<0$; thus x^2+x+3 is irreducible. Therefore the partial fraction representation of (8.4.2) has the form

$$\begin{aligned}\frac{k_1}{x+1}+\frac{k_2}{(x+1)^2}+\frac{k_3}{(x+1)^3}+\frac{k_4}{(x+1)^4}+\frac{k_5}{(x+1)^5}\\+\frac{k_6}{x-1}+\frac{k_7}{(x-1)^2}+\frac{r_1x+s_1}{x^2+x+3}+\frac{r_2x+s_2}{(x^2+x+3)^2}.\end{aligned}$$

Note that the number of unknown constants equals the degree of the denominator in (8.4.2). ◇

Finding the constants in Example 2 would be a major task if done by hand. It would involve solving a system of 11 equations for the 11 unknown constants. Fortunately, this is an ideal problem for a computer to solve.

Denominator Has Only Linear Factors, Each Appearing Only Once

We illustrate this case, which can be done without a computer, by an example.

EXAMPLE 3 Express $\frac{1}{(2x+1)(x+3)}$ in the form $\frac{k_1}{2x+1} + \frac{k_2}{x+3}$ and then find $\int \frac{dx}{(2x+1)(x+3)}$.

SOLUTION

$$\frac{1}{(2x+1)(x+3)} = \frac{k_1}{2x+1} + \frac{k_2}{x+3}. \tag{8.4.3}$$

To find k_1, multiply both sides of (8.4.3) by the denominator of the term that contains k_1, $2x+1$, getting

$$\frac{1}{x+3} = k_1 + \frac{k_2(2x+1)}{x+3}. \tag{8.4.4}$$

Equation (8.4.4) is valid for all values of x except $x = -3$, in particular for the value of x that makes $2x + 1 = 0$, namely $x = -1/2$. Evaluating (8.4.3) when $x = -1/2$ we get

$$\frac{1}{\left(\frac{-1}{2}\right) + 3} = k_1 + 0.$$

We have found that k_1 is $\frac{2}{5}$.

The same idea can be used to solve for k_2: multiply both sides of (8.4.3) by $(x+3)$, obtaining

$$\frac{1}{2x+1} = \frac{k_1(x+3)}{2x+1} + k_2.$$

Replace x by -3, the solution to $x + 3 = 0$, to obtain

$$\frac{1}{2(-3)+1} = 0 + k_2.$$

Thus $k_2 = \frac{-1}{5}$.

Since $k_1 = \frac{2}{5}$ and $k_2 = \frac{-1}{5}$, (8.4.3) takes the form

$$\frac{1}{(2x+1)(x+3)} = \frac{2/5}{2x+1} - \frac{1/5}{x+3}.$$

For a quicker, but not complete, check replace x in (8.4.3) by a convenient number and see if the resulting equation is correct. Try it, with, say, $x = 0$.

To verify this identity, check it by multiplying both sides by $(2x+1)(x+3)$, getting

$$1 = \frac{2}{5}(x+3) - \frac{1}{5}(2x+1) = \frac{2}{5}x + \frac{6}{5} - \frac{2}{5}x - \frac{1}{5} = \frac{5}{5}. \tag{8.4.5}$$

It checks.

Another way to solve for the unknown constants is to clear the denominator and equate coefficients of like powers of x. For instance, let us find k_1 and k_2 in (8.4.3). We obtain

$$1 = k_1(x+3) + k_2(x+3).$$

Collecting coefficients, we have

$$1 = (k_1 + 2k_2)x + (3k_1 + k_2). \tag{8.4.6}$$

Comparing coefficients on both sides of (8.4.6) we have

$$\begin{array}{rclll} 0 & = & k_1 + 2k_2 & & \text{[equating coefficients of } x\text{]} \\ 1 & = & 3k_1 + k_2 & & \text{[equating constant terms]} \end{array}$$

There are many ways to solve these simultaneous equations. One way is to use the first equation to express k_1 in terms of k_2: $k_1 = -2k_2$. Then replace k_1 by $-2k_2$ in the second, getting

$$1 = 3(-2k_2) + k_2 = -5k_2$$

from which it is seen that $k_2 = \frac{-1}{5}$. Then $k_1 = \frac{2}{5}$.

In general, in this method the number of equations always equals the number of unknowns, which is also equal to the degree of the denominator. If that degree is large, it is not realistic to do the calculations by hand. ◇

If the denominator is just a repeated linear factor, there are two options: "clearing the denominator and equate coefficients" or "substitution". For instance, the partial fraction representation of

$$\frac{7x+6}{(x+2)^2}$$

you could let $u = x+2$, hence $x = u-2$. Then

$$\begin{array}{rcl} \dfrac{7x+6}{(x+2)^2} & = & \dfrac{7(u-2)+6}{u^2} = \dfrac{7u}{u^2} - \dfrac{8}{u^2} \\ & = & \dfrac{7u}{u^2} - \dfrac{8}{u^2} = \dfrac{7}{u} - \dfrac{8}{u^2} = \dfrac{7}{x+2} - \dfrac{8}{(x+2)^2}. \end{array}$$

This method for representing

$$\frac{A(x)}{(ax+b)^n}$$

is practical if the degree of $A(x)$ is small. Here $u = ax+b$, hence $x = \frac{1}{a}(u-b)$. If the degree of $A(x)$ is not small, expressing a power of x, x^m, in terms of u

would best be done by the Binomial Theorem, which is proved in Exercise 32 in Section 5.5.

Binomial Theorem: $(u+v)^n = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} u^{n-k} v^k$

The next example illustrates one way of dealing with a denominator that has both first and second degree factors.

EXAMPLE 4 Obtain the partial-fraction representation of $\frac{x^2}{x^4-1}$.

SOLUTION First factor the denominator: $x^4 - 1 = (x^2+1)(x+1)(x-1)$. There are constants c_1, c_2, c_3, and c_4 such that

As a check, note that there are 4 constraints to find and $x^4 - 1$ has degree 4.

$$\frac{x^2}{x^4-1} = \frac{c_1}{x+1} + \frac{c_2}{x-1} + \frac{c_3 x + c_4}{x^2+1}.$$

Clear the denominator, but *then do not expand the right-hand side*:

$$x^2 = c_1(x-1)(x^2+1) + c_2(x+1)(x^2+1) + (c_3 x + c_4)(x-1)(x+1). \quad (8.4.7)$$

Instead, substitute $x = 1$ and $x = -1$ into (8.4.7) to obtain, respectively:

$$\begin{array}{rcll} 1 & = & 0 + 4c_2 + 0 & [\text{substitute } x = 1 \text{ in } (8.4.7)] \\ 1 & = & -4c_1 + 0 + 0 & [\text{substitute } x = -1 \text{ in } (8.4.7)]. \end{array}$$

Already we see that $c_1 = \frac{-1}{4}$ and $c_2 = \frac{1}{4}$.

Next, substitute 0 for x in (8.4.7), obtaining

Setting $x = 0$ compares the constant terms on both sides of (8.4.7).

$$0 \quad = \quad -c_1 + c_2 - c_4 \qquad [\text{substituting } x = 0 \text{ in } (8.4.7)].$$

Hence $c_4 = \frac{1}{2}$.

We still have to find c_3. We could substitute another number, say $x = 2$, or compare coefficients in (8.4.7). Let us compare coefficients of just the highest degree, x^3. Without going to the bother of multiplying (8.4.7) out in full, we can read off the coefficient of x^3 on both sides by sight, getting

$$0 = c_1 + c_2 + c_3.$$

Since $c_1 = \frac{-1}{4}$, $c_2 = \frac{1}{4}$, if follows that $c_3 = 0$. Hence

$$\frac{x^2}{x^4-1} = \frac{\frac{-1}{4}}{x+1} + \frac{\frac{1}{4}}{x-1} + \frac{\frac{1}{2}}{x^2+1}.$$

◇

Example 4 used a combination of two methods: substituting convenient values of x and equating coefficients. We could have just compared coefficients. There would be an equation corresponding to each power of x up to x^3. That would give 4 equations in 4 unknowns. The Exercises suggest how to solve such equations, if you must solve them by hand.

The constant term corresponds to the power x^0.

Summary

We described ways to integrate rational functions. The key idea is algebraic: express the function as the sum of functions that are easier to integrate.

The first step is to check that the integrand is a proper rational function, that is, the degree of the numerator is less than the degree of the denominator. If it isn't, use long division to express the function as the sum of a polynomial and a proper rational function. A flowchart for this process is presented in Figure 8.4.1.

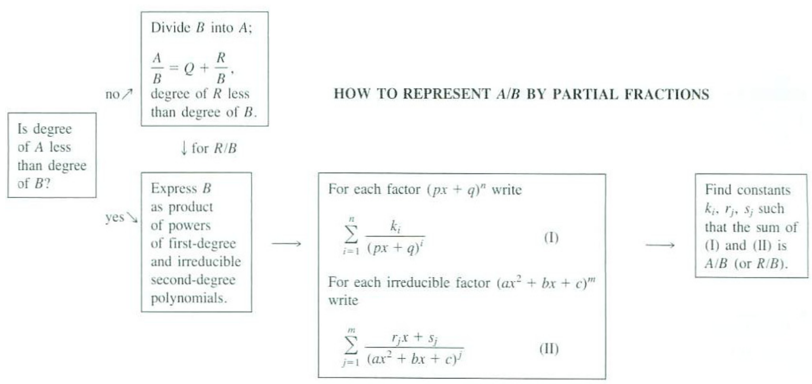


Figure 8.4.1

THE REAL WORLD

Say that you wanted to compute the definite integral

$$\int_1^2 \frac{x+3}{x^3+x^2+x+2}\,dx.$$

One way is by partial fractions, but this can be tedious. You would probably prefer to estimate the definite integral by one of the approximation techniques in Section 6.5. Alternatively, computers and many scientific calculators, can be programmed to estimate a definite integral. On many graphing calculators you would enter the integrand, the variable of integration, and the limits of integration. In a matter of seconds the TI-89 provides 0.49353 as an approximation with an error less than 0.00001.

As noted in Chapter 6, in some cases computers and calculators can even give the exact (symbolic) value of a definite integral by first finding an antiderivative. In practical applications, however, formal antidifferentiation is not that important. The present example could theoretically be computed by partial fractions, but modern computational tools can evaluate it accurately to as many decimal places as we may want. For example, Simpson's rule with only 8 sections gives 0.514393 as an approximate value for this definite integral.

In other situations some of the coefficients in either the numerator or denominator of the integrand may be given only as decimal approximations. In these situations, too, it often is easier and more appropriate to use a computational method to obtain a numerical answer.

EXERCISES for Section 8.4

In Exercises 1 to 10 give the form of the partial fraction representation, but do not find the unknowns. Each expression is already proper.

1. $\dfrac{3x^3+5x+2}{(x-1)(x-2)(x-3)(x-4)}$

2. $\dfrac{x^2-5x+3}{(x+1)^2(2x+3)}$

3. $\dfrac{2x^2+x+1}{(x+1)^3}$

4. $\dfrac{3x}{(x+1)(2x+2)}$

5. $\dfrac{x^2-x+3}{(x+1)(x^2+1)}$

6. $\dfrac{2x^2+3x+5}{(x-1)(x^2+x+1)}$

7. $\dfrac{5x^3+x^2+1}{(x^2+x+1)^2}$

8. $\dfrac{x^3+x+1}{(x^2+x+1)^3}$

9. $\dfrac{x^2+x+2}{x^3-x}$

10. $\dfrac{x^2+x+2}{x^4-1}$

11. The rational function $1/(a^2-x^2)$, where a is constant, commonly appears in applications. Represent this function in partial fractions.

Exercises 12 to 15 concern improper rational functions. In each case express the given function as the sum of a polynomial and a proper rational function.

12. $\dfrac{x^2}{x^2+x+1}$

13. $\dfrac{x^3}{(x+1)(x+2)}$

14. $\dfrac{x^5-2x+1}{(x+1)(x^2+1)}$

15. $\dfrac{x^5+x}{(x+1)^2(x-2)}$

In Exercises 16 to 19 find the partial fraction representation.

16. $\dfrac{5}{x^2-1}$

17. $\dfrac{x+3}{(x+1)(x+2)}$

18. $\dfrac{1}{(x-1)^2(x+2)}$

19. $\dfrac{6x^2-2}{(x-1)(x-2)(2x-3)}$

20. Show that $\dfrac{6+5e^{3x}+2e^{2x}+e^x}{5+e^{2x}+e^x}$ has an elementary antiderivative, but do not find it.

21. Solve Example 3 by clearing the denominator in (8.4.3) to get

$$1 = k_1(x+3) + k_2(2x+1).$$

Replace x by any number you please. That gives an equation in k_1 and k_2. Then replace x by another number of your choice, to obtain a second equation in k_1 and k_2. Solve the equations. Why are $x = -3$ and $x = -1/2$ the nicest choices?

22. Express each of these polynomials as the product of first degree polynomials.

(a) x^2+2x+1

(b) x^2+5x-3

(c) x^2-4x-6

(d) $2x^2+3x-4$

23. Which of these polynomials is irreducible:

(a) $3x^2+2x+1$

(b) $2x^2+4x+1$

In Exercises 24 to 33 express the rational function in terms of partial fractions.

24. $\dfrac{5x^2-x-1}{x^2(x-1)}$

25. $\dfrac{2x^2+3}{x(x+1)(x+2)}$

26. $\dfrac{5x^2 - 2x - 2}{x(x^2 - 1)}$

27. $\dfrac{5x^2 + 9x + 6}{(x+1)(x^2 + 2x + 2)}$

28. $\dfrac{5x^2 + 2x + 3}{x(x^2 + x + 1)}$

29. $\dfrac{x^3 - 3x^2 + 3x - 3}{x^2 - 3x + 2}$

30. $\dfrac{3x^3 + 2x^2 + 3x + 1}{x(x^2 + 1)}$

31. $\dfrac{x^5 + 2x^4 + 4x^3 + 2x^2 + x - 2}{x^4 - 1}$

32. $\dfrac{5x^2 + 6x + 10}{(x-2)(x^2 + 3x + 4)}$

33. $\dfrac{3x^2 - x - 2}{(x+1)(2x^2 + x + 1)}$

34.

(a) For which value of b is $3x^2 + bx + 2$ reducible? irreducible?

(b) For which value of b is $3x^2 + bx - 2$ reducible? irreducible?

35.

(a) For which value of c is $3x^2 + 5x + c$ reducible? irreducible?

(b) For which value of c is $3x^2 - 5x + c$ reducible? irreducible?

36.

Sam: I found this formula in my integral tables:

$$\int \frac{dx}{a^2 - b^2x^2} \quad = \quad \frac{1}{2ab} \ln \left| \frac{a + bx}{a - bx} \right| \qquad (a,\, b \text{ constants})$$

Jane: What's your point?

Sam: My instructor said you won't get any logs other than logs of linear and quadratic polynomials.

Jane: Maybe the table is wrong.

Sam: I took the derivative. It's correct. Can I sue my instructor for misleading the young?

Does Sam has a foundation for a case against his instructor? Explain.

We did not discuss the problem of factoring a polynomial $B(x)$ into linear and irreducible quadratic polynomials. Exercises 37 to 41 concern this problem when the degree of $B(x)$ is 2, 3, or 4.

37.

(a) Show that if $b^2 - 4ac > 0$, then $ax^2 + bx + c = a(x - r_1)(x - r_2)$, where r_1 and r_2 are the distinct roots of $ax^2 + bx + c$.

(b) Show that if $b^2 - 4ac = 0$, then $ax^2 + bx + c = a(x - r)(x - r)$, with r the only root of $ax^2 + bx + c - 0$.

The two parts show that if $b^2 - 4ac \geq 0$, then $ax^2 + bx + c$ is reducible. Compare with Exercise 38.

38.

(a) Show that if $ax^2 + bx + c$ is reducible, then it can be written in the form $a(x - s_1)(x - s_2)$ for some real numbers s_1 and s_2.

(b) Deduce that s_1 and s_2 are the roots of $ax^2 + bx + c = 0$.

(c) Deduce that $b^2 - 4ac \geq 0$.

It follows that if $ax^2 + bx + c$ is reducible, then $b^2 - 4ac \geq 0$. Compare with Exercise 37.

39. Factor each of these polynomials:

(a) $x^2 + 6x + 5$,

(b) $x^2 - 5$,

(c) $2x^2 + 6x + 3$.

40.

(a) Show that $x^2 + 3x - 5$ is reducible.

(b) Using (a), find $\int dx/(x^2 + 3x - 5)$ by partial fractions.

(c) Find $\int dx/(x^2 + 3x - 5)$ by using an integral table.

41. Compute as easily as possible.

(a) $\displaystyle\int \frac{x^3\,dx}{x^4+1}$

(b) $\displaystyle\int \frac{x\,dx}{x^4+1}$

(c) $\displaystyle\int \frac{dx}{x^4+1}$

42. Show that any rational function of e^x has an elementary antiderivative. That is, any function of the form $\frac{P(e^x)}{Q(e^x)}$ where P and Q are polynomials.

43. If ax^2+bx+c is irreducible must ax^2-bx+c also be irreducible? Must ax^2+bx-c?

44. Explain why every polynomial of odd degree has at least one linear factor. (Therefore, every polynomial of odd degree greater than one is reducible.)

45. In arithmetic every fraction can be written as an integer plus a proper fraction. For instance, $\frac{25}{3}=8+\frac{1}{3}$. Why?

46. In arithmetic, the analog of the partial fraction representation is this: Every fraction can be written as the sum of an integer and fractions of the form c/p^n, where p is a prime and $|c|$ is less than p. Check that this is true for 53/18.

47. Let a be a solution of the equation $P(x)=0$, where $P(x)$ is a polynomial. Prove that $x-a$ must be a factor of $P(x)$. (When you use long division to divide $P(x)$ by $x-a$, show why the remainder is 0.) This is the basis for the Factor Theorem.

48.

(a) Use the quadratic formula to find the roots of $x^2+7x+9=0$.

(b) With the aid of the Factor Theorem (Exercise 47), write x^2+7x+9 as the product of two linear polynomials.

(c) Check the factorization by multiplying it out.

49. Assume $x-c$ is a factor of $Q(x)$ and not of $P(x)$. Also assume $(x-c)^2$ is not a factor of $Q(x)$. The term $A/(x-c)$ therefore appears in the partial fraction representation of $P(x)/Q(x)$. Show that $A=P(c)/Q'(c)$. (First, multiply both sides of the partial fraction representation by $x-c$.)

8.5 Special Techniques

So far in this chapter you have met three techniques for computing integrals. The first, substitution, and the second, integration by parts, are used most often. Partial fractions applies to special rational functions and is used in solving some differential equations. In this section we compute some special integrals such as $\int \sin(mx)\cos(nx)\ dx$, $\int \sin^2(\theta)\ d\theta$, and $\int \sec(\theta)\ d\theta$, which you may meet in applications. Then we describe substitutions that deal with special classes of integrands.

Computing $\int \sin(mx)\sin(nx)\ dx$

The integrals

m and n are integers

$$\int \sin(mx)\sin(nx)\, dx, \qquad \int \cos(mx)\sin(nx)\, dx, \quad \text{and} \quad \int \cos(mx)\cos(nx)\, dx$$

are needed in the study of Fourier series, an important tool in the study of heat, sound, and signal processing. They can be computed with the aid of the identities:

Fourier series are discussed in Section 12.7

$$\begin{aligned}
\sin(A)\sin(B) &= \frac{1}{2}\cos(A-B) - \frac{1}{2}\cos(A+B);\\
\sin(A)\cos(B) &= \frac{1}{2}\sin(A+B) + \frac{1}{2}\sin(A-B);\\
\cos(A)\cos(B) &= \frac{1}{2}\cos(A-B) + \frac{1}{2}\cos(A+B).
\end{aligned}$$

These identities can be checked using the well-known identities for $\sin(A \pm B)$ and $\cos(A \pm B)$.

EXAMPLE 1 Find $\int_0^{\pi/4} \sin(3x)\sin(2x)\ dx$.

SOLUTION

$$\begin{aligned}
\int_0^{\pi/4} \sin(3x)\sin(2x)\ dx &= \int_0^{\pi/4} \left(\frac{1}{2}\cos(x) - \frac{1}{2}\cos(5x)\right)\ dx\\
&= \left.\left(\frac{1}{2}\sin(x) - \frac{1}{10}\sin(5x)\right)\right|_0^{\pi/4}\\
&= \left(\frac{\sqrt{2}}{4} + \frac{\sqrt{2}}{20}\right) - \left(\frac{0}{2} - \frac{0}{10}\right) = \frac{3\sqrt{2}}{10} \approx 0.42426.
\end{aligned}$$

◇

Computing $\int \sin^2(x)\ dx$ and $\int \cos^2(x)\ dx$

These integrals can be computed with the aid of the identities

$$\sin^2(x) = \frac{1-\cos(2x)}{2} \qquad \text{and} \qquad \cos^2(x) = \frac{1+\cos(2x)}{2}. \tag{8.5.1}$$

EXAMPLE 2 Find an antiderivative of $\sin^2(x)$:
SOLUTION

$$\int \sin^2(x)\ dx = \int \frac{1-\cos(2x)}{2}\ dx = \int \frac{dx}{2} - \int \frac{\cos(2x)}{2}\ dx = \frac{x}{2} - \frac{\sin(2x)}{4} + C.$$

◇

Computing $\int \tan(\theta)\ d\theta$ and $\int \tan^2(\theta)\ d\theta$

Antiderivatives of $\tan(\theta)$ and $\sec(\theta)$ are found using similar methods.

EXAMPLE 3 Find $\displaystyle\int \tan(\theta)\ d\theta$.

SOLUTION The approach is to rewrite the integrand in a form where the trigonometric functions can be eliminated with a substitution. In the present case, this is accomplished by writing $\tan(\theta) = \frac{\sin(\theta)}{\cos(\theta)}$ and using the substitution $u = \cos(\theta)$ and $du = -\sin(\theta)$ as follows:

$$\int \tan(\theta)\ d\theta = \int \frac{\sin(\theta)}{\cos(\theta)}\ d\theta = \int \frac{-du}{u} = -\ln(u) + C = -\ln|\cos(\theta)| + C. \tag{8.5.2}$$

◇

Most integral tables have the formula

$$\int \tan(\theta)\ d\theta = \ln|\sec(\theta)| + C. \tag{8.5.3}$$

Exercise 49 shows that this formula agrees with (8.5.2).

Finding $\int \tan^2(\theta)\ d\theta$ is easier. Using the trigonometric identity $\tan^2(\theta) = \sec^2(\theta) - 1$, we obtain

$$\int \tan^2(\theta)\ d\theta = \int (\sec^2(\theta) - 1)\ d\theta = \tan(\theta) - \theta + C.$$

Computing $\int \sec(\theta)\ d\theta$

EXAMPLE 4 Find $\displaystyle\int \sec(\theta)\ d\theta$, assuming $0 \le \theta < \pi/2$.

This integral is the key to Mercator maps, discussed in the CIE on page 858.

SOLUTION We begin by, once again, rewriting the integrand in a form where substitution can be used:

$$\int \sec(\theta)\ d\theta = \int \frac{1}{\cos(\theta)}\ d\theta = \int \frac{\cos(\theta)}{\cos^2(\theta)}\ d\theta = \int \frac{\cos(\theta)}{1-\sin^2(\theta)}\ d\theta.$$

The substitution $u = \sin(\theta)$ and $du = \cos(\theta)\ d\theta$ transforms this last integral into the integral of a rational function:

$$\begin{aligned} \int \frac{du}{1-u^2} &= \frac{1}{2}\int \left(\frac{1}{1+u} + \frac{1}{1-u}\right)\ du \\ &= \frac{1}{2}\left(\ln(1+u) - \ln(1-u)\right) + C \\ &= \frac{1}{2}\ln\left(\frac{1+u}{1-u}\right) + C. \end{aligned}$$

Since $u = \sin(\theta)$,

Because $\frac{1+u}{1-u}$ is positive for $-1 < u < 1$, absolute values are not needed.

$$\frac{1}{2}\ln\left(\frac{1+u}{1-u}\right) = \frac{1}{2}\ln\left(\frac{1+\sin(\theta)}{1-\sin(\theta)}\right).$$

Thus,

$$\int \sec(\theta)\ d\theta = \frac{1}{2}\ln\left(\frac{1+\sin(\theta)}{1-\sin(\theta)}\right) + C. \tag{8.5.4}$$

◇

Most integral tables have the formula

Another formula for $\int \sec(\theta) d\theta$.

$$\int \sec(\theta)\ d\theta = \ln|\sec(\theta) + \tan(\theta)| + C. \tag{8.5.5}$$

Exercise 48 shows that this formula agrees with (8.5.4).

In contrast to Example 4, $\int \sec^2(\theta)\ d\theta$ is easy, since it is simply $\tan(\theta) + C$.

The Substitution $u = \sqrt[n]{ax+b}$

The next example illustrates the use of the substitution $u = \sqrt[n]{ax+b}$. After the example we describe the integrands for which the substitution is appropriate.

EXAMPLE 5 Find $\displaystyle\int_4^7 x^2\sqrt{3x+4}\ dx$.

SOLUTION Let $u = \sqrt{3x+4}$, hence $u^2 = 3x+4$. Then $x = (u^2-4)/3$ and

$dx = (2u/3)\ du$. Moreover, as x goes from 4 to 7, u goes from $\sqrt{16} = 4$ to $\sqrt{25} = 5$. Thus

$$\begin{aligned}\int_4^7 x^2\sqrt{3x+4}\ dx &= \int_4^5 \underbrace{\left(\frac{u^2-4}{3}\right)^2}_{x^2} \underbrace{u}_{\sqrt{3x+4}} \underbrace{\frac{2u}{3}\ du}_{dx} = \frac{2}{27}\int_4^5 (u^2-4)^2u^2\ du \\ &= \frac{2}{27}\int_4^5 (u^6 - 8u^4 + 16u^2)\ du = \frac{1214614}{2835} \approx 428.43527.\end{aligned}$$

◇

Exercise 54 uses the substitution $u = \sqrt[n]{ax+b}$ to integrate any rational function of x and $u = \sqrt[n]{ax+b}$.

Three Trigonometric Substitutions

For the following substitutions we need the notion of a rational function in two variables, u and v. First, a polynomial in u and v is a sum of terms of the form $cu^m v^n$, where c is a number and m and n are nonnegative integers. The quotient of two such polynomials is called a **rational function in two variables**, and labeled $R(u,v)$. If one replaces u by x and v by $\sqrt{a^2-x^2}$ we obtain a rational function of x and $\sqrt{a^2-x^2}$, $R(x, \sqrt{a^2-x^2})$.

Any rational function of x and $\sqrt{a^2-x^2}$, where a is a constant, is transformed into a rational function of $\cos(\theta)$ and $\sin(\theta)$ by the substitution $x = a\sin(\theta)$. Similar substitutions are possible for integrands involving $\sqrt{a^2+x^2}$ or $\sqrt{x^2-a^2}$. In each case, one of the trigonometric identities $1-\sin^2(\theta) = \cos^2(\theta)$, $\tan^2(\theta)+1$, or $\sec^2(\theta)-1 = \tan^2(\theta)$ converts a sum or difference of squares into a perfect square.

How to integrate

If the integrand is a rational function of x and

$R(x, \sqrt{a^2-x^2})$ **Case 1** $\sqrt{a^2-x^2}$; let $x = a\sin(\theta)$ $\quad$ ($a > 0$, $-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$).

$R(x, \sqrt{a^2+x^2})$ **Case 2** $\sqrt{a^2+x^2}$; let $x = a\tan(\theta)$ $\quad$ ($a > 0$, $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$).

$R(x, \sqrt{x^2-a^2})$ **Case 3** $\sqrt{x^2-a^2}$; let $x = a\sec(\theta)$ $\quad$ ($a > 0$, $0 \le \theta \le \pi$, $\theta \ne \frac{\pi}{2}$).

The motivation is simple. Consider Case 1, for instance. If you replace x in $\sqrt{a^2-x^2}$ by $a\sin(\theta)$, you obtain

How to make the square root sign in $\sqrt{a^2-x^2}$ disappear

$$\sqrt{a^2-x^2} = \sqrt{a^2-(a\sin(\theta))^2} = \sqrt{a^2(1-\sin^2(\theta))} = \sqrt{a^2\cos^2(\theta)} = a\cos(\theta).$$

(Keep in mind that a and $\cos(\theta)$ are positive.) The important thing is that *the square root sign disappears.*

Case 3 raises a fine point. We have $a > 0$. However, whenever x is negative, θ is an angle in the second-quadrant, so $\tan(\theta)$ is *negative*. In that case,

If $c < 0$, $\sqrt{c^2} = -c$.

$$\sqrt{x^2 - a^2} = \sqrt{(a\sec(\theta))^2 - a^2} = a\sqrt{\sec^2(\theta) - 1} = a\sqrt{\tan^2(\theta)} = a(-\tan(\theta))$$

In the Examples and Exercises involving Case 3 it will be assumed that x varies through nonnegative values, so that θ remains in the first quadrant and $\sqrt{\sec^2(\theta) - 1} = \tan(\theta)$.

Note that for $\sqrt{a^2 - x^2}$ to be meaningful, $|x|$ must be no larger than a. On the other hand, for $\sqrt{x^2 - a^2}$ to be meaningful, $|x|$ must be at least as large as a. The quantity $\sqrt{a^2 + x^2}$ is meaningful for all values of x.

EXAMPLE 6 Compute $\int \sqrt{1 + x^2}\, dx$

SOLUTION The identity $\sqrt{1 + \tan^2(\theta)} = \sec(\theta)$ suggests the substitution

$$\begin{aligned} x &= \tan(\theta) \\ \text{so that} \qquad dx &= \sec^2(\theta)\, d\theta. \end{aligned}$$

(Figure 8.5.1 shows the geometry of this substitution.) Thus

$$\int \sqrt{1 + x^2}\, dx = \int \sec(\theta)\sec^2(\theta)\, d\theta = \int \sec^3(\theta)\, d\theta.$$

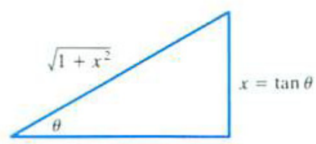


Figure 8.5.1

By Formula 51 from the integral table on the front cover,

$$\int \sec^3(\theta)\, d\theta = \frac{\sec(\theta)\tan(\theta)}{2} + \frac{1}{2}\ln|\sec(\theta) + \tan(\theta)| + C. \qquad (8.5.6)$$

To express the antiderivative just obtained in terms of x rather than θ, it is necessary to express $\tan\theta$ and $\sec\theta$ in terms of x. Starting with the definition $x = \tan(\theta)$, find $\sec(\theta)$ by means of the relation $\sec(\theta) = \sqrt{1 + \tan^2(\theta)} = \sqrt{1 + x^2}$, as in Figure 8.5.1. Thus

$$\int \sqrt{1 + x^2}\, dx = \frac{x\sqrt{1 + x^2}}{2} + \frac{1}{2}\ln\left(\sqrt{1 + x^2} + x\right) + C. \qquad (8.5.7)$$

◇

EXAMPLE 7 Compute $\int_4^5 \frac{dx}{\sqrt{x^2 - 9}}$.

SOLUTION Let $x = 3\sec(\theta)$; hence $dx = 3\sec(\theta)\tan(\theta)\, d\theta$. (See Fig-

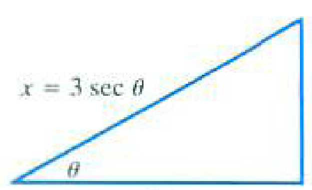
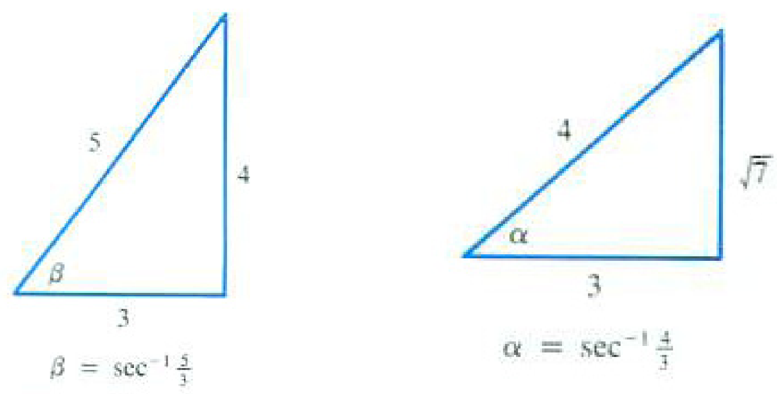


Figure 8.5.2

ure 8.5.2.) Thus, letting $\alpha = \text{arcsec}(4/3)$ and $\beta = \text{arcsec}(5/3)$, we obtain

$$\begin{aligned}
\int_4^5 \frac{dx}{\sqrt{x^2-9}} &= \int_\alpha^\beta \frac{2\sec(\theta)\tan(\theta)\ d\theta}{\sqrt{9\sec^2(\theta)-9}} = \int_\alpha^\beta \frac{\sec(\theta)\tan(\theta)\ d\theta}{\tan(\theta)} \\
&= \int_\alpha^\beta \sec(\theta)\ d\theta = \ln|\sec(\theta)+\tan(\theta)|\Big|_\alpha^\beta \\
&= \ln\left(\tfrac{5}{3}+\tfrac{4}{3}\right) - \ln\left(\tfrac{4}{3}+\tfrac{\sqrt{7}}{3}\right) = \ln(3) - \ln\left(\tfrac{4+\sqrt{7}}{3}\right) \\
&= 2\ln(3) - \ln(4+\sqrt{7}) = \ln\left(\tfrac{9}{4+\sqrt{7}}\right) \approx 0.30325.
\end{aligned}$$



(a) (b)

Figure 8.5.3
Figures 8.5.3(a) and (b) were used to find $\tan(\alpha) = \frac{\sqrt{7}}{3}$ and $\tan(\beta) = \frac{4}{3}$. $\diamond$

A Half-Angle Substitution for $R(\cos\theta, \sin\theta)$

Any rational function of $\cos(\theta)$ and $\sin(\theta)$ is transformed into a rational function of u by the substitution $u = \tan(\theta/2)$. This is sometimes useful after one of the three basic trigonometric substitutions has been used, leaving the integrand in terms of $\cos(\theta)$ and $\sin(\theta)$. The substitution $u = \tan(\theta/2)$ then yields an integral that can be treated by partial fractions. (See Exercises 56 and 57.)

Summary

We discussed some special integrals and integration techniques. First we saw how to evaluate the following common integrals:

$$\int \sin(mx)\sin(nx)\ dx, \quad \int \sin(mx)\cos(nx)\ dx, \quad \int \cos(mx)\cos(nx)\ dx,$$

$$\int \sin^2(x)\ dx, \quad \int \cos^2(x)\ dx,$$

$$\int \sec(\theta)\ d\theta, \quad \int \tan(\theta)\ d\theta, \quad \text{and} \int \tan^2(\theta)\ d\theta.$$

The integration of higher powers of the trigonometric functions is discussed in the exercises.

We also pointed out that the substitution $u = \sqrt[n]{ax+b}$ transforms a rational function in x and $\sqrt[n]{ax+b}$, $R(x, \sqrt[n]{ax+b})$, into a rational function of u. Similarly, $R(x, \sqrt[n]{a^2-x^2})$, $R(x, \sqrt[n]{x^2-a^2})$ and $R(x, \sqrt[n]{a^2+x^2})$ can be transformed into rational functions of $\cos(\theta)$ and $\sin(\theta)$ by trigonometric substitutions. $R(\cos(\theta), \sin(\theta))$ can be transformed into a rational function of u by the substitution $u = \tan(\theta/2)$, which can then be integrated by partial fractions.

EXERCISES for Section 8.5

Exercises 1 to 16 are related to Examples 1 to 3. In Exercises 1 to 14 find the integrals.

1. $\int \sin(5x)\sin(3x)\ dx$

2. $\int \sin(5x)\cos(2x)\ dx$

3. $\int \cos(3x)\sin(2x)\ dx$

4. $\int \cos(2\pi x)\sin(5\pi x)\ dx$

5. $\int \sin^2(3x)\ dx$

6. $\int \cos^2(5x)\ dx$

7. $\int (3\sin(2x) + 4\sin^2(5x))\ dx$

8. $\int (5\cos(2x) + \cos^2(7x))\ dx$

9. $\int (3\sin^2(\pi x) + 4\cos^2(\pi x))\ dx$

10. $\int \sec(3\theta)\ d\theta$

11. $\int \tan(2\theta)\ d\theta$

12. $\int \sec^2(4x)\ dx$

13. $\int \tan^2(5x)\ dx$

14. $\int \frac{dx}{\cos^2(3x)}$

15. Show that $\sin(A)\sin(B) = \frac{1}{2}\cos(A-B) - \frac{1}{2}\cos(A+B)$.

16. Show that $\sin(A)\cos(B) = \frac{1}{2}\sin(A+B) + \frac{1}{2}\sin(A-B)$.

Exercises 17 to 19 develop the formulas that are the foundation for Fourier series, discussed in more detail in Section 12.7.

17. Let m and k be distinct positive integers. Show that

(a) $\int\limits_{-L}^{L} \sin\left(\frac{k\pi x}{L}\right)\sin\left(\frac{k\pi x}{L}\right)\ dx = L.$

(b) $\displaystyle\int_{-L}^{L} \sin\left(\frac{k\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)\,dx = 0.$

18. Let m and k be distinct positive integers. Show that

(a) $\displaystyle\int_{-L}^{L} \cos\left(\frac{k\pi x}{L}\right)\cos\left(\frac{k\pi x}{L}\right)\,dx = L.$

(b) $\displaystyle\int_{-L}^{L} \cos\left(\frac{k\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)\,dx = 0.$

19. Let m and k be distinct positive integers. Show that $\displaystyle\int_{-L}^{L} \sin\left(\frac{k\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)\,dx =$ 0.

Exercises 20 to 29 concern the substitution $u = \sqrt[n]{ax+b}$. In each case evaluate the integral.

20. $\displaystyle\int x^2\sqrt{2x+1}\,dx$

21. $\displaystyle\int \frac{x^2\,dx}{\sqrt[3]{x+1}}$

22. $\displaystyle\int \frac{dx}{\sqrt{x+3}}$

23. $\displaystyle\int \frac{\sqrt{2x+1}}{x}\,dx$

24. $\displaystyle\int x\sqrt[3]{3x+2}\,dx$

25. $\displaystyle\int \frac{\sqrt{x}+3}{\sqrt{x}-2}\,dx$

26. $\displaystyle\int \frac{x\,dx}{\sqrt{x}+3}$

27. $\displaystyle\int x(3x+2)^{5/3}\,dx$

28. $\displaystyle\int \frac{dx}{\sqrt[3]{x}+\sqrt{x}}$ (Let $u = \sqrt[6]{x}$).

29. $\displaystyle\int (x+2)\sqrt[5]{x-3}\,dx$

In Exercises 30 to 40 find the integrals using trigonometric substitutions. (a is a positive constant.)

30. $\int \sqrt{4-x^2}\ dx$

31. $\int \frac{dx}{\sqrt{9+x^2}}$

32. $\int \frac{x^2\ dx}{\sqrt{x^2-9}}$

33. $\int x^3\sqrt{1-x^2}\ dx$

34. $\int \frac{\sqrt{4+x^2}}{x}\ dx$

35. $\int \sqrt{a^2-x^2}\ dx$

36. $\int \frac{dx}{\sqrt{a^2-x^2}}$

37. $\int \sqrt{a^2+x^2}\ dx$

38. $\int \sqrt{a^2-x^2}\ dx$

39. $\int \frac{dx}{\sqrt{25x^2-16}}$

40. $\int_{\sqrt{2}}^{2} \sqrt{x^2-1}\ dx$

Exercises 41 and 42 concern the recursion formulas for $\int \tan^n(\theta)\ d\theta$ and $\int \sec^n(\theta)\ d\theta$.

41. In Example 3 we found $\int \tan(\theta)\ d\theta$ and $\int \tan^2(\theta)\ d\theta$.

(a) Obtain the recursion

$$\int \tan^n(\theta)\ d\theta = \frac{\tan^{n-1}(\theta)}{n-1} - \int \tan^{n-2}(\theta)\ d\theta.$$

Begin by writing

$$\tan^n(\theta) \quad = \quad \tan^{n-2}(\theta)\tan^2(\theta) = \tan^{n-2}(\theta)(\sec^2(\theta)-1).$$

(b) Use the recursion formula to find $\displaystyle\int \tan^3(\theta)\ d\theta$.

(c) Find $\displaystyle\int \tan^4(\theta)\ d\theta$.

See Example 3.

42. In Example 4 we found $\int \sec(\theta)\ d\theta$ and $\int \sec^2(\theta)\ d\theta$.

(a) Obtain the recursion

$$\int \sec^n(\theta)\ d\theta = \frac{\sec^{n-2}(\theta)\tan(\theta)}{n-1} + \frac{n-2}{n-1}\int \sec^{n-2}(\theta)\ d\theta.$$

Begin by writing $\sec^n(\theta) = \sec^{n-2}(\theta)\sec^2(\theta)$, and integrating by parts. After the integration, $\tan^2(\theta)$ will appear in the integrand. Write it as $\sec^2(\theta)-1$.

(b) Evaluate $\displaystyle\int \sec^3(\theta)\ d\theta$.

(c) Evaluate $\displaystyle\int \frac{d\theta}{\cos^4(\theta)}$.

(d) Evaluate $\displaystyle\int \sec^2(2x)\ dx$.

See Example 4.

43. Find

(a) $\displaystyle\int \csc(\theta)\ d\theta$

(b) $\displaystyle\int \csc^2(\theta)\ d\theta$

44. Find

(a) $\displaystyle\int \cot(\theta)\ d\theta$

(b) $\displaystyle\int \cot^2(\theta)\ d\theta$

45. Consider $\displaystyle\int \sin^n(\theta)\cos^m(\theta)\ d\theta$, where m and n are nonnegative integers, and m is odd. To evaluate $\int \sin^n(\theta)\cos^m(\theta)\ d\theta$, write it as $\int \sin^n(\theta)\cos^{m-1}(\theta)\cos(\theta)\ d\theta$. Then, because $m-1$ is even, rewrite $\cos^{m-1}(\theta)$ as $(1-\sin^2(\theta))^{(m-1)/2}$ and use the substitution $u = \sin(\theta)$. Using this technique, find

(a) $\displaystyle\int \sin^3(\theta)\cos^3(\theta)\ d\theta$

(b) $\displaystyle\int \sin^4(\theta)\cos(\theta)\ d\theta$

(c) $\displaystyle\int_0^{\pi/2} \sin^4(\theta)\cos^3(\theta)\ d\theta$

(d) $\displaystyle\int \cos^5(\theta)\ d\theta$.

46. How would you integrate $\int \sin^n(\theta)\cos^m(\theta)\ d\theta$, where m and n are nonnegative integers and n is odd? Illustrate your techniques by three examples. See Exercise 45.

47. The techniques in Exercises 45 and 46 apply to $\int \sin^n(\theta)\cos^m(\theta)\,d\theta$ only when at least one of m and n is odd. If both are even, first use the identities

$$\sin^2(\theta) = \frac{1-\cos(2\theta)}{2} \qquad \text{and} \qquad \cos^2(\theta) = \frac{1+\cos(2\theta)}{2}.$$

You will get a polynomial in $\cos(2\theta)$. If $\cos(2\theta)$ appears only to odd powers, the technique of Exercise 45 suffices. To treat an even power of $\cos(2\theta)$, use the identity $\cos^2(2\theta) = (1+\cos(4\theta))/2$ and continue. Using this method find

(a) $\int \cos^2(\theta)\sin^4(\theta)\;d\theta$

(b) $\int_0^{\pi/4} \cos^2(\theta)\sin^2(\theta)\;d\theta$

Antiderivatives of $\sec(\theta)$ and $\tan(\theta)$ were found in Examples 4 and 3. Exercises 48 to 50 explore some other antiderivatives of these functions.

48. Let $0 \le \theta < \pi/2$.

(a) Show that $\displaystyle\int \sec(\theta)\;d\theta = \ln|\sec(\theta)+\tan(\theta)|+C$, by differentiating $\ln|\sec(\theta)+\tan(\theta)|$.

(b) Does (a) contradict the formula given in Example 4?

49. Show that $-\ln(\cos(\theta))$ and $\ln(\sec(\theta))$ are both antiderivatives for $\tan(\theta)$.

50. In 1645, Henry Bond conjectured from experimental data that $\int_0^\theta \sec(t)\;dt = \ln\left(\tan\left(\frac{\theta}{2}+\frac{\pi}{4}\right)\right)$ While Bond's conjecture was originally verified well before the advent of calculus, today we can verify Bond's conjecture by (i) checking that this formula holds for $\theta = 0$ and (ii) checking that the right-hand side is an antiderivative of $\sec(\theta)$. Bond's conjecture is related to Mercator's projection (discussed in the CIE on page 858. Reference: `http://www.math.ubc.ca/~israel/m103/mercator/mercator.html`. [Does permission need to be requested from Robert Israel?]

51. The region R under $y = \sin(x)$ and above $[0,\pi]$ is revolved about the x-axis to produce a solid S.

(a) Draw R.

(b) Draw S.

(c) Set up a definite integral for the area of R.

(d) Set up a definite integral for the volume of S.

(e) Evaluate the integrals in (c) and (d).

52. Transform the following integrals into integrals of rational functions of $\cos(\theta)$ and $\sin(\theta)$. Do *not* evaluate the integrals.

(a) $\int \frac{x+\sqrt{9-x^2}}{x^3}\,dx$

(b) $\int \frac{x^3\sqrt{5-x^2}}{1+\sqrt{5x^2}}\,dx$

53. Transform the following integrals into integrals of rational functions of $\cos(\theta)$ and $\sin(\theta)$. Do *not* evaluate the integrals.

(a) $\int \frac{x^2+\sqrt{x^2-9}}{x}\,dx$

(b) $\int \frac{x^3\sqrt{5+x^2}}{x+2}]dx$

54. Let $R(x,y)$ be a rational function of x and y. Let n be an integer greater than or equal to 2. Then $R(x, \sqrt[n]{ax+b})$ is a "rational function of x and $\sqrt[n]{ax+b}$." Let $R(x,y) = \frac{x+y^2}{2x-y}$.

(a) Evaluate $R(x, \sqrt[3]{4x+5})$.

(b) Use the substitution $u = \sqrt[3]{4x+5}$ to show that

$$\int \frac{x+(4x+5)^{2/3}}{2x-(4x+5)^{1/3}}\,dx = \frac{3}{8}\int \frac{(u^3+4u^2-5)u^2}{u^3-2u-5}\,du$$

Do not attempt to evaluate this integral. The partial fraction decomposition of this integrand is messy.

55. Transform the following integrals into integrals of rational functions of u. Do *not* evaluate the integrals.

(a) $$\int \frac{\sqrt[3]{x+2}}{x^2+(x+2)^{2/3}}\,dx$$

(b) $$\int \frac{\sqrt{x}+x+x^{3/2}}{\sqrt{x}+2}\,dx$$

Exercises 56 to 58 concern $\int R(\cos(\theta),\sin(\theta))\,d\theta$.

56. Let $-\pi < \theta < \pi$ and $u = \tan(\theta/2)$. (See Figure 8.5.4(a).) The following steps show that this substitution transforms $\int R(\cos\theta,\sin\theta)d\theta$ into the integral of a rational function of u (which can be integrated by partial fractions).

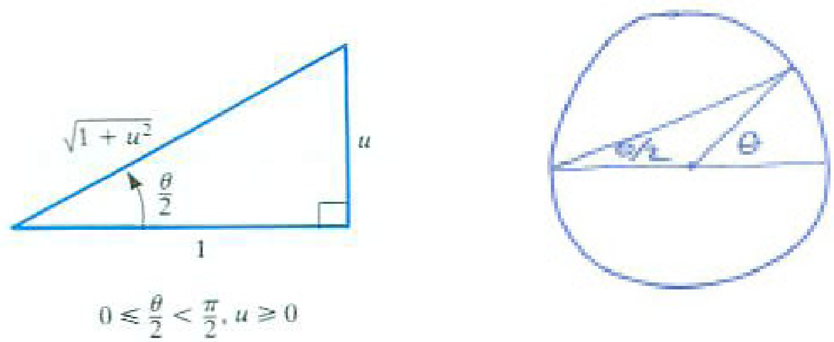
(a) Show that $\cos\left(\frac{\theta}{2}\right) = \frac{1}{\sqrt{1+u^2}}$ and $\sin\left(\frac{\theta}{2}\right) = \frac{u}{\sqrt{1+u^2}}$.

(b) Using (a), show that $\cos(\theta) = \frac{1-u^2}{1+u^2}$.

(c) Show that $\sin(\theta) = \frac{2u}{1+u^2}$.

(d) Show that $d\theta = \frac{2\,du}{1+u^2}$. (Note that $\theta = 2\arctan(u)$.)

Combining (b), (c), and (d) shows that the substitution $u = \tan(\theta/2)$ transforms $\int R(\cos(\theta), \sin(\theta))\,d\theta$ into $\int R\left(\frac{1-u^2}{1+u^2}, \frac{2u}{1+u^2}\right)\,\frac{2}{1+u^2}\,du$, an integral of a rational function of u.



(a) (b)

Figure 8.5.4

57. Using the substitution $u = \tan(\theta/2)$, transform the following integrals into integrals of rational functions. (Refer to Figure 8.5.4(b).) (Do not evaluate them.)

(a) $\displaystyle\int \frac{1+\sin(\theta)}{1+\cos^2(\theta)}\,d\theta$

(b) $\displaystyle\int \frac{5+\cos(\theta)}{(\sin(\theta))^2+\cos(\theta)}\,d\theta$

(c) $\displaystyle\int_0^{\pi/2} \frac{5\,d\theta}{2\cos(\theta)+3\sin(\theta)}$ (Be sure to transform the limits of integraton also.)

58. Compute $\displaystyle\int_0^{\pi/2} \frac{d\theta}{4\sin(\theta)+3\cos(\theta)}$.

59. Explain why any rational function of $\tan(\theta)$ and $\sec(\theta)$ has an elementary antiderivative.

60. Show that any rational function of x, $\sqrt{x+a}$, and $\sqrt{x+b}$ has an elementary antiderivative. (Use the substitution $u=\sqrt{x+a}$.)
However, it is not the case that every rational function of $\sqrt{x+a}$, $\sqrt{x+b}$, and $\sqrt{x+c}$ has an elementary antiderivative. For instance,

$$\int \frac{dx}{\sqrt{x}\sqrt{x+1}\sqrt{x-1}} = \int \frac{dx}{\sqrt{x^3-x}}$$

is not an elementary function.

61. Every rational function of x and $\sqrt[n]{(ax+b)/(cx+d)}$ has an elementary antiderivative. Explain why.

Exercise 62 is known as the tractrix problem. While typically discussed in a differential equations course, only integration is needed to find the solution.
62. A point P is dragged across the xy-plane by a string SP of length a. Let S start at the origin and move to the right along the positive x axis. Assume, as in Figure 8.5.5, P starts at $(0,a)$.

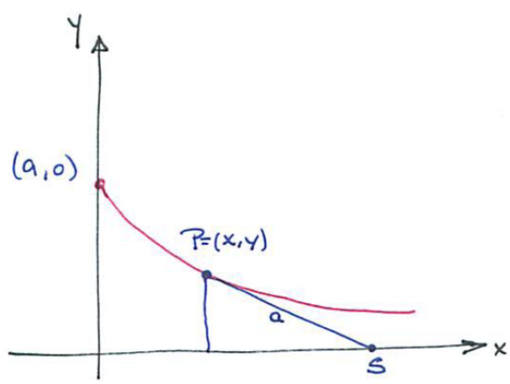


Figure 8.5.5

(a) Find an equation involving dy/dx that the function describing the tractrix satisfies.

(b) Rewrite the equation in the form dx/dy equal to an expression involving y and a (but not x).

(c) Find x explicitly in terms of y.

The **tractrix** can also be visualized as the path of the rear wheel of a scooter when the front wheel follows a straight path. The case when the front wheel follows a circular path is analyzed in CIE 22 at the end of Chapter 15. Revolving the tractrix about the x-axis creates a surface that in non-Euclidean geometry.

8.6 What to do When Confronted with an Integral

Since the exercises in each section of this chapter focus on the techniques of that section, it is usually clear what technique to use on a given integral. But what if an integral is met "in the wild," where there is no clue how to evaluate it? This section suggests what to do in this typical situation.

The more integrals you compute, the more quickly you will be able to choose an appropriate technique. Moreover, such practice will put you at ease in using integral tables or computer software. Besides, it may be quicker to find an integral by hand.

This table summarizes the techniques and shortcuts emphasized in this chapter.

General	Substitution	Section 8.2
	Integration by Parts	Section 8.3
	Partial Fractions	Sections 8.4 and 8.2
Special	if f is odd, then $\int_{-a}^{a} f(x)\ dx = 0$	Section 8.1
	if f is even, then $\int_{-a}^{a} f(x)\ dx = 2\int_0^a f(x)\ dx$	Section 8.1
	$\int_0^a \sqrt{a^2-x^2}\ dx = \frac{\pi a^2}{4}$	Section 8.1
	$\int \sin(mx)\sin(nx)\ dx$, etc.	Section 8.5
	$\int \sin^2(\theta)\ d\theta$, etc.	Section 8.5
	$\int \tan(\theta)\ d\theta$, $\int \sec(\theta)\ d\theta$, etc.	Section 8.5
	$\int R(x, \sqrt[n]{ax+b})\ dx$	Section 8.5
	$\int R(x, \sqrt{a^2-x^2})\ dx$, etc.	Section 8.5
	$\int R(\cos(\theta), \sin(\theta))\ dx$, etc.	Section 8.5

Table 8.6.1

Exercises in Section 8.5 develop other specialized techniques, but they will not be required in this section.

A few examples will illustrate how to choose a method for computing an antiderivative.

EXAMPLE 1

$$\int \frac{x\ dx}{1+x^4}$$

See Exercise 57 in Section 7.5

SOLUTION DISCUSSION: Since the integrand is a rational function of x, partial fractions would work. This requires factoring x^4+1 and then representing $x/(1+x^4)$ as a sum of partial fractions. With some struggle it can be found that

$$x^4+1 = (x^2+\sqrt{2}x+1)(x^2-\sqrt{2}x+1)$$

The constants A, B, C, and D will have to be found such that

$$\frac{x}{1+x^4} = \frac{Ax+B}{x^2+\sqrt{2}x+1} + \frac{Cx+D}{x^2-\sqrt{2}x+1}$$

The method would work but would certainly be tedious.

Try another attack. The numerator x is almost the derivative of x^2. The substitution $u = x^2$ is at least worth testing. With $u = x^2$ we find $du = 2x\ dx$ and so

$$\int \frac{x\ dx}{1+x^4} = \int \frac{du/2}{1+u^2},$$

Check by differentiating.

which is easy:

$$\int \frac{x\ dx}{1+x^4} = \frac{1}{2}\arctan(u) + C = \frac{1}{2}\arctan(x^2) + C.$$

◇

EXAMPLE 2

$$\int \frac{1+x}{1+x^2}\ dx.$$

SOLUTION DISCUSSION: This is a rational function of x, but partial fractions will not help, since the integrand is already in its partial-fraction form.

The numerator is not the derivative of the denominator, but it comes close enough to persuade us to break the integrand into two summands:

$$\int \frac{1+x}{1+x^2}\ dx = \int \frac{dx}{1+x^2} + \int \frac{x\ dx}{1+x^2}.$$

Both the latter integrals can be done in your head. The first is $\arctan(x)$, and the second is $(1/2)\ln(1+x^2)$. So

$$\int \frac{1+x}{1+x^2}\ dx = \arctan(x) + \frac{1}{2}\ln(1+x^2) + C.$$

◇

EXAMPLE 3

$$\int \frac{e^{2x}}{1+e^x}\ dx.$$

SOLUTION DISCUSSION: At first glance, this integral looks so peculiar that it may not even be elementary. However, e^x is a fairly simple function, with $d(e^x) = e^x\ dx$. This suggests trying the substitution $u = e^x$ and seeing what happens:

$$u = e^x \qquad\qquad du = e^x\ dx$$

Thus

$$dx = \frac{du}{e^x} = \frac{du}{u}.$$

It is essential to express dx completely in terms of u and du.

But what will be done to e^{2x}? Recalling that $e^{2x} = (e^x)^2 = u^2$, we anticipate there will be no difficulty:

$$\int \frac{e^{2x}}{1+e^x}\ dx = \int \frac{u^2}{1+u}\frac{du}{u} = \int \frac{u\ du}{1+u}.$$

which can be integrated quickly:

Long division of $u/(u+1)$ also works.

$$\begin{aligned} \int \frac{u\ du}{1+u} &= \int \frac{u+1-1}{1+u}\ du = \int \left(1 - \frac{1}{1+u}\right)\ du \\ &= u - \ln(|1+u|) + C = e^x - \ln(1+e^x) + C. \end{aligned}$$

The same substitution could have been done more elegantly:

$$\int \frac{e^{2x}}{1+e^x}\ dx = \int \frac{e^x(e^x\ dx)}{1+e^x} = \int \frac{u\ du}{1+u}.$$

◇

EXAMPLE 4

$$\int \frac{x^3\ dx}{(1-x^2)^5}.$$

SOLUTION DISCUSSION: Partial fractions would work, but the denominator, when factored, would be $(1+x)^5(1-x)^5$. There would be 10 unknown constants to find. Look for an easier approach.

Since the denominator is the obstacle, try $u = x^2$ or $u = 1 - x^2$ to see if the integrand gets simpler. Let us examine what happens in each case. Try $u = x^2$ first. Assume that we are interested only in getting an antiderivative for positive x, $x = \sqrt{u}$:

$$u = x^2 \qquad\qquad du = 2x\ dx \qquad\qquad dx = \frac{du}{2x} = \frac{du}{2\sqrt{u}}.$$

Then

$$\int \frac{x^3\ dx}{(1-x^2)^5} = \int \frac{u^{3/2}}{(1-u)^5}\frac{du}{2\sqrt{u}} = \frac{1}{2}\int \frac{u\ du}{(1-u)^5}.$$

The same substitution could be carried out as follows:

$$\int \frac{x^3\ dx}{(1-x^2)^5} = \int \frac{x^2 x\ dx}{(1-x^2)^5} = \int \frac{u(du/2)}{(1-u)^5} = \frac{1}{2}\int \frac{u\ du}{(1-u)^5}.$$

Verify this claim for yourself.

The substitution $v = 1-u$ then results in an easy integral.

Observe that the two substitutions $u = x^2$ and $v = 1-u$ are equivalent to the single substitution $w = 1-x^2$. So, let us apply the substitution $w = 1-x^2$ to the original integral. Then $dw = -2x\ dx$; thus

$$\int \frac{x^3\ dx}{(1-x^2)^5} = \int \frac{x^2(x\ dx)}{(1-x^2)^5} = \int \frac{(1-w)(-dw/2)}{w^5} = \int \frac{1}{2}(w^{-4} - w^{-5})\ dw,$$

an integral that can be computed without further substitution. So $w = 1-x^2$ is quicker than $u = x^2$. ◇

EXAMPLE 5

$$\int x^3 e^{x^2}\ dx.$$

SOLUTION DISCUSSION: Integration by parts may come to mind, since if $u = x^3$, then $du = 3x^2\ dx$ is simpler. However, dv must then be $e^{x^2}\ dx$ and force v to be non-elementary. This is a dead end.

If we can raise an exponent, we should be able to lower it.

So try integration by parts with $u = e^{x^2}$ and $dv = x^3\ dx$. What will $v\ du$ be? We have $v = x^4/4$ and $du = 2xe^{x^2}\ dx$, which is worse than the original $u\ dv$. The exponent of x has been raised by 2, from 3 to 5.

This time try $u = x^2$ and $dv = xe^{x^2}\ dx$; thus $du = 2x\ dx$ and $v = e^{x^2}/2$. Integration by parts yields

$$\begin{aligned}\int x^3 e^{x^2}\ dx &= \int \underbrace{x^2}_{u}\underbrace{xe^{x^2}\ dx}_{dv} = \underbrace{x^2}_{u}\underbrace{\frac{e^{x^2}}{2}}_{v} - \int \underbrace{\frac{e^{x^2}}{2}}_{v}\underbrace{2x\ dx}_{du} \\ &= \frac{x^2 e^{x^2}}{2} - \frac{e^{x^2}}{2} + C.\end{aligned}$$

See Exercise 71.

Another approach is to use the substitution $u = x^2$ followed by an integration by parts. ◇

EXAMPLE 6

$$\int \frac{1-\sin(\theta)}{\theta + \cos(\theta)} d\theta.$$

See Exercise 72.

SOLUTION DISCUSSION: The numerator is the derivative of the denominator, so the integral is simply $\ln|\theta + \cos\theta| + C$. ◇

EXAMPLE 7

$$\int \frac{1-\sin(\theta)}{\cos(\theta)}\, d\theta.$$

SOLUTION DISCUSSION: Break the integrand into two summands:

$$\begin{aligned} \int \frac{1-\sin(\theta)}{\cos(\theta)}\, d\theta &= \int \left(\frac{1}{\cos(\theta)} - \frac{\sin(\theta)}{\cos(\theta)}\right)\, d\theta \\ &= \int (\sec(\theta) - \tan(\theta))\, d\theta \\ &= \int \sec\theta\, d\theta - \int \tan(\theta)\, d\theta \\ &= \ln|\sec(\theta) + \tan(\theta)| + \ln|\cos(\theta)| + C. \end{aligned}$$

Since $\ln(A) + \ln(B) = \ln(AB)$, the answer can be simplified to

$$\ln\left(|\sec(\theta) + \tan(\theta)|\, |\cos(\theta)|\right) + C.$$

But $\sec(\theta)\cos(\theta) = 1$ and $\tan(\theta)\cos(\theta) = \sin(\theta)$. The result becomes even simpler:

$$\int \frac{1-\sin(\theta)}{\cos(\theta)}\, d\theta = \ln(1 + \sin(\theta)) + C.$$

The absolute values are not needed because $1 + \sin(\theta) \geq 0$

◇

EXAMPLE 8

$$\int \frac{\ln x\, dx}{x}.$$

SOLUTION DISCUSSION: Integration by parts, with $u = \ln(x)$ and $dv = dx/x$, may come to mind. In that case, $du = dx/x$ and $v = \ln(x)$; thus

$$\int \underbrace{\ln(x)}_{u} \underbrace{\frac{dx}{x}}_{dv} = \underbrace{(\ln(x))}_{u} \underbrace{(\ln(x))}_{v} - \int \underbrace{\ln(x)}_{v} \underbrace{\frac{dx}{x}}_{du}.$$

Bringing $\int \ln(x)\, dx/x$ all to one side produces the equation

$$2\int \ln(x) \frac{dx}{x} = (\ln x)^2,$$

from which it follows that

$$\int \ln(x)\frac{dx}{x} = \frac{(\ln(x))^2}{2} + C.$$

The integration by parts approach worked, but is not the easiest one to use. Since $1/x$ is the derivative of $\ln(x)$, we could have used the substitution $u = \ln(x)$, which means $du = dx/x$. Thus

$$\int \frac{\ln(x)\ dx}{x} = \int u\ du = \frac{u^2}{2} + C = \frac{(\ln(x))^2}{2} + C.$$

◇

EXAMPLE 9

$$\int_0^{3/5} \sqrt{9 - 25x^2}\ dx.$$

SOLUTION DISCUSSION: This integral reminds us of $\int_0^a \sqrt{a^2 - x^2}\ dx = \pi a^2/4$, the area of a quadrant of a circle of radius a. This resemblance suggests a substitution u such that $25x^2 = 9u^2$ or $u = \frac{5}{3}x$, hence $dx = \frac{3}{5}\ du$. Then substitution gives

$$\begin{aligned} \int_0^{3/5} \sqrt{9 - 25x^2}\ dx &= \int_0^1 \sqrt{9 - 9u^2}\ \frac{3}{5}\ du = \frac{9}{5}\int_0^1 \sqrt{1 - u^2}\ du \\ &= \frac{9}{5} \cdot \frac{\pi}{4} = \frac{9\pi}{20} \approx 1.41372. \end{aligned}$$

◇

EXAMPLE 10

$$\int \sin^5(2x)\cos(2x)\ dx.$$

SOLUTION DISCUSSION: We could try integration by parts with $u = \sin^5(2x)$ and $dv = \cos(2x)\ dx$. (See Exercise 73.)

However, $\cos(2x)$ is almost the derivative of $\sin(2x)$. For this reason make the substitution

$$u = \sin(2x) \qquad\qquad du = 2\cos(2x)\ dx;$$

This means that

$$\cos(2x)\ dx = \frac{du}{2}.$$

and so

$$\int \sin^5(2x)\cos(2x)\ dx = \int u^5\ \frac{du}{2} = \frac{1}{2}\frac{u^6}{6} + C = \frac{\sin^6(2x)}{12} + C.$$

◇

EXAMPLE 11

$$\int_{-3}^{3} x^3\cos(x)\ dx.$$

SOLUTION DISCUSSION: Since the integrand is of the form $P(x)\cos(x)$, where P is a polynomial, repeated integration by parts would work. On the other hand, x^3 is an odd function and $\cos(x)$ is an even function. The integrand is therefore an odd function and the integral over $[-3, 3]$ is 0. ◇

EXAMPLE 12

$$\int \sin^2(3x)\ dx.$$

SOLUTION DISCUSSION: You could rewrite this integral as $\int \sin(3x)\sin(3x)\ dx$ and use integration by parts. However, it is easier to use the trigonometric identity $\sin^2(\theta) = (1 - \cos 2(\theta))/2$:

$$\int \sin^2(3x)\ dx = \int \frac{1-\cos(6x)}{2}\ dx = \int \frac{dx}{2} - \int \frac{\cos(6x)}{2}\ dx = \frac{x}{2} - \frac{\sin(6x)}{12} + C.$$

◇

EXAMPLE 13

$$\int_{1}^{2} \frac{x^3 - 1}{(x+2)^2}\ dx.$$

SOLUTION DISCUSSION: Partial fractions would certainly work. (The first step would be division of $x^3 - 1$ by $x^2 + 4x + 4$.) However, the substitution $u = x + 2$ is easier because it makes the denominator simply u^2. We have

$$u = x + 2 \qquad du = dx \quad \text{and} \quad x = u - 2.$$

Note the new limits for u.

Thus

$$
\begin{aligned}
\int_1^2 \frac{x^3-1}{(x+2)^2}\,dx &= \int_3^4 \frac{(u-2)^3-1}{u^2}\,du = \int_3^4 \frac{u^3-6u^2+12u-8-1}{u^2}\,du \\
&= \int_3^4 \left(u-6+\frac{12}{u}-\frac{9}{u^2}\right)\,du = \left(\frac{u^2}{2}-6u+12\ln|u|+\frac{9}{u}\right)\Big|_3^4 \\
&= \left(8-24+12\ln(4)+\frac{9}{4}\right)-\left(\frac{9}{2}-18+12\ln(3)+3\right) \\
&= -(\frac{13}{4})+12\ln(4)-12\ln(3) = 12\ln\left(\frac{4}{3}\right)-\frac{13}{4} \approx 0.20218.
\end{aligned}
$$

◇

Summary

One word: PRACTICE.

Reading worked examples is a first step to mastering integration, but doesn't offer the challenge of having to decide which approach is promising and which will only lead to a dead end. The more you practice, the more comfortable you will be when facing an integral in the wild and using integral tables or software programs that find antiderivatives.

Many integrals can be evaluated in several different ways, but one method is usually the easiest.

It is also important to learn to recognize integrals that can be evaluated without finding an antiderivative or are known to not have an elementary antiderivative.

EXERCISES for Section 8.6

All the integrals in Exercises 1 to 59 are elementary. In each case, list the technique or techniques that could be used to evaluate the integral. If there is a preferred technique, state what it is (and why). Do *not* evaluate the integrals.

1. $\displaystyle\int \frac{1+x}{x^2}\,dx$

2. $\displaystyle\int \frac{x^2}{1+x}\,dx$

3. $\displaystyle\int \frac{dx}{x^2+x^3}$

4. $\displaystyle\int \frac{x+1}{x^2+x^3}\,dx$

5. $\displaystyle\int \arctan(2x)\,dx$

6. $\displaystyle\int \arcsin(2x)\,dx$

7. $\displaystyle\int x^{10}e^x\,dx$

8. $\displaystyle\int \frac{\ln(x)}{x^2}\,dx$

9. $\displaystyle\int \frac{\sec^2(\theta)\,d\theta}{\tan(\theta)}$

10. $\displaystyle\int \frac{\tan(\theta)\,d\theta}{\sin^2(\theta)}$

11. $\displaystyle\int \frac{x^3}{\sqrt[3]{x+2}}\,dx$

12. $\displaystyle\int \frac{x^2}{\sqrt[3]{x^3+2}}\,dx$

13. $\displaystyle\int \frac{2x+1}{(x^2+x+1)^5}\,dx$

14. $\displaystyle\int \sqrt{\cos(\theta)}\sin(\theta)\,d\theta$

15. $\displaystyle\int \tan^2(\theta)\,d\theta$

16. $\displaystyle\int \frac{d\theta}{\sec^2(\theta)}$

17. $\displaystyle\int e^{\sqrt{x}}\,dx$

18. $\displaystyle\int \sin\sqrt{x}\,dx$

19. $\displaystyle\int \frac{dx}{(x^2-4x+3)^2}$

20. $\displaystyle\int \frac{x+1}{x^5}\ dx$

21. $\displaystyle\int \frac{x^5}{x+1}\ dx$

22. $\displaystyle\int \frac{\ln(x)}{x(1+\ln(x))}\ dx$

23. $\displaystyle\int \frac{e^{3x}\ dx}{1+e^x+e^{2x}}$

24. $\displaystyle\int \frac{\cos(x)\ dx}{(3+\sin(x))^2}$

25. $\displaystyle\int \ln(e^x)\ dx$

26. $\displaystyle\int \ln(\sqrt[3]{x})\ dx$

27. $\displaystyle\int \frac{x^4-1}{x+2}\ dx$

28. $\displaystyle\int \frac{x+2}{x^4-1}\ dx$

29. $\displaystyle\int \frac{dx}{\sqrt{x}(3+\sqrt{x})^2}$

30. $\displaystyle\int \frac{dx}{(3+\sqrt{x})^3}$

31. $\displaystyle\int (1+\tan(\theta))^3 \sec^2(\theta)\ d\theta$

32. $\displaystyle\int \frac{e^{2x}+1}{e^x-e^{-x}}\ dx$

33. $\displaystyle\int \frac{e^x+e^{-x}}{e^x-e^{-x}}\ dx$

34. $\displaystyle\int \frac{(x+3)(\sqrt{x+2}+1)}{\sqrt{x+2}-1}\ dx$

35. $\displaystyle\int \frac{(\sqrt[3]{x+2}-1\ dx}{\sqrt{x+2}+1}$

36. $\displaystyle\int \frac{dx}{x^2-9}$

37. $\displaystyle\int \frac{x+7}{(3x+2)^{10}}\ dx$

38. $\displaystyle\int \frac{x^3\ dx}{(3x+2)^7}$

39. $\displaystyle\int \frac{2^x+3^x}{4^x}\ dx$

40. $\displaystyle\int \frac{2^x}{1+2^x}\ dx$

41. $\displaystyle\int \frac{(x+\arcsin(x))\ dx}{\sqrt{1-x^2}}$

42. $\displaystyle\int \frac{x+\arctan(x)}{1+x^2}\ dx$

43. $\displaystyle\int x^3\sqrt{1+x^2}\ dx$

44. $\displaystyle\int x(1+x^2)^{3/2}\ dx$

45. $\displaystyle\int \frac{x\ dx}{\sqrt{x^2-1}}$

46. $\displaystyle\int \frac{x^3}{\sqrt{x^2-1}}\ dx$

47. $\displaystyle\int \frac{x\ dx}{(x^2-9)^{3/2}}$

48. $\displaystyle\int \frac{\arctan(x)}{1+x^2}\ dx$

49. $\displaystyle\int \frac{\arctan(x)}{x^2}\ dx$

50. $\displaystyle\int \frac{\sin(\ln(x))}{x}\ dx$

51. $\displaystyle\int \cos(x)\ln(\sin(x))\ dx$

52. $\displaystyle\int \frac{x\ dx}{\sqrt{x^2+4}}$

53. $\displaystyle\int \frac{dx}{x^2+x+5}$

54. $\displaystyle\int \frac{x\ dx}{x^2+x+5}$

55. $\displaystyle\int \frac{x+3}{(x+1)^5}\ dx$

56. $\displaystyle\int \frac{x^5+x+\sqrt{x}}{x^3}\ dx$

57. $\displaystyle\int (x^2+9)^{10}x\ dx$

58. $\displaystyle\int (x^2+9)^{10}x^3\ dx$

59. $\displaystyle\int \frac{x^4\ dx}{(x+1)^2(x-2)^3}$

In Exercises 60 to 62, (a) decide which positive integers n yield integrals you can evaluate and (b) evaluate them.

60. $\displaystyle\int \sqrt{1+x^n}\ dx$

61. $\int (1+x^2)^{1/n}\ dx$

62. $\int (1+x)^{1/n}\sqrt{1-x}\ dx$

63. Find $\int \frac{dx}{\sqrt{x+2}-\sqrt{x-2}}$.

64. Find $\int \sqrt{1-\cos(x)}\ dx$.

In Exercises 65 to 70, evaluate the integrals.

65. $\int \frac{x\ dx}{(\sqrt{9-x^2}^5}$

66. $\int \frac{dx}{\sqrt{9-x^2}}$

67. $\int \frac{dx}{x\sqrt{x^2+9}}$

68. $\int \frac{x\ dx}{\sqrt{x^2+9}}$

69. $\int \frac{dx}{x+\sqrt{x^2+25}}$

70. $\int (x^3+x^2)\sqrt{x^2-5}\ dx$

71.

(a) Evaluate $\int x^3 e^{x^2}$ using the substitution $u = x^2$ followed by an application of integration by parts.

(b) How does this approach compare with the one used in Example 5?

72. In Example 6 it is found that

$$\int \frac{1-\sin(\theta)}{\theta+\cos(\theta)} d\theta = \ln|\theta+\cos\theta| + C.$$

Check this result by differentiation.

73.

(a) Use integration parts to evaluate $\int \sin^5(2x)\cos(2x)\ dx$.

(b) How does this approach compare with the one used in Example 10?

8.S Chapter Summary

The previous section reviewed the techniques discussed in the chapter. Here we will offer some general comments on finding antiderivatives.

First of all, while the derivative of an elementary function is again elementary, that is not necessarily the case with antiderivatives. Moreover, it isn't easy to predict whether an antiderivative will be elementary. For instance $\ln(x)$ and $\frac{\ln(x)}{x}$ have elementary antiderivatives but $\frac{x}{\ln(x)}$ does not. Also, $x\sin(x)$ does, but $\frac{\sin(x)}{x}$ does not. Remembering that some elementary functions lack elementary antiderivatives can save you lots of time and frustration.

The substitution method is the one that will come in handy most often, to reduce an integral to an easier one or to something listed in an integral table.

When an integrand involves a product or quotient, integration by parts may be of use.

A common partial fraction decomposition is

$$\frac{1}{a^2-x^2} = \frac{1}{2a}\left(\frac{1}{a-x}+\frac{1}{a+x}\right).$$

While it is comforting to know that every rational function has an elementary antiderivative, finding it can be a daunting task except for the simplest denominators. First, factoring the denominator into first and second degree polynomials may be a major hurdle. Second, finding the unknown coefficients in the representation could require lots of computation. In such cases, it may be simpler just to use Simpson's approximation (Section 6.5) — unless one needs to know the antiderivative. In such cases it might be best to take advantage of an automated integrator available through your calculator or computer.

As we will see in Chapter 12, approximating an integrand by a polynomial offers another way to estimate an integral.

Some definite integrals over intervals of the form $[-a, a]$ can be simplified before evaluation by using properties of even and odd functions. If $f(x)$ is an even function, then $\int_{-a}^{a} f(x)\ dx = 2\int_0^a f(x)\ dx$; if f is an odd function, then

$\int_{-a}^{a} f(x)\ dx = 0$. (For instance, $\int_{-1}^{1} xe^{x^2}\ dx = 0$.)

Method	*Description*
Substitution (Section 8.2)	Introduce $u = h(x)$. If $f(x)\ dx = g(u)\ du$, t $\int f(x)\ dx = \int g(u)\ du$.
Substitution in a definite integral (Section 8.2)	If $u = h(x)$ with $f(x)\ dx = g(u)\ du$, t $\int_a^b f(x)\ dx = \int_{h(a)}^{h(b)} g(u)\ du$.
Table of Integrals (Section 8.1)	Obtain and become familiar with a table of in grals. Remember to use substitution to put in grands into the proper form.
Integration by Parts (Section 8.3)	$\int u\ dv = uv - \int v\ du$. Choose u and dv $u\ dv = f(x)\ dx$ and $\int v\ du$ is easier to integr than $\int u\ dv$.
Partial Fractions (applies to any rational function of x) (Section 8.4 (and Section 8.2))	This is an algebraic method in which the tegrand is written as a sum of a polynon (which can be zero)) plus terms of the t $\frac{k_i}{(ax+b)^i}$ and $\frac{r_j x + s_j}{(ax^2+bx+c)^j}$.
Certain Trigonometric Integrands (Section 8.5)	$\int \sin(mx)\cos(nx)\ \ dx$, $\int \sin(mx)\sin(nx)$ $\int \cos(mx)\cos(nx)\ dx$ $\int \sin^2(x)\ dx$, $\int \cos^2(x)$ $\int \tan(x)\ dx$, $\int \tan^2(x)\ dx$ $\int \sec(x)\ dx$,
Rational Functions of x and one of $\sqrt{a^2-x^2}$, $\sqrt{a^2+x^2}$, $\sqrt{x^2-a^2}$ (Section 8.5)	For $\sqrt{a^2-x^2}$, let $x = a\sin(\theta)$. For $\sqrt{a^2+x^2}$, let $x = a\tan(\theta)$. For $\sqrt{x^2-a^2}$, let $x = a\sec(\theta)$.
Rational Functions of x and $\sqrt[n]{ax+b}$ (Section 8.5)	Let $u = \sqrt[n]{ax+b}$.
Rational Functions of $\cos(\theta)$ and $\sin(\theta)$ (Section 8.5)	Let $u = \tan(\theta/2)$.

Table 8.S.1 Summary of integration techniques

Integrand	Method of Integration
$\frac{1}{(ax+b)^n}$	substitute $u = ax + b$
$\frac{1}{ax^2+c}$, a, $c > 0$	substitute $cu^2 = ax^2$: $u = \sqrt{\frac{a}{c}}x$
$\frac{1}{ax^2+bx+c}$, $b^2 - 4ac < 0$	factor out a, complete the square, then substitute
$\frac{x}{ax^2+bx+c}$, $b^2 - 4ac < 0$	first, write x in numerator as $\frac{1}{2a}(2ax+b)-\frac{b}{2a}$, then break into two parts. (That is, get $2ax+b$ into the numerator.)
$\frac{1}{(ax^2+bx+c)^n}$ $b^2 - 4ac < 0$, $n \geq 2$	use a recursive formula from the integral tables
$\frac{x}{(ax^2+bx+c)^n}$ $b^2 - 4ac < 0$, $n \geq 2$	express in terms of the previous type by the method in Example 7.

Table 8.S.2 Antiderivatives of common forms that appear in partial fraction representations.

EXERCISES for 8.S

1.

(a) By an appropriate substitution, transform this definite integral into a simpler definite integral:

$$\int_0^{\pi/2} \sqrt{(1+\cos(\theta))^3}\sin(\theta)\ d\theta.$$

(b) Evaluate the new integral found in (a).

2. Two of these antiderivatives are elementary functions; evaluate them.

(a) $\displaystyle\int \ln(x)\ dx$

(b) $\displaystyle\int \frac{\ln(x)}{x}\ dx$

(c) $\displaystyle\int \frac{dx}{\ln(x)}$

3. Evaluate

(a) $\int_1^2 (1+x^3)^2\ dx$

(b) $\int_1^2 (1+x^3)^2 x^2 \; dx$

4. Use a table of integrals to compute

(a) $\displaystyle\int \frac{e^x \; dx}{5e^{2x}-3}$

(b) $\displaystyle\int \frac{dx}{\sqrt{x^2-3}}$

5. Compute

(a) $\displaystyle\int \frac{dx}{x^3}$

(b) $\displaystyle\int \frac{dx}{\sqrt{x+1}}$

(c) $\displaystyle\int \frac{e^x}{1+5e^x} \; dx$

6. Compute $\displaystyle\int \frac{5x^4-5x^3+10x^2-8x+4}{(x^2-1)(x-1)} \; dx.$

7. Transform the definite integral

$$\int_0^3 \frac{x^3}{\sqrt{x+1}} \; dx$$

into another definite integral in the following ways (and evaluate each transformed integral).

(a) by the substitution $u = x+1$

(b) by the substitution $u = \sqrt{x+1}$.

(c) Which method was easier to apply?

8.

(a) Transform the definite integral

$$\int_{-1}^{4} \frac{x+2}{\sqrt{x+3}}\,dx$$

into an easier definite integral by a substitution.

(b) Evaluate the integral obtained in (a).

9. Compute $\int x^2 \ln(1+x)\,dx$ (a) without an integral table, (b) with an integral table.

10. Verify that the following factorizations into irreducible polynomials are correct.

(a) $x^3 - 1 = (x-1)(x^2+x+1)$

(b) $x^4 - 1 = (x-1)(x+1)(x^2+1)$

(c) $x^3 + 1 = (x+1)(x^2-x+1)$

Express each expression in Exercises 11 to 17 as a sum of partial fractions. (Do not integrate.) Exercise 10 may be helpful.

11. $\dfrac{2x^2+3x+1}{x^3-1}$

12. $\dfrac{x^4+2x^2-2x+2}{x^3-1}$

13. $\dfrac{2x-1}{x^3+1}$

14. $\dfrac{x^4+3x^3-2x^2+3x-1}{x^4-1}$

15. $\dfrac{2x+5}{x^2+3x+2}$

16. $\dfrac{5x^3+11x^2+6x+1}{x^2+x}$

17. $\dfrac{5x^3+6x^2+8x+5}{(x^2+1)(x+1)}$

18. The Fundamental Theorem of Calculus can be used to evaluate one of these definite integrals, but not the other. Evaluate that integral using the FTC.

(a) $\int_0^1 \sqrt[3]{x}\sqrt{x}\, dx$

(b) $\int_0^1 \sqrt[3]{1-x}\sqrt{x}\, dx$

19. Compute $\int \frac{x^3}{(x-1)^2}\, dx$

(a) using partial fractions

(b) using the substitution $u = x - 1$

(c) which method, (a) or (b), is easier in this case?

20.

(a) Compute $\int \frac{x^{2/3}}{x+1}\, dx$.

(b) What does a table of integrals say about the integral in (a)?

21. Compute $\int x\sqrt[3]{x+1}\, dx$ using

(a) the substitution $u = \sqrt[3]{x+1}$

(b) the substitution $u = x + 1$

In Exercises 22 to 25 evaluate the integrals.

22. $\int_0^1 (e^x + 1)^3 e^x\, dx$

23. $\int_0^1 (x^4 + 1)^5 x^3\, dx$

24. $\int_1^e \frac{\sqrt{\ln(x)}}{x}\, dx$

25. $\displaystyle\int_{9}^{\pi/2} \frac{\cos(\theta)}{\sqrt{1+\sin(\theta)}}\, dx$

26.

(a) Without an integral table, evaluate

$$\int \sin^5(\theta)\, d\theta \qquad \text{and} \qquad \int \tan^6(\theta)\, d\theta.$$

(b) Evaluate each integral with an integral table.

(c) Resolve any differences in the appearance of the antiderivatives found in (a) and (b).

27. Two of these three antiderivatives are elementary. Find them, and explain why you know they are elementary (without necessarily evaluating the integral).

(a) $\displaystyle\int \sqrt{1-4\sin^2(\theta)}\, d\theta$

(b) $\displaystyle\int \sqrt{4-4\sin^2(\theta)}\, d\theta$

(c) $\displaystyle\int \sqrt{1+\cos(\theta)}\, d\theta$

28. Find $\displaystyle\int \cot(3\theta)\, d\theta$.

29. Find $\displaystyle\int \csc(5\theta)\, d\theta$.

30. Compute

(a) $\displaystyle\int \sec^5(x)\tan(x)\, dx$

(b) $\displaystyle\int \frac{\sin(x)}{\cos^3(x)}\, dx$

31. Compute $\displaystyle\int \frac{x^3\, dx}{(1+x^2)^4}$ in two different ways:

(a) by the substitution $u = 1 + x^2$,

(b) by the substitution $x = \tan(\theta)$.

32. Find $\int \frac{x\ dx}{\sqrt{9x^4 + 16}}$

(a) without an integral table,

(b) with an integral table.

33. Transform $\int \frac{x^2\ dx}{\sqrt{1+x}}$ by each of the substitutions

(a) $u = \sqrt{1+x}$

(b) $y = 1 + x$

(c) $x = \tan^2(\theta)$

(d) Evaluate the easiest of the above three reformulations.

34. Compute $\int x\sqrt{1+x}\ dx$ in three ways:

(a) $u = \sqrt{1+x}$,

(b) $x = \tan^2(\theta)$,

(c) by parts, with $u = x$ and $dv = \sqrt{1+x}\ dx$.

35. Find $\int x\sqrt{(1-x^2)^5}\ dx$ using the substitutions

(a) $u = x^2$,

(b) $u = 1 - x^2$,

(c) $x = \sin(\theta)$.

In Exercises 36 to 48, evaluate the definite integral appearing in the given exercise.

36. Exercise 21 in Section 7.1.

37. Exercise 22 in Section 7.1.

38. Exercise 23 in Section 7.1.

39. Exercise 24 in Section 7.1.

40. Exercise 25 in Section 7.1.

41. Exercise 26 in Section 7.1.

42. Exercise 27 in Section 7.1.

43. Exercise 28 in Section 7.1.

44. Exercise 30 in Section 7.1.

45. Exercise 1 in Section 7.5.

46. Exercise 2 in Section 7.5.

47. Exercise 3 in Section 7.5.

48. Exercise 4 in Section 7.5.

49. The region $\mathcal{R}$ below the line $y = e$, above $y = e^x$, and to the right of the y-axis is revolved around the y-axis to form a solid $\mathcal{S}$. In Example 1 in Section 7.5 it is shown that the definite integral for the volume of $\mathcal{S}$ using disks is

$$\int_1^e \pi\left(\ln(y)\right)^2 \ dy$$

and the volume of $\mathcal{S}$ using shells is

$$\int_0^1 2\pi x\left(e - e^x\right) \ dx.$$

Evaluate each integral. Which integral is easier to evaluate?

50. The region $\mathcal{R}$ below the line $y = \frac{\pi}{2} - 1$, to the right of the y-axis, and above the curve $y = x - \sin(x)$ is revolved around the y-axis to form a solid $\mathcal{S}$. In Example 2 in Section 7.5 it is shown that the definite integral for the volume of $\mathcal{S}$ using disks cannot be evaluated in terms of elementary functions, and that the volume of $\mathcal{S}$ using shells is

$$\int_0^{\pi/2} 2\pi x\left(\frac{\pi}{2} - 1 - (x - \sin(x))\right) \ dx.$$

Evaluate the value of this integral.

51.

(a) Evaluate $\int \frac{x+1}{x^2} e^{-x} \ dx$.

(b) Evaluate $\int \frac{ax-1}{ax^2} e^{ax} \ dx$, $a \neq 0$

52. In Example 1 in Section 7.6 the total force on a submerged circular tank is found to be

$$\int_{-5}^{5}(0.036)(x+17)\sqrt{100-4x^2}\,dx = 0.036\int_{-5}^{5} x\sqrt{100-4x^2}\,dx + 0.036\int_{-5}^{5} 17\sqrt{100-4x^2}\,dx \text{ pou}$$

At that time, the value of this integral was found using the fact that the first integral has an odd integrand over an interval symmetric about the origin and by relating the second integral to the area of a quarter circle.

(a) Evaluate the first integral using the substitution $u = 100 - 4x^2$.

(b) Evaluate the second integral using the substitution $x^2 = 25\sin^2(\theta)$.

(c) Which approach is easier?

53. Find $\int \frac{dx}{\sin(2x)}$ by first writing $\sin(2x)$ as $2\sin(x)\cos(x)$.

54.

(a) Show that $\int_0^\infty \frac{\sin(kx)}{x}\,dx = \int_0^\infty \frac{\sin(x)}{x}\,dx$, where k is a positive constant.

(b) Show that $\int_0^\infty \frac{\sin(x)\cos(x)}{x}\,dx = \int_0^\infty \frac{\sin(x)}{x}\,dx$.

(c) If k is negative, what is the relation between $\int_0^\infty \frac{\sin kx}{x}\,dx$ and $\int_0^\infty \frac{\sin x}{x}\,dx$?

55. Evaluate $\int_0^\infty e^{-x}\sin^2(x)\,dx$.

56. Evaluate $\int_0^\infty e^{-x}\sin(x)\,dx$. This integral was first encountered in Example 4 on page 663.

In statistics a function $F(x)$ defined on $[0,\infty)$ is called a **probability distribution** if $F(0) = 0$, $\lim_{x\to\infty} F(x) = 1$, and F has a nonnegative derivative f. The function f is called a **probability density**. The integral $\int_0^\infty xf(x)\,dx$ is called the **expected value** or **average value** of x. Exercises 57 and 58 show that if one of the integrals $\int_0^\infty xf(x)\,dx$ and $\int_0^\infty (1-F(x))\,dx$ is convergent, so is the other one and these two integrals are equal.

57. Assume $\int_0^\infty xf(x)\,dx$ is finite.

(a) Show that $\int_k^\infty xf(x)\,dx$ approach zero as k approaches ∞.

(b) Using the fact that $\int_k^\infty xf(x)\ dx \geq \int_k^\infty kf(x)\ dx$, show that $\lim_{k\to\infty} k(1 - F(k)) = 0$.

(c) Show that
$$\int_0^k xf(x)\ dx = k(F(k) - 1) + \int_0^k (1 - F(x))\ dx.$$
(Use integration by parts and $d(F(x) - 1) = f(x)\ dx$.)

(d) From (c) show that
$$\int_0^\infty xf(x)\ dx = \int_0^\infty (1 - F(x))\ dx.$$

58. Assume that $\int_0^\infty (1 - F(x))\ dx$ is finite.

(a) Show that $\int_0^k f(x)\ dx = kF(k) - \int_0^k F(x)\ dx$. (Use integration by parts with $dF(x) = f(x)\ dx$.)

(b) Show $kF(k) - \int_0^k F(x)\ dx \leq \int_0^k (1 - F(x))\ dx$.

(c) Show that $\int_0^\infty xf(x)\ dx$ is finite.

(d) Show that $\int_0^\infty xf(x)\ dx = \int_0^\infty (1 - F(x))\ dx$. (Review Exercise 57.)

Exercises 59 to 62 are related.

59. Show that $\int_1^\infty (\cos(x))/x^2\ dx$ is convergent.

60. Show that $\int_1^\infty (\sin(x))/x\ dx$ is convergent. (Start with integration by parts.)

61. Show that $\int_0^\infty (\sin(x))/x\ dx$ is convergent.

62. Show that $\int_0^\infty \sin(e^x)\ dx$ is convergent.

63. In a statistics text it is asserted that for $\lambda > 0$ and n a positive integer
$$\int_0^\infty 1 - \left(1 - e^{-\lambda t}\right)^n\ dt = \frac{1}{\lambda}\sum_{k=1}^{n}\frac{1}{k}.$$

(a) Check this assertion for $n = 1$.

(b) Check this assertion for $n = 2$.

(c) Show that for all n the integral is convergent.

(For (c), use the Binomial Theorem (see Exercise 32 in Section 5.5).)

64. Let $\int_{-\infty}^{\infty} f(x)\ dx$ be a convergent integral with value A.

(a) Express $\int_{-\infty}^{\infty} f(x+2)\ dx$ in terms of A.

(b) Express $\int_{-\infty}^{\infty} f(2x)\ dx$ in terms of A.

65. Find the error in the following computations: The substitution $x = y^2$, $dx = 2y\ dy$, yields

$$\begin{aligned}\int_0^1 \frac{1}{x}\ dx &= \int_0^1 \frac{2y}{y^2}\ dy = \int_0^1 \frac{2}{y}\ dy \\ &= 2\int_0^1 \frac{1}{y}\ dy = 2\int_0^1 \frac{1}{x}\ dx.\end{aligned}$$

Hence

$$\int_0^1 \frac{1}{x}\ dx = 2\int_0^1 \frac{1}{x}\ dx;$$

from which it follows that $\int_0^1 (1/x)\ dx = 0$.

Laplace Transforms were introduced in Exercises 51 to 55 in Section 8.3. Exercises 66 to 68 develop properties of Laplace Transforms.

66. Let f and its derivative f' both have Laplace transforms. Let P be the Laplace transform of f, and let Q be the Laplace transform of f'. Show that

$$Q(r) = -f(0) + rP(r).$$

67. Assume that $f(t) = 0$ for $t < 0$ and that f has a Laplace transform. Let a be a positive constant. Define $g(t)$ to be $f(t-a)$. Show that the Laplace transform of g is e^{-ar} times the Laplace transform of f. The graph of g is the graph of f shifted to the right by a.

68. Let P be the Laplace transform of f. Let a be a positive constant, and let $g(t) = f(at)$. Let P be the Laplace transform of f, and let Q be the Laplace transform of g. Show that $Q(r) = (1/a)P(r/a)$.

69.

(a) Estimate $\int_0^1 \frac{\sin(x)}{x} dx$ by using the Maclaurin polynomial $P_6(x;0)$ associated with $\sin(x)$ to approximate $\sin(x)$.

(b) Use the Lagrange form of the error to put an upper bound on the error in (a).

70.

(a) Estimate $\int_{-1}^{1} \frac{e^x}{x+2} dx$ by using the Maclaurin polynomial $P_3(x;-2)$ associated with e^x to approximate e^x.

(b) Use the Lagrange form of the error to put an upper bound on the error in (a).

71.

(a) Estimate $\int_{-1}^{1} \frac{e^x}{x-2} dx$ by using the Taylor polynomial $P_3(x;2)$ associated with e^x to approximate e^x.

(b) Use the Lagrange form of the error to put an upper bound on the error in (a).

72. Find $\displaystyle\int \frac{\ln(x^2)}{x^2} dx$.

73. If a is a constant, show that $\int_{-\infty}^{\infty} e^{-(x-a)^2}\, dx = \int_{-\infty}^{\infty} e^{-x^2}\, dx = 2\int_0^{\infty} e^{-x^2}\, dx$.

74. When studying the normal distribution in statistics one will meet an equation that amounts to

$$\frac{\int_{-\infty}^{\infty} x \exp(-(x-\mu)^2)\, dx}{\int_{-\infty}^{\infty} \exp(-(x-\mu)^2)\, dx} = \mu,$$

where μ is a constant. Show that the equation is correct. (Make the substitution $t = x - \mu$.)

75. Show that $\int_1^{\infty} x \exp(-x^2)\, dx$ is less than $\int_0^1 x \exp(-x^2)\, dx$. This implies that the probability of a large disaster, compared to the long tail of the bell curve, is smaller than what must be planned for in spite of the growth of the coefficient x. As a result, economic predictions based on the bell curve may downplay the likelihood of rare events. This bias may have been one of the several factors that combined to produce the credit crisis and recession that began in 2008.

76. For which values of the positive constant k is $\displaystyle\int\limits_e^{\infty} \frac{dx}{x(\ln(x))^k}$ convergent? divergent?

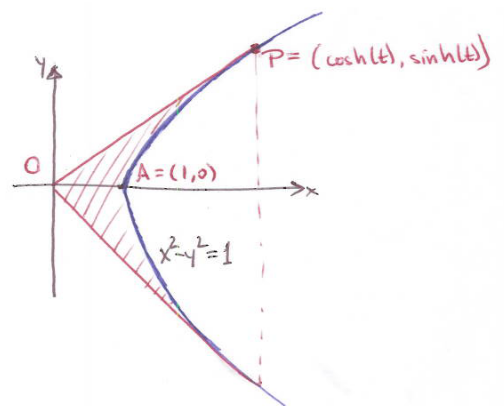


Figure 8.S.1

77. The formula for the area of region OAP in Figure 8.S.1 was found, in Exercise 64 in Section 6.5, to be

$$\frac{1}{2}\cosh(t)\sinh(t) - \int_{1}^{\cosh(t)} \sqrt{x^2-1}dx$$

Use the substitution $x = \cosh(u)$ to evaluate the definite integral. See also Exercises 64 in Section 6.5 and 8 in Section 15.4.

The molecules in a gas move at various speeds. In 1859 James Maxwell developed a formula for the distribution of the speeds of a gas consisting of N molecules. The formula is

$$f(v) = 4\pi N\left(\frac{m}{2\pi kT}\right)^{3/2} v^2 e^{\frac{-1}{2}\frac{mv^2}{kT}}$$

This means that for small values, dv, the number of molecules with speeds between v and $v + dv$ is approximately $f(v)\ dv$. In the formula k is a physical constant, T is the absolute temperature, and m is the mass of a molecule. The only variable is v. Exercises 78 to 80 investigate Maxwell's model.

78. Show that $\int_0^\infty f(v)\ dv = N$.

79. (continuation of Exercise 78) The average speed of the molecules is

$$\frac{\int_0^\infty vf(v)\ dv}{N}.$$

Show that this equals $\sqrt{8kT/\pi m} \approx 1.5958\sqrt{kT/m}$.

80. (continuation of Exercise 79) The "most probable speed" occurs where $f(v)$ has a maximum. Show that this speed is $\sqrt{2kT/m} \approx 1.4142\sqrt{kT/m}$. So the most likely speed is a bit less than the average speed.

81. In the study of heat capacity of a crystal one meets

$$\int_0^b \frac{x^4e^x}{(e^x-1)^2}\ dx.$$

(a) Show that the integral is convergent.

(b) Is $\int_0^b \frac{xe^x}{(e^x-1)^2}\, dx$ convergent?

82. Show that $\int_{-\infty}^{\infty} \frac{dt}{(1+t^2)^{3/2}} = 2$.

83.

(a) Show that $\int_0^{\infty} \frac{x^2}{(x^2+1)^{5/2}}$ is convergent.

(b) Show that the value of this improper integral is 1/3.

84. In the theory of probability one meets the equation

$$\int_0^{\infty} e^{-\lambda x} R(x)\, dx = \frac{1}{\lambda} \int_0^{\infty} e^{-\lambda x} R'(x)\, dx + \frac{1}{\lambda} R(0)$$

Assuming the integrals are convergent, explain how the equation is obtained.

85. The velocity of a particle at time t seconds is $e^{-t}\sin(\pi t)$ meters per second. Find how far it travels in the first second, from time $t = 0$ to $t = 1$,

(a) using the integral table in the front of the book,

(b) using Simpson's method with $n = 4$, expressing your answer to four decimal places.

(Notice that the particle changes direction at $t = 1/2$ second.)

86. Assume that f is continuous on $[0, \infty)$ and has period one, that is, $f(x) = f(x + 1)$ for all x in $[0, \infty]$. Assume also that $\int_0^{\infty} e^{-x} f(x)\, dx$ is convergent. Show that

$$\int_0^{\infty} e^{-x} f(x)\, dx = \frac{e}{e-1} \int_0^{1} e^{-x} f(x)\, dx.$$

87. Assume that f is continuous on $[0, \infty)$ and has period $p > 0$. Let s be a positive number and assume $\int_0^{\infty} e^{-st} f(t)\, dt$ converges. Show that this improper integral equals

$$\frac{1}{1-e^{-sp}} \int_0^{p} e^{-st} f(t)\, dt.$$

88. The integral $\int_0^\infty x^{2n}e^{-kx^2}\ dx$ appears in the kinetic theory of gases. In Chapter 16, we will show that $\int_0^\infty e^{-x^2}\ dx = \sqrt{\pi}/2$. With the aid of this information, evaluate

(a) $\int_0^\infty e^{-kx^2}\ dx$,

(b) $\int_0^\infty x^2e^{-kx^2}\ dx$.

89. (continuation of Exercise 4 in the Chapter 7 Summary) This exercise presents an alternate approach to evaluating the integral in Exercise 4 in the Chapter 7 Summary. Express the integral as the Laplace transform of an appropriate function. Then, use a table of Laplace transforms to find the value of the integral.

90. James Maxwell's "On the Geometric Mean Distance of Two Figures in a Plane," written in 1872, begins "There are several problems of great practical importance in electro-magnetic measurements, in which the value of the quantity has to be calculated by taking the sum of the logarithms of the distances of a system of parallel wires from a given point."
This leads him to several problems, of which this is the first.

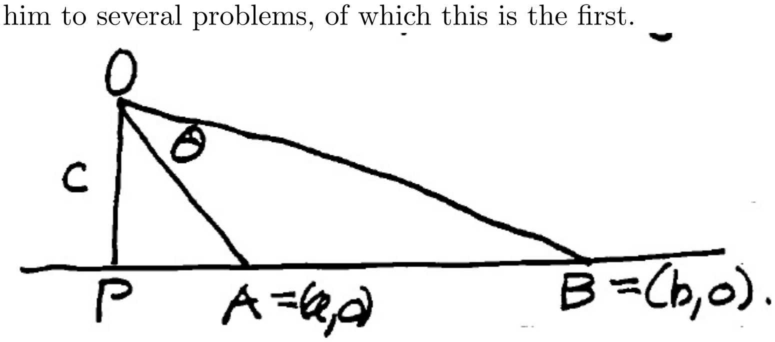


Figure 8.S.2
A point $\mathcal{O}$ is a distance c from the line that contains the line segment AB. Let P be the point on that line nearest $\mathcal{O}$, as in Figure 8.S.2. Introduce a coordinate system in which P is the origin, AB lies on the x-axis, and OP lies on the y-axis.
Let $f(x)$ be the distance from $\mathcal{O}$ to $(x,0)$.
Show that the average value of $\ln(f(x))$ for x in $[a,b]$ is

$$\frac{b\ln(b) - a\ln(a) - (b-a) + c\theta}{b-a},$$

where θ is the angle AOB in radians. This theme continues in Exercises 58 in Section 12.5, and 5 to 7 in the Summary for Chapter 12.

91. Evaluate $\int \frac{\cos(\theta)}{(b^2 + c^2\cos^2(\theta))^{1/2}}\,d\theta$. This integral appears in Exercise 19 in the Summary of Chapter 18. (Let $u = c\cos(\theta)$.)

92. Show that $\int \sqrt{x}e^x\,dx$ is not elementary. (Use the fact that $\int e^{x^2}\,dx$ is not elementary.)

93. We have seen that $\int e^{x^2}\,dx$ is not elementary.

(a) Show that for positive odd integers n, $\int x^n e^{x^2}\,dx$ is elementary.

(b) Show that for positive even integers n, $\int x^n e^{x^2}\,dx$ is not elementary.

94. We have seen that $\int e^{x^2}\,dx$ and $\int \frac{e^x}{x}\,dx$ are not elementary.

(a) Show that $\int \frac{e^{x^2}}{x}\,dx$ is not elementary.

(b) Show that $\int \frac{e^{x^2}}{x^2}\,dx$ is not elementary.

(c) Show that for any positive integer n, $\int \frac{e^{x^2}}{x^n}\,dx$ is not elementary.

95. We have seen that $\int \frac{e^x}{x}\,dx$ is not elementary.

(a) Show that for positive integers n, $\int x^n e^x\,dx$ is elementary.

(b) Show that for positive integers n, $\int \frac{e^x}{x^n}\,dx$ is not elementary

96.

(a) Show that $\displaystyle\int x^2 e^{x^2}\,dx$ is not elementary.

(b) Show that $\displaystyle\int x^4 e^{x^2}\,dx$ is not elementary.

(c) Find non-zero values for a and b such that $\displaystyle\int (ax^4+bx^2)e^{x^2}\,dx$ is an elementary function.

97. Show that $\int x^n e^{x^2}$ is elementary only when n is an odd positive integer.

98. Let n be an integer. Show that $\int x^n e^x$ is elementary only when n is not negative.

99.

Sam: I understand what a definite integral is — the limit of certain sums. I accept on faith that for a continuous function the limit exists. I agree that it is a handy idea, with many uses, but I don't see why I have to learn all those ways to compute it: antiderivatives, trapezoids, Simpson's method. My trusty calculator evaluates integrals to eight decimal places and a computer algebra system can often give me the exact expression.

Jane: What's your point?

Sam: I would make this text much shorter by omitting this chapter. This would allow us more time to spend on the stuff at the end.

Does Sam have a valid argument, for a change?

Exercises 100 to 105 all relate to the famous **bell curve** that arises in statistics.

100. Use the fact that $\int_{-\infty}^{\infty} e^{-x^2}\ dx = \sqrt{\pi}$ (see Exercise 32 in Section 17.3) to show that

$$\int\limits_{-\infty}^{\infty} x^2 e^{-x^2}\ dx = \frac{1}{2}\sqrt{\pi}.$$

101. Let σ (lower case Greek sigma corresponds to our letter s) be a positive constant. The famous **bell curve** is the graph of the function

$$f(x) = \frac{\exp\left(\frac{-x^2}{2\sigma^2}\right)}{\sigma\sqrt{2\pi}}.$$

Show that $\int_{-\infty}^{\infty} f(x)\ dx = 1$.

102. Show that f has inflection points at points where $x = \sigma$ and at $x = -\sigma$.

103. Show that $\int_{-\infty}^{\infty} x^2 f(x)\ dx = \sigma^2$. Thus σ^2 measures the discrepancy from 0. It is called the **variance**.

104. The mean value of x is defined as $\int_{-\infty}^{\infty} x f(x)\ dx$. Show that it is 0. (Avoid labor.)

105. Assume that $\int_{-\infty}^{\infty} g(x)\ = 1$ and $\int_{-\infty}^{\infty} xg(x)\ dx = k$. Let $h(x) = g(x-k)$. Show that $\int_{-\infty}^{\infty} h(x)\ dx = 1$, $\int_{-\infty}^{\infty} xh(x)\ dx = k$, and $\int_{-\infty}^{\infty}(x-k)^2 h(x)\ dx = \int_{-\infty}^{\infty} x^2 g(x)\ dx$.

106. If $f(x)$ and $g(x)$ have elementary antiderivatives, which of the following necessarily do also? (a) $f(x)g(x)$, (b) $f(g(x))$, and (c) $f(x) + g(x)$. Justify each answer.

107.

(a) Show that $e^{x^{1/2}}$ has an elementary antiderivative.

(b) Show that $e^{x^{1/3}}$ has an elementary antiderivative.

(c) Show that for every positive integer n, $e^{x^{1/n}}$ has an elementary antiderivative.

108. When a curve situated above the x-axis is revolved around the x-axis, the area of the resulting surface of revolution is 31. When the curve is revolved around the line $y = -2$, the surface area of this solid is 75. How long is the curve?

109. In a letter dated May 24, 1872 Maxwell wrote: "It is strange ... that W. Weber could not correctly integrate

$$\int_0^\pi \cos(\theta)\sin(\phi)\, d\phi \qquad \text{where} \qquad \tan(\theta) = \frac{A\sin(\phi)}{B + A\cos(\phi)},$$

but that everyone should have copied such a wild result as

$$\frac{B}{\sqrt{A^2+B^2}} \cdot \frac{B^4 + \frac{7}{6}A^2B^2 + \frac{2}{3}A^2}{B^4 + A^2B^2 + A^4}.$$

Of course there are two forms of the result according as A or B is greater." Assuming that A and B are positive, find the correct value of the integral. (Begin by expressing $\cos(\theta)$ in terms of the constants ϕ, A, and B.)

110. The following calculation appears in *Electromagnetic Fields*, 2nd ed., Roald K. Wangsness, Wiley, 1986. (See also Exercise 4 in the Chapter 12 Summary.)

(a) The substitution $\frac{\pi}{2}\cos(\theta) = \frac{1}{2}(\pi - v)$, turns $\int_0^\pi \frac{\cos^2\left(\frac{\pi}{2}\cos(\theta)\right)}{\sin(\theta)}\, d\theta$ into

$$\frac{1}{4}\left(\int_0^{2\pi} \frac{1-\cos(v)}{v}\, dv + \int_0^{2\pi} \frac{1-\cos(v)}{2\pi - v}\, dv\right).$$

(b) Introducing $w = 2\pi - v$ shows that the two integrals with respect to v are equal.

(c) So we must find $\frac{1}{2}\int_0^{2\pi} \frac{1-\cos(v)}{v}\, dv$. The integrand does not have an elementary antiderivative. However, its value (2.438) is listed in integral tables. Reference: *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed., Dover, 1964 (online version available at `http://www.math.sfu.ca/~cbm/aands/`.)

111. Which has the larger absolute value $\int_0^{\sqrt{\pi}} \sin\left(x^2\right)\ dx$ or $\int_{\sqrt{\pi}}^{\sqrt{2\pi}} \sin\left(x^2\right)\ dx$? (Substitute $u = x^2$ and, in the second integral obtained, $u = v + \pi$.)

In Exercises 112 and 113 a, b, c, m, and p are constants. In each case verify that the derivative of the first function is the second function.

112. $\frac{e^{ax}(a\sin(px) - p\cos(px))}{a^2+p^2}$; $e^{ax}\sin(px)$.

113. $\sec(x) + \ln\left(\tan\left(\frac{x}{2}\right)\right)$; $\frac{1}{\sin(x)\cos^2(x)}$.

Calculus is Everywhere # 11
An Improper Integral in Economics

Both business and government frequently face the question, "How much money do I need today to have one dollar t years in the future?"

Implicit in this question are such considerations as the present value of a business being dependent on its future profit and the cost of a toll road being weighed against its future revenue. We determine the present value of a business which depends on the future rate of profit.

To begin the analysis, assume that the annual interest rate r remains constant and that 1 dollar deposited today is worth e^{rt} dollars t years from now. This assumption corresponds to continuously compounded interest or to natural growth. Thus A dollars today will be worth Ae^{rt} dollars t years from now. What is the present value of the promise of 1 dollar t years from now? In other words, what amount A invested today will be worth 1 dollar t years from now? To find out, solve the equation $Ae^{rt} = 1$ for A. The solution is

t need not be an integer

The present value of \$1 t years from now is \$ e^{-rt}

$$A = e^{-rt}. \tag{C.11.1}$$

Now consider the present value of the future profit of a business (or future revenue of a toll road). Assume that the profit flow t years from now is at the rate $f(t)$. This rate may vary within the year; consider f to be a continuous function of time. The profit in the small interval of time dt, from time t to time $t + dt$, would be approximately $f(t)dt$. The total future profit, $F(T)$, from now, when $t = 0$, to some time T in the future is therefore

$$F(T) = \int_0^T f(t)dt. \tag{C.11.2}$$

But the **present value** of the future profit is *not* given by (C.11.2). It is necessary to consider the present value of the profit earned in a typical short interval of time from t to $t + dt$. According to (C.11.1), its present value is approximately

$$e^{-rt} f(t)dt.$$

Hence the present value of future profit from $t = 0$ to $t = T$ is given by

$$\int_0^T e^{-rt} f(t)dt. \tag{C.11.3}$$

The present value of all future profit is, therefore, the improper integral $\int_0^\infty e^{-rt} f(t)dt$.

To see what influence the interest rate r has, denote by $P(r)$ the present value of all future revenue when the interest rate is r; that is,

$$P(r) = \int_0^\infty e^{-rt} f(t)dt. \tag{C.11.4}$$

If the interest rate r is raised, then according to (C.11.4) the present value of a business declines. An investor choosing between investing in a business or placing the money in a bank account may find the bank account more attractive when r is raised.

A proponent of a project, such as a toll road, will argue that the interest rate r will be low in the future. An opponent will predict that it will be high. Of course, neither knows what the inscrutable future will do to the interest rate. Even so, the prediction is important in a cost-benefit analysis.

Equation (C.11.4) assigns to a profit function f (which is a function of time t) a present-value function P, which is a function of r, the interest rate. In the theory of differential equations, P is called the **Laplace transform** of f. This transform can replace a differential equation by a simpler equation that looks quite different.

The Laplace transform was first encountered in Exercises 51 to 55 in Section 8.3 and reappeared in Exercises 66 to 68 in Section 8.6.

EXERCISES

In Exercises 1 to 8 $f(t)$ is defined on $[0, \infty)$ and is continuous. Assume that for $r > 0$, $\int_0^\infty e^{-rt} f(t)dt$ converges and that $e^{-rt} f(t) \to 0$ as $t \to \infty$. Let $P(r) = \int_0^\infty e^{-rt} f(t)dt$. Find $P(r)$, the Laplace transform of $f(t)$, in Exercises 1 to 5.

1. $f(t) = t$

2. $f(t) = e^t$, assume $r > 1$

3. $f(t) = t^2$

4. $f(t) = \sin(t)$

5. $f(t) = \cos(t)$

6. Let P be the Laplace transform of f, and let Q be the Laplace transform of f'. Show that $Q(r) = -f(0) + rP(r)$.

7. Let P be the Laplace transform of f, a a positive constant, and $g(t) = f(at)$. Let Q be the Laplace transform of g. Show that $Q(t) = \frac{1}{a} P\left(\frac{r}{a}\right)$.

8. Which is worth more today, \$100, 8 years from now or \$80, five years from now? (a) Assume $r = 4\%$. (b) Assume $r = 8\%$. (c) For which interest rate are the two equal?

Chapter 9

Polar Coordinates and Plane Curves

This chapter presents further applications of the derivative and integral. Section 9.1 describes polar coordinates. Section 9.2 shows how to compute the area of a region described in polar coordinates. Section 9.3 introduces a method of describing a curve which is especially useful in the study of motion.

The speed of an object moving along a curved path is developed in Section 9.4, where we show how to express the length of a curve as a definite integral. The area of a surface of revolution is expressed as a definite integral in Section 9.5.

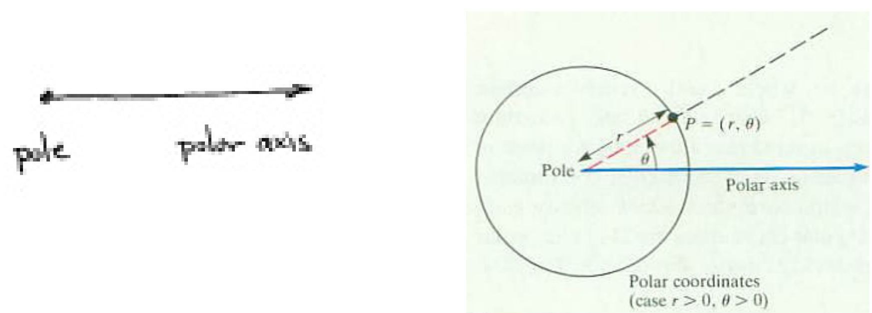
Section 9.6 shows how the derivative and second derivative determine the curvature of a curve.

9.1 Polar Coordinates

Rectangular coordinates provide one way to describe points in the plane by pairs of numbers. This section describes another coordinate system, called polar coordinates.

Polar Coordinates

The rectangular coordinates x and y describe a point P in the plane as the intersection of two perpendicular lines. Polar coordinates describe P as the intersection of a circle and a ray from its center.



(a) (b)

Figure 9.1.1

When we say "The storm is 10 miles northeast," we are using polar coordinates: $r = 10$ and $\theta = \pi/4$.

Select a point in the plane and a ray emanating from it. The point is called the **pole**, and the ray the **polar axis**. (See Figure 9.1.1(a).) Measure positive angles θ counterclockwise from the polar axis and negative angles clockwise. To plot the point P that corresponds to the pair of numbers r and θ:

> If r is positive, P is the intersection of the circle of radius r whose center is at the pole and the ray of angle θ from the pole. (See Figure 9.1.1(b).)
>
> If r is 0, P is the pole, no matter what θ is.
>
> If r is negative, P is at a distance $|r|$ from the pole on the ray directly opposite the ray of angle θ, that is, on the ray of angle $\theta + \pi$.

The pair r and θ are called **polar coordinates** of P. The point (r, θ) is on the circle of radius $|r|$ whose center is the pole. The point $(-r, \theta + \pi)$ is the same as the point (r, θ). Changing the angle by 2π does not change the point; that is, $(r, \theta) = (r, \theta + 2\pi) = (r, \theta + 4\pi) = \cdots = (r, \theta + 2k\pi)$ for any integer k.

EXAMPLE 1 Plot the points $(3, \pi/4)$, $(2, -\pi/6)$, $(-3, \pi/3)$ in polar coordinates. See Figure 9.1.2.
SOLUTION

- To plot $(3, \pi/4)$, go out a distance 3 on the ray of angle $\pi/4$, shown in Figure 9.1.2.
- To plot $(2, -\pi/6)$, go out a distance 2 on the ray of angle $-\pi/6$.
- To plot $(-3, \pi/3)$, draw the ray of angle $\pi/3$, and then go a distance 3 in the direction opposite from the pole.

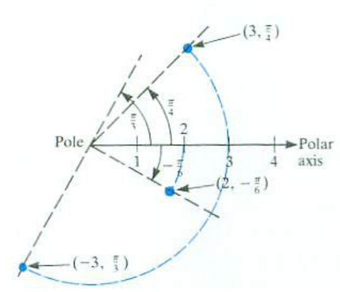


Figure 9.1.2

◇

It is customary to have the polar axis coincide with the positive x-axis as in Figure 9.1.3. The diagram shows the relation between the rectangular coordinates (x, y) and the polar coordinates of P:

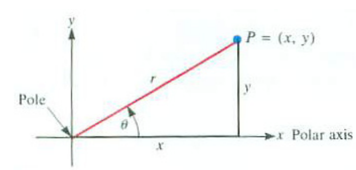


Figure 9.1.3 The relation between polar and rectangular coordinates.

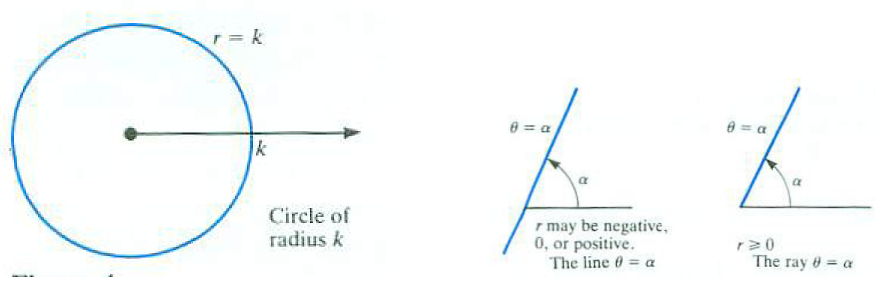
$$x = r\cos(\theta) \qquad y = r\sin(\theta)$$
$$r^2 = x^2 + y^2 \qquad \tan(\theta) = \frac{y}{x}$$

They hold even if r is negative. If r is positive, then $r = \sqrt{x^2 + y^2}$. If $-\pi/2 < \theta < \pi/2$, then $\theta = \arctan(y/x)$.

Graphing $r = f(\theta)$

Just as we may graph the set of points (x, y), where x and y satisfy an equation, we may graph the set of points (r, θ), where r and θ satisfy an equation. Although a point in the plane is specified by a unique ordered pair (x, y) in rectangular coordinates, there are many ordered pairs (r, θ) in polar coordinates that specify each point. For instance, the point whose rectangular coordinates are $(1, 1)$ has polar coordinates $(\sqrt{2}, \pi/4)$, $(\sqrt{2}, \pi/4 + 2\pi)$, $(\sqrt{2}, \pi/4 + 4\pi)$, or $(-\sqrt{2}, \pi/4 + \pi)$ and so on.

The simplest equation in polar coordinates is $r = k$, where k is a positive constant. Its graph is the circle of radius k, centered at the pole. (See Figure 9.1.4(a).) The graph of $\theta = \alpha$, where α is a constant, is the line through the pole with inclination α. If we restrict r to be nonnegative, then $\theta = \alpha$ describes the ray (half-line) through the pole with angle α. (See Figure 9.1.4(b).)



(a) (b)

Figure 9.1.4

EXAMPLE 2 Graph $r = 1 + \cos\theta$. Since $\cos(\theta)$ has period 2π, we consider only θ in $[0, 2\pi]$.

SOLUTION Tabulate some values:

θ	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$	π	$\frac{5\pi}{4}$	$\frac{3\pi}{2}$	$\frac{7\pi}{4}$	2π
r	2	$1 + \frac{\sqrt{2}}{2} \approx 1.7$	1	$1 - \frac{\sqrt{2}}{2} \approx 0.3$	0	$1 - \frac{\sqrt{2}}{2} \approx 0.3$	1	$1 + \frac{\sqrt{2}}{2} \approx 1.7$	2

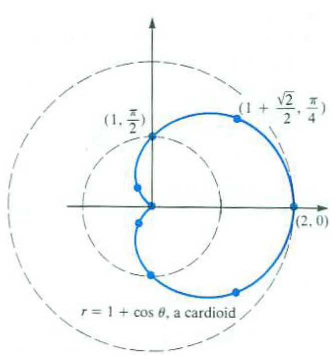


Figure 9.1.5 A cardioid is not shaped like a real heart, only like the conventional image of a heart.

As θ goes from 0 to π, r decreases, and as θ goes from π to 2π, r increases. The point with $\theta = 0$ is the same as the one with $\theta = 2\pi$. The graph begins to repeat itself. This heart-shaped curve, shown in Figure 9.1.5, is called a **cardioid**. ◇

Spirals are quite easy to describe in polar coordinates. One is illustrated by the graph of $r = 2\theta$ in the next example.

EXAMPLE 3 Graph $r = 2\theta$ for $\theta \geq 0$.

SOLUTION Make a table:

θ	0	$\frac{\pi}{2}$	π	$\frac{3\pi}{2}$	2π	$\frac{5\pi}{2}$	$\cdots$
r	0	π	2π	3π	4π	5π	$\cdots$

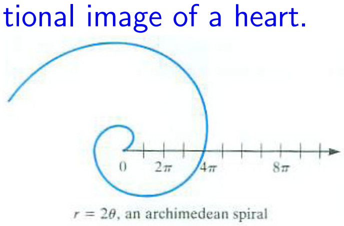


Figure 9.1.6

Increasing θ by 2π does not produce the same value of r. As θ increases, r increases. The graph for $\theta \geq 0$ is a sprial, going infinitely often around the pole, as indicated in Figure 9.1.6. ◇

If a is a nonzero constant, the graph of $r = a\theta$ is called an **Archimedean spiral** because Archimedes was the first person to study it, finding the area within it up to any angle and also its tangent lines. The spiral with $a = 2$ is sketched in Example 3.

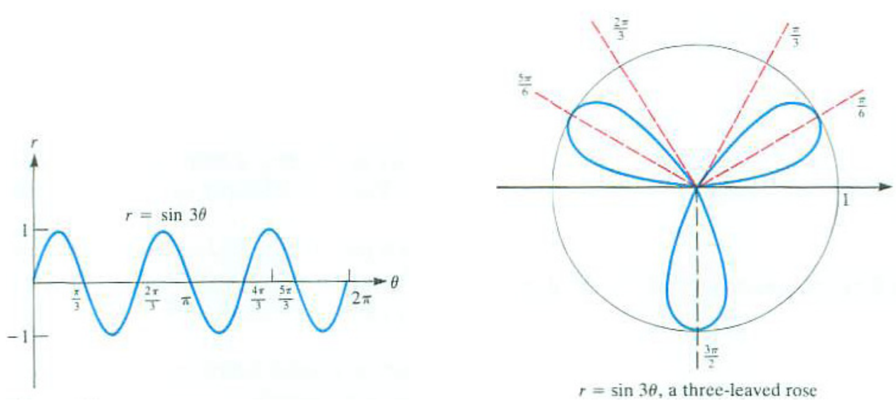
Polar coordinates are also convenient for describing loops arranged like the petals of a flower, as Example 4 shows.

EXAMPLE 4 Graph $r = \sin(3\theta)$.
SOLUTION The values of $\sin(3\theta)$ range from -1 to 1. For instance, when $3\theta = \pi/2$, $\sin(3\theta) = \sin(\pi/2) = 1$. That tells us that when $\theta = \pi/6$, $r = \sin(3\theta) = \sin(3(\pi/6)) = \sin(\pi/2) = 1$. This suggests that we calculate r at integer multiples of $\pi/6$, as in Table 9.1.1. The variation of r as a function of

θ	0	$\frac{\pi}{18}$	$\frac{\pi}{12}$	$\frac{\pi}{9}$	$\frac{\pi}{6}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{5\pi}{6}$	π	$\frac{3\pi}{2}$	2π
3θ	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	π	$\frac{3\pi}{2}$	2π	$\frac{5\pi}{2}$	3π	$\frac{9\pi}{2}$	6π
$r = \sin(3\theta)$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1	0	-1	0	1		1	0

Table 9.1.1

θ is shown in Figure 9.1.7(a). Because $\sin(\theta)$ has period 2π, $\sin(3\theta)$ has period $2\pi/3$.



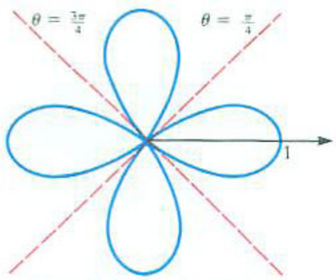
(a) (b)

Figure 9.1.7

As θ increases from 0 to $\pi/3$, 3θ increases from 0 to π. Thus r, which is $\sin(3\theta)$, starts at 0 (for $\theta = 0$) increases to 1 (for $\theta = \pi/6$) and then decreases to 0 (for $\theta = \pi/3$). This gives one of the three loops that make up the graph of $r = \sin(3\theta)$. For θ in $[\pi/3, 2\pi/3]$, $r = \sin(3\theta)$ is negative (or 0), which is the lower loop in Figure 9.1.7(b). For θ in $[2\pi/3, \pi]$, r is again positive, and we obtain the upper left loop. Further choices of θ repeat these three loops. $\diamond$

The graph of $r = \sin(n\theta)$ or $r = \cos(n\theta)$is often described as a rose. It has

n loops when n is an odd integer and $2n$ loops when n is an even integer. The next example illustrates the graph when n is even.



$r = \cos 2\theta$, a four-leaved rose

Figure 9.1.8

EXAMPLE 5 Graph the four-leaved rose, $r = \cos(2\theta)$.
SOLUTION To isolate one loop, find the two smallest nonnegative values of θ for which $\cos(2\theta) = 0$. They are the θ that satisfy $2\theta = \pi/2$ and $2\theta = 3\pi/2$; so $\theta = \pi/4$ and $\theta = 3\pi/4$. One leaf is described by letting θ go from $\pi/4$ to $3\pi/4$. For θ in $[\pi/4, 3\pi/4]$, 2θ is in $[\pi/2, 3\pi/2]$. Since 2θ is then a second- or third-quadrant angle, $r = \cos(2\theta)$ is negative or 0. In particular, when $\theta = \pi/2$, $\cos(2\theta)$ reaches its smallest value, -1. This loop is the bottom one in Figure 9.1.8. The other loops are obtained similarly. We could also sketch the graph by making a table of values. ◇

EXAMPLE 6 Transform the equation $y = 2$, which describes a horizontal straight line, into polar coordinates.
SOLUTION Since $y = r\sin\theta$, $r\sin\theta = 2$, or

$$r = \frac{2}{\sin(\theta)} = 2\csc(\theta).$$

This is more complicated than the rectangular version of this equation, but is still sometimes useful. ◇

EXAMPLE 7 Transform the equation $r = 2\cos(\theta)$ into rectangular coordinates and graph it.
SOLUTION Since $r^2 = x^2 + y^2$ and $r\cos\theta = x$, multiply the equation $r = 2\cos\theta$ by r, obtaining

$$r^2 = 2r\cos(\theta).$$

Hence

$$x^2 + y^2 = 2x$$

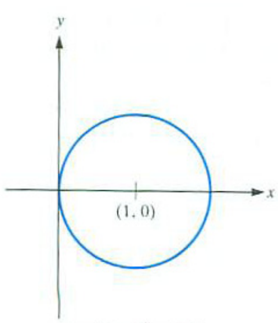
or

$$x^2 - 2x + y^2 = 0$$

and complete the square, obtaining

$$(x-1)^2 + y^2 = 1.$$

The graph is a circle of radius 1 with center at $(1, 0)$ in rectangular coordinates. It is graphed in Figure 9.1.9. ◇



$r = 2\cos\theta$, a circle

Figure 9.1.9

The step in Example 7 where we multiply by r deserves some attention. If $r = 2\cos(\theta)$, then $r^2 = 2r\cos(\theta)$. However, if $r^2 = 2r\cos(\theta)$, it does not follow that $r = 2\cos(\theta)$. We can cancel the r only when r is not 0. If $r = 0$, it is true that $r^2 = 2r\cos(\theta)$, but it not necessarily true that $r = 2\cos(\theta)$. Since $r = 0$ satisfies the equation $r^2 = 2r\cos\theta$, the pole is on the curve $r^2 = 2r\cos\theta$. Luckily, it is also on the original curve $r = 2\cos(\theta)$, since $\theta = \pi/2$ makes $r = 0$. Hence the graphs of $r^2 = 2r\cos(\theta)$ and $r = 2\cos(\theta)$ are the same.

However, as you may check, the graphs of $r = 2 + \cos(\theta)$ and $r^2 = r(2 + \cos(\theta))$ are not the same. The origin lies on the second curve, but not on the first.

The Intersection of Two Curves

Finding the intersection of two curves in polar coordinates is complicated because a point has many descriptions in polar coordinates.

EXAMPLE 8 Find the intersection of the curve $r = 1 - \cos(\theta)$ and the circle $r = \cos(\theta)$.

SOLUTION Graph the curves. The curve $r = \cos(\theta)$ is a circle half the size of the one in Example 7. Both curves are shown in Figure 9.1.10. The curve $r = 1 - \cos(\theta)$ is a cardioid, being congruent to the graph of $r = 1 + \cos(\theta)$. It appears that there are three points of intersection.

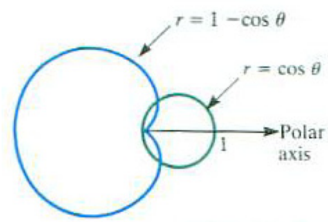


Figure 9.1.10

A point of intersection is produced when one value of θ yields the same value of r in both equations, that is, when

$$1 - \cos(\theta) = \cos(\theta).$$

Hence $\cos(\theta) = \frac{1}{2}$. Thus $\theta = \pi/3$ or $\theta = -\pi/3$ or any angle differing from these by $2n\pi$, n an integer. This gives two of the three points, but it fails to give the origin. Why?

How does the origin get to be on the circle $r = \cos(\theta)$? Because when $\theta = \pi/2$, $r = 0$. How does it get to be on the cardioid $r = 1 - \cos(\theta)$? Because when $\theta = 0$, $r = 0$. The origin lies on both curves, but we do not learn this by simply equating $1 - \cos(\theta)$ and $\cos(\theta)$. ◇

When looking for intersections of two curves, $r = f(\theta)$ and $r = g(\theta)$ in polar coordinates, examine the origin separately. The curves may also intersect at other points not obtainable by setting $f(\theta) = g(\theta)$. This possibility is due to the fact the point (r, θ) is the same as the points $(r, \theta + 2n\pi)$ and $(-r, \theta + (2n + 1)\pi)$ for any integer n. The safest procedure is to graph the two curves first, identify the intersections in the graph, and then see why the curves intersect there.

Summary

We introduced polar coordinates and showed how to graph curves given with equation $r = f(\theta)$. Some common polar curves are listed below.

Equation	Curve	
$r = a$, $a > 0$	circle of radius a, center at pole	
$r = 1 + \cos(\theta)$	cardioid	
$r = a\theta$, $a > 0$	Archimedean spiral (traced clockwise)	
$r = \sin(n\theta)$, n odd	n-leafed rose (one loop symmetric about $\theta = \pi/n$)	
$r = \sin(n\theta)$, n even	$2n$-leafed rose	5
$r = \cos(n\theta)$, n odd	n-leafed rose (one loop symmetric about $\theta = 0$)	
$r = \cos(n\theta)$, n even	$2n$-leafed rose	
$r = a\csc(\theta)$	the line $y = a$	
$r = a\sec(\theta)$	the line $x = a$	
$r = a\cos(\theta)$, $a > 0$	circle of radius $a/2$ through pole and $(a, 0)$	
$r = a\sin(\theta)$, $a > 0$	circle of radius $a/2$ through pole and $(a, \pi/2)$	

Table 9.1.2

To find the intersection of two curves in polar coordinates, first graph them.

EXERCISES for Section 9.1

1. Plot the points whose polar coordinates are

(a) $(1, \pi/6)$

(b) $(2, \pi/3)$

(c) $(2, -\pi/3)$

(d) $(-2, \pi/3)$

(e) $(2, 7\pi/3)$

(f) $(0, \pi/4)$

2. Find the rectangular coordinates of the points in Exercise 1.

3. Give at least three pairs of polar coordinates (r, θ) for the point whose polar coordinates are $(3, \pi/4)$,

(a) with $r > 0$

(b) with $r < 0$

4. Find polar coordinates (r, θ) with $0 \leq \theta < 2\pi$ and r positive, for the points whose rectangular coordinates are

(a) $(\sqrt{2}, \sqrt{2})$

(b) $(-1, \sqrt{3})$

(c) $(-5, 0)$

(d) $(-\sqrt{2}, -\sqrt{2})$

(e) $(0, -3)$

(f) $(1, 1)$

In Exercises 5 to 8 transform the equation into one in rectangular coordinates.

5. $r = \sin(\theta)$

6. $r = \csc(\theta)$

7. $r = 4\cos(\theta) + 5\sin(\theta)$

8. $r = 3/(4\cos(\theta) + 5\sin(\theta))$

In Exercises 9 to 12 transform the equation into one in polar coordinates.

9. $x = -2$

10. $y = x^2$

11. $xy = 1$

12. $x^2 + y^2 = 4x$

In Exercises 13 to 22 graph the given equations.

13. $r = 1 + \sin\theta$

14. $r = 3 + 2\cos(\theta)$

15. $r = e^{-\theta/\pi}$

16. $r = 4^{\theta/\pi}, \theta > 0$

17. $r = \cos(3\theta)$

18. $r = \sin(2\theta)$

19. $r = 2$

20. $r = 3$

21. $r = 3\sin(\theta)$

22. $r = -2\cos(\theta)$

23. Suppose $r = 1/\theta$ for $\theta > 0$.

(a) What happens to the y-coordinate of (r, θ) as $\theta \to \infty$?

(b) What happens to the x-coordinate of (r, θ) as $\theta \to \infty$?

(c) Sketch the curve.

24. Suppose $r = 1/\sqrt{\theta}$ for $\theta > 0$.

(a) What happens to the y coordinate of (r, θ) as $\theta \to \infty$?

(b) What happens to the x coordinate of (r, θ) as $\theta \to \infty$?

(c) Sketch the curve.

In Exercises 25 to 30, find the intersections of the curves after drawing them.

25. $r = 1 + \cos(\theta)$ and $r = \cos(\theta) - 1$

26. $r = \sin(2\theta)$ and $r = 1$

27. $r = \sin(3\theta)$ and $r = \cos(3\theta)$

28. $r = 2\sin(2\theta)$ and $r = 1$

29. $r = \sin(\theta)$ and $r = \cos(2\theta)$

30. $r = \cos(\theta)$ and $r = \cos(2\theta)$

A curve $r = 1 + a\cos(\theta)$ (or $r = 1 + a\sin(\theta)$) is called a **limaçon**. Its shape depends on the choice of a. For $a = 1$ we have the cardioid of Example 2. Exercises 31 to 33 concern other choices of a.

31. Graph $r = 1 + 2\cos(\theta)$. (If $|a| > 1$, then the graph of $r = 1 + a\cos\theta$ crosses itself and forms two loops.)

32. Graph $r = 1 + \frac{1}{2}\cos(\theta)$.

33. Let $r = 1 + a\cos(\theta)$, where $0 \le a \le 1$.

(a) Relative to the same polar axis, graph the curves corresponding to $a = 0$, $1/4$, $1/2$, $3/4$, 1

(b) For $a = 1/4$ the graph in (a) is convex, but not for $a = 1$. Show that for $1/2 < a \le 1$ the curve is not convex. (Convex is defined in Section 2.5. Find the points on the curve farthest to the left and compare them to the point on the curve corresponding to $\theta = \pi$.)

34.

(a) Graph $r = 3 + \cos(\theta)$

(b) Find the point on the graph in (a) that has the maximum y coordinate.

35. Find the y coordinate of the highest point on the right-hand leaf of the four-leaved rose $r = \cos(2\theta)$.

36. Graph $r^2 = \cos(2\theta)$. If $\cos(2\theta)$ is negative, r is not defined, and if $\cos(2\theta)$ is positive, there are two values of r, $\sqrt{\cos(2\theta)}$ and $-\sqrt{\cos(2\theta)}$. The curve is called a **lemniscate**.

The graph of $r = 1/(1 + e\cos(\theta))$ is a parabola if $e = 1$, an ellipse if $0 \le e < 1$, and a hyperbola if $e > 1$. (Here e denotes eccentricity, not Euler's number.) Exercises 37 to 38 concern such graphs.

37.

(a) Graph $r = \frac{1}{1+\cos(\theta)}$.

(b) Find an equation in rectangular coordinates for the curve in (a).

38.

(a) Graph $r = \frac{1}{1-(1/2)\cos(\theta)}$.

(b) Find an equation in rectangular coordinates for the curve in (a).

39. Where do the spirals $r = \theta$ and $r = 2\theta$, intersect?

9.2 Computing Area in Polar Coordinates

In Section 6.1 we saw how to compute the area of a region if the lengths of parallel cross sections are known. Sums based on rectangles led to the formula

$$\text{Area} = \int_a^b c(x)\ dx,$$

where $c(x)$ denoted the cross-sectional length. In polar coordinates sectors of circles, not rectangles, provide an estimate of area.

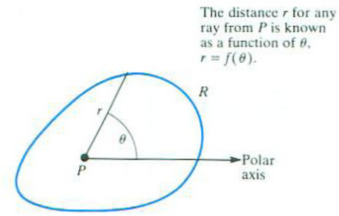


Figure 9.2.1

Let R be a region in the plane and P a point inside it that we take as the pole of a polar coordinate system. Assume that the distance r from P to a point on the boundary of R is known as a function $r = f(\theta)$. Also, assume that any ray from P meets the boundary of R just once, as in Figure 9.2.1.

The cross sections made by the rays from P are not parallel. Like the spokes in a wheel, they meet at the point P. It would be unnatural to use rectangles to estimate the area, but it is reasonable to use sectors of circles that have P as a common vertex.

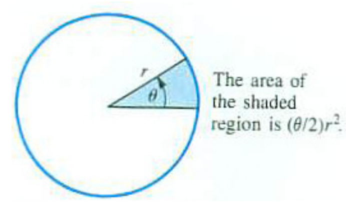


Figure 9.2.2

In a circle of radius r a sector of central angle θ has area $\frac{\theta}{2}r^2$. (See Figure 9.2.2.) This formula plays the same role now as the formula for the area of a rectangle did in Section 6.1.

Area in Polar Coordinates

Let R be the region bounded by the rays $\theta = \alpha$ and $\theta = \beta$ and by the curve $r = f(\theta)$, as shown in Figure 9.2.3. To obtain a **local estimate** for the area of R, consider the portion of R between the rays corresponding to the angles θ and $\theta + d\theta$, where $d\theta$ is a small positive number. (See Figure 9.2.4(a).) The area of the narrow wedge shaded in Figure 9.2.4(a) is approximately that of a sector of a circle of radius $r = f(\theta)$ and angle $d\theta$, shown in Figure 9.2.4(b), whose area is

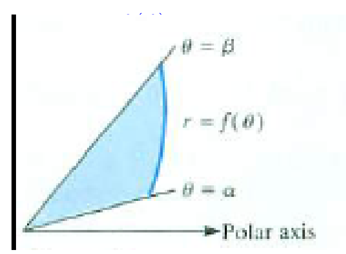


Figure 9.2.3 How to find area in polar coordinates.

$$\frac{f(\theta)^2}{2} d\theta. \tag{9.2.1}$$

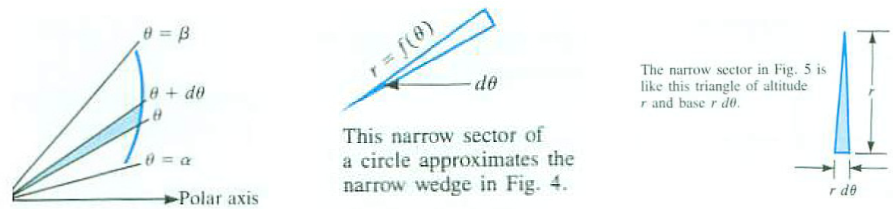
Having found the local estimate of area (9.2.1), we conclude that the area of R is

$$\int_\alpha^\beta \frac{f(\theta)^2}{2}\ d\theta \qquad \text{or} \qquad \int_\alpha^\beta \frac{r^2}{2} d\theta. \tag{9.2.2}$$

Formula 9.2.2 is applied in Section 15.1 (and in a CIE) to the motion of satellites and planets.

It may seem surprising to find $(f(\theta))^2$ in the integrand. But area has the dimension $(\text{length})^2$. Since θ is dimensionless, because it is the length of a

Area has dimensions of length squared.



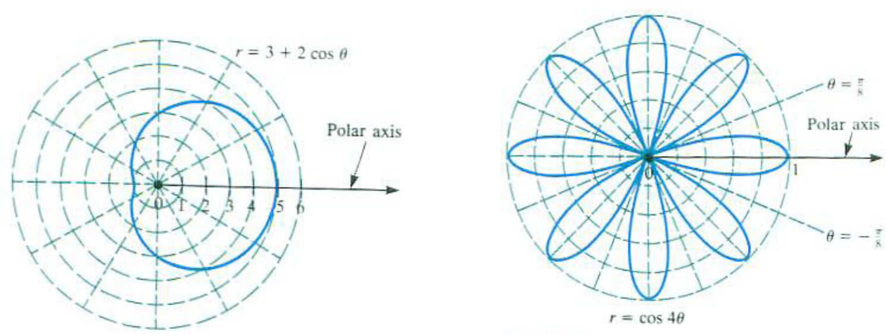
(a) (b) (c)

Figure 9.2.4

circular arc divided by the length of the radius, $d\theta$ is also dimensionless. Hence $f(\theta)\,d\theta$ has the dimension of length and $\frac{1}{2}(f(\theta))^2\,d\theta$ has the dimension of area. For rectangular coordinates, in $f(x)\,dx$, both $f(x)$ and dx have the dimension of length, one along the y-axis, the other along the x-axis, so $f(x)\,dx$ has the dimension of area. To remember the area of the sector in Figure 9.2.4(b), think of it as a triangle of height r and base $r\,d\theta$, as shown in Figure 9.2.4(c). Its area is

Memory device

$$\frac{1}{2}\cdot\underbrace{r}_{\text{height}}\cdot\underbrace{rd\theta}_{\text{base}}=\frac{r^2d\theta}{2}.$$



(a) (b)

Figure 9.2.5 (a) Graph of $r = 3 + 2\cos(\theta)$ for Example 1. (b) Graph of $r = \cos(4\theta)$ for Example 2.

EXAMPLE 1 Find the area of the region bounded by the polar curve $r = 3 + 2\cos(\theta)$ shown in Figure 9.2.5(a).

SOLUTION The cardioid is traced once for $0 \le \theta \le 2\pi$. By the formula just obtained, its area is

$$\begin{aligned}\int_0^{2\pi} \frac{1}{2}(3+2\cos(\theta))^2 d\theta &= \frac{1}{2}\int_0^{2\pi}(9+12\cos(\theta)+4\cos^2(\theta))d\theta \\ &= \frac{1}{2}\int_0^{2\pi}(9+12\cos(\theta)+2(1+\cos(2\theta))\ d\theta \\ &= \frac{1}{2}(9\theta+12\sin(\theta)+2\theta+\sin(2\theta))\Big|_0^{2\pi} = 11\pi.\end{aligned}$$

◇

EXAMPLE 2 Find the area of the region inside one of the eight loops of the eight-leaved rose $r = \cos(4\theta)$.

SOLUTION To graph one of the loops, start with $\theta = 0$. For that angle, $r = \cos(4\cdot 0) = \cos 0 = 1$. The point $(r,\theta) = (1,0)$ is the outer tip of a loop. As θ increases from 0 to $\pi/8$, $\cos(4\theta)$ decreases from $\cos(0) = 1$ to $\cos(\pi/2) = 0$. One of the eight loops is therefore bounded by the rays $\theta = \pi/8$ and $\theta = -\pi/8$, as shown in Figure 9.2.5(b). The area of this loop, which is bisected by the polar axis, is

$$\begin{aligned}\int_{-\pi/8}^{\pi/8} \frac{r^2}{2}d\theta &= \int_{-\pi/8}^{\pi/8} \frac{\cos^2(4\theta)}{2}d\theta = \int_0^{\pi/8}\cos^2(4\theta)\ d\theta = \frac{1}{2}\int_0^{\pi/8}(1+\cos(8\theta))\ d\theta \\ &= \frac{1}{2}\left(\theta+\frac{\sin(8\theta)}{4}\right)\Big|_0^{\pi/8} = \frac{1}{2}\left(\frac{\pi}{8}+\frac{\sin(\pi)}{8}\right)-0 = \frac{\pi}{16} \approx 0.19635.\end{aligned}$$

That the integrand is an even function simplified the calculation. ◇

The Area between Two Curves

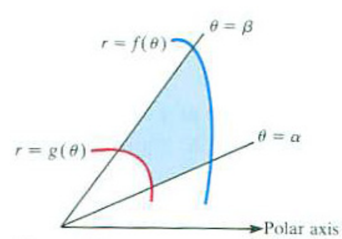


Figure 9.2.6

Assume that $r = f(\theta)$ and $r = g(\theta)$ describe two curves in polar coordinates and that $f(\theta) \geq g(\theta) \geq 0$ for θ in $[\alpha,\beta]$. Let R be the region between them and the rays $\theta = \alpha$ and $\theta = \beta$, as shown in Figure 9.2.6.

The area of R is obtained by subtracting the area within the inner curve, $r = g(\theta)$, from the area within the outer curve, $r = f(\theta)$.

EXAMPLE 3 Find the area of the top half of the region inside the cardioid $r = 1+\cos(\theta)$ and outside the circle $r = \cos(\theta)$.

SOLUTION The region is shown in Figure 9.2.7.

The top half of the cardioid is swept out by $r = 1 + \cos(\theta)$ as θ goes from 0 to π so its area is

$$\begin{aligned}
\frac{1}{2}\int_0^{\pi}(1+\cos(\theta))^2 d\theta &= \frac{1}{2}\int_0^{\pi}\left(1+2\cos(\theta)+\cos^2(\theta)\right)d\theta \\
&= \frac{1}{2}\int_0^{\pi}\left(1+2\cos(\theta)+\frac{1+\cos(2\theta)}{2}\right)d\theta \\
&= \frac{1}{2}\int_0^{\pi}\left(\frac{3}{2}+2\cos(\theta)+\frac{\cos(2\theta)}{2}\right)d\theta \\
&= \frac{1}{2}\left(\frac{3\theta}{2}+2\sin(\theta)+\frac{\sin(2\theta)}{4}\right)\Bigg|_0^{\pi} \\
&= \frac{3\pi}{4}.
\end{aligned}$$

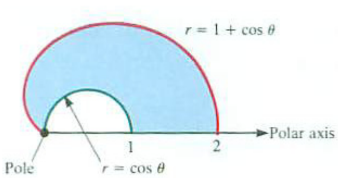


Figure 9.2.7 We must integrate over two different intervals to find the area between the two curves.

The top half of the circle $r = \cos(\theta)$ is half the area of a circle of radius 1/2:

$$\frac{1}{2}\int_0^{\pi/2}\cos^2(\theta)\ d\theta = \frac{1}{2}\pi\left(\frac{1}{2}\right)^2 = \frac{\pi}{8}.$$

Thus the area is

$$\frac{3\pi}{4} - \frac{\pi}{8} = \frac{5\pi}{8} \approx 1.96349.$$

◇

Summary

In this section we saw how to find the area within a curve $r = f(\theta)$ and the rays $\theta = \alpha$ and $\theta = \beta$. The method uses the local approximation by a narrow sector of radius r and angle $d\theta$, which has area $\frac{1}{2}r^2\ d\theta$. This approximation leads to the formula,

$$\text{Area} = \int_{\alpha}^{\beta}\frac{1}{2}r^2\ d\theta.$$

It resembles a triangle of height r and base $rd\theta$. It is prudent to remember the triangle rather than the area formula. This makes it easier to remember the factor of 1/2 in the integrand.

EXERCISES for Section 9.2

In Exercises 1 to 6, draw the region enclosed by the curve and rays and then find its area.

1. $r = 2\theta$, $\alpha = 0$, $\beta = \frac{\pi}{2}$

2. $r = \sqrt{\theta}$, $\alpha = 0$, $\beta = \pi$

3. $r = \frac{1}{1+\theta}$, $\alpha = \frac{\pi}{4}$, $\beta = \frac{\pi}{2}$

4. $r = \sqrt{\sin(\theta)}$, $\alpha = 0$, $\beta = \frac{\pi}{2}$

5. $r = \tan(\theta)$, $\alpha = 0$, $\beta = \frac{\pi}{4}$

6. $r = \sec(\theta)$, $\alpha = \frac{\pi}{6}$, $\beta = \frac{\pi}{4}$

In each of Exercises 7 to 16 draw the region bounded by the curve(s) and find its area.

7. $r = 2\cos(\theta)$

8. $r = e^{\theta}$, $\theta = 0$, and $\theta = 2\pi$

9. inside the cardioid $r = 3 + 3\sin(\theta)$ and outside the circle $r = 3$.

10. $r = \sqrt{\cos(2\theta)}$

11. one loop of $r = \sin(3\theta)$

12. one loop of $r = \cos(2\theta)$

13. inside one loop of $r = 2\cos(2\theta)$ and outside $r = 1$

14. inside $r = 1 + \cos(\theta)$ and outside $r = \sin(\theta)$

15. inside $r = \sin(\theta)$ and outside $r = \cos(\theta)$

16. inside $r = 4 + \sin(\theta)$ and outside $r = 3 + \sin(\theta)$

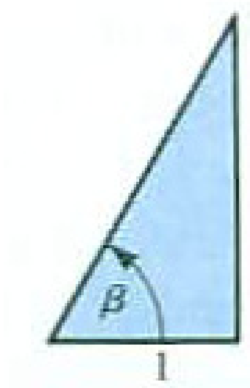
17. Sketch the graph of $r = 4 + \cos(\theta)$. Is it a circle?

18.

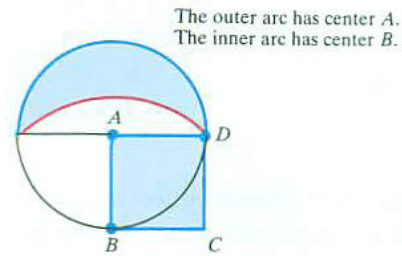
(a) Show that the area of the triangle in Figure 9.2.8(a) is $\int_0^{\beta} \frac{1}{2}\sec^2(\theta)d\theta$.

(b) From (a) and the fact that the area of a triangle is $\frac{1}{2}$(base)(height), show that $\tan(\beta) = \int_0^{\beta} \sec^2(\theta)d\theta$.

(c) Using the equation in (b), obtain another proof that $(\tan(x))' = \sec^2(x)$.



(a)



(b)

Figure 9.2.8

19. Show that the area of the crescent between the two circular arcs, shaded in Figure 9.2.8(b), is equal to the area of square $ABCD$. This type of result encouraged mathematicians from the time of the Greeks to try to find a method using only straightedge and compass for constructing a square whose area equals that of a circle. This was not proved impossible until the end of the nineteenth century when it was shown that π is not the root of a non-zero polynomial with integer coefficients.

20.

(a) Graph $r = 1/\theta$ for $0 < \theta \leq \pi/2$.

(b) Is the area of the region bounded by the curve drawn in (a) and the rays $\theta = 0$ and $\theta = \pi/2$ finite or infinite?

21.

(a) Sketch the curve $r = 1/(1 + \cos(\theta))$.

(b) What is its equation in rectangular coordinates?

(c) Find the area of the region bounded by the curve in (a) and the rays $\theta = 0$ and $\theta = 3\pi/4$, using polar coordinates.

(d) Solve (c) using rectangular coordinates and the equation in (b).

22. Use Simpson's method to estimate the area of the region between $r = \sqrt[3]{1 + \theta^2}$, $\theta = 0$, and $\theta = \pi/2$, correct to three decimal places.

23. Estimate the area of the region bounded by $r = e^{\theta}$, $r = 2\cos(\theta)$ and $\theta = 0$. (You may need to approximate a limit of integration.)

24. Figure 9.2.9 shows a point P inside a convex region $\mathcal{R}$.

(a) Assume that P cuts each chord through it into two intervals of equal length. Does each chord through P cut $\mathcal{R}$ into two regions of equal area?

(b) Assume that each chord through P cuts $\mathcal{R}$ into two regions of equal area. Must P cut each chord through P into two intervals of equal length?

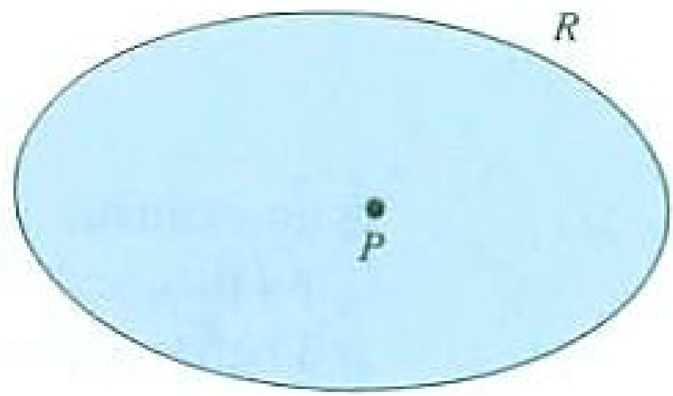


Figure 9.2.9

25. Let $\mathcal{R}$ be a convex region in the plane and P a point on its boundary. Assume that every chord of $\mathcal{R}$ that has an end at P has length not more than 1.

(a) Draw several examples of such an $\mathcal{R}$.

(b) Make a general conjecture about the area $\mathcal{R}$.

(c) Prove it.

26.

(a) Show that a line through the origin intersects the region bounded by the curve in Example 1 in a segment of length 6.

(b) A line through the center of a disk of radius 3 also intersects the disk in a segment of length 6. Does it follow that the disk and the region in Example 1 have the same areas?

27. Let P be a point inside the convex region $\mathcal{R}$. Assume that each chord through P has length 1. How small can the area of $\mathcal{R}$ be? How large?

28. Given a convex region $\mathcal{R}$ in the plane and a point P inside it, if you know the length of each chord that passes through P, can you then determine the area of $\mathcal{R}$

(a) if P is on the border of $\mathcal{R}$?

(b) if P is in the interior of $\mathcal{R}$?

Exercises 29 to 31, contributed by Rick West, are related.

29. The graph of $r = \cos(n\theta)$ has $2n$ loops when n is even. Find the total area within them.

30. The graph of $r = \cos(n\theta)$ has n loops when n is odd. Find the total area within them.

31. Find the total area of the petals within the curve $r = \sin(n\theta)$, where n is a positive integer. (Take the cases n even or odd separately.)

9.3 Parametric Equations

We have considered curves described in three forms: y is a function of x, x and y are related implicitly, and r is a function of θ. Sometimes a curve is described by giving x and y as functions of a third variable. We now look at this description, which arises in the study of motion. It was the basis for the CIE on the Uniform Sprinkler in Chapter 5.

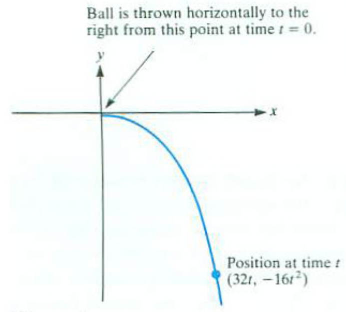


Figure 9.3.1

Two Examples

EXAMPLE 1 A ball is thrown horizontally out of a window with a speed of 32 feet per second falls in a curved path. Air resistance disregarded, its position after t seconds is given by $x = 32t$, $y = -16t^2$ relative to the coordinate system in Figure 9.3.1. The curve is completely described, not by expressing y as a function of x, but by expressing x and y as functions of a third variable t. The third variable is called a **parameter**. The equations $x = 32t$, $y = -16t^2$ are called **parametric equations** for the curve.

para meaning "together," *meter* meaning "measure".

In this example it is easy to eliminate t and so find a direct relation between x and y:

$$t = \frac{x}{32},$$

so

$$y = -16\left(\frac{x}{32}\right)^2 = -\frac{16}{(32)^2}x^2 = -\frac{1}{64}x^2.$$

The path is part of the parabola $y = -\frac{1}{64}x^2$. ◇

In Example 2 elimination of the parameter would lead to a complicated equation involving x and y. An advantage of parametric equations is that they can provide a description of a curve that has no simple representation as $y = f(x)$.

EXAMPLE 2 As a bicycle wheel of radius a rolls along, a tack stuck in its circumference traces out a curve called a **cycloid**, which consists of a sequence of arches, one arch for each revolution of the wheel. (See Figure 9.3.2.) Find the position of the tack as a function of the angle θ through which the wheel turns.

SOLUTION Assume that the tack is initially at the bottom of the wheel. The x coordinate of the tack, corresponding to θ, is

$$\overline{AF} = \overline{AB} - \overline{ED} = a\theta - a\sin(\theta)$$

and the y coordinate is

$$\overline{EF} = \overline{BC}| - \overline{CD} = a - a\cos(\theta).$$

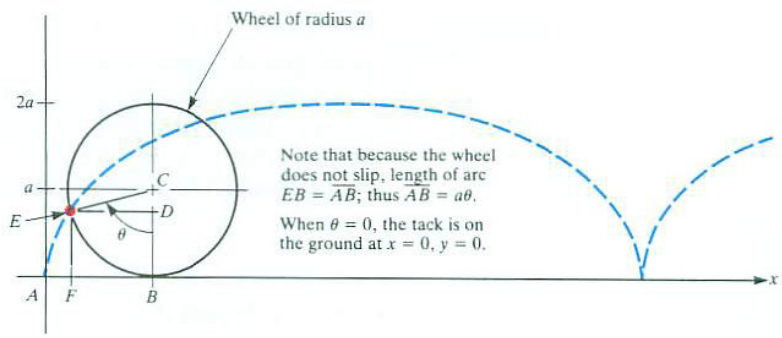


Figure 9.3.2

Thus the position of the tack as a function of the parameter θ is

$$x = a\theta - a\sin(\theta), \quad y = a - a\cos(\theta).$$

See Exercise 35. Eliminating θ leads to a complicated relation between x and y. ◇

Any curve given in the form $y = f(x)$ can be described parametrically. For instance, for $y = e^x + x$ we can introduce a parameter t equal to x and write

$$x = t, \quad y = e^t + t.$$

This may seem artificial, but it will be useful in the next section in order to apply results for curves expressed by means of parametric equations to curves given in the form $y = f(x)$.

How to Find $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$

How can we find the slope of a curve described parametrically as

$$x = g(t), \quad y = h(t)?$$

An often difficult, perhaps impossible, approach is to solve $x = g(t)$ for t as a function of x and substitute it into the equation $y = h(t)$, thus expressing y explicitly in terms of x and then differentiating the result to find dy/dx. Fortunately, there is an easier way. Assume that y is a differentiable function of x. Then, by the Chain Rule,

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt},$$

from which it follows that

Slope of a parameterized curve

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}}. \tag{9.3.1}$$

EXAMPLE 3 At what angle does the arch of the cycloid in Example 2 meet the x-axis at the origin?

SOLUTION The parametric equations of the cycloid are

$$x = a\theta - a\sin(\theta) \qquad \text{and} \qquad y = a - a\cos(\theta).$$

Then

$$\frac{dx}{d\theta} = a - a\cos(\theta) \qquad \text{and} \qquad \frac{dy}{d\theta} = a\sin(\theta).$$

Consequently,

$$\frac{dy}{dx} = \frac{dy/d\theta}{dx/d\theta} = \frac{a\sin(\theta)}{a - a\cos(\theta)} = \frac{\sin(\theta)}{1-\cos(\theta)}.$$

When $\theta = 0$, $(x, y) = (0, 0)$ and $\frac{dy}{dx}$ is not defined because $\frac{dx}{d\theta} = 0$. But, when θ is near 0, (x, y) is near the origin and the slope of the cycloid at $(0, 0)$ can be found by looking at the limit of the slope, which is $\sin\theta/(1 - \cos(\theta))$, as $\theta \to 0^+$. L'Hôpital's Rule applies, and we have

$$\lim_{\theta\to 0^+} \frac{\sin(\theta)}{1-\cos(\theta)} = \lim_{\theta\to 0^+} \frac{\cos(\theta)}{\sin(\theta)} = \infty.$$

Thus the cycloid comes in vertically at the origin, as shown in Figure 9.3.2. ◇

We assume that in (9.3.1) dx/dt is not 0. To obtain d^2y/dx^2 just replace y in (9.3.1) by dy/dx, obtaining

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{\frac{d}{dt}\left(\frac{dy}{dx}\right)}{\frac{dx}{dt}}.$$

EXAMPLE 4 Find d^2y/dx^2 for the cycloid of Example 2.

SOLUTION In Example 3 we found

$$\frac{dy}{dx} = \frac{\sin(\theta)}{1-\cos(\theta)}.$$

As shown in Example 3, $dx/d\theta = a - a\cos(\theta)$. To find $\frac{d^2y}{dx^2}$ we compute

$$\frac{d}{d\theta}\left(\frac{dy}{dx}\right) = \frac{(1-\cos(\theta))\cos(\theta) - \sin(\theta)(\sin(\theta))}{(1-\cos(\theta))^2} = \frac{\cos(\theta)-1}{(1-\cos(\theta))^2} = \frac{-1}{1-\cos(\theta)}.$$

Thus

$$\frac{d^2y}{dx^2} = \frac{\frac{d}{d\theta}\left(\frac{dy}{dx}\right)}{\frac{dx}{d\theta}} = \frac{\frac{-1}{1-\cos(\theta)}}{a - a\cos(\theta)} = \frac{-1}{a(1-\cos(\theta))^2}.$$

Since the denominator is positive (or 0), the quotient is negative, when it is defined. This agrees with Figure 9.3.2, which shows each arch of the cycloid is concave down. ◇

Summary

This section described parametric equations, where x and y are given as functions of a third variable, often time (t) or angle (θ). We also showed how to compute dy/dx and d^2y/dx^2:

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt}$$

and, replacing y by $\frac{dy}{dx}$,

$$\frac{d^2y}{dx^2} = \frac{\frac{d}{dt}\left(\frac{dy}{dx}\right)}{\frac{dx}{dt}}.$$

EXERCISES for Section 9.3

1. For the parametric equations $x = 2t + 1$, $y = t - 1$,

(a) fill in the table:

t	-2	-1	0	1	2
x					
y					

(b) plot the points (x, y) obtained in (a).

(c) graph the curve.

(d) eliminate t to find an equation for the curve in terms of x and y.

2. For the parametric equations $x = t + 1$, $y = t^2$.

(a) fill in the table:

t	-2	-1	0	1	2
x					
y					

(b) plot the five points (x, y) obtained in (a).

(c) graph the curve.

(d) eliminate t to find an equation for the curve in terms of x and y.

3. Consider the parametric equations $x = t^2$, $y = t^2 + t$.

(a) fill in the table:

t	-3	-2	-1	0	1	2	3
x							
y							

(b) plot the seven points (x, y) obtained in (a).

(c) graph the curve.

(d) eliminate t to find an equation for the curve in terms of x and y.

4. Consider the parametric equations $x = 2\cos(t)$, $y = 3\sin(t)$.

(a) fill in the table, expressing the entries as decimals:

t	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$	π	$\frac{5\pi}{4}$	$\frac{3\pi}{2}$	$\frac{7\pi}{4}$	2π
x									
y									

(b) plot the eight distinct points in (a).

(c) Graph the curve given by $x = 2\cos(t)$, $y = 3\sin(t)$.

(d) Using the identity $\cos^2(t) + \sin^2(t) = 1$, eliminate t.

In Exercises 5 to 8 express the curves parametrically with parameter t.

5. $y = \sqrt{1 + x^3}$

6. $y = \tan^{-1}(3x)$

7. $r = \cos^2(\theta)$

8. $r = 3 + \cos(\theta)$

In Exercises 9 to 14 find dy/dx and d^2y/dx^2.

9. $x = t^3 + t$, $y = t^7 + t + 1$

10. $x = \sin(3t)$, $y = \cos(4t)$

11. $x = 1 + \ln(t)$, $y = t\ln(t)$

12. $x = e^{t^2}$, $y = \tan(t)$

13. $r = \cos(3\theta)$ (Introduce x and y.)

14. $r = 2 + 3\sin(\theta)$

In Exercises 15 to 16 find the equation of the tangent line to the curve at the point.

15. $x = t^3 + t^2$, $y = t^5 + t$; $(2, 2)$

16. $x = \frac{t^2+1}{t^3+t^2+1}$, $y = \sec 3t$; $(1, 1)$

In Exercises 17 and 18 find d^2y/dx^2.

17. $x = t^3 + t + 1$, $y = t^2 + t + 2$

18. $x = e^{3t} + \sin(2t)$, $y = e^{3t} + \cos(t^2)$

19. For which values of t is the curve in Exercise 17 concave up? concave down?

20. Let $x = t^3 + 1$ and $y = t^2 + t + 1$. For which values of t is the curve concave up? concave down?

21. Find the slope of the three-leaved rose, $r = \sin(3\theta)$, at the point $(r, \theta) = (\sqrt{2}/2, \pi/12)$.

22.

(a) Find the slope of the cardioid $r = 1 + \cos(\theta)$ at (r, θ).

(b) What happens to the slope as θ approaches π from the left?

(c) What does (b) tell us about the graph of the cardioid? (Show it on the graph.)

23. Obtain parametric equations for the circle of radius a and center (h, k), using as parameter the angle θ shown in Figure 9.3.3(a).

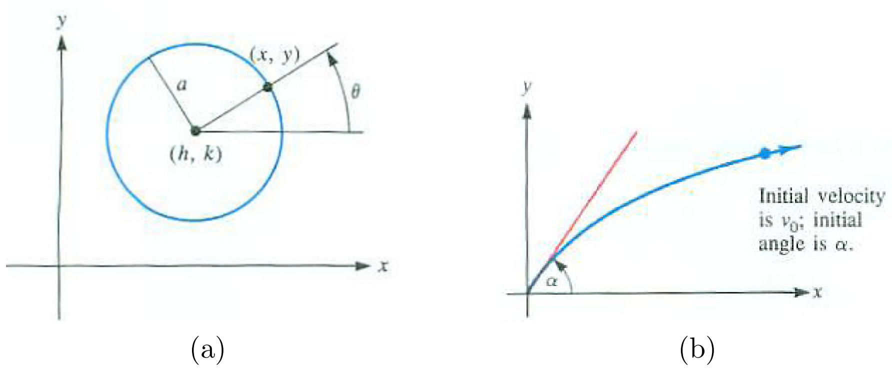


Figure 9.3.3

Exercises 24 to 26 analyze the trajectory of a ball thrown from the origin at an angle α and initial velocity v_0, as sketched in Figure 9.3.3(b). The results are used in the CIE on the Uniform Sprinkler in Chapter 5.

24. It can be shown that if time is in seconds and distance is in feet, then t seconds later the ball is at (x, y) with

$$x = (v_0 \cos(\alpha))t, \quad y = (v_0 \sin(\alpha))t - 16t^2.$$

(a) Express y as a function of x. (Eliminate t.)

(b) What type of curve does the ball follow?

(c) Find the coordinates of its highest point on the curve.

25. Eventually the ball in Exercise 24 falls back to the ground.

(a) Show that the horizontal distance it travels is proportional to $\sin(2\theta)$.

(b) Use (a) to determine the angle that maximizes the horizontal distance traveled.

(c) Show that the horizontal distance traveled in (a) is the same when the ball is thrown at an angle θ or at an angle $\pi/2 - \theta$.

26. Is it possible to extend the horizontal distance traveled by throwing the ball in Exercise 24 from the top of a hill? Assume the hill has height d. (Work with the horizontal distance traveled, x, not the distance along the hill.)

27. The spiral $r = e^{2\theta}$ meets the ray $\theta = \alpha$ at an infinite number of points.

(a) Graph the spiral.

(b) Find the slope of the spiral at each intersection with the ray.

(c) Show that at all the intersections the slopes are the same.

(d) Show that the analog of (c) is not true for the spiral $r = \theta$.

28. The spiral $r = \theta$, $\theta > 0$ meets the ray $\theta = \alpha$ at an infinite number of points (α, α), $(\alpha + 2\pi, \alpha)$, $(\alpha + 4\pi, \alpha)$, ... What happens to the angle between the spiral and the ray at the point $(\alpha + 2\pi n, \alpha)$ as $n \to \infty$?

29. Let a and b be positive numbers and a curve be given parametrically by the equations

$$x = a\cos(t) \qquad y = b\sin(t).$$

(a) Show that the curve is the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$.

(b) Find the area of the region bounded by the ellipse by making a substitution that expresses $4\int_0^a y\,dx$ in terms of an integral in which the variable is t and the range of integration is $[0, \pi/2]$.

30. For the curve given parametrically by

$$x = t^2 + e^t \qquad y = t + e^t$$

for t in $[0, 1]$.

(a) Plot the points corresponding to $t = 0$, $1/2$, and 1.

(b) Find the slope of the curve at $(1, 1)$.

(c) Find the area of the region under the curve and above the interval $[1, e + 1]$. (See Exercise 29(b).)

31. What is the slope of the cycloid in Figure 9.3.2 when it first has height a? See Example 1.

32. The region under the arch of the cycloid

$$x = a\theta - a\sin(\theta), \quad y = a - a\cos(\theta) \qquad (0 \leq \theta \leq 2\pi)$$

and above the x-axis is revolved around the x-axis. Find the volume of the solid of revolution produced.

33. Find the volume of the solid of revolution obtained by revolving the region in Exercise 32 about the y-axis.

34. Let a be a positive constant. For the curve given parametrically by the equations $x = a\cos^3(t)$, $y = a\sin^3(t)$.

(a) sketch the curve.

(b) express the slope of the curve in terms of the parameter t.

35. Solve the parametric equations for the cycloid, $x = a\theta - a\sin(\theta)$, $y = a - a\cos(\theta)$, for x as a function of y. See 2.

36. For a tangent line to the curve in Exercise 34 at a point P in the first quadrant, show that the length of the segment of that line intercepted by the coordinate axes is a.

37. L'Hôpital's rule asserts that if $\lim_{t\to 0} f(t) = 0$, $\lim_{t\to 0} g(t) = 0$, and $\lim_{t\to 0}(f'(t)/g'(t))$ exists, then $\lim_{t\to 0}(f(t)/g(t)) = \lim_{t\to 0}(f'(t)/g'(t))$. Interpret it in terms of the parameterized curve $x = g(t)$, $y = f(t)$. (Make a sketch of the curve near $(0, 0)$ and show on it the geometric meaning of the quotients $f(t)/g(t)$ and $f'(t)/g'(t)$.)

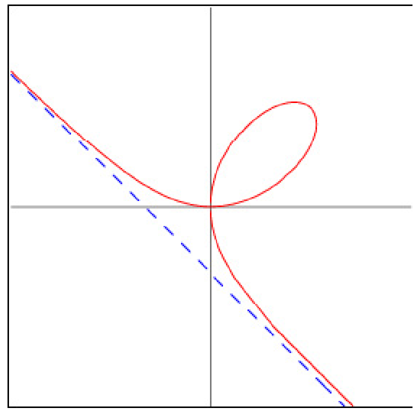


Figure 9.3.4

38. The **Folium of Descartes**, shown in Figure 9.3.4, is the graph of

$$x^3 + y^3 = 3xy.$$

It consists of a loop and two infinite pieces asymptotic to the line $x + y + 1 = 0$. Parameterize the curve by the slope t of the line joining the origin with (x, y). Thus for the point (x, y) on the curve, $y = xt$.

(a) show that

$$x = \frac{3t}{1+t^3} \qquad \text{and} \qquad y = \frac{3t^2}{1+t^3}.$$

(b) find the highest point on the loop.

(c) find the point on the loop furthest to the right.

(d) the loop is parameterized by t in $[0, \infty)$. which values of t parameterize the part in the fourth quadrant?

(e) which values of t parameterize the part in the second quadrant? Show that the Folium of Descartes is symmetric with respect to the line $y = x$.

Search for "Folium Descartes" to see its history that goes back to 1638. See also Exercise 34.

9.4 Arc Length and Speed on a Curve

In Section 4.2 we studied the motion of an object moving on a line. If at time t its position is $x(t)$ then its velocity is $\frac{dx}{dt}$ and its speed is $\left|\frac{dx}{dt}\right|$. Now we will examine the velocity and speed of an object moving along a curved path.

Arc Length and Speed in Rectangular Coordinates

Suppose an object is moving on a path given parametrically by

$$\begin{aligned} x &= g(t) \\ y &= h(t) \end{aligned}$$

where g and h have continuous derivatives. If we think of t as time we can find a formula for its speed.

Let $s(t)$ be the arc length covered from the initial time to time t. In an interval of time of length Δt it travels a distance Δs along the path. We want to find

$$\lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t}.$$

During the time interval $[t, t+\Delta t]$ the object goes from P to Q on the path, covering a distance Δs, as shown in Figure 9.4.1. Its x-coordinate changes by Δx and its y-coordinate by Δy. The chord PQ has length $\sqrt{(\Delta x)^2 + (\Delta y)^2}$.

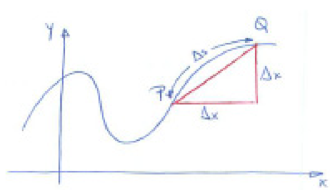


Figure 9.4.1

We assume then that the curve is well behaved in the sense that $\lim_{\Delta t \to 0} \frac{\Delta s}{\overline{PQ}} = 1$. Then

$$\begin{aligned} \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} &= \lim_{\Delta t \to 0} \left(\frac{\Delta s}{\overline{PQ}} \frac{\overline{PQ}}{\Delta t} \right) = \lim_{\Delta t \to 0} \frac{\Delta s}{\overline{PQ}} \lim_{\Delta t \to 0} \frac{\overline{PQ}}{\Delta t} \\ &= 1 \cdot \lim_{\Delta t \to 0} \frac{\overline{PQ}}{\Delta t} = \lim_{\Delta t \to 0} \frac{\sqrt{(\Delta x)^2 + (\Delta y)^2}}{\Delta t} \\ &= \lim_{\Delta t \to 0} \sqrt{\left(\frac{\Delta x}{\Delta t}\right)^2 + \left(\frac{\Delta y}{\Delta t}\right)^2} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}. \end{aligned}$$

so

$$\boxed{\frac{ds}{dt} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}}$$

or, in terms of differentials,

$$ds = \sqrt{(dx)^2 + (dy)^2} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\ dt$$

The rates at which x and y change determine how fast the arc length s changes, as shown in Figure 9.4.2.

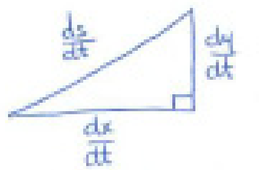


Figure 9.4.2

Now that we have a formula for ds/dt, we integrate it to get the distance along the path covered during a time interval $[a, b]$:

$$\text{arc length} = \int_a^b ds = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\ dt. \qquad (9.4.1)$$

If the curve is given in the form $y = f(x)$, we can use x as the parameter. A parametric representation of the curve then is

$$x = x, \qquad y = f(x)$$

and (9.4.1) becomes

$$\text{arc length} = \int_a^b \sqrt{1 + (f'(x))^2}\ dx.$$

The arc length function is, by definition, an non-decreasing function. This means ds/dt is never negative. In most applications ds/dt will never be zero either.

Three examples will show how these formulas are applied. The first goes back to the year 1657, when the 20-year old Englishman, William Neil, found the length of an arc on the graph of $y = x^{3/2}$. His method was more complicated. Earlier, Thomas Harriot had found the length of an arc of the spiral $r = e^{\theta}$, but his work was not widely known.

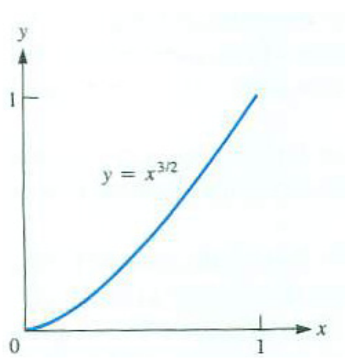


Figure 9.4.3

EXAMPLE 1 Find the arc length of the curve $y = x^{3/2}$ for x in $[0, 1]$. (See Figure 9.4.3.)

SOLUTION By (9.4.1),

$$\text{arc length} = \int_0^1 \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\ dx.$$

Since $y = x^{3/2}$, we differentiate to find $dy/dx = \frac{3}{2}x^{1/2}$. Thus

$$\begin{aligned}
\text{arc length} &= \int_0^1 \sqrt{1+\left(\tfrac{3}{2}x^{1/2}\right)^2}\,dx = \int_0^1 \sqrt{1+\tfrac{9}{4}x}\,dx \\
&= \int_1^{13/4} \sqrt{u}\cdot\tfrac{4}{9}\,du \qquad\qquad (u = 1+\tfrac{9}{4}x,\ du = \tfrac{9}{4}dx) \\
&= \tfrac{4}{9}\cdot\tfrac{2}{3}u^{3/2}\Big|_1^{13/4} = \tfrac{8}{27}\left(\left(\tfrac{13}{4}\right)^{3/2} - 1^{3/2}\right) \\
&= \tfrac{8}{27}\left(\tfrac{13^{3/2}}{8} - 1\right) = \tfrac{13^{3/2}-8}{27} \approx 1.43971.
\end{aligned}$$

◇

The arc length of the curve $y = x^a$ where a is a non-zero rational number, usually cannot be computed using the Fundamental Theorem of Calculus. The only cases in which it can be computed by the FTC are $a = 1$ (the graph of $y = x$) and $a = 1 + \frac{1}{n}$ where n is an integer. Exercise 32 treats this question.

EXAMPLE 2 In Section 9.3 the parametric equations for the motion of a ball thrown horizontally with a speed of 32 feet per second ($\approx$ 21.8 mph) were found to be $x = 32t$, $y = -16t^2$. (See Example 1 and Figure 9.3.1 in Section 9.3.) How fast is the ball moving at time t? Find the distance s that the ball travels during the first b seconds.

SOLUTION From $x = 32t$ and $y = -16t^2$ we compute $\frac{dx}{dt} = 32$ and $\frac{dy}{dt} = -32t$. Its speed at time t is

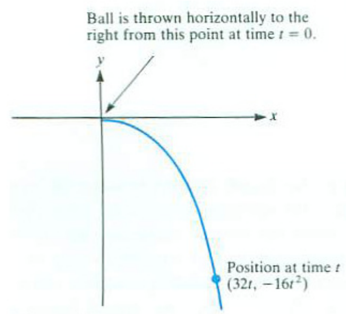


Figure 9.4.4

$$\text{Speed} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} = \sqrt{(32)^2 + (-32t)^2} = 32\sqrt{1+t^2} \text{ feet per second.}$$

The distance traveled is the arc length from $t = 0$ to $t = b$. By (9.4.1),

$$\text{arc length} = \int_0^b \sqrt{(32)^2 + (-32t)^2}\,dt = 32\int_0^b \sqrt{1+t^2}\,dt.$$

The integral can be evaluated with an integration table or with the trigonometric substitution $x = \tan(\theta)$. An antiderivative of $\sqrt{1+t^2}$ is

See Formula 31 in the integral table.

$$\frac{1}{2}\left(t\sqrt{1+t^2} + \ln\left(t + \sqrt{1+t^2}\right)\right)$$

and the distance traveled is

$$16b\sqrt{1+b^2} + 16\ln\left(b + \sqrt{1+b^2}\right).$$

◇

EXAMPLE 3 Find the length of one arch of the cycloid found in Example 2 of Section 9.3.

SOLUTION The curve can be parameterized as $x = a\theta - a\sin(\theta)$, and $y = a - a\cos(\theta)$ with θ as the parameter. Over one arch of the cycloid, θ varies from 0 to 2π.

We compute

$$\frac{dx}{d\theta} = a - a\cos(\theta) \qquad \text{and} \qquad \frac{dy}{d\theta} = a\sin(\theta).$$

The square of the speed is

$$\begin{aligned}
(a - a\cos(\theta))^2 + (a\sin(\theta))^2 &= a^2\left((1-\cos(\theta))^2 + (\sin(\theta))^2\right) \\
&= a^2\left(1 - 2\cos(\theta) + (\cos(\theta))^2 + (\sin(\theta))^2\right) \\
&= a^2\left(2 - 2\cos(\theta)\right) \\
&= 2a^2(1-\cos(\theta)).
\end{aligned}$$

Using formula (9.4.1) and the trigonometric identity $1 - \cos(\theta) = 2\sin^2(\theta/2)$, we have

$$\begin{aligned}
\text{arc length of one arch} &= \int_0^{2\pi} \sqrt{2a^2(1-\cos(\theta))}\ d\theta = a\sqrt{2}\int_0^{2\pi}\sqrt{1-\cos(\theta)}\ d\theta \\
&= a\sqrt{2}\int_0^{2\pi}\sqrt{2}\sin\left(\frac{\theta}{2}\right)\ d\theta = 2a\int_0^{2\pi}\sin\left(\frac{\theta}{2}\right)\ d\theta \\
&= 2a\left(-2\cos\left(\frac{\theta}{2}\right)\Big|_0^{2\pi}\right) = 2a\left(-2(-1) - (-2)(1)\right) = 8a.
\end{aligned}$$

So, as θ varies from 0 to 2π, a bicycle travels a distance of $2\pi a \approx 6.28318a$ and a tack in the tread of the tire travels a distance $8a$. ◇

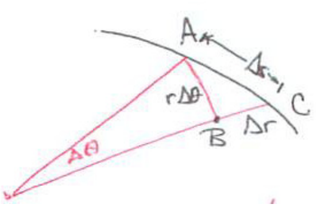


Figure 9.4.5

Arc Length and Speed in Polar Coordinates

So far in this section curves have been described in rectangular coordinates. Now we consider a curve given in polar coordinates by $r = f(\theta)$.

We will estimate the length of arc Δs corresponding to small changes $\Delta\theta$ and Δr in polar coordinates, as shown in Figure 9.4.5. The region bounded by the circular arc AB, the straight segment BC, and AC, part of the curve, resembles a right triangle whose legs have lengths $r\Delta\theta$ and Δr. Because Δs

is well approximated by its hypotenuse, $\sqrt{(r\Delta\theta)^2 + (\Delta r)^2}$. We expect

$$\begin{aligned}
\frac{ds}{d\theta} = \lim_{\Delta\theta\to 0}\frac{\Delta s}{\Delta\theta} &= \lim_{\Delta\theta\to 0}\frac{\sqrt{(r\Delta\theta)^2 + (\Delta r)^2}}{(\Delta\theta)} \\
&= \lim_{\Delta\theta\to 0}\sqrt{r^2 + \left(\frac{\Delta r}{\Delta\theta}\right)^2} \\
&= \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2}.
\end{aligned}$$

That is,

For a curve given in polar coordinates

$$\frac{ds}{d\theta} = \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2} \quad \text{or} \quad ds = \sqrt{(r\ d\theta)^2 + (dr)^2} = \sqrt{r^2 + (r')^2}\ d\theta.$$

This formula can also be obtained from the rectangular coordinate formula by using $x = r\cos(\theta)$ and $y = r\sin(\theta)$. (See Exercise 20.) However, we prefer the geometric approach because it is more direct, more intuitive, and easier to remember.

See Exercise 20.

Arc Length of a Polar Curve $r = f(\theta)$

The length of the curve $r = f(\theta)$ for θ in $[\alpha, \beta]$ is $s = \int_\alpha^\beta\ ds$ where

$$ds = \sqrt{r^2 + (r')^2}\ d\theta = \sqrt{(f(\theta))^2 + (f'(\theta))^2}\ d\theta.$$

EXAMPLE 4 Find the length of the spiral $r = e^{-3\theta}$ for θ in $[0, 2\pi]$.

SOLUTION From

$$r' = \frac{dr}{d\theta} = -3e^{-3\theta},$$

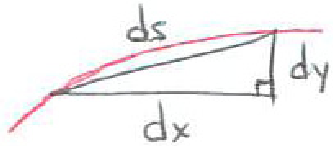
the formula gives

$$\begin{aligned}
\text{arc length} &= \int_{\alpha}^{\beta} \sqrt{r^2 + (r')^2}\, d\theta = \int_0^{2\pi} \sqrt{\left(e^{-3\theta}\right)^2 + \left(-3e^{-3\theta}\right)^2}\, d\theta \\
&= \int_0^{2\pi} \sqrt{e^{-6\theta} + 9e^{-6\theta}}\, d\theta = \sqrt{10} \int_0^{2\pi} \sqrt{e^{-6\theta}}\, d\theta \\
&= \sqrt{10} \int_0^{2\pi} e^{-3\theta}\, d\theta = \sqrt{10}\, \frac{e^{-3\theta}}{-3}\Bigg|_0^{2\pi} \\
&= \sqrt{10} \left(\frac{e^{-3\cdot 2\pi}}{-3} - \frac{e^{-3\cdot 0}}{-3} \right) = \sqrt{10} \left(\frac{e^{-6\pi}}{-3} + \frac{1}{3} \right) \\
&= \frac{\sqrt{10}}{3} \left(1 - e^{-6\pi}\right) \approx 1.0541.
\end{aligned}$$

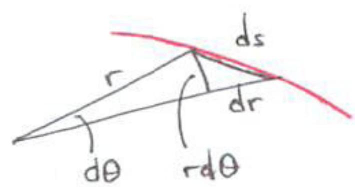
◇

Summary

This section concerned speed along a parametric path and the length of the path. If the path is described in rectangular coordinates, then Figure 9.4.6(a) conveys the key ideas. If in polar coordinates, Figure 9.4.6(b) is the key. It is easier to recall the diagrams than the formulas for speed and arc length. Everything depends on the Pythagorean Theorem.



(a)



(b)

Figure 9.4.6 (a) $ds = \sqrt{(dx)^2 + (dy)^2}$ (b) $ds = \sqrt{(rd\theta)^2 + (dr)^2}$

EXERCISES for Section 9.4

In Exercises 1 to 8 find the arc lengths of the curves over the intervals.

1. $y = x^{3/2}$, x in $[1, 2]$

2. $y = x^{2/3}$, x in $[0, 1]$

3. $y = (e^x + e^{-x})/2$, x in $[0, b]$

4. $y = x^2/2 - \ln(x)/4$, x in $[2, 3]$

5. $x = \cos^3(t)$, $y = \sin^3(t)$, t in $[0, \pi/2]$

6. $r = e^\theta$, θ in $[0, 2\pi]$

7. $r = 1 + \cos(\theta)$, θ in $[0, \pi]$

8. $r = \cos^2(\theta/2)$, θ in $[0, \pi]$

In each of Exercises 9 to 12 find the speed of a particle at time t, given the parametric equations for its path.

9. $x = 50t$, $y = -16t^2$

10. $x = \sec(3t)$, $y = \sin^{-1}(4t)$

11. $x = t + \cos(t)$, $y = 2t - \sin(t)$

12. $x = \csc(\theta/2)$, $y = \tan^{-1}(\sqrt{t})$

13.

(a) Graph $x = t^2$, $y = t$ for $0 \le t \le 3$.

(b) Estimate its arc length from $(0, 0)$ to $(9, 3)$ by an inscribed polygon whose vertices have x-coordinates 0, 1, 4, and 9.

(c) Set up a definite integral for the arc length of the curve in question.

(d) Estimate the definite integral by using a partition of $[0, 3]$ into three sections, each of length 1, and the trapezoid method.

(e) Estimate the definite integral by Simpson's method with six sections.

(f) If your calculator has a program to evaluate definite integrals, use it to evaluate the definite integral in (c) to four decimal places.

14.

(a) Graph $y = 1/x^2$ for x in $[1, 2]$.

(b) Estimate the length of the arc by using an inscribed polygon whose vertices are $(1, 1)$, $(\frac{5}{4}, \left(\frac{4}{5}\right)^2)$, $(\frac{3}{2}, \left(\frac{2}{3}\right)^2)$, and $(2, \frac{1}{4})$.

(c) Set up a definite integral for the arc length of the curve.

(d) Estimate the definite integral by the trapezoid method, using four equal length sections.

(e) Estimate the definite integral by Simpson's method with four sections.

(f) If your calculator has a program to evaluate definite integrals, use it to evaluate the definite integral to four decimal places.

15. How long is the spiral $r = e^{-3\theta}$, $\theta \geq 0$?

16. How long is the spiral $r = 1/\theta$, $\theta \geq 2\pi$?

17. Suppose that a curve has equation $x = f(y)$ in rectangular coordinates. Show that

$$\text{arc length} = \int_c^d \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy$$

where y ranges over the interval $[c, d]$, using a triangle whose sides have length dx, dy, and ds.

18. Consider the arc length of $y = x^{2/3}$ for x in the interval $[1, 8]$.

(a) Set up a definite integral for the arc length using x as the parameter.

(b) Set up a definite integral for the arc length using y as the parameter.

(c) Evaluate the easier of the two integrals found in parts (a) and (b).

See Exercise 17.

19. At time $t \geq 0$ a ball is at the point $(24t, -16t^2 + 5t + 3)$.

(a) Where is it at time $t = 0$?

(b) What is its horizontal speed at that time?

(c) What is its vertical speed at that time?

20. We obtained the formula $\frac{ds}{d\theta} = \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2}$ geometrically.

(a) Obtain it by calculus, starting with $\left(\frac{ds}{d\theta}\right)^2 = \left(\frac{dx}{d\theta}\right)^2 + \left(\frac{dy}{d\theta}\right)^2$ and using the relations $x = r\cos(\theta)$ and $y = r\sin(\theta)$, where $r = f(\theta)$.

(b) Which derivation do you prefer? Why?

21. Let $P = (x, y)$ depend on θ as shown in Figure 9.4.7.

(a) Sketch the curve that P sweeps out.

(b) Show that $P = (2\cos(\theta), \sin(\theta))$.

(c) Set up a definite integral for the length of the curve in(a). Do not evaluate the integral.

(d) Eliminate θ and show that P is on the ellipse

$$\frac{x^2}{4} + \frac{y^2}{1} = 1.$$

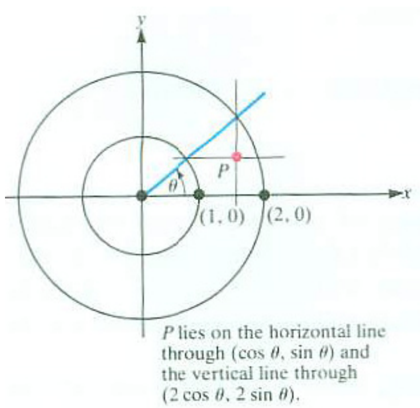


Figure 9.4.7

22.

(a) At time t a particle has polar coordinates $r = g(t)$, $\theta = h(t)$. How fast is it moving?

(b) Use (a) to find the speed of a particle which at time t is at the point $(r, \theta) = (e^t, 5t)$.

23.

(a) How far does a bug travel from time $t = 1$ to time $t = 2$ if at time t it is at the point $(x, y) = (\cos \pi t, \sin \pi t)$?

(b) How fast is it moving at time t?

(c) Graph its path relative to an xy coordinate system. Where is it at time $t = 1$? At $t = 2$?

(d) Eliminate t to find a relation between x and y.

24. Let a be a positive number. Find the arc length of the Archimedean spiral $r = a\theta$ for θ in $[0, 2\pi]$.

25. If $r = 1 + \cos\theta$ for θ in $[0, \pi]$ we may consider r as a function of θ or as a function of s, arc length along the curve, measured, say, from $(2, 0)$.

(a) Find the average of r with respect to θ in $[0, \pi]$.

(b) Find the average of r with respect to s. (Express all quantities appearing in this average in terms of θ.)

(See also Exercises 12 and 13 in the Chapter 9 Summary.)

26. Let $r = f(\theta)$ describe a curve in polar coordinates. Assume that $df/d\theta$ is continuous. Let θ be a function of time t and let $s(t)$ be the length of the curve corresponding to the time interval $[a, t]$.

(a) What definite integral is equal to $s(t)$?

(b) What is the speed ds/dt?

27. The function $r = f(\theta)$ describes a curve in polar coordinates, for θ in $[0, 2\pi]$. Assume r' is continuous and $f(\theta) > 0$. Prove that the average of r as a function of arc length is at least as large as $2A/s$, where A is the area swept out by the radius and s is the arc length of the curve. For what curves is the average equal to $2A/s$?

28. The equations $x = \cos(t)$, $y = 2\sin(t)$, t in $[0, \pi/2]$ describe a quarter of an ellipse. Draw the arc. Describe at least two different ways of estimating its length. Compare the advantages and challenges each method presents. Use the method of your choice to estimate the length of this arc.

29. When a curve is given in rectangular coordinates, its slope is $\frac{dy}{dx}$. To find the slope of the tangent line to the curve given in polar coordinates involves more work. Assume that $r = f(\theta)$. Use the relation

$$\frac{dy}{dx} = \frac{dy/d\theta}{dx/d\theta},$$

which comes from the Chain Rule ($\frac{dy}{d\theta} = \frac{dy}{dx}\frac{dx}{d\theta}$).

(a) Using $y = r\sin(\theta)$ and $x = r\cos(\theta)$ find $\frac{dy}{d\theta}$ and $\frac{dx}{d\theta}$.

(b) Show that the slope is

$$\frac{r\cos(\theta) + \frac{dr}{d\theta}\sin(\theta)}{-r\sin(\theta) + \frac{dr}{d\theta}\cos(\theta)}. \tag{9.4.2}$$

30. Use (9.4.2) to find the slope of the cardioid $r = 1 + \sin(\theta)$ at $\theta = \frac{\pi}{3}$.

31. Let $y = f(x)$ for x in $[0, 1]$ describe a curve that starts at $(0, 0)$, ends at $(1, 1)$, and lies in the square with vertices $(0, 0)$,$(1, 0)$,$(1, 1)$, and $(0, 1)$. Assume f has a continuous derivative.

(a) What can be said about the arc length of the curve? How small and how large can it be?

(b) Answer (a) if we assumed also that $f'(x) \geq 0$ for x in $[0, 1]$.

32. Consider the length of the curve $y = x^m$, $0 \leq x \leq 1$, where m is a rational number. Show that the Fundamental Theorem of Calculus is of aid in computing this length only if $m = 1$ or if $m = 1 + 1/n$ for some integer n. (It is known that $\int x^p(1 + x)^q\, dx$ is elementary for rational numbers p and q only when at least one of p, q, and $p + q$ is an integer.)

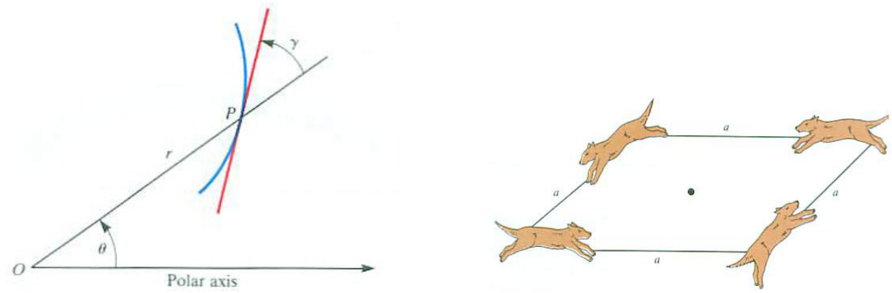
33. If one convex polygon P_1 lies inside another polygon P_2 is the perimeter of P_1 necessarily less than the perimeter of P_2? What if P_1 is not convex?

34. One part of the cardioid $r = 1 + \sin(\theta)$ is traced as θ increases from $\frac{-\pi}{2}$ to $\frac{\pi}{2}$. Find its highest point in polar coordinates.

Exercises 35 and 36 form a unit.

35. Figure 9.4.8(a) shows the angle between the radius and tangent line to the curve $r = f(\theta)$. Using $\gamma = \alpha - \theta$ and $\tan(A - B) = \frac{\tan(A) - \tan(B)}{1 + \tan(A)\tan(B)}$, show that $\tan(\gamma) = \dfrac{r}{r'}$. See Exercise 36 for an intuitive derivation of $\tan(\gamma)$.

36. The formula $\tan(\gamma) = r/r'$ in Exercise 35 is so simple one would expect a simple geometric explanation. Use the triangle in Figure 9.4.5 that we used to obtain the formula for $\frac{ds}{d\theta}$ to show that $\tan(\gamma)$ should be r/r'. See Exercise 35.



(a) (b)

Figure 9.4.8

37. Four dogs are chasing each other counterclockwise at the same speed as shown in Figure 9.4.8(b). Initially they are at the vertices of a square of side a. As they chase each other, each running directly toward the dog in front, they approach the center of the square in spiral paths. How far does each dog travel?

(a) Find the equation of the spiral path each dog follows and use calculus to answer this question.

(b) Answer the question without using calculus.

38. We assumed that a chord AB of a smooth curve is a good approximation of the arc AB when B is near to A. Show that the formula for arc length is consistent with this assumption. That is, if $y = f(x)$ has a continuous derivative, $A = (a, f(a))$, $B = (x, f(x))$, then

$$\frac{\int_a^x \sqrt{1 + f'(t)^2}\ dt}{\sqrt{(x-a)^2 + (f(x) - f(a))^2}}$$

approaches 1 as x approaches a. (L'Hôpital's Rule does not help. For simplicity, assume $a = 0 = f(0)$.)

39. In some approaches to arc length and speed on a curve the arc length is found first, then the speed. We outline this method now.
Let $x = g(t)$, $y = h(t)$ where g and h have continuous derivatives. Let $a = t_0 < t_1 < t_2 < \cdots < t_n = b$ be a partition of $[a, b]$ into n equal sections of length $\Delta t = (b-a)/n$. Let $P_i = (g(t_i), h(t_i))$, which we write as (x_i, y_i). Then the polygon $P_0P_1P_2\cdots P_n$ is inscribed in the curve. We assume that as $n \to \infty$, the length of the polygon, $\sum_{i=1}^n \overline{P_{i-1}P_i}$ approaches the length of the curve from $(g(a), h(a))$ to $(g(b), h(b))$.

(a) Show that the length of the polygon is $\sum_{i=1}^n \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}$.

(b) Show that the sum can be written as

$$\sum_{i=1}^{n} \sqrt{(g'(t_i^*))^2 + (h'(t_i^{**}))^2} \cdot \Delta t \qquad (9.4.3)$$

for some t_i^* and t_i^{**} in $[t_{i-1}, t_i]$.

(c) Why would you expect the limit of (9.4.3) as $n \to \infty$ to be $\int_a^b \sqrt{(g'(t))^2 + h'(t))^2}\, dt$? This result is typically proved in advanced courses, and is true even though t_i^* and t_i^{**} may be different.

(d) From (c) deduce that the speed is $\sqrt{(g'(t))^2 + h'(t))^2}$.

9.5 The Area of a Surface of Revolution

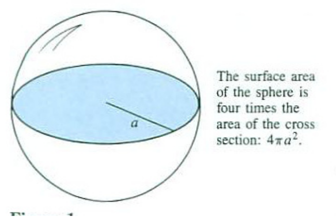
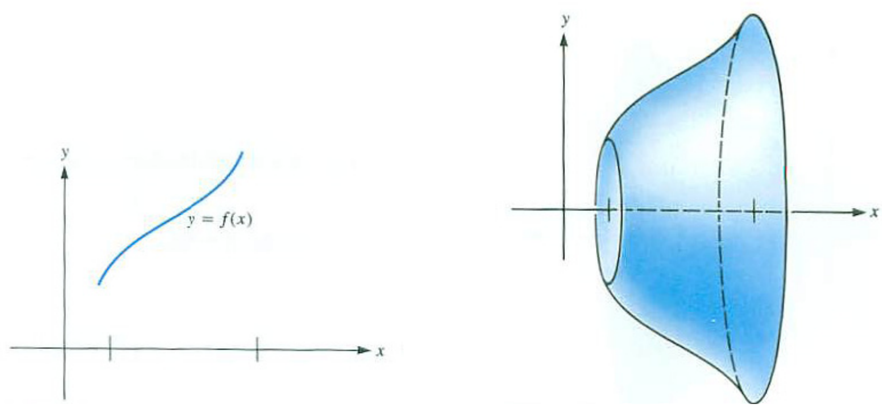


Figure 9.5.1

In this section we develop a formula for the surface area of a solid of revolution as a definite integral. We will show that the surface area of a sphere is four times the area of a cross section through its center. (See Figure 9.5.1.) This was one of the great discoveries of Archimedes in the third century B.C.

The Local Estimate of Surface Area

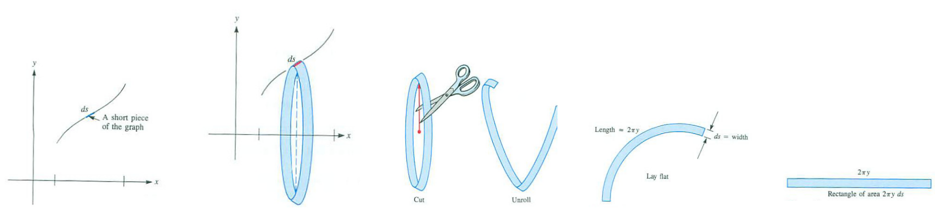
Let $y = f(x)$ have a continuous derivative for x in some interval. Assume that $f(x) \geq 0$ on it. When the graph fo f is revolved about the x-axis it sweeps out a surface, as shown in Figures 9.5.2. To develop a definite integral for this



(a) (b)

Figure 9.5.2

surface area, we use an informal approach.



(a) (b) (c) (d) (e)

Figure 9.5.3

A short section of the graph $y = f(x)$ is almost straight. We approximate it by a short line segment of length ds, a small number. When the segment is revolved about the x-axis it sweeps out a narrow band. (See Figures 9.5.3(a) and (b).)

If we can estimate the area of the band, then we will have a local approximation of the surface area, from which we get a definite integral for the entire surface area.

Cut the band with scissors and lay it flat, as in Figures 9.5.3(c) and (d). The area of the flat band in Figure 9.5.3(d) is close to the area of a flat rectangle of length $2\pi y$ and width ds, as in Figure 9.5.3(e). (See Exercises 28 and 29.)

The gives us

$$\text{local approximation of the surface area of one slice} = 2\pi y\ ds.$$

The Key Integral for Surface Area

From the local estimate for surface area we obtain the following integral for the total area of the curved surface:

$$\text{surface area} = \int_{s_0}^{s_1} 2\pi y\ ds. \tag{9.5.1}$$

In (9.5.1), $[s_0, s_1]$ describes the appropriate interval on the s-axis. Since s is a clumsy parameter, for computations we will use one of the forms for ds to change (9.5.1) into more convenient integrals.

Say that the section of the graph $y = f(x)$ that was revolved corresponds to the interval $[a, b]$ on the x-axis, as in Figure 9.5.4. Then

Assume that $y \geq 0$ and that dy/dx is continuous.

$$ds = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\ dx$$

and the surface area integral $\int_{s_0}^{s_1} 2\pi y\ ds$ becomes

$$\text{surface area} = \int_a^b 2\pi y \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\ dx. \tag{9.5.2}$$

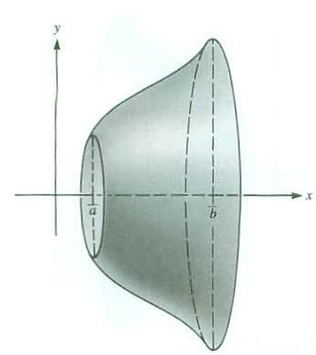


Figure 9.5.4

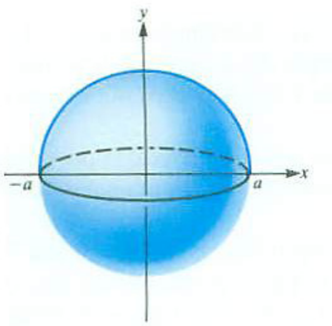


Figure 9.5.5

EXAMPLE 1 Find the surface area of a sphere of radius a.

SOLUTION The circle of radius a has the equation $x^2 + y^2 = a^2$. The top half has the equation $y = \sqrt{a^2 - x^2}$. The sphere of radius a is formed by revolving it about the x-axis. (See Figure 9.5.5.) We have

$$\text{surface area of sphere} = \int_{-a}^{a} 2\pi y\ ds.$$

Because $dy/dx = -x/\sqrt{a^2 - x^2}$ we find that

$$\begin{aligned} ds &= \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\ dx = \sqrt{1 + \left(\frac{-x}{\sqrt{a^2 - x^2}}\right)^2}\ dx \\ &= \sqrt{1 + \frac{x^2}{a^2 - x^2}}\ dx = \sqrt{\frac{a^2}{a^2 - x^2}}\ dx = \frac{a}{\sqrt{a^2 - x^2}}\ dx. \end{aligned}$$

Thus,

$$\begin{aligned} \text{surface area of sphere} &= \int_{-a}^{a} 2\pi y\ ds = \int_{-a}^{a} 2\pi \sqrt{a^2 - x^2} \frac{a}{\sqrt{a^2 - x^2}}\ dx \\ &= \int_{-a}^{a} 2\pi a\ dx = 2\pi a x|_{-a}^{a} = 4\pi a^2. \end{aligned}$$

The surface area of a sphere is 4 times the area of its equatorial cross section. ◇

If the graph is given parametrically, $x = g(t)$, $y = h(t)$, where g and h have continuous derivatives and $h(t) \geq 0$, then it is natural to express the integral $\int_{s_0}^{s_1} 2\pi y\ ds$ as an integral over an interval on the t-axis. If t varies in the interval $[a, b]$, then

$$ds = \sqrt{(dx)^2 + (dy)^2} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\ dt$$

which leads to

$$\begin{array}{c}\text{Surface area for} \\ \text{a parametric curve}\end{array} = \int_{a}^{b} 2\pi y \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\ dt. \qquad (9.5.3)$$

9.5.2 is the special case of (9.5.3) when the parameter is x.

The formulas may seem to refer only to surfaces obtained by revolving a curve about the x-axis. In fact, they can apply to revolution about any line. The factor y in the integrand, $2\pi y\ ds$, is the distance from a point on the curve to the axis of revolution. We replace y by R (for radius) to free ourselves from coordinate systems. (We use capital R to avoid confusion with polar coordinates.) The simplest way to write the formula for surface area of revolution is then

$$\boxed{\text{surface area} = \int_c^d 2\pi R\ ds,}$$

where the interval $[c, d]$ refers to the parameter s. However, in practice arc length, s, is seldom a convenient parameter. Instead, x, y, t or θ is used and the interval of integration describes the interval over which the parameter varies.

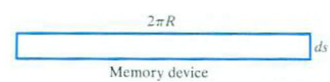


Figure 9.5.6 The key to this section.

To remember the formula, think of a narrow circular band of width ds and radius R as having an area close to the area of the rectangle shown in Figure 9.5.6.

EXAMPLE 2 Find the area of the surface obtained by revolving around the y-axis the part of the parabola $y = x^2$ that lies between $x = 1$ and $x = 2$. (See Figure 9.5.7.)

R is found by inspection of a diagram.

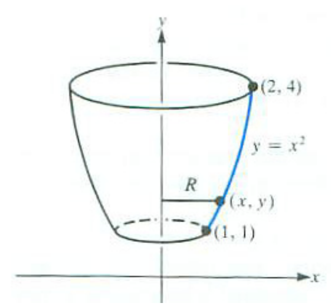


Figure 9.5.7

SOLUTION The surface area is $\int_a^b 2\pi R\ ds$. Since the curve is described as a function of x, choose x as the parameter. From Figure 9.5.7, $R = x$. Because

$$ds = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\ dx = \sqrt{1 + 4x^2}\ dx$$

the surface area is

$$\int_1^2 2\pi x\sqrt{1 + 4x^2}\ dx.$$

To evaluate the integral, use the substitution

$$u = 1 + 4x^2 \qquad du = 8x\ dx.$$

Hence $x\,dx = du/8$. The new limits of integration are $u = 5$ and $u = 17$. Thus

$$\begin{aligned}\text{surface area} &= \int_5^{17} 2\pi\sqrt{u}\frac{du}{8} = \frac{\pi}{4}\int_5^{17}\sqrt{u}\,du \\ &= \frac{\pi}{4}\cdot\frac{2}{3}u^{3/2}\Big|_5^{17} = \frac{\pi}{6}(17^{3/2} - 5^{3/2}) \approx 30.846.\end{aligned}$$

◇

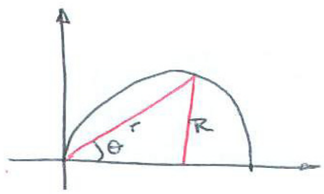


Figure 9.5.8

EXAMPLE 3 Find the surface area when the curve $r = \cos(\theta)$, θ in $[0, \pi/2]$ is revolved around (a) the x-axis and (b) the y-axis.

SOLUTION The curve shown in Figure 9.5.8 is a semicircle with radius 1/2 and center (1/2, 0).

(a) We need to find both R and $ds/d\theta$. $R = r\sin(\theta) = \cos(\theta)\sin(\theta)$ and using the formula for $\frac{ds}{d\theta}$ for a polar curve from Section 9.4

$$\frac{ds}{d\theta} = \sqrt{r(\theta)^2 + r'(\theta)^2} = \sqrt{(\cos(\theta))^2 + (-\sin(\theta))^2} = 1.$$

Then

$$\begin{aligned}\text{surface area} &= \int_0^{\pi/2} 2\pi R\frac{ds}{d\theta}\,d\theta = \int_0^{\pi/2} 2\pi\cos(\theta)\sin(\theta)(1)\,d\theta \\ &= \int_0^{\pi/2} 2\pi\sin(\theta)\cos(\theta)\,d\theta = 2\pi\frac{\sin^2(\theta)}{2}\Big|_0^{\pi/2} = \pi.\end{aligned}$$

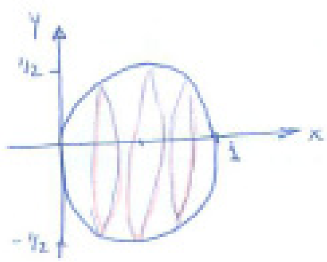


Figure 9.5.9

This is expected since the surface is a sphere of radius 1/2. See Figure 9.5.9.

(b) In this case $R = r\cos(\theta) = \cos^2(\theta)$. Thus

$\int_0^{\pi/2}\cos^2(\theta)\,d\theta$ was evaluated in Section 8.5.

$$\begin{aligned}\text{surface area} &= \int_0^{\pi/2} 2\pi R\frac{ds}{d\theta}\,d\theta = \int_0^{\pi/2} 2\pi\cos^2(\theta)(1)\,d\theta \\ &= 2\pi\int_0^{\pi/2}\cos^2(\theta)\,d\theta = 2\pi(\frac{\pi}{4}) = \frac{\pi^2}{2}.\end{aligned}$$

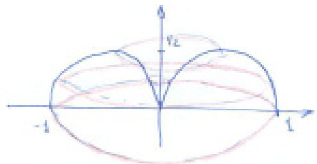


Figure 9.5.10

This surface is the top half of a doughnut whose hole has vanished, more like the top half of a modern bagel. See Figure 9.5.10. ◇

Summary

This section developed a definite integral for the area of a surface of revolution. It rests on the use of $2\pi R\, ds$ as a local estimate of the area swept out by a short segment of length ds revolved around a line L at a distance R from the segment. (See Figure 9.5.11.) We gave an informal argument for it. Exercises 28 and 29 offer a more formal treatment.

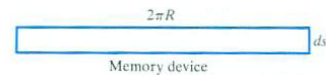


Figure 9.5.11

EXERCISES for Section 9.5

In Exercises 1 to 4 set up a definite integral for the area of the surface using the suggested parameter. Show the radius R on a diagram. Do not evaluate the definite integrals.

1. The graph of $y = x^3$, x in the interval $[1, 2]$, revolved about the x-axis with parameter x.

2. The graph of $y = x^3$, x in the interval $[1, 2]$, revolved about the line $y = -1$ with parameter x.

3. The graph of $y = x^3$, x in the interval $[1, 2]$, revolved about the y-axis with parameter y.

4. The graph of $y = x^3$, x in the interval $[1, 2]$, revolved about the y-axis with parameter x.

5. Find the area of the surface obtained by rotating about the x-axis that part of the curve $y = e^x$ that lies above $[0, 1]$.

6. Find the area of the surface formed by rotating one arch of the curve $y = \sin(x)$ about the x-axis.

7. One arch of the cycloid given parametrically by $x = \theta - \sin(\theta)$, $y = 1 - \cos(\theta)$ is revolved around the x-axis. Find the area of the surface produced.

8. The curve given parametrically by $x = e^t \cos(t)$, $y = e^t \sin(t)$ $(0 \le t \le \pi/2)$ is revolved around the x-axis. Find the area of the surface produced.

In Exercises 9 to 16 find the area of the surface formed by revolving the curve about an axis. Leave the answer as a definite integral, but indicate how it could be evaluated by the Fundamental Theorem of Calculus.

9. $y = 2x^3$ for x in $[0, 1]$, about the x-axis.

10. $y = 1/x$ for x in $[1, 2]$, about the x-axis.

11. $y = x^2$ for x in $[1, 2]$, about the x-axis.

12. $y = x^{4/3}$ for x in $[1, 8]$, about the y-axis.

13. $y = x^{2/3}$ for x in $[1, 8]$, about the line $y = 1$.

14. $y = x^3/6 + 1/(2x)$ for x in $[1, 3]$, about the y-axis.

15. $y = x^3/3 + 1/(4x)$ for x in $[1, 2]$, about the line $y = -1$.

16. $y = \sqrt{1 - x^2}$ for x in $[-1, 1]$, about the line $y = -1$.

Exercise 17 was solved by Archimedes more than 2300 years ago. He considered it his greatest accomplishment. About two centuries after Archimedes' death, Cicero wrote

> I shall call up from the dust [the ancient equivalent of a blackboard] and his measuring-rod an obscure, insignificant person belonging to the same city [Syracuse], who lived many years after, Archimedes. When I was quaestor I tracked out his grave, which was unknown to the Syracusans (as they totally denied its existence), and found it enclosed all round and covered with brambles and thickets; for I remembered certain doggerel lines inscribed, as I had heard, upon his tomb, which stated that a sphere along with a cylinder had been set up on the top of his grave. Accordingly, after taking a good look around (for there are a great quantity of graves at the Agrigentine Gate), I noticed a small column rising a little above the bushes, on which there was the figure of a sphere and a cylinder. And so I at once said to the Syracusans (I had their leading men with me) that I believed it was the very thing of which I was in search. Slaves were sent in with sickles who cleared the ground of obstacles, and when a passage to the place was opened we approached the pedestal fronting us; the epigram was traceable with about half the lines legible, as the latter portion was worn away. [Cicero, *Tusculan Disputations*, vol. 23, translated by J. E. King, Loef Classical Library, Harvard Univeristy, Cambridge, 1950.]

Archimedes was killed by a Roman soldier in 212 B.C. Cicero was quaestor in 75 B.C.

17. Consider the smallest tin can that contains a given sphere. (The height and diameter of the tin can equal the diameter of the sphere.)

(a) Compare the volume of the sphere with the volume of the tin can.

(b) Compare the surface area of the sphere with the total surface area of the can.

See also Exercise 32.

18.

(a) Compute the area of the portion of a sphere of radius a that lies between two parallel planes at distances c and $c+h$ from the center of the sphere ($0 \le c \le c+h \le a$).

(b) The result in (a) depends only on h, not on c. What does this mean geometrically? (See Figure 9.5.12.)

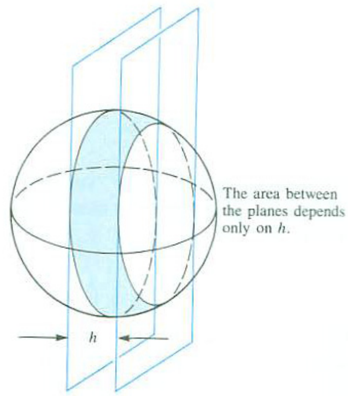


Figure 9.5.12

In Exercises 19 and 20 estimate the surface area obtained by revolving the arc about the given line. Find a definite integral for the surface area and then use either Simpson's method with six sections or a programmable calculator or computer to approximate the value of the integral.

19. $y = x^{1/4}$, x in $[1, 3]$, about the x-axis.

20. $y = x^{1/5}$, x in $[1, 3]$, about the line $y = -1$.

Exercises 21 to 24 are concerned with the area of a surface obtained by revolving a curve given in polar coordinates.

21. Show that the area of the surface obtained by revolving the curve $r = f(\theta)$, $\alpha \leq \theta \leq \beta$, around the polar axis is

$$\int_{\alpha}^{\beta} 2\pi r \sin\theta \sqrt{r^2 + (r')^2}\, d\theta.$$

(Use a local approximation informally.)

22. Use Exercise 21 to find the surface area of a sphere of radius a.

23. Find the area of the surface formed by revolving the portion of the curve $r = 1 + \cos(\theta)$ in the first quadrant about (a) the x-axis, (b) the y-axis. (The identity $1 + \cos(\theta) = 2\cos^2(\theta/2)$ may help in (b).)

24. The curve $r = \sin(2\theta)$, θ in $[0, \pi/2]$, is revolved around the polar axis. Set up an integral for the surface area.

25. The portion of the curve $x^{2/3} + y^{2/3} = 1$ in the first quadrant is revolved around the x-axis. Find the area of the surface produced.

26. Although the Fundamental Theorem of Calculus is of no use in computing the perimeter of the ellipse $x^2/a^2 + y^2/b^2 = 1$, it is useful in computing the surface area of the football-shaped surface formed when the ellipse is rotated about one of its axes.

(a) Assuming that $a > b$ and that the ellipse is revolved around the x-axis, find the area.

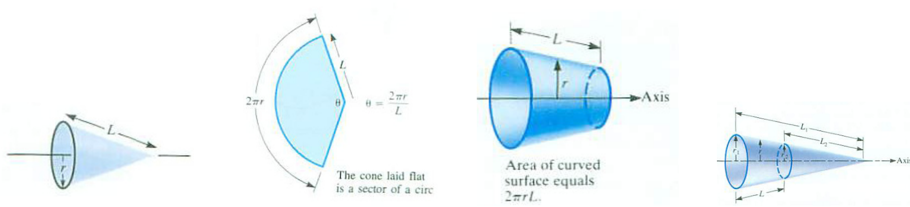
(b) Does your answer give the correct formula for the surface area of a sphere of radius a, $4\pi a^2$? (Let b approach a from the left.)

27. The unbounded region bounded by $y = 1/x$ and the x-axis and situated to the right of $x = 1$ is revolved around the x-axis to produce a solid region S.

(a) Show that the volume of S is finite but its surface area is infinite.

(b) Does this mean that an infinite surface area can be painted by pouring a finite amount of paint into the solid region it bounds?

Exercises 28 and 29 obtain the formula for the area of the surface obtained by revolving a line segment about a line that does not meet it. (The area was estimated in the text.)



(a) (b) (c) (d)

Figure 9.5.13

28. A right circular cone has slant height L and radius r, as shown in Figure 9.5.13(a). If it is cut along a line through its vertex and laid flat, it becomes a sector of a circle of radius L, as shown in Figure 9.5.13(b). By comparing Figure 9.5.13(b) to a complete disk of radius L find the area of the sector and thus the area of the cone.

29. Consider a line segment of length L in the plane that does not meet a line in the plane, called the axis. (See Figure 9.5.13(c).) When the segment is revolved around the axis, it sweeps out a curved surface. Show that its area equals $2\pi r L$ where r is the distance from the midpoint of the line segment to the axis. The surface in Figure 9.5.3 is called a **frustum of a cone**. Follow these steps:

(a) Complete the cone by extending the frustum as shown in Figure 9.5.13(d). Label the radii and lengths as in the figure. Show that $\frac{r_1}{r_2} = \frac{L_1}{L_2}$, so $r_1 L_2 = r_2 L_1$.

(b) Show that the surface area of the frustum is $\pi r_1 L_1 - \pi r_2 L_2$.

(c) Express L_1 as $L_2 + L$ and, using the result of (a), show that

$$\pi r_1 L_1 - \pi r_2 L_2 \quad = \quad \pi r_2 (L_1 - L_2) + \pi r_1 L = \pi r_2 L + \pi r_1 L.$$

(d) Show that the surface area of the frustum is $2\pi r L$, where $r = (r_1 + r_2)/2$. This justifies our approximation $2\pi R\ ds$.

30. The derivative with respect to r of the volume of a sphere is its surface area: $\frac{d}{dr}\left(4\pi r^3/3\right) = 4\pi r^2$. Is this a coincidence?

31. For some continuous functions $f(x)$ the definite integral $\int_a^b f(x)\ dx$ depends only on the width of the interval $[a, b]$. That is, there is a function $g(x)$ such that, for all a and b, $a < b$, in some interval,

$$\int_a^b f(x)\ dx = g(b - a). \tag{9.5.4}$$

(a) Show that a constant function $f(x)$ satisfies (9.5.4).

(b) Prove that if $f(x)$ satisfies (9.5.4), then it must be constant.

See Exercise 18.

32. The Mercator map discussed in the CIE of this chapter preserves angles. A **Lambert azimuthal equal-area projection** preserves areas, but not angles. It is made by projecting a sphere on a cylinder tangent at the equator by rays parallel to the equatorial plane and having one end on the diameter that joins the north and south poles, as shown in Figure 9.5.14.
Explain why a Lambert map preserves areas. (See Exercise 17.)

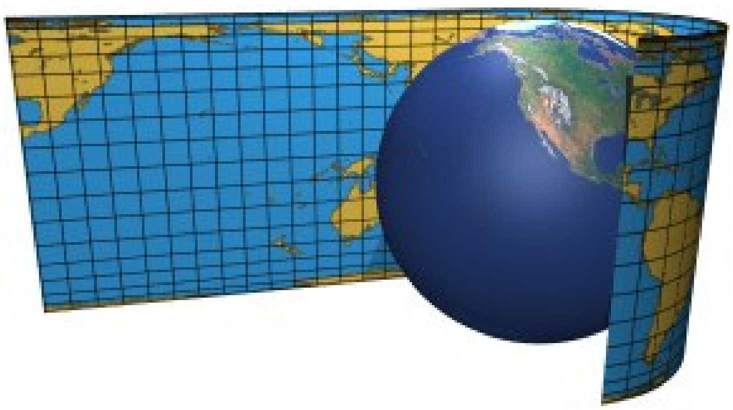


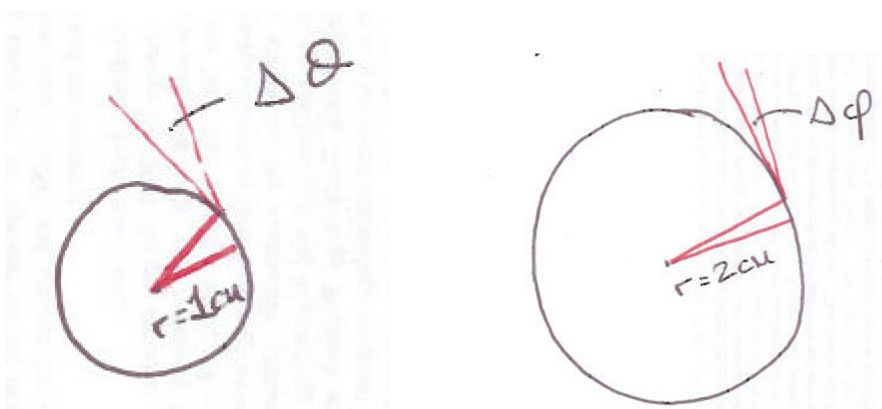
Figure 9.5.14

9.6 Curvature

In this section we use calculus to obtain a measure of the curviness or curvature of a curve. This concept will be generalized in Section 15.2 in the study of motion along a curved path in space.

Introduction

Imagine a bug crawling around a circle of radius one centimeter, as in Figure 9.6.1(a). As it walks a small distance, say 0.5 cm, it notices that its direction, measured by angle θ, changes. Another bug walks around a larger circle, as in Figure 9.6.1(b). Whenever it goes 0.5 cm, its direction, measured by ϕ, changes less. The first bug feels that his circle is curvier than the circle of the second bug. We will provide a measure of **curvature**. A straight line will have zero curvature everywhere. A circle of radius a will turn out to have curvature $1/a$ everywhere. For other curves, the curvature varies from point to point.



(a) (b)

Figure 9.6.1 The circle in (b) has twice the radius as the circle in (a). But the change in $\Delta\phi$ in (b) is half that in (a).

Definition of Curvature

Curvature measures how rapidly the direction changes as we move a small distance along a curve. We have a way of assigning a numerical value to direction, namely, the angle of the tangent line. The rate of change of this angle with respect to arc length will be our measure of curvature.

DEFINITION *(Curvature)* Assume that a curve is given parametrically, with the parameter of P being s, the distance along the curve from a fixed P_0 to P. Let ϕ be the angle between the tangent line at P and the positive part of the x-axis. The **curvature** κ at P is the absolute value of the derivative, $\frac{d\phi}{ds}$, whenever the derivative exists. (See Figure 9.6.2.)

κ is the Greek letter "kappa".

$$\text{curvature} = \kappa = \left|\frac{d\phi}{ds}\right|$$

A straight line has zero curvature everywhere, since ϕ is constant.

The next theorem shows that curvature of a small circle is large and the curvature of a large circle is small, in agreement with the bugs' experience.

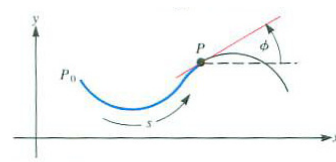


Figure 9.6.2

Theorem 9.6.1. *(Curvature of Circles) For a circle of radius a, the curvature $\left|\frac{d\phi}{ds}\right|$ is constant and equals $1/a$, the reciprocal of the radius.*

Proof

We needy to express ϕ as a function of arc length s on a circle of radius a. In Figure 9.6.3 measure arc length s counterclockwise from the point P_0 on the x-axis. Then $\phi = \frac{\pi}{2} + \theta$, as Figure 9.6.3 shows. By definition of radian measure, $s = a\theta$, so that $\theta = s/a$. So $\phi = \frac{\pi}{2} + \frac{s}{a}$. Differentiating with respect to arc length yields:

$$\frac{d\phi}{ds} = \frac{1}{a},$$

as claimed. •

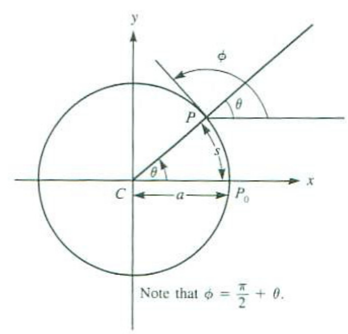


Figure 9.6.3

Computing Curvature

When a curve is given in the form $y = f(x)$, its curvature can be expressed in terms of the first and second derivatives, $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$.

Theorem 9.6.2. *(Curvature of $y = f(x)$) Let arc length s be measured along the curve $y = f(x)$ from a point P_0. Assume that x increases as s increases and that y' and y'' are continuous. Then*

$$\text{curvature} = \kappa = \frac{\left|\frac{d^2y}{dx^2}\right|}{\left(1+\left(\frac{dy}{dx}\right)^2\right)^{3/2}}.$$

Proof

The Chain Rule, $\frac{d\phi}{dx} = \frac{d\phi}{ds}\frac{ds}{dx}$, implies

$$\frac{d\phi}{ds} = \frac{\frac{d\phi}{dx}}{\frac{ds}{dx}}.$$

As was shown in Section 9.3,

$$\frac{ds}{dx} = \left(1+\left(\frac{dy}{dx}\right)^2\right)^{1/2}.$$

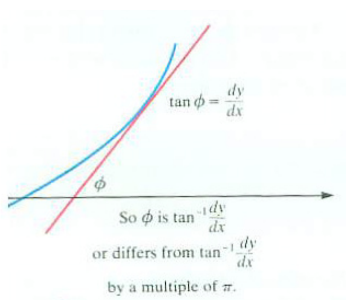


Figure 9.6.4

All that remains is to express $\frac{d\phi}{dx}$ in terms of $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$. In Figure 9.6.4,

$$\frac{dy}{dx} = \text{slope of tangent line to the curve} = \tan(\phi).$$

We find $\frac{d\phi}{dx}$ by differentiating both sides of $\frac{dy}{dx} = \tan(\phi)$ with respect to x. Thus

$$\frac{d^2y}{dx^2} = \frac{d}{dx}(\tan(\phi)) = \sec^2(\phi)\cdot\frac{d\phi}{dx} = \left(1+\tan^2(\phi)\right)\frac{d\phi}{dx} = \left(1+\left(\frac{dy}{dx}\right)^2\right)\frac{d\phi}{dx}.$$

Solving for $d\phi/dx$, we get

$$\frac{d\phi}{dx} = \frac{\frac{d^2y}{dx^2}}{1+\left(\frac{dy}{dx}\right)^2}.$$

Consequently,

$$\frac{d\phi}{ds} = \frac{\frac{d\phi}{dx}}{\frac{ds}{dx}} = \frac{\frac{d^2y}{dx^2}}{\left(1+\left(\frac{dy}{dx}\right)^2\right)\sqrt{1+\left(\frac{dy}{dx}\right)^2}} = \frac{\frac{d^2y}{dx^2}}{\left(1+\left(\frac{dy}{dx}\right)^2\right)^{3/2}},$$

and the theorem is proved. •

Geometry of curvature

One might have expected the curvature to depend only on the second derivative, $\frac{d^2y}{dx^2}$, since it measures the rate at which the slope changes. This is correct only when $\frac{dy}{dx} = 0$, that is, at critical points in the graph of $y = f(x)$. (See also Exercise 28.)

EXAMPLE 1 Find the curvature of $y = x^2$.
SOLUTION We have $\frac{dy}{dx} = 2x$ and $\frac{d^2y}{dx^2} = 2$ so

$$\kappa = \frac{\left|\frac{d^2y}{dx^2}\right|}{\left(1 + \left(\frac{dy}{dx}\right)^2\right)^{3/2}} = \frac{2}{(1 + (2x)^2)^{3/2}}.$$

The maximum curvature occurs when $x = 0$. The curvatures at (x, x^2) and at $(-x, x^2)$ are equal. As $|x|$ increases, the curve becomes straighter and the curvature approaches 0. (See Figure 9.6.5.) ◇

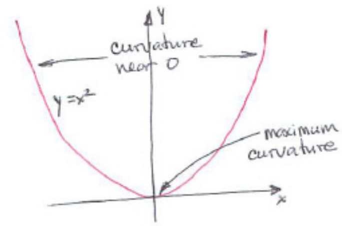


Figure 9.6.5

Curvature of a Parameterized Curve

Theorem 9.6.2 tells how to find the curvature if y is given as a function of x. It holds also when the curve is described parametrically, where x and y are functions of a parameter. Use

Theorem 9.6.2 applies also to curves given parametrically.

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} \qquad \text{and} \qquad \frac{d^2y}{dx^2} = \frac{\frac{d}{dt}\left(\frac{dy}{dx}\right)}{\frac{dx}{dt}}. \tag{9.6.1}$$

As mentioned in Section 9.3, both equations in (9.6.1) are consequences of the chain rule.

EXAMPLE 2 The cycloid determined by a wheel of radius 1 has the parametric equations

$$x = \theta - \sin(\theta) \qquad \text{and} \qquad y = 1 - \cos(\theta),$$

as shown in Figure 9.6.6. Find its curvature.

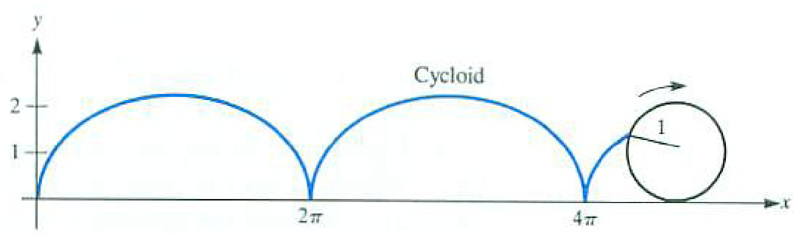


Figure 9.6.6

SOLUTION We find $\frac{dy}{dx}$ in terms of θ:

$$\frac{dx}{d\theta} = 1 - \cos(\theta) \qquad \text{and} \qquad \frac{dy}{d\theta} = \sin(\theta).$$

So

$$\frac{dy}{dx} = \frac{\sin(\theta)}{1-\cos(\theta)}.$$

Similar calculations show that

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{\frac{d}{d\theta}\left(\frac{dy}{dx}\right)}{\frac{dx}{d\theta}} = \frac{\frac{d}{d\theta}\left(\frac{\sin(\theta)}{1-\cos(\theta)}\right)}{1-\cos(\theta)} = \frac{-1}{(1-\cos(\theta))^2}.$$

Thus the curvature is

$$\kappa = \frac{\left|\frac{d^2y}{dx^2}\right|}{\left(1+\left(\frac{dy}{dx}\right)^2\right)^{3/2}} = \frac{\left|\frac{-1}{(1-\cos(\theta))^2}\right|}{\left(\frac{2}{1-\cos(\theta)}\right)^{3/2}} = \frac{1}{2^{3/2}\sqrt{1-\cos(\theta)}}.$$

Since $y = 1 - \cos(\theta)$ and $2^{3/2} = \sqrt{8}$, the curvature equals $1/\sqrt{8y}$. ◇

Radius of Curvature

As Theorem 9.6.1 shows, a circle with curvature κ has radius $1/\kappa$. This suggests the definition

A large radius of curvature implies a small curvature.

DEFINITION *(Radius of Curvature)* The **radius of curvature** of a curve at a point is the reciprocal of the curvature there:

$$\text{radius of curvature} = \frac{1}{\text{curvature}} = \frac{1}{\kappa}.$$

As can be checked, the radius of curvature of a circle of radius a is a.

The cycloid in Example 2 has radius of curvature at the point (x, y) equal to $\sqrt{8y}$. The higher the point on the curve, the straighter the curve. And the cycloid is nearly vertical at points near the x-axis. See Exercise 29.

The Osculating Circle

The line through a point P on a curve that looks most like the curve near P is the tangent line. The circle through P that looks most like the curve near P has the same slope at P as the curve and a radius equal to the radius of curvature at P. It is called the **osculating circle**, from the Latin osculum, meaning kiss.

At a given point P on a curve, the osculating circle at P is defined to be that circle that passes through P, has the same slope at P as the curve does, and has the same curvature there.

For the parabola $y = x^2$ of Example 1, when $x = 1$, the curvature is $2/5^{3/2}$ and the radius of curvature is $5^{3/2}/2 \approx 5.5902$. The osculating circle at $(1, 1)$ is shown in Figure 9.6.7. It crosses the parabola as it passes through the point $(1, 1)$. Although this may seem surprising, a little reflection will show why it is to be expected.

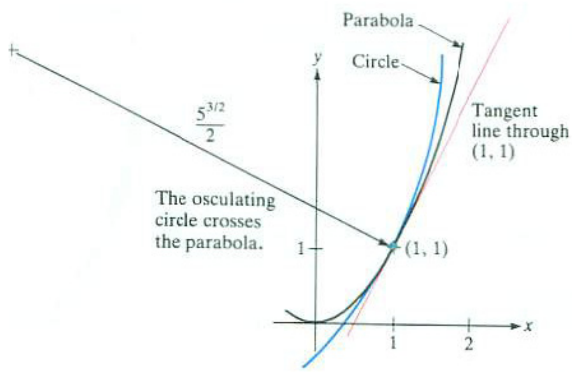


Figure 9.6.7

Think of driving along the parabola $y = x^2$. If you start at $(1, 1)$ and drive up along the parabola, the curvature diminishes. It is smaller than that of the circle of curvature at $(1, 1)$. The curve would be staighter than the osculating circle at $(1, 1)$ and you would be outside that circle. If you start at $(1, 1)$ and move toward the origin (to the left) on the parabola, the curvature increases and is greater than that of the osculating circle at $(1, 1)$, so you would be driving inside the osculating circle at $(1, 1)$. This shows why the osculating circle crosses the curve. The only osculating circle that does not cross the curve $y = x^2$ at its point of tangency is the one that is tangent at $(0, 0)$, where the curvature is a maximum.

Summary

The curvature κ of a curve was defined as the absolute value of the rate at which the angle between the tangent line and the x-axis changes as a function of arc length: curvature equals $\left|\frac{d\phi}{ds}\right|$. If the curve is the graph of $y = f(x)$,

then

$$\kappa = \frac{\left|\frac{d^2y}{dx^2}\right|}{\left(1 + \left(\frac{dy}{dx}\right)^2\right)^{3/2}}.$$

If the curve is given in terms of a parameter t then compute $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$ with the aid of this version of the Chain Rule,

$$\frac{d(\ )}{dx} = \frac{\frac{d(\ )}{dt}}{\frac{dx}{dt}}, \tag{9.6.2}$$

the empty parentheses enclosing first y, then $\frac{dy}{dx}$.

Radius of curvature is the reciprocal of curvature.

EXERCISES for Section 9.6

In Exercises 1 to 6 find the curvature and radius of curvature of the curve at the point.

1. $y = x^2$ at $(1, 1)$

2. $y = \cos(x)$ at $(0, 1)$

3. $y = e^{-x}$ at $(1, 1/e)$

4. $y = \ln(x)$ at $(e, 1)$

5. $y = \tan(x)$ at $(\frac{\pi}{4}, 1)$

6. $y = \sec(2x)$ at $(\frac{\pi}{6}, 2)$

In Exercises 7 to 10 find the curvature of the curve for the value of the parameter.

7. $x = 2\cos(3t)$, $y = 2\sin(3t)$ at $t = 0$

8. $x = 1 + t^2$, $y = t^3 + t^4$ at $t = 2$

9. $x = e^{-t}\cos(t)$, $y = e^{-t}\sin(t)$ at $t = \frac{\pi}{6}$

10. $x = \cos^3(\theta)$, $y = \sin^3(\theta)$ at $\theta = \frac{\pi}{3}$

11.

(a) Compute the curvature and radius of curvature for the curve $y = (e^x + e^{-x})/2$.

(b) Show that the radius of curvature at (x, y) is y^2.

12. Find the radius of curvature along the curve $y = \sqrt{a^2 - x^2}$, where a is a constant. (Since the curve is part of a circle of radius a, the answer should be a.)

13. For what value of x is the radius of curvature of $y = e^x$ smallest?
Hint: How does one find the minimum of a function?

14. For what value of x is the radius of curvature of $y = x^2$ smallest?

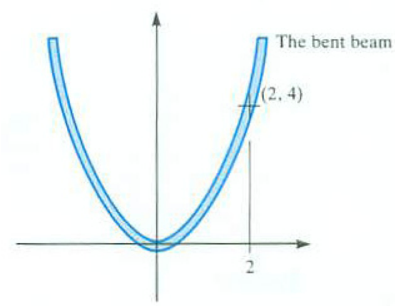
15.

(a) Show that at a point where a curve has its tangent parallel to the x-axis its curvature is the absolute value of the second derivative d^2y/dx^2.

(b) Show that the curvature is never larger than the absolute value of d^2y/dx^2.

16. An engineer lays out a railroad track as indicated in Figure 9.6.8(a). BC is part of a circle and AB and CD are straight and tangent to the circle. After a train runs over the track, the engineer is fired because the curvature is not a continuous function. Why should the curvature be continuous?

(a)



(b)

Figure 9.6.8

17. Railroad curves are banked to reduce wear on the rails and flanges. The greater the radius of curvature, the less the curve must be banked. The best bank angle A satisfies $\tan(A) = v^2/(32R)$, where v is speed in feet per second and R is radius of curvature in feet. A train travels in the elliptical track

$$\frac{x^2}{1000^2} + \frac{y^2}{500^2} = 1$$

at 60 miles per hour. Find the best angle A at $(1000, 0)$ and $(0, 500)$. x and y are measured in feet and 60 mph = 88 fps.

18. The flexure formula in the theory of beams asserts that the bending moment M required to bend a beam is proportional to the curvature, $M = c/R$, where c is a constant depending on the beam and R is the radius of curvature. A beam is bent to form the parabola $y = x^2$. What is the ratio between the moments required at (a) $(0, 0)$ and (b) $(2, 4)$? (See Figure 9.6.8(b).)

Exercises 19 to 21 are related.

19. Find the radius of curvature at a point on the curve whose parametric equations are

$$x = a\cos\theta, \qquad y = b\sin\theta.$$

20.

(a) Show, by eliminating θ, that the curve in Exercise 19 is the ellipse $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$.

(b) What is the radius of curvature of the ellipse at $(a, 0)$? at $(0, b)$?

21. An ellipse has major axis of length 6 and minor axis of length 4. Draw the circles that most closely approximate the ellipse at the four points that lie at the extremities of its axes. (See Exercises 19 and 20.)

In each of Exercises 22 to 24 a curve is given in polar coordinates. To find its curvature write it in rectangular coordinates with parameter θ, using $x = r\cos(\theta)$ and $y = r\sin(\theta)$.

22. Find the curvature of $r = a\cos(\theta)$.

23. Show that at (r, θ) the cardioid $r = 1 + \cos(\theta)$ has curvature $3\sqrt{2}/(4\sqrt{r})$.

24. Find the curvature of $r = \cos(2\theta)$.

25. If $dy/dx = y^3$, express the curvature in terms of y.

26. As is shown in physics, the larger the radius of curvature of a turn, the faster a car can travel around it. The required radius of curvature is proportional to the square of the maximum speed. That says that the maximum speed around a turn is proportional to the square root of the radius of curvature. If a car moving on the path $y = x^3$ (x and y measured in miles) can go 30 miles per hour at $(1, 1)$ without sliding off, how fast can it go at $(2, 8)$?

27. Find the local extrema of the curvature of (a) $y = x + e^x$, (b) $y = e^x$, (c) $y = \sin(x)$, and (d) $y = x^3$.

28. Sam says, "I don't like the definition of curvature. It should be the rate at which the slope changes as a function of x. That is $\frac{d}{dx}\left(\frac{dy}{dx}\right)$, which is the second derivative, $\frac{d^2y}{dx^2}$." Give an example of a curve that would have constant curvature according to Sam's definition but whose changing curvature is obvious to the eye.

29. In Example 2 some of the steps were omitted in finding that the cycloid given by $x = \theta - \sin(\theta)$, $y = 1 - \cos(\theta)$ has curvature $\kappa = 1/(2^{3/2}\sqrt{1 - \cos(\theta)}) = 1/\sqrt{8y}$. In this exercise you are asked to fill in the details.

(a) Verify that $\dfrac{dy}{dx} = \dfrac{\sin(\theta)}{1 - \cos(\theta)}$.

(b) Verify that $\dfrac{d}{d\theta}\left(\dfrac{dy}{dx}\right) = \dfrac{-1}{1 - \cos(\theta)}$.

(c) Verify that $\dfrac{d^2y}{dx^2} = \dfrac{-1}{(1 - \cos(\theta))^2}$.

(d) Verify that $1 + \left(\dfrac{dy}{dx}\right)^2 = \dfrac{2}{1 - \cos(\theta)}$.

(e) Compute the curvature, κ, in terms of θ.

(f) Express it in terms of x and y.

(g) At what points on the cycloid is the curvature largest?

(h) At what points on the cycloid is the curvature smallest?

30. Assume that g and h are functions with continuous second derivatives. and that as we move along the curve $x = g(t)$, $y = h(t)$, the arc length s from a point P_0 increases as t increases. Show that

$$\kappa = \frac{|\dot{x}\ddot{y} - \dot{y}\ddot{x}|}{(\dot{x}^2 + \dot{y}^2)^{3/2}}.$$

Newton's dot notation for derivatives shortens the formula: $\dot{x} = \frac{dx}{dt}$, $\ddot{x} = \frac{d^2x}{dt^2}$, $\dot{y} = \frac{dy}{dt}$, and $\ddot{y} = \frac{d^2y}{dt^2}$.

31. Use the result of Exercise 30 to find the curvature of the cycloid of Example 2, which has parametric equations $x = \theta - \sin(\theta)$, $y = 1 - \cos(\theta)$

32. (Contributed by G.D. Chakerian) Parts (a), (b), and (c) refer to curves in general, where R may not be constant. Part (d) treats the special case, where R is constant.
If a planar curve has a constant radius of curvature must it be part of a circle? That the answer is "yes" is important in experiments conducted with a cyclotron. Physical assumptions imply that the path of an electron entering a uniform magnetic field at right angles to the field has constant curvature. Show that it follows that the path is part of a circle. View the curve as parameterized by the angle ϕ of the tangent. Here s denotes arc length.

(a) Show that $\frac{ds}{d\phi} = R$, the radius of curvature.

(b) Show that $\frac{dx}{d\phi} = R\cos(\phi)$.

(c) Show that $\frac{dy}{d\phi} = R\sin(\phi)$.

(d) Now assume the curvature is constant. Use (b) and (c) to find an equation that involves x, y, and κ.

(For (b) and (c) draw the little triangle whose hypotenuse is like a short piece of arc length ds on the curve and whose legs are parallel to the axes. For (d), think about antiderivatives.)

33. At the top of the cycloid in Example 2 the radius of curvature is twice the diameter of the rolling circle. What would you have guessed the radius of curvature to be at this point? Why is it not the diameter of the wheel, since the wheel at each moment is rotating about its point of contact with the ground?

34. A smooth convex curve with no straight edges bounds a region. Assume it has length L and continuous curvature.

(a) Show that the average of its curvature, as a function of arc length, is $2\pi/L$.

(b) Check that this is correct for a circle of radius a.

(c) Must there be a point on the curve where the curvature is $2\pi/L$?

9.S Chapter Summary

This chapter dealt mostly with curves described in polar coordinates and curves given parametrically. Table 9.S.1 is a list of reminders for most of the ideas in the chapter.

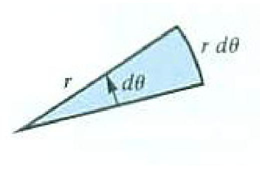
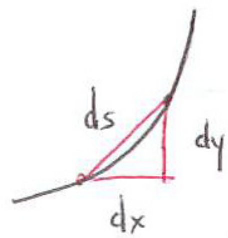
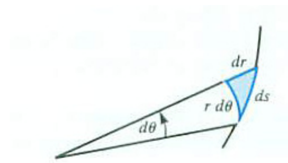
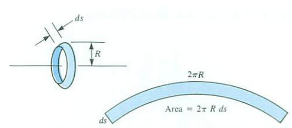
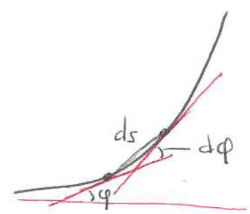
Concept	*Memory Aid*	*Comme*
Area $= \int_\alpha^\beta \frac{r^2}{2}\, d\theta$		The narrow sector res gle of base $r\ d\theta$ and h $\frac{1}{2}(r\ d\theta)(r) = \frac{1}{2}r^2\ d\theta$.
Arc length $= \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\ dt$ $= \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\ dx$		A short part of the straight, suggesting ($(dy)^2$.
Arc length $= \int_\alpha^\beta \sqrt{r^2 + (r')^2}\ d\theta$ $= \int_\alpha^\beta \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2}\ d\theta$ Speed $= \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}$ $= \sqrt{\left(r\frac{d\theta}{dt}\right)^2 + \left(\frac{dr}{dt}\right)^2}$		The shaded area with looks like a right tria $(ds)^2 = (rd\theta)^2 + (dr)^2$.
Area of surface of revolution $= \int_a^b 2\pi R\ ds$		
Curvature $= \kappa = \left\lvert\frac{d\phi}{ds}\right\rvert$		Using the chain rule $\left\lvert\frac{(d\phi/dx)}{(ds/dx)}\right\rvert$ one gets the $\frac{\lvert y''\rvert}{(1+(y')^2)^{3/2}}$

Table 9.S.1 Summary of key points for the concepts in Chapter 9.

For a curve given parametrically, its curvature can be found by replacing $\frac{dy}{dx}$ by $\frac{dy/dt}{dx/dt}$, and, similarly, $\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right)$ by $\frac{\frac{d}{dt}\left(\frac{dy}{dx}\right)}{dx/dt}$.

EXERCISES for 9.S

1. When driving along a curved road, which is more important in avoiding car sickness, $d\phi/ds$ or $d\phi/dt$, where t is time.

2. The solution to Example 3 (Section 9.2) requires the evaluation of the definite integrals $\int_0^{\pi/2} \cos^2(\theta)\, d\theta$ and $\int_0^{\pi} (1+\cos(\theta))^2\, d\theta$. Evaluate them as simply as possible.

3. A physics midterm includes the following information:

When $r = \sqrt{x^2+y^2}$, and y is a constant, recall that

$$\text{(a)} \int \frac{dx}{r} = \ln(x+r), \qquad \text{(b)} \int \frac{x}{r}\, dx = r, \qquad \text{(c)} \int \frac{dx}{r^3} = \frac{x}{y^2 r}.$$

Show by differentiating that the equations are correct.

The inequality

$$\int\limits_0^{2\pi} f(\theta)g(\theta)\, d\theta \le \left(\int\limits_0^{2\pi} f(\theta)^2\, d\theta \right)^{1/2} \left(\int\limits_0^{2\pi} g(\theta)^2\, d\theta \right)^{1/2}$$

is a special case of the **Cauchy-Schwarz inequality** used in a CIE at the end of Chapter 7 (see page 679) and proved in Exercise 29 in Section 16.7. It will be of use in Exercises 4 and 5.

4. Let P be a point inside a region in the plane bounded by a smooth convex curve. ("Smooth" means it has a continuously defined tangent line.) Place the pole of a polar coordinate system at P. Let $c(\theta)$ be the length of the chord of angle θ through P. Show that $\int_0^{2\pi} c(\theta)^2\, d\theta \le 8A$, where A is the area of the region.

5. (This continues Exercise 4.) Show that if $\int_0^{2\pi} c(\theta)^2\, d\theta = 8A$ then P is the midpoint of each chord through P.

6. Let L be the line $3x + 4y = 1$. Consider the function that assigns to the point with polar coordinates (r, θ), r not equal to 0, the point $(1/r, \theta)$.

(a) Plot L and at least four images of points on L.

(b) Sketch what you suspect is the image of L.

(c) Find the equation, in rectangular coordinates, of the image of L. (Using polar coordinates may help.)

(d) What kind of curve is the image of L?

7. Let $r = f(\theta)$ describe a convex curve surrounding the origin.

(a) Show that $\int_0^{2\pi} r \, d\theta \leq$ arc length of the boundary.

(b) Show that if the equality holds in (a), the curve is a circle with center at the origin.

8.

Sam: I've discovered an easy formula for the length of a closed curve that encloses the origin.

Jane: Well?

Sam: First of all, $\int_0^{2\pi} \sqrt{r^2 + (r')^2} \, d\theta$ is obviously greater than or equal to $\int_0^{2\pi} r \, d\theta$.

Jane: I'll grant you that much, because $(r')^2$ is never negative.

Sam: Now, if a and b are not negative, $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$.

Jane: Why?

Sam: Just square both sides. So $\sqrt{r^2 + (r')^2} \leq \sqrt{r^2} + \sqrt{(r')^2} = r + r'$.

Jane: Looks all right.

Sam: Thus

$$\int_0^{2\pi} \sqrt{r^2 + (r')^2} \, d\theta \leq \int_0^{2\pi} (r + r') \, d\theta = \int_0^{2\pi} r \, d\theta + \int_0^{2\pi} r' \, d\theta.$$

But $\int_0^{2\pi} r' \, d\theta$ equals $r(2\pi) - r(0)$, which is 0. So, putting it all together, I get

$$\int_0^{2\pi} r \, d\theta \leq \int_0^{2\pi} \sqrt{r^2 + (r')^2} \, d\theta \leq \int_0^{2\pi} r \, d\theta.$$

So the arc length is $\int_0^{2\pi} r \, d\theta$.

Jane: That can't be right. If it were, it would be an Exercise.

Sam: They like to keep a few things secret to surprise us on a mid-term.

Who is right, Sam or Jane? Explain.

9.

(a) Let the radius of Earth be r miles. What fraction of the Earth's surface can be seen from an object h miles above it?

(b) What is $\lim_{h\to\infty} f(h)$?

(c) What is $f(h)$ when h is 100 miles? Earth's radius is 3960 miles.

(d) What is $f(h)$ when h is 22,240 miles, the altitude of a geosynchronous satellite?

In Exercises 10 and 11, a, b, c, m, and p are constants. Verify that the derivative of the first function is the second function.

10. $\frac{1}{\sqrt{c}} \arcsin\left(\frac{cx-b}{\sqrt{b^2+ac}}\right)$, $\sqrt{\frac{c}{a+2bx-cx^2}}$.

11. $\frac{1}{c}\sqrt{a+2bx+cx^2} - \frac{b}{\sqrt{c}} \ln\left(b+cx+\sqrt{c}\sqrt{a+2bx+cx^2}\right)$, $\frac{x}{a+bx+cx^2}$ (assume c is positive).

In Exercises 12 and 13, L is the length of a smooth, closed curve C and P is a point in the region R bounded by C.

12.

(a) Let the area of R be A. Show that the average distance from P to points on the curve, averaged with respect to arc length, is greater than or equal to $2A/L$.

(b) Give an example when equality holds.

13.

(a) Show that the average distance from P to points on the curve, averaged with respect to the polar angle, is less than or equal to $L/(2\pi)$.

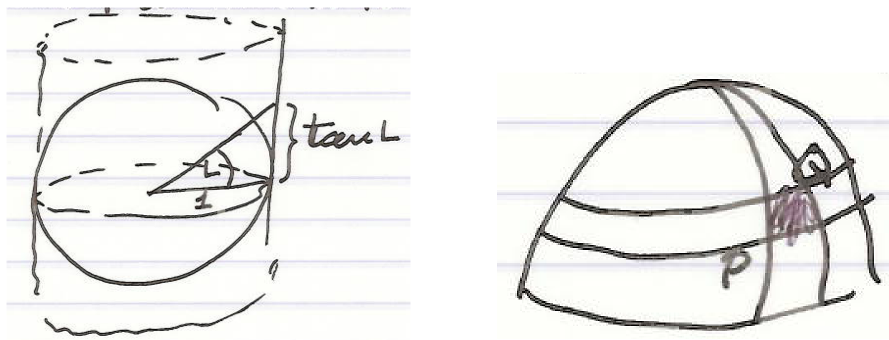
(b) Give an example when equality holds.

(See Exercise 25 in Section 9.4.)

Calculus is Everywhere # 12
The Mercator Map

A web search for "map projection" leads to detailed information about this and other projections. The US Geological Society has some particularly good resources.

One way to make a map of a sphere is to wrap a paper cylinder around the sphere and project points on the sphere onto the cylinder by rays from the center of the sphere. This **central cylindrical projection** is illustrated in Figure C.12.1(a).



(a) (b)

Figure C.12.1

Points at latitude L project onto points at height $\tan(L)$ from the equatorial plane.

A **meridian** is a great circle passing through the north and south poles. It corresponds to a fixed longitude. A short segment on a meridian at latitude L of length dL projects onto the cylinder in a segment of length approximately $d(\tan(L)) = \sec(L)^2\, dL$. This tells us that the map magnifies short vertical segments at latitude L by the factor $\sec^2(L)$.

Points on the sphere at latitude L form a circle of radius $\cos(L)$. Its image on the cylinder is a circle of radius 1, so the projection magnifies horizontal distances at latitude L by $1/\cos(L) = \sec(L)$.

Consider the effect on a small "almost rectangular" patch bordered by two meridians and two latitude lines, shaded in Figure C.12.1(b). The vertical edges are magnified by $\sec^2(L)$, but the horizontal edges by only $\sec(L)$. The image on the cylinder will not resemble the patch, for it is stretched more vertically than horizontally.

The path a ship sailing from P to Q makes an angle with the latitude line through P. The map distorts it. The ship's captain would prefer a map without distortion, one that preserves direction. That way, to chart a voyage from A to B on the sphere corresponding to Earth's surface at a fixed compass heading, simply draw a straight line from A to B on the map to determine the compass setting.

In 1569, Gerhardus Mercator designed a map that preserves direction by making the vertical magnification the same as the horizontal distortion, $\sec(L)$.

Let y be the height on the map that represents latitude L_0. Then Δy should be approximately $\sec(L)\Delta L$. Taking the limit of $\Delta y/\Delta L$ as ΔL approaches 0, we see that $dy/dL = \sec(L)$. Thus

$$y = \int_0^{L_0} \sec(L)\ dL. \qquad \text{(C.12.1)}$$

Mercator, working a century before the invention of calculus, did not have the integral or the Fundamental Theorem of Calculus. Instead, he broke the interval $[0, L_0]$ into several short sections of length ΔL, computed $(\sec(L))\Delta L$ for each, and summed to estimate y in (C.12.1).

Using calculus, we see

$$y = \int_0^{L_0} \sec(L)\, dL = \ln|\sec(L) + \tan(L)|\ \Big|_0^{L_0} = \ln(\sec(L_0)+\tan(L_0)) \qquad \text{for } 0 \le L_0 \le \pi/2.$$

In 1645, Henry Bond conjectured from numerical evidence that $\int_0^\alpha \sec(\theta)\, d\theta = \ln(\tan(\alpha/2 + \pi/4))$ but offered no proof. In 1666, Nicolaus Mercator (no relation to Gerhardus) offered the royalties on one of his inventions to the mathematician who could prove Bond's conjecture was right. Within two years James Gregory provided the proof.

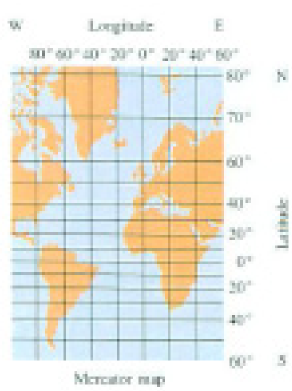


Figure C.12.2

Figure C.12.2 shows a Mercator map. Though it preserves angles, it greatly distorts areas: Greenland looks bigger than South America even though it is only one eighth its size. The first map we described distorts areas even more than does a Mercator map.

EXERCISES

1. Draw a clear diagram showing why segments at latitude L are magnified vertically by the factor $\sec(L)$.

2. Explain why a short arc of length dL in Figure C.12.1(a) projects onto a short interval of length approximately $\sec^2(L)\ dL$.

3. On a Mercator map, what is the ratio of the distance between the lines representing latitudes 60° and 50° to the distance between the lines representing latitudes 40° and 30°?

4. What magnifying effect does a Mercator map have on areas?

5. If the distance on a Mercator map is 3 inches from latitude 0° to latitude 20° how far is it on the map from (a) 60° to 80°, (b) 75° to 85°.

6. Bond's conjecture was first encountered in Exercise 50 in Section 8.5. Show that it is correct. That is, that $\int_0^{\alpha} \sec(\theta)\ d\theta = \ln(\tan(\alpha/2 + \pi/4))$

Chapter 10

Sequences and Their Applications

When trying to write $1/3$ as a decimal, we meet the following sequence of numbers:

$$0.3, 0.33, 0.333, 0.3333, \ldots$$

The more 3s we write, the closer the numbers are to $1/3$.

When estimating a definite integral $\int_a^b f(x)\,dx$, we pick a positive integer n, divide the interval $[a, b]$ into n equal pieces of length $\Delta x = (b-a)/n$, pick a number c_i in the i^{th} interval and form the sum $E_n = \sum_{i=1}^{n} f(c_i)\Delta x$. In this way we obtain a sequence of estimates,

$$E_1, E_2, E_3, \ldots, E_n, \ldots.$$

As n increases the estimates approach $\int_a^b f(x)\,dx$, if $f(x)$ is continuous.

In the analysis of APY (annual percentage yield on an account at a bank), in CIE #3 in Chapter 2 (page 159) we encounter the sequence

$$(1+1/1)^1, (1+1/2)^2, (1+1/3)^3, \ldots, (1+1/n)^n, \ldots.$$

As n increases, these numbers approach e.

What happens to the numbers

$$S_1 = 1, S_2 = 1 + \frac{1}{2}, S_3 = 1 + \frac{1}{2} + \frac{1}{3}, S_4 = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}, \ldots, S_n = \sum_{k=1}^{n} \frac{1}{k}, \ldots$$

as we add more and more reciprocals of integers? Do the S_n get arbitrarily large or do they approach some finite number? When students, neither author guessed right.

Chapters 10, 11, and 12 concern the behavior of endless sequences of numbers. Such sequences arise in estimating a solution of an equation. They also provide a way to estimate such important functions as e^x, $\sin(x)$, and $\ln(x)$, and therefore a way to estimate such integrals as $\int_0^1 e^{x^2}\,dx$, for which the fundamental theorem of calculus is of no help. They also offer another way to evaluate indeterminate limits.

10.1 Introduction to Sequences

A **sequence** of numbers,

$$a_1, a_2, a_3, \ldots, a_n, \ldots$$

is a function that assigns to each positive integer n a number a_n. The number a_n is called the n^{th} **term** of the sequence. For example, the sequence

$$\left(1+\frac{1}{1}\right)^1, \left(1+\frac{1}{2}\right)^2, \left(1+\frac{1}{3}\right)^3, \ldots, \left(1+\frac{1}{n}\right)^n, \ldots$$

was first seen in Section 2.2 and was later shown to be related to the number e. In this case, the n^{th} term of the sequence is

$$a_n = \left(1+\frac{1}{n}\right)^n.$$

n	a_n
1	2.0000
2	2.2500
3	2.3704
4	2.4414
5	2.4883
10	2.5937
100	2.7048
1000	2.7169
10000	2.7181

For example, $a_1 = (1+1)^1 = 2$, $a_2 = \left(1+\frac{1}{2}\right)^2 = \frac{9}{4} = 2.25$, $a_{10} = \left(1+\frac{1}{10}\right)^{10} \approx 2.5937$, and $a_{100} = \left(1+\frac{1}{100}\right)^{10} \approx 2.7048$.

The notation $\{a_n\}$ is an abbreviation for the sequence $a_1, a_2, \ldots, a_n, \ldots$.

Read a_1 as "a sub 1" and a_n as "a sub n."

The "sub" stands for "subscript."

If, as n gets larger, a_n approaches a number L, then L is called the **limit** of the sequence $\{a_n\}$. When the sequence $a_1, a_2, \ldots, a_n, \ldots$ has a limit L we say it is convergent and write

$$\lim_{n\to\infty} a_n = L.$$

For instance, we write

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e.$$

A sequence need not begin with the term a_1. Later, sequences of the form $a_0, a_1, a_2, \ldots$ will be considered. In such a case, a_0 is called the **zeroth term**. In other instances we consider sequences $a_k, a_{k+1}, a_{k+2}, \ldots$ that begin with a_k for $k > 1$. These sequences are called a "tail" of the sequence $a_1, a_2, a_3, \ldots$. Two important features of any sequence are i) the terms of a sequence are defined only for integers and ii) the sequence never ends.

The Sequence $\{r^n\}$

The next example introduces a simple but important type of sequence called a **geometric sequence**.

EXAMPLE 1 A certain (small) piece of equipment depreciates in value over the years. In fact, at the end of any year it has only 80% of the value it had at the beginning of the year. What happens to its value in the long run if its value when new is \$1?

SOLUTION Let a_n be the value of the equipment at the end of the n^{th} year. Call the initial value $a_0 = 1$. At the end of year 1 the value is $a_1 = 0.8(1)$. Similarly, $a_2 = 0.8(0.8) = 0.8^2 = 0.64$ and $a_3 = 0.8(0.8^2) = 0.8^3$. After n years the value is $a_n = 0.8^n$. This question is asking about the limit of the sequence $\{0.8^n\}$. After 5 years, the value is just under \$0.33. In another five years the value is reduced to about \$0.11, and at the end of year 20, the value is roughly \$0.01. This is strong evidence that

n	0.8^n
0	$0.8^0 = 1$
1	$0.8^1 = 0.8$
2	$0.8^2 = 0.64$
3	$0.8^3 = 0.512$
4	$0.8^4 = 0.4096$
5	$0.8^5 = 0.3277$
10	$0.8^{10} = 0.1074$
20	$0.8^{20} = 0.0115$

$$\lim_{n\to\infty} 0.8^n = 0.$$

$\diamond$

Even if the piece of equipment in Example 1 retained 99% of its value each year, in the long run it would still be worth less than a dime, then less than a penny, etc. The data in Table 10.1.1 indicates that 0.99^n approaches 0 as $n \to \infty$, but much more slowly than 0.8^n does.

n	0	1	2	3	4	5	10	20	100	200
0.99^n	1	0.99	0.9801	0.9703	0.9606	0.9510	0.9044	0.8179	0.3660	0.1340

Table 10.1.1

On the basis of Example 1, it is plausible that if $0 \leq r < 1$, then $\lim_{n\to\infty} r^n = 0$. To show that this is the case, we introduce a property of the real numbers which we will use often. It concerns **monotone sequences**. A sequence is **monotone** either if it is nondecreasing ($a_1 \leq a_2 \leq a_3 \leq \cdots \leq a_n \leq \dots$) or nonincreasing ($a_1 \geq a_2 \geq a_3 \geq \cdots \geq a_n \geq \dots$).

Every bounded and monotone sequence converges.

Theorem 10.1.1. *Let $\{a_n\}$ be a nondecreasing sequence with the property that there is a number B such that $a_n \leq B$ for all n. That is, $a_1 \leq a_2 \leq a_3 \leq a_4 \leq \cdots \leq a_n \leq a_{n+1} \leq \dots$ and $a_n \leq B$ for all n. Then the sequence $\{a_n\}$ is convergent and a_n approaches a number L less than or equal to B.*

Similarly, if $\{a_n\}$ is a nonincreasing sequence and there is a number B such that $a_n \geq B$ for all n, then the sequence $\{a_n\}$ is convergent and its limit is greater than or equal to B.

Figure 10.1.1 suggests the first part of Theorem 10.1.1 is plausible. The monotonicity prevents the terms from backtracking or entering a cycle. Without the bound on the terms, the sequence could continue to approach ∞. Any sequence that is both bounded and monotone has to converge to a limit.

Theorem 10.1.1 is proved in advanced calculus.

The next theorem shows the power of Theorem 10.1.1.

Figure 10.1.1

Theorem 10.1.2. *If* $0 < r < 1$ *then* $\{r^n\}$ *converges to* 0.

Proof

Let r be a number between 0 and 1. The sequence r^1, r^2, r^3, $\dots r^n$, $\dots$ is decreasing and each term is greater than 0. By Theorem 10.1.1, the sequence has a limit, L, and $L \geq 0$.

The sequence r^2, r^3, $\dots$, r^{n+1}, $\dots$ also approaches L. We then have

$$L = \lim_{n\to\infty} r^{n+1} = \lim_{n\to\infty} rr^n = r \lim_{n\to\infty} r^n = rL.$$

In short,

$$L = rL.$$

Thus $(1-r)L = 0$. So either $1-r=0$ or $L=0$. Because $0<r<1$, $1-r$ is not zero, L has to be 0, which shows that $\lim_{n\to\infty} r^n = 0$. •

The behavior of $\{r^n\}$ for other values of r is much more easily obtained:

1. If $r = 1$, then $r^n = 1$ for all n. So $\lim_{n\to\infty} r^n = 1$.
2. If $r > 1$, then r^n gets arbitrarily large as $n \to \infty$. Hence is divergent.
3. If $r < -1$, then $|r|^n$ gets arbitrarily large. Thus $\lim_{n\to\infty} r^n$ does not exist.
4. If $r = -1$, then the sequence is $-1, 1, -1, 1, \dots.$ which is divergent.
5. If $-1 < r < 0$, then $\lim_{n\to\infty} r^n = 0$.
6. If $r = 0$, then $r^n = 0$ for all n. So $\lim_{n\to\infty} r^n = 0$.

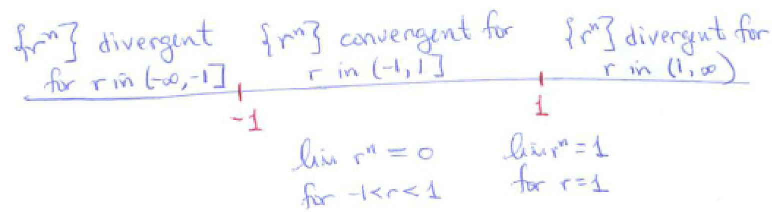


Figure 10.1.2

Figure 10.1.2 records this information.

We prove (2) and (5). First, (2). If $r > 1$, the sequence r, r^2, r^3, r^4, ..., r^n, ... is monotone increasing. The terms either approach a limit, L, or they get arbitrarily large. In the first case we would have, as before, $(1 - r)L = 0$, which implies $L = 0$ (because $1 - r$ is not zero). That's impossible since every term is greater than or equal to r.

To prove (5), let $-1 < r < 0$ and note that $|r^n| = |r|^n$ approaches zero as $n \to \infty$ (by Theorem 10.1.2). Since the absolute value of r^n approaches 0, so must r^n.

The terms of a convergent sequence usually never equal their limit, L, but merely get closer to it as the index, n, increases.

Informal definition of $\lim_{n\to\infty} a_n = \infty$.

If a_n becomes and remains arbitrarily large and positive as n gets larger, the sequence diverges and we write $\lim_{n\to\infty} a_n = \infty$. For instance, $\lim_{n\to\infty} 2^n = \infty$. Similarly, $\lim_{n\to\infty}(-2^n) = -\infty$. But, for $\lim_{n\to\infty}(-2)^n$ all we can say is that the sequence diverges because the values alternate between positive and negative values and $\lim_{n\to\infty} |2^n| = \lim_{n\to\infty} 2^n = \infty$.

The Sequence $\{k^n/n!\}$

Example 2 introduces a type of sequence that occurs in the study of $\sin(x)$, $\cos(x)$, and e^x in Chapter 12.

EXAMPLE 2 Does the sequence defined by $a_n = 3^n/n!$ converge or diverge?

SOLUTION The first terms of this sequence are recorded (to four decimal places) in Table 10.1.2. Although a_2 is larger than a_1 and a_3 is equal to a_2, from a_4 through a_8, as Table 10.1.2 shows, the terms decrease.

The numerator 3^n becomes large as $n \to \infty$, influencing a_n to grow large. But the denominator $n!$ also becomes large as $n \to \infty$, influencing the quotient a_n to shrink toward 0. For $n = 1$ and $n = 2$ the first influence dominates, but then, as the table shows, the denominator $n!$ grows faster than the numerator 3^n, forcing a_n toward 0.

n	1	2	3	4	5	6	7	8
3^n	3	9	27	81	243	729	2,187	6,561
$n!$	1	2	6	24	120	720	5,040	40,320
$a_n = 3^n/n!$	3.0000	4.5000	4.500	3.3750	2.0250	1.0125	0.4339	0.1627

Table 10.1.2

To see why the denominator grows so fast that the quotient $3^n/n!$ approaches 0, consider a_{10}. This term can be expressed as the product of 10 fractions:

$$a_{10} = \frac{3^{10}}{10!} = \frac{3}{1}\frac{3}{2}\frac{3}{3}\frac{3}{4}\frac{3}{5}\frac{3}{6}\frac{3}{7}\frac{3}{8}\frac{3}{9}\frac{3}{10}.$$

The first three fractions are greater than or equal to 1, but the seven remaining fractions are all less than or equal to $\frac{3}{4}$. Thus

$$a_{10} < \frac{3}{1}\frac{3}{2}\frac{3}{3}\left(\frac{3}{4}\right)^7.$$

Similarly,

$$a_{100} < \frac{3}{1}\frac{3}{2}\frac{3}{3}\left(\frac{3}{4}\right)^{97}.$$

More generally, for $n > 4$,

$$a_n < \frac{3}{1}\frac{3}{2}\frac{3}{3}\left(\frac{3}{4}\right)^{n-3}.$$

By Theorem 10.1.2,

$$\lim_{n\to\infty}\left(\frac{3}{4}\right)^n = 0,$$

from which it follows that $\lim_{n\to\infty} a_n = 0$. ◇

Reasoning like that in Example 2 shows that for any fixed number k,

$$\boxed{\lim_{n\to\infty}\frac{k^n}{n!} = 0.}$$

This limit will be used often.

This means that the factorial grows faster than any exponential k^n.

Properties of Limits of Sequences

The limits of sequences $\{a_n\}$ behave like the limits of functions $f(x)$, as discussed in Section 2.4. The most important properties are summarized in Theorem 10.1.3 without proof.

Remember that A and B are numbers (not $\pm\infty$).

Theorem 10.1.3. *If* $\lim_{n\to\infty} a_n = A$ *and* $\lim_{n\to\infty} b_n = B$, *then*

- $\lim_{n\to\infty}(a_n + b_n) = A + B$.
- $\lim_{n\to\infty}(a_n - b_n) = A - B$.
- $\lim_{n\to\infty}(a_n b_n) = AB$.
- $\lim_{n\to\infty}(\frac{a_n}{b_n}) = \frac{A}{B}$ $(B \neq 0)$.
- *If* k *is a constant,* $\lim_{n\to\infty} ka_n = kA$. *In particular,* $\lim_{n\to\infty}(-a_n) = -\lim_{n\to\infty} a_n$.
- *If* f *is continuous on an open interval that contains* A, *then* $\lim_{n\to\infty} f(a_n) = f(A)$.

For instance,

$$\begin{aligned}\lim_{n\to\infty}\left(\frac{3}{n} + \left(\frac{1}{2}\right)^n\right) &= 3\lim_{n\to\infty}\frac{1}{n} + \lim_{n\to\infty}\left(\frac{1}{2}\right)^n \\ &= 3 \cdot 0 + 0 \\ &= 0.\end{aligned}$$

Techniques for dealing with $\lim_{x\to\infty} f(x)$ can often help to determine the limit of a sequence. The essential point is

$$\text{if } \lim_{x\to\infty} f(x) = L \quad \text{then} \quad \lim_{n\to\infty} f(n) = L.$$

EXAMPLE 3 Find $\lim_{n\to\infty} \frac{n}{2^n}$.

SOLUTION Consider the function $f(x) = \frac{x}{2^x}$. By l'Hôpital's Rule (∞-over-∞ case),

$$\lim_{x\to\infty}\frac{x}{2^x} = \lim_{x\to\infty}\frac{1}{2^x\ln(2)} = 0.$$

Thus

$$\lim_{n\to\infty}\frac{n}{2^n} = 0.$$

◇

WARNING *(On Limits of Sequences and Limits of Functions)* The converse of the statement "if $\lim_{x\to\infty} f(x) = L$, then $\lim_{n\to\infty} f(n) = L$" is not true. For example, take $f(x) = \sin(\pi x)$. Then $\lim_{n\to\infty} f(n) = 0$, but $\lim_{x\to\infty} f(x)$ does not exist.

The Precise Definition of $\lim_{n\to\infty} a_n = L$

In Sections 3.8 and 3.9 limit concepts were given precise (as opposed to informal) definitions. The following definition is in the same spirit.

Precise definition of limit of a sequence.

DEFINITION *(Limit of a sequence.)* The number L is the **limit of the sequence** $\{a_n\}$ if for each $\epsilon > 0$ there is an integer N such that

$$|a_n - L| < \epsilon \quad \text{for all integers } n > N.$$

EXAMPLE 4 Use the precise definition of the limit of a sequence to show that $\lim_{n\to\infty} \frac{1}{n} = 0$.

SOLUTION Given $\epsilon > 0$ we want to show that there is an integer N such that

$$\left|\frac{1}{n} - 0\right| < \epsilon \quad \text{for all integers } n > N.$$

For instance, if $\epsilon = 0.01$, we want

$$\begin{aligned} \left|\frac{1}{n} - 0\right| &< 0.01 \\ \text{or simply} \qquad \frac{1}{n} &< 0.01 = \frac{1}{100}. \end{aligned}$$

This inequality holds for $n > 100$. Hence $N = 100$ suffices. (So does any integer greater than 100.)

The general case is similar. We wish to have

$$\begin{aligned} \left|\frac{1}{n} - 0\right| &< \epsilon \\ \text{or} \qquad \frac{1}{n} &< \epsilon \\ \text{Hence,} \qquad 1 &< n\epsilon \\ \text{and finally} \qquad n &> \frac{1}{\epsilon}. \end{aligned}$$

Any integer $N > 1/\epsilon$ suffices. ◇

k^n and Energy from the Atom

In a particular nuclear chain reaction, when a neutron strikes the nucleus of an atom of uranium or plutonium, on the average a certain number of neutrons split off. Call this number k. These k neutrons then strike further atoms. Since each splits off k neutrons, in this second generation there are k^2 neutrons. In the third generation there are k^3 neutrons, and so on. Each generation is born in a fraction of a second and produces energy.

If k is less than 1, then the chain reaction dies out, since $k^n \to 0$ as $n \to \infty$. A successful chain reaction — whether in a nuclear reactor or an atomic bomb — requires that k be greater than 1, since then $k^n \to \infty$ as $n \to \infty$.

In September 1941, Enrico Fermi and Leo Szilard achieved $k = 0.87$ with a uranium pile at Columbia University. In 1942, they obtained an encouraging $k = 0.918$. in the meantime, Samuel Allison at the Univeristy of Chicago, Fermi and Szilard attained $k = 1.0006$. With this k the neutron intensity doubled every 2 minutes. They had achieved the first controlled, sustained, chain reaction, producing energy from the atom. Fermi let the pile run for 4.5 minutes. Had he let it go on much longer, the atomic pile, the squash court, the university, and part of Chicago might have disappeared.

Eugene Wigner, one of the scientists present, wrote, "We felt as, I presume, everyone feels who has done something that he knows will have very far-reaching consequences which he cannot foresee."

Szilard had a different reaction: "There was a crowd there and then Fermi and I stayed there alone. I shook hands with Fermi and I said I thought this day would go down as a black day in the history of mankind."

However it may be regarded, December 2, 1942, is a historic date. Before that date k was less than 1, and $\lim_{n\to\infty} k^n = 0$. After that date, k was larger than 1, and $\lim_{n\to\infty} k^n = 0$.

Based on Richard Rhodes, *The Making of the Atomic Bomb*, Simon and Schuster, New York, 1986.

Summary

We defined convergent sequences and their limits and divergent sequences, which have no limit. The sequences $\{r^n\}$ and $\{k^n/n!\}$ will be used often in Chapters 10, 11, and 12. We have

$$\lim_{n\to\infty} r^n = 0 \quad (|r| < 1) \qquad \lim_{n\to\infty} \frac{k^n}{n!} = 0 \quad (k \text{ any constant}).$$

Also, a bounded monotone sequence converges, even though we may not be able to find its limit exactly.

EXERCISES for Section 10.1

In Exercises 1 to 18 write out the first three terms of the given sequence and state whether it converges or diverges. If it converges, give its limit.

1. $\{0.999^n\}$

2. $\{1.001^n\}$

3. $\{1^n\}$

4. $\{(-0.8)^n\}$

5. $\{(-1)^n\}$

6. $\{(-1.1)^n\}$

7. $\{n!\}$

8. $\left\{\frac{10^n}{n!}\right\}$

9. $\left\{\frac{3n+5}{5n-3}\right\}$

10. $\left\{\frac{(-1)^n}{n}\right\}$

11. $\left\{\frac{\cos(n)}{n}\right\}$

12. $\{n\sin(1/n)\}$ (A limit in Section 2.2 will help.)

13. $\left\{n(a^{1/n}-1)\right\}$ (A limit in Section 2.2 will help.)

14. $\left\{\frac{n}{2^n}+\frac{3n+1}{4n+2}\right\}$

15. $\left\{\left(1+\frac{2}{n}\right)^n\right\}$

16. $\left\{\left(\frac{n-1}{n}\right)^n\right\}$

17. $\left\{\left(1+\frac{1}{n^2}\right)^n\right\}$ (Write $f(n)^{g(n)}$ as $e^{g(n)\ln(f(n))}$.)

18. $\left\{\left(1+\frac{1}{n}\right)^{n^2}\right\}$

19. Assume that each year inflation eats away 2 percent of the value of a dollar. Let a_n be the value of one dollar after n years.

(a) Find a_4.

(b) Find $\lim_{n\to\infty} a_n$.

20. Let $a_n = 6^n/n!$.

(a) Fill in this table:

n	1	2	3	4	5	6	7	8
a_n								

(b) Plot the points (n, a_n) corresponding to each column in the table above. Let the n-axis be the horizontal axis.

(c) What is the largest value of a_n? What is the corresponding n?

(d) What is $\lim_{n\to\infty} a_n$?

21. What is the largest value of $(11.8)^n/n!$? Explain.

22. Find an index n such that 0.999^n is less than 0.0001

(a) by experimenting with the aid of your calculator

(b) by solving the equation $0.999^x = 0.0001$

23. Find the first index n such that 1.0006^n is larger than 2

(a) by experimenting with the aid of your calculator

(b) by solving the equation $1.0006^x = 2$.

In Exercises 24 and 27 determine the limits of the given sequences by first identifying each limit as a definite integral, $\int_a^b f(x)\,dx$, for a suitable interval $[a, b]$ and function $f(x)$. (Review Section 6.2)

24.

$$a_n = \sum_{k=1}^{n} \left(\frac{k}{n}\right)^2 \frac{1}{n}$$

25.

$$a_n = \sum_{k=1}^{n} \frac{3}{n} + \frac{k}{n^2}$$

26.

$$a_n = \sum_{k=1}^{n} \frac{1}{n} \cos\left(\frac{k\pi}{n}\right)$$

27.

$$a_n = \sum_{k=1}^{n} \frac{n}{n^2 + k^2}$$

28. For each integer $n \geq 1$, let

$$a_n = \frac{1}{n} + \frac{1}{n+1} + \cdots + \frac{1}{2n} = \sum_{k=n}^{2n} \frac{1}{k}.$$

For example, $a_3 = 1/3 + 1/4 + 1/5 + 1/6 = 0.95$.

(a) Compute decimal approximations to a_n for $n = 1$, 2, 3, 4, and 5.

(b) Show that $\{a_n\}$ is a monotone and bounded sequence.

(c) Show that it has a limit and that the limit is at least $1/2$.

29. We showed that for $-1 < r < 0$, $\lim_{n\to\infty} r^n = 0$ by considering $|r^n|$. Here is a more direct argument.

(a) Let $r = -s$, $0 < s < 1$. Show that for even n, $r^n = s^n$ and for odd n, $r^n = -(s^n)$.

(b) Show that the sequence $\{r^{2n}\}$ converges to 0.

(c) Show that the sequence $\{r^{2n-1}\}$ converges to 0.

(d) Conclude that $\lim_{n\to\infty} r^n = 0$.

30. The binomial theorem asserts that if n is a positive integer, then $(1+x)^n$ is equal to $1 + nx$ plus other terms that are positive if $x > 0$. Use this to show that if $r > 1$, then $\lim_{n\to\infty} r^n = \infty$.

31. Exercise 30 makes use of the binomial theorem. It was not necessary to use the binomial theorem, as this exercise shows. Assume that $x > 0$.

(a) Show that $(1+x)^n \geq 1 + nx$ for $n = 1$.

(b) Assume that you know that $(1+x)^n \geq 1 + nx$ when n is 100. Show that it follows that $(1+x)^n \geq 1 + nx$ when n is 101.

(c) Explain why $(1+x)^n \geq 1 + nx$ for all positive integers n.

32. The sequence $\{a_n\}$ with $a_n = \sum_{k=n}^{2n} \frac{1}{k}$ was shown to be convergent in Exercise 28. Show that the limit of this sequences is $\ln(2)$ by expressing it as a certain definite integral and evaluating that integral.

33. Let $a_n = \sum_{k=2n}^{3n} \frac{1}{k}$. Does $\{a_n\}$ converge or diverge? If it converges, find its limit.

34. Using the precise definition of $\lim_{n\to\infty} a_n = L$, prove that if $\lim_{n\to\infty} |a_n| = 0$, then $\lim_{n\to\infty} a_n = 0$.

35. Use the precise definition of $\lim_{n\to\infty} a_n = L$ to prove $\lim_{n\to\infty} \frac{\sin(n)}{n} = 0$.

36. Use the precise definition of $\lim_{n\to\infty} a_n = L$ to prove $\lim_{n\to\infty} \frac{3}{n^2} = 0$.

37. Use the precise definition of $\lim_{n\to\infty} a_n = L$ to prove that the statement $\lim_{n\to\infty}(-1)^n = 0$ is false.

38.

(a) What would be the precise definition of $\lim_{n\to\infty} a_n = \infty$?

(b) Use the precise definition in (a) and the precise definition of $\lim_{n\to\infty} a_n = L$ to show that:

$$\text{if } \lim_{n\to\infty} a_n = \infty, \text{ then } \lim_{n\to\infty} 1/a_n = 0.$$

39. Using the same approach as in the proof of Theorem 10.1.2, show that if r is greater than 1, then r^n gets arbitrarily large as n increases.

40.

Sam: I have a better proof of Theorem 10.1.2.

Jane: You always do.

Sam: Say that r is positive and less than 1, and r, r^2, r^3, r^4, ... approaches L. Then the sequence r^2, r^4, r^6, r^8, ... must approach L^2. Right?

Jane: Absolutely.

Sam: So L must equal L^2.

Jane: Yes, I see it.

Sam: If $L = L^2$, then L is either 0 or 1, but it can't be 1. So it's 0.

Jane: Brilliant.

Sam: I think so too.

Is Sam's argument correct? Explain.

10.2 Recursively-Defined Sequences and Fixed Points

The terms in each sequence considered in Section 10.1 were given by an explicit formula$a_n = f(n)$. Often a sequence is not given explicitly. Instead, each term (after the first) may be expressed in terms of earlier terms. For instance, the sequence of powers $a_0 = r^0 = 1$, $a_1 = r^1 = r$, $a_2 = r^2$, ..., $a_n = r^n$, ... can be described this way:

> The first term, a_0, is 1.
> For $n \geq 1$, $a_n = ra_{n-1}$.

That is, each term after a_0 is r times the preceding term. We will describe a technique for finding the limit of such sequences, defined indirectly, if they are convergent.

Sequences Defined Recursively

A sequence given by a formula that describes the n^{th} term in terms of previous terms is said to be given **recursively**. If a_n depends only on its immediate predessor, we would have $a_n = f(a_{n-1})$, for some function f. If a_n depends on both a_{n-1} and a_{n-2}, then there would be a function f such that $a_n = f(a_{n-1}, a_{n-2})$.

EXAMPLE 1 Let $a_0 = 1$ and $a_n = na_{n-1}$ for $n \geq 1$. Give an explicit definition of $\{a_n\}$.
SOLUTION $a_1 = 1a_0 = 1$; $a_2 = 2a_1 = 2 \cdot 1$; $a_3 = 3a_2 = 3 \cdot 2 \cdot 1$; $a_4 = 4a_3 = 4 \cdot 3 \cdot 2 \cdot 1$. Evidently, a_n is $n!$, "n factorial," the product of all integers from 1 to n. ◇

EXAMPLE 2 Let $b_0 = 1$ and $b_1 = 1$ and $b_n = b_{n-1} + b_{n-2}$ for $n \geq 2$. Compute b_2, b_3, b_4, and b_5.
SOLUTION $b_2 = b_1 + b_0 = 1 + 1 = 2$; $b_3 = b_2 + b_1 = 2 + 1 = 3$; $b_4 = b_3 + b_2 = 3 + 2 = 5$; $b_5 = b_4 + b_3 = 5 + 3 = 8$. This sequence, which appears often in both pure and applied mathematics, is called the **Fibonacci sequence**.

The terms in the Fibonacci sequence are positive and become arbitrarily large as n gets larger. The Fibonacci sequence diverges (to ∞). ◇

The Fibonacci sequence appears in the following problem from Chapter XII of the *Liber abaci* of Leonard Fibonacci. This book appeared in 1202 (hand copied) and was revised in 1228.

> A man put a pair of rabbits in a place surrounded by a wall. How many pairs of rabbits can be produced from that pair in a year if every month each pair produces a new pair which from the second month on can produce another pair?

For a discussion of the Fibonacci sequence and the Golden Ratio and the myths that surround it, see S. Stein, "Strength in Numbers," John Wiley and Sons, New York, 1996 (p. 39).

Finding the Limit of a Recursive Sequence

Assume that a sequence satisfies the relation $a_n = f(a_{n-1})$ and has a limit L. Since $a_n \to L$ as $n \to \infty$, we also have $a_{n-1} \to L$ and $n \to \infty$. Now assume also that f is continuous. Then we have, because f is continuous,

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} f(a_{n-1}) = f\left(\lim_{n\to\infty} a_{n-1}\right).$$

Hence,

$$L = f(L) \tag{10.2.1}$$

There could be other solutions.

According to (10.2.1), L is a solution to the equation $x = f(x)$. A number L such that $f(L) = L$ is called a **fixed point** of f.

EXAMPLE 3 Let $f(n) = rf(n-1)$ where $0 < r < 1$. Let $a_1 = 1$. Use (10.2.1) to find $\lim_{n\to\infty} a_n$.

This is the same argument as in Section 10.1, but obtained now directly from (10.2.1).

SOLUTION We recognize this recursion as giving the sequence $1, r, r^2, \ldots$. Since this is a monotonic sequence bounded below by 0, it has a limit L. Thus

$$L = f(L) = rL.$$

Since r is not 1, $L = 0$. ◇

Exercises 42 to 45 provide a proof that the sequence of ratios of the Fibonacci sequence converges.

EXAMPLE 4 Define c_n to be the ratio of successive terms in the Fibonacci sequence $\{b_n\}$: $c_n = \frac{b_n}{b_{n-1}}$ for all $n \geq 2$. Assuming this sequence converges, find its limit.

SOLUTION $c_2 = \frac{b_2}{b_1} = \frac{1}{1} = 1$. For $n \geq 3$ the definition of the Fibonacci sequence can be used to obtain a formula relating c_n to c_{n-1}:

$$c_n = \frac{b_n}{b_{n-1}} = \frac{b_{n-1}+b_{n-2}}{b_{n-1}} = 1 + \frac{b_{n-2}}{b_{n-1}} = 1 + \frac{1}{c_{n-1}}.$$

So

$$c_n = 1 + \frac{1}{c_{n-1}} \qquad \text{for all } n \geq 3. \tag{10.2.2}$$

Thus, $c_n = f(c_{n-1})$ where $f(x) = 1 + \frac{1}{x}$.

The table showing the first few terms of this sequence suggests that this sequence converges. Note that the sequence is neither increasing nor decreasing, so Theorem 10.1.1 does not apply.

n	c_n
2	1.000000
3	2.000000
4	1.500000
5	1.666667
6	1.600000
7	1.625000
8	1.615385
9	1.619048
10	1.617647
15	1.618037
25	1.618034

Assume that $\lim_{n\to\infty} c_n$ exists and call that limit L. Then, by (10.2.2),

$$\begin{aligned}\lim_{n\to\infty} c_n &= \lim_{n\to\infty}\left(1 + \frac{1}{c_{n-1}}\right) = 1 + \frac{1}{\lim_{n\to\infty} c_{n-1}}\\ \text{So,} \qquad L &= 1 + \frac{1}{L}.\\ L^2 - L - 1 &= 0.\end{aligned}$$

The two solutions to $L^2 - L - 1 = 0$ are

$$L = \frac{1}{2}\left(1+\sqrt{5}\right) \qquad L = \frac{1}{2}\left(1-\sqrt{5}\right).$$

Since every term in this sequence is positive, the limit cannot be negative. The only possible limit is

$$L = \frac{1}{2}\left(1+\sqrt{5}\right) \approx 1.61803.$$

◇

$\frac{1}{2}\left(1+\sqrt{5}\right) \approx 1.618034$ is known as the **Golden Ratio**.

A Famous Recursion

The recursion $p_{n+1} = kp_n$, where k is a constant greater than 1, describes a population growing at a rate proportional to the amount present. If the initial population is p_1, then $p_2 = kp_1$, $p_3 = k^2p_1$, $p_4 = k^3p_1$, ... and the population would increase without bound. But a population cannot do that. Instead, let us assume it approaches a limiting population, which we will say is 1. As it approaches this size, the struggle to find food slows its growth. Taking this into consideration, we assume that $\{p_n\}$ satisfies the **logistic equation**:

$$p_{n+1} = kp_n(1-p_n).$$

The behavior of this equation, considered on its own is surprising. For instance, if k is near 3.5699456 the behavior of the sequence changes a great deal even when k is changed only a little.

EXAMPLE 5 Examine the sequence given by $p_{n+1} = kp_n(1 - p_n)$ for $0 \leq k \leq 1$.

SOLUTION For $p_0 = 0$ or 1, $p_n = 0$ for all $n \geq 1$. For $0 < p_0 < 1$, $p_1 = kp_0(1 - p_0)$ is at most $p_0(1 - p_0)$, which is less than p_0. Similarly, p_2 is less than p_1, and, in general we have $p_{n+1} < p_n$. The sequence $\{p_n\}$ decreases but stays above 0. Therefore it has a limit L, and $L \geq 0$. Taking limits on both sides of (10.2) shows that $L = kL(1 - L)$. Either $L = 0$ or $1 = k(1 - L)$, hence $L = 0$ or $L = 1 - 1/k$. But $1 - 1/k$ is either negative (if $0 < k < 1$) or 0 (if $k = 1$). So $L = 0$. ⋄

Summary

A recursive sequence is one whose n^{th} term is given in terms of previous terms. If a_n depends only on its immediate predecessor, then $a_n = f(a_{n-1})$. If a_1, a_2, $\ldots$, a_{n-1}, a_n, $\ldots$ converges to L, then $f(L) = L$. Thus L is a root of the equation $f(x) = x$. It is called a **fixed point** of F.

EXERCISES for Section 10.2

In Exercises 1 to 6 give an explicit formula for a_n as a function of n.

1. $a_0 = 1$, $a_n = -a_{n-1}$ for $n \geq 1$

2. $a_0 = 3$, $a_n = a_{n-1}/n$ for $n \geq 1$

3. $a_0 = 2$, $a_n = 3 + a_{n-1}$ for $n \geq 1$

4. $a_0 = 5$, $a_n = -a_{n-1}/2$ for $n \geq 1$

5. $a_1 = 1$, $a_n = a_{n-1} + 1/n$ for $n \geq 2$

6. $a_1 = 1$, $a_n = -a_{n-1} + (-1)^n/n$ for $n \geq 2$

In Exercises 7 to 12 describe a_n in terms of a_{n-1} and an initial term a_0.

7. $a_n = 3^n$, $n = 0, 1, 2, \ldots$

8. $a_n = 5^n/n!$, $n = 0, 1, 2, \ldots$

9. $a_n = 3\,(n!)$, $n = 0, 1, 2, \ldots$

10. $a_n = 2n + 5$, $n = 0, 1, 2, \ldots$

11. $a_n = 1 + 1/2^2 + 1/3^2 + \cdots + 1/n^2$, $n = 1, 2, 3, \ldots$

12. $a_n = 1/2 + 1/4 + 1/8 + \cdots + 1/2^{n-1}$, $n = 0, 1, 2, \ldots$

13. Define $\{b_n\}$ by $b_0 = 2$ and $b_1 = 1/b_{n-1}$ for $n \geq 1$.

(a) Find b_1, b_2, $\ldots$, b_5.

(b) Show that if $\{b_n\}$ converges, its limit is 1 or -1.

(c) Does $\{b_n\}$ converge?

(d) For which choices of b_0 does $\{b_n\}$ converge to 1?

(e) For which choices of b_0 does $\{b_n\}$ converge to -1?

(f) For which choices of b_0 does $\{b_n\}$ diverge?

14. Consider the logistic recursion (10.2) with $k = 2$, that is $p_{n+1} = 2p_n(1 - p_n)$.

(a) Choose p_0 between 0 and 1/2. Find enough p_n to be able to conjecture if the sequence converges.

(b) Repeat (a) for another value of p_0 between 0 and 1/2.

(c) Repeat (a) with p_0 between 1/2 and 1.

(d) Repeat (a) for another value of p_0 between 1/2 and 1.

(e) What happens to the sequence $\{p_n\}$ if p_0 is 0 or 1?

(f) What happens to the sequence $\{p_n\}$ if p_0 is $1/2$?

(g) For which values of p_0 does $\{p_n\}$ converge? And, in those cases, to what limit?

15. For which values of x does $\left\{\frac{x^n}{n!}\right\}$ converge?

16. For which values of x does $\left\{\frac{x^n}{2^n}\right\}$ converge?

17. For which values of x does $\left\{\frac{x^n}{n^2}\right\}$ converge?

18. For which values of x does $\left\{\frac{x^n}{\sqrt{n}}\right\}$ converge?

19. Let $a_{n+2} = a_n + 2a_{n+1}$ with $a_0 = 1 = a_1$ and $c_n = a_n/a_{n-1}$. Examine $\{c_n\}$ numerically, deciding whether it converges and, if so, what its limit might be.

20. Consider the logistic recursion (10.2) with $0 < k \leq 4$. Show that if p_0 is in the interval $[0, 1]$, then p_n is also in $[0, 1]$ for all $n \geq 0$.

21. Let $a_{n+2} = (a_n + 3a_{n+1})/4$, with $a_0 = 0$ and $a_1 = 1$.

(a) Compute enough terms of $\{a_n\}$ to guess the limit, L.

(b) When you take limits of both sides of the recursion equation, what equation do you get for L?

22. Consider the recursion $a_{n+2} = (1 + a_{n+1})/a_n$.

(a) Starting with $a_1 = 1$ and $a_2 = 2$, compute a_3, a_4, a_5, a_6, a_7, and a_8.

(b) Repeat (a) with $a_1 = 3$ and $a_2 = 3$.

(c) Repeat (a) with a_1 and a_2 of your choice.

(d) Explain what is going on.

23. Let k and p be positive numbers and define the sequence $\{f_n\}$ as follows: given f_1, define $f_{n+1} = k(f_n)^p$ for $n \geq 1$.

(a) Assuming this sequence converges, find its limit.

(b) Explain how to choose k so that the sequence converges to 2.

24. Show that if $0 \le k \le 4$, $0 \le p_0 \le 1$, and $p_{n+1} = kp_n(1-p_n)$, then $0 \le p_n \le 1$.

25. Explore the sequence $\{a_n\}$ where $a_{n+1} = a_n - a_{n-1}$ for $n \ge 2$ if

(a) $a_0 = 3$ and $a_1 = 4$,

(b) $a_0 = 1$ and $a_1 = 0$,

(c) the general case, $a_0 = a$, $a_1 = b$.

In each case, try to find the shortest interval containing all a_n. Does the sequence have a limit? Explain.

26.

(a) Investigate the logistic sequence $\{p_n\}$ for $k = 2$.

(b) Make a conjecture based on (a).

(c) Let $p_n = \frac{1}{2} + q_n$. Show that $q_{n+1} = -2q_n^2$.

(d) Use (c) to discuss your conjecture in (b).

27. A path that is $1'$ by n' is to be tiled with $1'$ by $1'$ tiles and $1'$ by $2'$ tiles. Let a_n be the number of ways this can be done.

(a) Obtain a recursive formula for a_n.

(b) Use your formula found in (a) to find a_{10}.

28. Repeat Exercise 27 with $1'$ by $1'$ and $1'$ by $3'$ tiles.

29. Repeat Exercise 27 with $1'$ by $2'$ and $1'$ by $3'$ tiles.

30. Let $u(n)$ be the number of ways of tiling a 2 by n rectangle with 1 by 2 dominoes.

(a) Find $u(1)$, $u(2)$, and $u(3)$.

(b) Find a recursive definition of the function u.

Exercises 31 to 34 illustrate some of the characteristics that make the logistic recursion $p_{n+1} = kp_n(1 - p_n)$ so interesting. In each case, create two sequences corresponding to two values of k in the indicated range and with different values for the initial value, p_0.

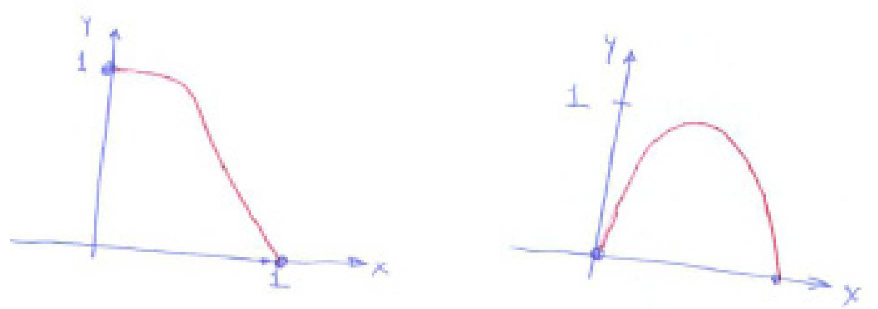
31. $1 < k < 3$

32. $3 < k < 3.4$

33. $3.4 < k < 3.5$

34. $3.6 < k < 4$

35. Figure 10.2.1(a) shows the graph of a decreasing continuous function f such that $f(0) = 1$ and $f(1) = 0$.



(a) (b)

Figure 10.2.1

(a) Show that f has exactly one fixed point in the interval $[0, 1]$. That is, show that there is one number a with $0 \le a \le 1$ that satisfies $f(a) = a$. (Draw the line $y = x$ on the graph of $y = f(x)$.)

(b) If $0 < x < a$, in what interval does $f(x)$ lie?

(c) If $a < x < 1$, in what interval does $f(x)$ lie?

(d) Use the graphs of $y = f(x)$ and $y = x$ to find all values of x for which $f(f(x)) > x$ and all values of x for which $f(f(x)) < x$.

36. Let f be a decreasing function such that $f(0) = 1$ and $f(1) = 0$ and the graph of f is symmetric with respect to the line $y = x$. Examine the sequence x, $f(x)$, $f(f(x))$, ... for x in $[0, 1]$. What can you say about the convergence of this sequence?

37. Let k, c_1, and c_2 be positive numbers. Define the sequence $\{c_n\}$ as follows: given c_1, c_2, define $c_n = (1 + kc_{n-1})/c_{n-2}$ for $n \geq 3$. Assuming this sequence converges, find the possible limits.

38. Examine the sequence $\{x_n\}$ determined by $x_{n+1} = f(x_n)$ with $f(x) = 1 - x^2$ for various inputs in $[0, 1]$. Does f have a fixed point?

39. Let $f(x) = 1-x$, $g(x) = 1-1.1x$, and $h(x) = 1-0.9x$. Let $x_0 = 0.4$. Examine what happens to the sequences determined by each function.

40. Assume that f is a decreasing function for x in $[0, 1]$, $f(1) = 0$, and $-1 < f'(x) < 0$.

(a) What can be said about $f(0)$?

(b) Show that f has a unique fixed point.

(c) Assume a is the fixed point of f, that is $f(a) = a$. Show that if $1 \geq x > a$, then $f(x) < a$ and if $0 \leq x < a$, then $f(x) > a$.

(d) Let $g(x) = f(f(x))$. Examine the sequence x, $g(x)$, $g(g(x))$, $\ldots$ for x in $[0, 1]$. Show that this sequence is monotone.

(e) Show that for all x in $[0, 1]$ the sequence x, $f(x)$, $f(f(x))$, $\ldots$, approaches a.

41. Figure 10.2.1(b) is the graph of a function for which $f(0) = 0$, $f(1) = 0$, $f''(x) \leq 0$, and $0 \leq f(x) \leq 1$.

(a) Show that f has at least one fixed point.

(b) Show that if $f'(0) \geq 1$, then f has only one fixed point.

(c) Show that if $f'(0) < 1$, it has exactly two fixed points.

Exercises 42 to 45 show all of the steps in the proof that the sequence introduced in Example 4 converges. Recall that $c_2 = 1$ and $c_n = 1 + \frac{1}{c_{n-1}}$ for all $n \geq 3$.

42. Let $\{d_n\}$ be the sequence formed from the terms of $\{c_n\}$ with an odd index. That is, $d_n = c_{2n-1}$ for all $n \geq 2$.

(a) Show that $d_n \leq 2$ for all $n \geq 2$.

(b) Show that $\{d_n\}$ is a decreasing sequence.

(c) Explain why you know $\{d_n\}$ converges.

(d) What is $\lim_{n\to\infty} d_n$?

43. Let $\{e_n\}$ be the sequence formed from the terms of $\{c_n\}$ with an even index. That is, $e_n = c_{2n}$ for all $n \geq 1$.

(a) Show that $e_n \geq 1$ for all $n \geq 1$.

(b) Show that $\{e_n\}$ is an increasing sequence.

(c) Explain why you know $\{e_n\}$ converges.

(d) What is $\lim_{n\to\infty} e_n$?

44. Let $\{x_n\}$ be a sequence with the property that the (sub)sequence of odd terms converges to L, $\lim_{n\to\infty} x_{2n-1} = L$, and the (sub)sequence of even terms converges to M, $\lim_{x\to\infty} x_{2n} = M$. Show:

(a) if $L \neq M$ then $\{x_n\}$ diverges

(b) if $L = M$ then $\{x_n\}$ converges and $\lim_{n\to\infty} x_n = L$.

45. Use Exercises 42 to 44 to prove that $\{c_n\}$ converges. Hence its limit is the Golden Ratio.

46. Let k be a number and define the sequence $\{d_n\}$ as follows: given d_0, define $d_n = kd_{n-1}^2$ for $n \geq 1$.

(a) Assuming the sequence converges, find its limit.

(b) Explain how to choose k so that this sequence converges to $3/2$.

10.3 Bisection Method for Solving $f(x) = 0$

One way to estimate the solution of an equation $f(x) = 0$ is to zoom in on it with a graphing calculator. However, precision is limited by the resolution of the display. This section and the next describe techniques for estimating a root to as many decimal places as you may need. The technique in this section is based on the fact that a continuous function that is positive at one input and negative at another has a root between them.

Bisection Method for Solving $f(x) = 0$

Let $f(x)$ be a function. A *solution* or *root* of the equation $f(x) = 0$ is a number r such that $f(r) = 0$. The graph of $y = f(x)$ passes through the point $(r, 0)$, as shown in Figure 10.3.1.

A **root** of f is a solution to $f(x) = 0$.

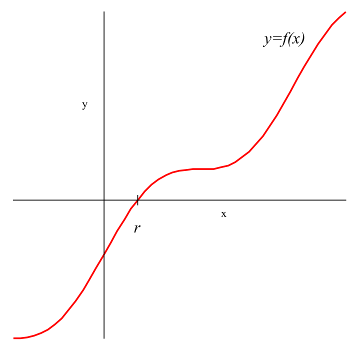


Figure 10.3.1

Let $f(x)$ be a continuous function defined at least on an interval $[a_0, b_0]$, with $a_0 < b_0$. Assume that $f(a_0)$ and $f(b_0)$ have opposite signs, one negative, the other positive. By the Intermediate Value Theorem, $f(x)$ has at least one root in $[a_0, b_0]$.

Not knowing where in $[a_0, b_0]$ a root lies, evaluate f at the midpoint, $m_0 = (a_0 + b_0)/2$. If, by chance, $f(m_0) = 0$, one has found a root and the search is over. Otherwise, the sign of $f(m_0)$ is opposite the sign of one (and only one) of $f(a_0)$ and $f(b_0)$.

If $f(a_0)$ and $f(m_0)$ have opposite signs, then a root must be in the interval $[a_0, m_0]$, which is half the width of $[a_0, b_0]$. On the other hand, if $f(m_0)$ and $f(b_0)$ have opposite signs, a root lies in $[m_0, b_0]$, again half the width of $[a_0, b_0]$.

In either case we have trapped a root in an interval half the width of $[a_0, b_0]$. Call this shorter interval $[a_1, b_1]$. Figure 10.3.2 shows the two possibilities for $[a_1, b_1]$ in the case when $f(a_0) > 0$ and $f(b_0) < 0$.

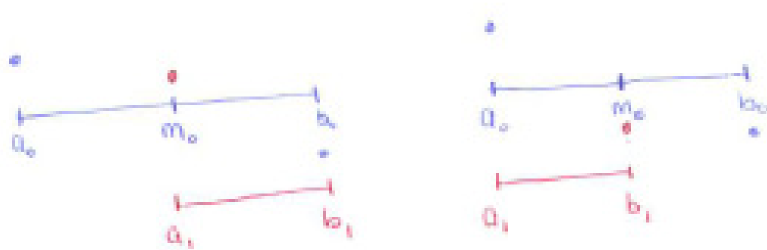


Figure 10.3.2

The Bisection Method is a recursive algorithm.

Then repeat the process, starting at $[a_1, b_1]$. In this way you obtain a sequence of shorter and shorter intervals $[a_0, b_0]$, $[a_1, b_1]$, $[a_2, b_2]$, ..., each half as long as its predecessor. Thus the length of $[a_n, b_n]$ is $(b_0 - a_0)/2^n$.

An Illustration of the Bisection Method

The bisection method is so named because at each step an interval is bisected.

When x is large and positive $f(x) = x + \sin(x) - 2$ is positive. When x is large and negative, $f(x)$ is negative. Therefore $f(x) = 0$ has at least one solution. The derivative of $f(x)$ is $1+\cos(x)$, which is positive except at odd multiples of π, when it is zero. Thus, $f(x)$ is an increasing function, which implies that it cannot have more than one root. Let r be the unique root of $x+\sin(x)-2 = 0$.

The larger $[a_o, b_0]$ is the longer this process will take.

Begin the search for the root by finding an interval on which we can be certain the root will lie.

Since $f(0) = -2$, the root must be positive. Using $\sin(x) \geq -1$ we know $f(x) = x + \sin(x) - 2 \geq x - 1 - 2 = x - 3$ and so $f(4)$ must be positive. Let $a_0 = 0$ and $b_0 = 4$. The root will be found in the interval $[a, b] = [0, 4]$.

The middle of this interval is $m_0 = (a_0 + b_0)/2 = 2$. Evaluate $y_0 = f(m_0) = f(2) \approx 0.909297$. Because $y_0 > 0$ we now know the root is in the interval $[a_1, b_1] = [0, 2]$.

The middle of the new interval is $m_1 = (a_1 + b_1)/2 = 1$. Then $y_1 = f(m_1) = f(1) \approx -0.15829$. Now $y_1 < 0$ so the root is trapped in the interval $[a_2, b_2] = [1, 2]$.

The third iteration of this process yields $m_2 = 1.5$ and $y_2 = f(1.5) \approx 0.497495$. Then, $[a_3, b_3] = [1, 1.5]$.

Rounding to three decimal digits, every number in $[1.105957, 1.106445]$ rounds to 1.106.

Ren more are shown in Table 10.3.1. After 13 iterations the root is known to exist on the interval $[a_{13}, b_{13}] = [1.105957, 1.106445]$. The midpoint of this interval, $m_{13} = 1.106201$, differs from r by at most half the width of $[a_{13}, b_{13}]$, that is, by at most 0.000244.

If the iterations were continued without end, this process defines sequences $\{a_n\}$ and $\{b_n\}$. Of course, one stops when the length of the interval containing r is short enough.

EXAMPLE 1 Use the bisection method to estimate the square root of 3 to three decimal places.

SOLUTION The square root of 3 is the positive number whose square is 3: $x^2 = 3$ or $x^2 - 3 = 0$. We are looking for the positive root of $f(x) = x^2 - 3$.

The function f is continuous. We know $\sqrt{3}$ is between 1 and 2. This suggests using the bisection method with initial interval $[1, 2]$.

The first 11 iterations of the bisection method are displayed in Table 10.3.2. After 7 iterations the approximation $\sqrt{3} \approx m_7 = 1.730469$ is accurate to two decimal places: $\sqrt{3} \approx 1.73$. After another 4 iterations the approximation is accurate to three decimal places: $\sqrt{3} \approx 1.732$. ⋄

n	a_n	b_n	m_n	y_n	$b_n - a_n$
0	0.000000	4.000000	2.000000	0.909297	4.000000
1	0.000000	2.000000	1.000000	−0.158529	2.000000
2	1.000000	2.000000	1.500000	0.497495	1.000000
3	1.000000	1.500000	1.250000	0.198985	0.500000
4	1.000000	1.250000	1.125000	0.027268	0.250000
5	1.000000	1.125000	1.062500	−0.063925	0.125000
6	1.062500	1.125000	1.093750	−0.017895	0.062500
7	1.093750	1.125000	1.109375	0.004796	0.031250
8	1.093750	1.109375	1.101562	−0.006522	0.015625
9	1.101562	1.109375	1.105469	−0.000857	0.007812
10	1.105469	1.109375	1.107422	0.001971	0.003906
11	1.105469	1.107422	1.106445	0.000558	0.001953
12	1.105469	1.106445	1.105957	−0.000149	0.000977
13	1.105957	1.106445	1.106201	0.000204	0.000488

Table 10.3.1

n	a_n	b_n	m_n	y_n	$b_n - a_n$
0	1.000000	2.000000	1.500000	−0.750000	1.000000
1	1.500000	2.000000	1.750000	0.062500	0.500000
2	1.500000	1.750000	1.625000	−0.359375	0.250000
3	1.625000	1.750000	1.687500	−0.152344	0.125000
4	1.687500	1.750000	1.718750	−0.045898	0.062500
5	1.718750	1.750000	1.734375	0.008057	0.031250
6	1.718750	1.734375	1.726562	−0.018982	0.015625
7	1.726562	1.734375	1.730469	−0.005478	0.007812
8	1.730469	1.734375	1.732422	0.001286	0.003906
9	1.730469	1.732422	1.731445	−0.002097	0.001953
10	1.731445	1.732422	1.731934	−0.000406	0.000977
11	1.731934	1.732422	1.732178	0.000440	0.000488

Table 10.3.2

The bisection method is known as a "bracketing method" because the two sequences bracket the solution.

Why the Bisection Method Works

The bisection method applied to $f(x)$ produces two sequences, $a_0 \le a_1 \le a_2 \le \cdots$ and $b_0 \ge b_1 \ge b_2 \ge \cdots$. If no a_n or b_n is a root of f, the sequences do not end. The sequence of left endpoints, $\{a_n\}$, is monotone increasing and the sequence of right endpoints is monotone decreasing. Moreover, since every a_n is less than or equal to b_0, $\{a_n\}$ is bounded. Thus $\{a_n\}$, being bounded and monotone, has a limit, $A \le b_0$. Similarly, $\{b_n\}$ also has a limit, $B \ge a_0$.

Recall that the length of the interval $[a_n, b_n]$ is $b_n - a_n = (b_0 - a_{)})/2^n$. This means that $\{b_n - a_n\}$ is a geometric sequence with ratio $1/2$, which is less than 1. Thus, $\lim_{n\to\infty}(b_n - a_n) = 0$, and we have

$$0 = \lim_{n\to\infty}(b_n - a_n) = \lim_{n\to\infty} b_n - \lim_{n\to\infty} a_n = B - A.$$

Consequently, $A = B$.

But, why is A a root of f?

Consider the sequence

$$f(a_0), f(b_0), f(a_1), f(b_1), f(a_2), f(b_2), \cdots f(a_n), f(b_n), \cdots . \tag{10.3.1}$$

Since f is continuous, (10.3.1) has a limit, $f(A)$. Moreover, the fact that one of $f(a_n)$ and $f(b_n)$ is positive means the limit, $f(A)$, cannot be negative. Similarly, because one of each pair of entries in (10.3.1) is negative, $f(A)$ cannot be positive. Thus, $f(A) = 0$ and A is a root of f.

Summary

In the bisection method for finding a root of a function f, one first finds two inputs a_0 and b_0 for which $f(a_0)$ and $f(b_0)$ have opposite signs. Then one evaluates f at the midpoint m_0. The function f will have opposite signs at the endpoints of exactly one of the intervals: $[a_0, m_0]$ or $[m_0, b_0]$. Call this new interval $[a_1, b_1]$, then repeat the process on this new interval. Continue the process until the interval is short enough to assure an estimate of the root that meets the desired accuracy.

EXERCISES for Section 10.3

In Exercises 1 and 2, use the bisection method to find a_1 and b_1.

1. $a_0 = 2$, $b_0 = 6$, $f(2) = 0.3$, $f(4) = 1.5$, $f(6) = -1.2$

2. $a_0 = 1$, $b_0 = 3$, $f(1) = -4$, $f(2) = -1.5$, $f(3) = 1$

3. In this exercise use the bisection method to approximate $\sqrt{2}$. Let $a_0 = 1$, $b_0 = 2$, and $f(x) = x^2 - 2$. Fill in the following table as you carry out the first five steps of the bisection method.

n	a_n	b_n
0	0	2
1		
2		
3		
4		
5		

4. Use the bisection method to estimate $\sqrt{5}$ with the bisection method.

(a) Use $f(x) = x^2 - 5$ and start with $a_0 = 2$, and $b_0 = 3$. Continue until the interval $[a_n, b_n]$ is shorter than 0.01, that is, $b_n - a_n < 0.01$.

(b) How many more steps of the bisection method are needed to reduce the interval by another factor of 10, that is, $b_n - a_n < 0.001$? (This can be answered *without* computing every a_n and b_n.)

5. Estimate $\sqrt[3]{2}$ by the bisection method.

(a) Use $f(x) = x^3 - 2$ and start with $a_0 = 1$, and $b_0 = 2$. Continue until the interval $[a_n, b_n]$ is shorter than 0.01, that is, $b_n - a_n < 0.01$.

(b) How many more steps of the bisection method are needed to reduce the interval by another factor of 10, that is, $b_n - a_n < 0.001$?

In Exercises 6 to 9 estimate the given numbers to the indicated number of decimal places.

6. $\sqrt{15}$ to 3 decimal places (Use $f(x) = x^2 - 15$ with $a_0 = 3$ and $b_0 = 4$.)

7. $\sqrt{19}$ to 2 decimal places

8. $\sqrt[3]{7}$ to 4 decimal places

9. $\sqrt[3]{25}$ to 3 decimal places

10. Let $f(x) = x^5 + x - 1$.

(a) Show that there is a root of the function $f(x)$ in the interval $[0, 1]$.

(b) Apply five steps of the bisection method with $a_0 = 0$ and $b_0 = 1$.

(c) Why is the root unique?

11. Let $f(x) = x^4 + x - 19$.

(a) Show that $f(2) < 0 < f(3)$ and that there is only one root of $f(x)$ between 2 and 3? What property of f assures that there is exactly one root between 2 and 3?

(b) Using the bisection method with $[a_0, b_0] = [2, 3]$, find an interval of length no more than 0.01 where this root must be found.

(c) The second real root of $f(x)$ is negative. Find an interval of length one in which this root must exist.

(d) Repeat (b) using the interval found in (c) as the initial interval.

12. In estimating $\sqrt{3}$ with the bisection method, Sam imprudently chooses the initial interval to be $[0, 10]$.

(a) How many steps of the bisection method will Sam have to execute before he has an interval shorter than 0.005?

(b) Jane started with $[1, 2]$. How many steps of the bisection method will she need to execute before she has an interval shorter than 0.0005?

13. Let $f(x) = 2x^3 - x^2 - 2$.

(a) Show that there is exactly one root of the equation $f(x) = 0$ in the interval $[1, 2]$.

(b) Using $[a_0, b_0] = [1, 2]$ as a first interval, apply two steps of the bisection method.

14.

(a) Graph $y = x$ and $y = \cos(x)$ relative to the same axes.

(b) Using the graph in (a), find an interval of length no more than 0.25 that contains the positive solution of the equation $x = \cos(x)$. Is there a negative solution?

(c) Using your estimate in (b) as $[a_0, b_0]$, apply the bisection method until the interval is shorter than 0.001.

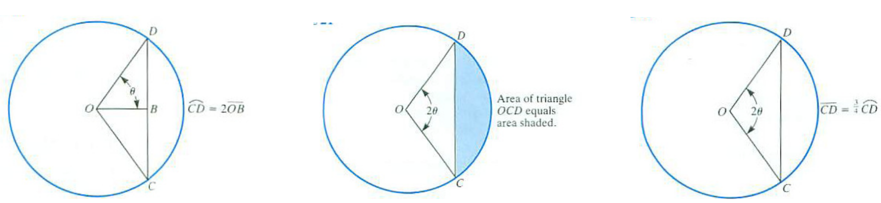
15.

(a) Graph $y = \cos(x)$ and $y = 2\sin(x)$ relative to the same axes.

(b) Without using the graph in (a), explain how you know the graphs intersect exactly once in $[0, \pi/2]$.

(c) Using $[a_0, b_0] = [0, \pi/2]$, apply the bisection method until the length of the interval is no more than 0.001.

In Exercises 16 to 18 (Figure 10.3.3) use the bisection method to estimate θ (to two decimal places). Angles are in radians. Also show that there is only one answer if $0 < \theta < \pi/2$.



(a) (b) (c)

Figure 10.3.3

16. Figure 10.3.3(a)

17. Figure 10.3.3(b)

18. Figure 10.3.3(c)

19.

(a) Graph $y = x\sin(x)$ for x in $[0, \pi]$.

(b) Using the first and second derivatives, show that the function has a unique relative maximum in the interval $[0, \pi]$.

(c) Show that the maximum value of $x\sin(x)$ occurs when $x\cos(x)+\sin(x)=0$.

(d) Use the bisection method, with $[a_0,b_0]=[0,\pi/2]$, to find an estimate for a root of $x\cos(x)+\sin(x)=0$ that is accurate to at least two decimal digits.

20.

(a) Graph $y=x\cos x$ for x in $[0,\pi]$.

(b) Using the first and second derivatives, show that there is a unique relative maximum in the interval $[0,\pi/2]$.

(c) Show that the maximum value of $x\cos x$ occurs when $\cos x - x\sin x = 0$.

(d) Use the bisection method, with $[0,\pi/2]$, to find an interval $[a_n,b_n]$ with length no more than 0.01 that contains a solution of $\cos x - x\sin x = 0$.

21. Use the bisection method to estimate the maximum value of $y=2\sin(x)-x^2$ over the interval $[0,\pi/2]$.

22. Use the bisection method to estimate the maximum value of $y=x^3+\cos(x)$ over the interval $[0,\pi/2]$.

23. We can show that the error in the bisection method diminishes rather slowly. Let $[a_0,b_0]$ be the initial interval containing the root r and let $[a_1,b_1]$ be the next estimate, obtained by the bisection method.

(a) Show that $b_1-a_1=\frac{1}{2}(b_0-a_0)$.

(b) Let $[a_2,b_2]$ be the next interval obtained by the bisection method. Show that $b_2-a_2=\frac{1}{2}(b_1-a_1)=\frac{1}{4}(b_0-a_0)$.

(c) Explain why, in general, $b_n-a_n=\frac{1}{2}(b_{n-1}-a_{n-1})=\frac{1}{2^n}(b_0-a_0)$.

(d) How many steps of the bisection method are needed to obtain an interval no longer than L ($L>0$) containing the given root.

24. The equation $x\tan(x)=1$ occurs in the theory of vibrations.

(a) How many roots does it have in $[0,\pi/2]$?

(b) Use the bisection method to estimate each root to two decimal places.

25. Use the bisection method to approximate all local extrema of $g(x) = 2x - (x+1)e^{-x}$ to three decimal places. How do you know you have found all extrema? See also Example 3 in Section 10.4.

26.

(a) Show that a critical number of the function $f(x) = (\sin x)/x$ for $x \neq 0$ and $f(0) = 1$ satisfies the equation $\tan x = x$.

(b) Show that $(\sin(x))/x$ is an even function. Thus we will consider only positive x.

(c) Graph the function $\tan(x)$ and x relative to the same axes. How often do they cross for x in $[\pi/2, 3\pi/2]$? for x in $[3\pi/2, 5\pi/2]$? Base your answer on your graphs.

(d) Show that $\tan(x) - x$ is an increasing function for x in $[\pi/2, 3\pi/2]$. What does that tell us about the number of solutions of the equation $\tan(x) = x$ for x in $[\pi/2, 3\pi/2]$?

(e) How many critical numbers does the function $f(x)$ have?

(f) Use the bisection method with $[a_0, b_0] = [\pi/2, 3\pi/2]$ to estimate the critical number in $[\pi/2, 3\pi/2]$ to at least two decimal places.

See also Exercise 20 in Section 10.4.

27. Examine the solutions of the equation $2x + \sin(x) = 2$. How many are there? Use the bisection method with appropriate initial intervals to evaluate each solution to two decimal places. Explain the steps in your solution in complete sentences.

28. How many solutions does the equation $\sin(x) = x$ have? Explain how you could use the bisection method to estimate each solution.

29. Explain how you could use the bisection method to estimate $\sqrt[5]{a}$.

30.

Sam: I have a better way than the bisection method.

Jane: What do you propose?

Sam: I trisect the interval into three equal intervals using two points.

Jane: What's so good about that?

Sam: I cut the error by a factor of 3 each step.

Jane: But you have to compute two points and evaluate the function there. That's four calculations instead of two.

Sam: But my method cuts the error so fast, it's still better, so the gain outweighs the cost.

Is Sam right?
Assume the initial interval is $[0, 1]$ and estimate the "cost" to reduce the length of the interval containing the root go the small number E.

31.

Sam: I have a better way than the bisection method.

Jane: What is it?

Sam: I break the interval into four equal intervals by three points.

Jane: Then?

Sam: I find on which of the four intervals the root must lie. I do two of the bisection steps in one step. So it must be more efficient.

Jane: That all depends. I'll think about it.

Think about it and offer your opinion.

10.4 Newton's Method for Solving $f(x) = 0$

This section presents another way to find a sequence of approximations to a solution of $f(x) = 0$. **Newton's Method** uses information about f and its derivative to produce estimates that usually converge much faster than the sequences obtained by the bisection method.

The Idea Behind Newton's Method

Figure 10.4.1 shows the graph of a function f which has a root r and initial estimate x_0. (You may make the initial estimate by looking at a graph, or doing some calculations on your calculator.)

To get a (hopefully) better estimate of r, find where the tangent at $P = (x_0, f(x_0))$ crosses the x-axis. Call the new estimate x_1, as shown in Figure 10.4.1.

Then repeat the process using x_1, instead of x_0, as the estimate of the root r. This produces an estimate x_2. Repeating the process produces a sequence x_0, x_1, x_2, ..., x_n, ... However, in practice, you stop Newton's Method when two successive estimates are sufficiently close together.

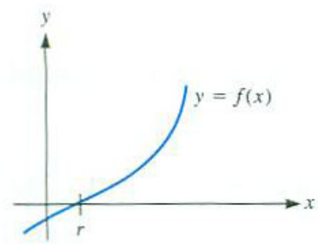


Figure 10.4.1 ARTIST: Please add labels for x_0 and $f(x_0)$.

The Key Formula

To obtain a formula for x_1 in terms of x_0, observe that the slope of the tangent at P in Figure 10.4.1 is $f'(x_0)$ and also $(f(x_0) - 0)/(x_0 - x_1)$. We assume $f'(x_0)$ is not zero, that is, the tangent at P is not parallel to the x-axis. Thus

What difficulties arise if $f'(x_0) = 0$?

$$f'(x_0) = \frac{f(x_0) - 0}{x_0 - x_1}$$

or

$$x_0 - x_1 = \frac{f(x_0)}{f'(x_0)}.$$

Consequently, we have the key formula for applying Newton's Method:

Newton's Recursion

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \tag{10.4.1}$$

The same idea gives $x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$ and so on for x_3, x_4, ... In general, we have the recursive definition,

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{10.4.2}$$

Before we examine whether the sequence converges, we illustrate the technique with some examples.

EXAMPLE 1 In the previous section, 13 iterations of the bisection method were needed to estimate the unique solution to $f(x) = x + \sin(x) - 2 = 0$ to 3 decimal places. Let's see how Newton's Method deals with the same problem.
SOLUTION A reasonable initial estimate is $x_0 = 2$, because it cancels the -2 in $x+\sin(x)-2$. The derivative of $x+\sin(x)-2$ is $1+\cos(x)$. The Newton recursion formula, (10.4.1), reads

$$x_{n+1} = x_n - \frac{x_n + \sin(x_n) - 2}{1 + \cos(x_n)}.$$

The first six iterations of Newton's Method are shown in Table 10.4.1.

Note that $f(x_5) = 0$. As a result, all subsequent estimates will be identical to x_5. We conclude that $r \approx x_5 = 1.106060$ and that this estimate is accurate to six decimal places.

n	x_n	$f(x_n)$	$f'(x_n)$
0	2.000000	0.909297	0.583853
1	0.442592	−1.129124	1.903644
2	1.035731	−0.104034	1.509898
3	1.104632	−0.002069	1.449463
4	1.106060	−0.000001	1.448188
5	1.106060	0.000000	1.448187
6	1.106060	0.000000	1.448187

Table 10.4.1

Each iteration of the bisection method is much easier to implement than Newton's method. However, Newton's Method needs only 5 steps to obtain an approximation of the root to f accurate to (at least) six decimal places while after 13 iterations the bisection method yields an approximation, $p_{13} \approx 1.106201$, accurate to only three decimal places. ◇

EXAMPLE 2 Use Newton's method to estimate the square root of 3, that is, the positive root of the equation $x^2 - 3 = 0$.
SOLUTION Here $f(x) = x^2 - 3$ and $f'(x) = 2x$. According to (10.4.1), if the initial estimate is x_0, then the next estimate x_1 is

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = x_0 - \frac{x_0^2 - 3}{2x_0} = \frac{1}{2}\left(x_0 + \frac{3}{x_0}\right).$$

For our initial estimate, let us use $x_0 = 2$. Its square is 4, which isn't far from 3. Then

$$x_1 = \frac{1}{2}\left(x_0 + \frac{3}{x_0}\right) = \frac{1}{2}\left(2 + \frac{3}{2}\right) = 1.75.$$

Repeat, using $x_1 = 1.75$ to obtain the next estimate:

$$x_2 = \frac{1}{2}\left(x_1 + \frac{3}{x_1}\right) = \frac{1}{2}\left(1.75 + \frac{3}{1.75}\right) \approx 1.73214.$$

One more step of the process yields (to five decimals) $x_3 \approx 1.73205$, which is close to $\sqrt{3}$. The decimal expansion of $\sqrt{3}$ begins 1.7320508. See Figure 10.4.2, which shows x_0, x_1 and the graph of $f(x) = x^2 - 3$, and Table 10.4.2, the numerical values used in these computations.

In fact, x_3 agrees with $\sqrt{3}$ to seven decimals.

n	x_n	$f(x_n)$	$f'(x_n)$
0	2.000000	1.000000	4.000000
1	1.750000	0.062500	3.500000
2	1.732143	0.000319	3.464286
3	1.732051	0.000000	3.464102
4	1.732051	0.000000	3.464102

Table 10.4.2

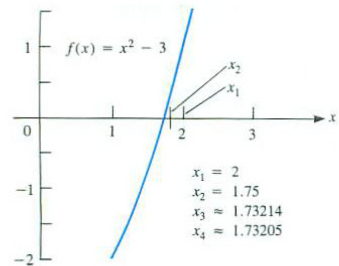


Figure 10.4.2 NOTE: Renumber indices.

When the same problem was solved using the bisection method in Example 1, after 11 iterations the best approximation to r is $p_{11} = 1.732178$. This approximation to $\sqrt{3}$ is accurate to only three decimal places. ◇

Compare with Table 10.3.2.

In practice, stop the process when either $|f(x_n)|$ or the difference between successive estimates, $|x_n - x_{n-1}|$, become sufficiently small.

Compare with Exercise 25.

EXAMPLE 3 Use Newton's method to approximate the location of the local extrema of $g(x) = 2x - (x+1)e^{-x}$.

SOLUTION This problem, which was first solved in Exercise 25 in Section 10.3 is equivalent to asking for all roots of $f(x) = g'(x) = 2 + xe^{-x}$.

To find an initial guess to start Newton's method, notice that $f(0) = 2$ and $f(x) > 0$ for all positive numbers x. Looking for a negative value of x that makes $f(x)$ negative, we see that $f(-2) = 2 + (-2)e^2 = 2 - 2e^2 < 0$ because $e > 1$.

The first few iterations of Newton's method with $x_0 = -1$ are shown in Table 10.4.3. After four steps the process is stopped because $f(x_3) = 0$. The critical number of g is approximately $x^* \approx x_3 = -0.852606$. This is correct to all six decimal places shown.

Because $g'(x)$ is negative to the immediate left of x^* and is positive to the immediate right of x^* we conclude that x^* is a local minimum of $g(x) =$

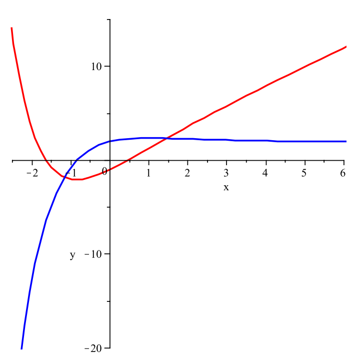


Figure 10.4.3 ARTIST: Label functions in graph.

n	x_n	$f(x_n)$	$f'(x_n)$	$\lvert x_n - x_{n-1}\rvert$
0	-1.000000	-0.718282	5.436564	
1	-0.867879	-0.067163	4.449017	0.132121
2	-0.852783	-0.000773	4.346941	0.015096
3	-0.852606	0.000000	4.345751	0.000177
4	-0.852606	0.000000	4.345751	0.000000

Table 10.4.3

$2x-(x+1)e^{-x}$. The graphs of g and $g'=f$ are shown in Figure 10.4.3. Observe the only local extremum is the local minimum near $x=-0.85$. ◇

Remarks on Newton's Method

The assumption that f'' exists implies f' (and f) are continuous.

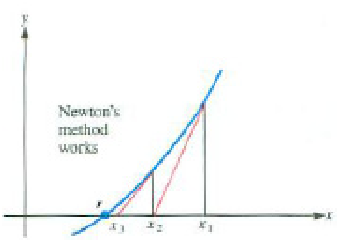


Figure 10.4.4

In an interval where $f''(x)$ is positive, the graph of $y=f(x)$ is concave up, and lies above its tangents, as shown in Figure 10.4.4. If x_1 is to the right of r, the sequence x_1, x_2, x_3, …is monotone and is bounded below by r. Thus, the sequence converges to a limit $L \geq r$. To show that L is r, take limits of both sides of the Newton recursion formula, (10.4.2):

$$\lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty}\left(x_n - \frac{f(x_n)}{f'(x_n)}\right) \tag{10.4.3}$$

obtaining (10.4.4)

$$L = L - \frac{f(L)}{f'(L)} \tag{10.4.5}$$

Thus, $0=-f(L)/f'(L)$, so $f(L)=0$, and L is a root of f.

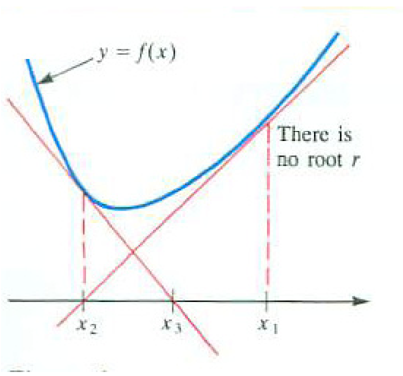
The reasoning that obtained (10.4.5) from (10.4.3) shows, more generally, if the sequence produced by Newton's Method converges, its limit is a root.

The equation $f(x)=0$ may not have a solution. In that case the sequence of estimates produced by Newton's method does not approach a specific number but may wander all over the place, as in Figure 10.4.5(a).

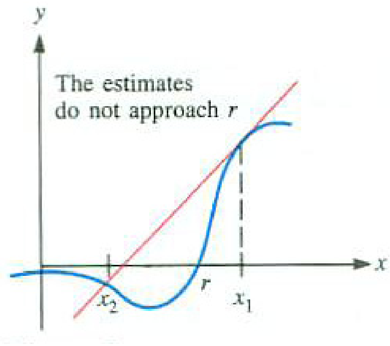
It is also possible that there is a root r, but your initial guess x_0 is so far from r that the sequence of estimates does not approach r. See Figure 10.4.5(b).

The tangent at $(x_n, f(x_n))$ is horizontal and does not intersect the x-axis.

Of course, if x_n is a number where $f'(x_n)=0$, then the Newton recursion, which has $f'(x_n)$ in the denominator, makes no sense.



(a)



(b)

Figure 10.4.5

How Good is Newton's Method

When you use Newton's method, you produce a sequence of estimates x_0, x_1, x_2, ... of a root r. How quickly does the sequence approach r? In other words, how rapidly does the difference between the estimate x_n and the root r, $|x_n - r|$, approach 0?

Newton's method for solving $x^2 - 3 = 0$ revisited from a different point of view.

To get a feel for the rate at which $|x_n - r|$ shrinks as we keep using Newton's method, take the case in Example 2, where we were estimating $\sqrt{3}$ using the recursion $x_{n+1} = \frac{1}{2}\left(x_1 + \frac{3}{x_n}\right)$.

In the following table, we list, x_1, x_2, x_3, x_4 to seven decimal places and compare to $\sqrt{3} \approx 1.7320508$:

Estimate	Value	Agreement with $\sqrt{3}$
x_1	2.000000000	Initial guess
x_2	1.750000000	First two digits
x_3	1.732142857	First four digits
x_4	1.732050810	First eight digits

At each stage the number of correct digits tends to *double*. This means the error at one step is roughly the square of the error of the previous guess,

$$|x_n - r| \leq M\,|x_{n-1} - r|^2$$

for an appropriate constant M. This constant depends on the maximum of the absolute values of the first and second derivatives. By contrast, the iterates for the bisection method tend to cut the error $|x_n - r|$ in half at each step. Because $2^3 < 10 < 2^4$, it generally takes 3 or 4 steps to gain one more decimal place accuracy.

This difference is evident in the number of iterations needed in each algorithm to achieve the same accuracy.

Summary

This section developed Newton's method for estimating a root of an equation, $f(x) = 0$. You start with an estimate x_0 of the root, then compute

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Then repeat the process, obtaining the sequence

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \qquad \text{for all } n = 1,\ 2,\ 3,\ \ldots.$$

When $f'(r) \neq 0$ and f' is continuous, the iterates in Newton's Method converge to r provided the initial guess is sufficiently close to r.

The Newton iterates converge quickly to the root: there is a constant M such that

$$|x_n - r| \leq M\,|x_{n-1} - r|^2$$

while the iterates computed by the bisection method converge slowly:

$$|x_n - r| \leq \frac{1}{2}\,|x_{n-1} - r|\,.$$

While, in general, Newton's method converges faster than the bisection method the actual performance depends on f and the initial estimates.

Iterative methods for finding a root generally stop when either $|f(x_n)|$ or $|x_{n+1} - x_n|$ becomes small enough.

EXERCISES for Section 10.4

In Exercises1 and 2, use Newton's method to find x_1.

1. $x_0 = 2$, $f(2) = 0.3$, $f'(2) = 1.5$

2. $x_0 = 3$, $f(3) = 0.06$, $f'(3) = 0.3$

3. Let a be a positive number. Show that the Newton recursion formula for estimating $\sqrt{a}$ is given by

$$x_{i+1} = \frac{1}{2}\left(x_i + \frac{a}{x_i}\right).$$

The sequence defined in Exercise 3 was the Babylonian method for estimating $\sqrt{a}$. If the guess x_0 is smaller than $\sqrt{a}$, then a/x_0 is larger than $\sqrt{a}$. So x_1 is the average of two numbers between which $\sqrt{a}$ lies.

4. Use the formula of Exercise 3 to estimate $\sqrt{15}$. Choose $x_0 = 4$ and compute x_1 and x_2 to three decimals.

5. Use the formula of Exercise 3 to estimate $\sqrt{19}$. Choose $x_0 = 4$ and compute x_1 and x_2 to three decimals.

6. Use Newton's Method to estimate $\sqrt[3]{7}$. Choose $x_0 = 2$ and compute x_1 and x_2 to three decimals.

7. Use Newton's Method to estimate $\sqrt[3]{25}$. Choose $x_0 = 3$ and compute x_1 and x_2 to three decimals.

8. In this exercise the ideas in Exercise 3 are used to estimate $\sqrt{5}$ with Newton's method.

(a) Use $f(x) = x^2 - 5$ and start with $x_0 = 2$. Continue until the consecutive estimates differ by at most 0.01, that is, $x_{n+1} - x_n < 0.01$.

(b) How many more steps of Newton's method are needed to reduce the interval by another factor of 10, that is, $x_{n+1} - x_n < 0.001$?

9. Estimate $\sqrt[3]{2}$ with Newton's method.

(a) Use $f(x) = x^3 - 2$ and start with $x_0 = 1$. Continue until the consecutive estimates differ by at most 0.01, that is, $x_{n+1} - x_n < 0.01$.

(b) How many more steps of Newton's method are needed to reduce the interval by another factor of 10, that is, $x_{n+1} - x_n < 0.001$?

10. Let $f(x) = x^5 + x - 1$.

(a) Using $x_0 = \frac{1}{2}$ as a first estimate, apply Newton's method to find a second estimate x_1.

(b) Show that there is a root of the function $f(x)$ in the interval $[0, 1]$.

(c) Why is the root unique?

11. Let $f(x) = x^4 + x - 19$.

(a) Apply Newton's method, starting with $x_0 = 2$. Compute x_1 and x_2.

(b) Show that $f(2) < 0 < f(3)$. What additional property of f assures that there is exactly one root r between 2 and 3?

(c) The second real root of $f(x)$ is negative. Find an interval of length one in which this root must exist.

(d) Use the left endpoint of the interval in (c) as the initial guess for Newton's method. Compute x_1 and x_2.

12. In estimating $\sqrt{3}$ with Newton's method, Sam imprudently chooses $x_0 = 10$. What does Newton's method give for x_1, x_2, and x_3?

13. Let $f(x) = 2x^3 - x^2 - 2$.

(a) Show that there is exactly one root of the equation $f(x) = 0$ in the interval $[1, 2]$.

(b) Using $x_0 = \frac{3}{2}$ as a first estimate, apply Newton's method to find x_2 and x_3.

14.

(a) Graph $y = x$ and $y = \cos(x)$ relative to the same axes.

(b) Using the graph in (a), estimate the positive solution of the equation $x = \cos(x)$. Is there a negative solution?

(c) Using your estimate in (b) as x_0, apply Newton's method until consecutive estimates agree to four decimal places.

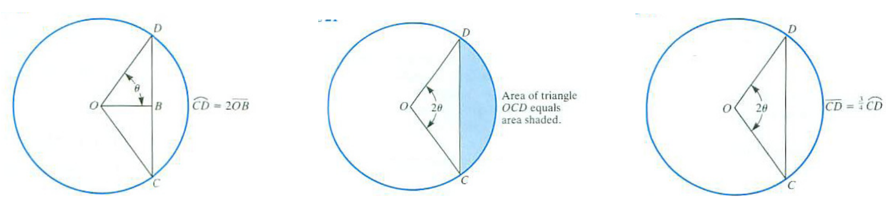
15.

(a) Graph $y = \cos(x)$ and $y = 2\sin(x)$ relative to the same axes.

(b) Using the graph in (a), estimate the solution that lies in $[0, \pi/2]$.

(c) Using your estimate in (b) as x_0, apply Newton's method until consecutive estimates agree to four decimal places.

In Exercises 16 to 18 (Figure 10.4.6) use Newton's method to estimate θ (to two decimal places). Angles are in radians. Also show that there is only one answer if $0 < \theta < \pi/2$.



(a) (b) (c)

Figure 10.4.6

16. Figure 10.4.6(a)

17. Figure 10.4.6(b)

18. Figure 10.4.6(c)

19. The equation $x\tan(x) = 1$ occurs in the theory of vibrations.

(a) How many roots does it have in $[0, \pi/2]$?

(b) Use Newton's method to estimate each root to two decimal places.

20. Repeat parts (a)–(e) of Exercise 26 in Section 10.3. For part (f), use Newton's method instead of the bisection method.

21. Examine the solutions of the equation $2x + \sin(x) = 2$. How many are there? Use Newton's method to evaluate each solution to two decimal places. Explain the steps in your solution in complete sentences.

22. How many solutions does the equation $\sin(x) = x$ have? Explain how you could use Newton's method to estimate each solution.

23. Explain how you could use Newton's method to obtain a formula for estimating $\sqrt[5]{a}$.

Exercises 24 and 25 show that care should be taken in applying Newton's method.
24. Let $f(x) = 2x^3 - 4x + 1$.

(a) Show that there must be a root r of $f(x) = 0$ in $[0, 1]$.

(b) Take $x_0 = 1$, and apply Newton's method to obtain x_1 and x_2.

(c) Graph f, and show what is happening in the sequence of estimates.

25. Apply Newton's method to the function $f(x) = x^3 - x$, starting with $x_0 = 1/\sqrt{5}$.

(a) Compute x_1 and x_2 exactly (not as decimal approximations).

(b) Graph $x^3 - x$ and explain why Newton's method fails in this case.

26. Let $f(x) = x^2 + 1$

(a) Using Newton's method with $x_0 = 2$, compute x_1, x_2, x_3, and x_4 to two decimal places.

(b) Using the graph of f, show geometrically what is happening in (a).

(c) Using Newton's method with $x_1 = \sqrt{3}/3$, compute x_2 and x_3. What happens to x_n as $n \to \infty$?

(d) What happens when you use Newton's method, startng with $x_1 = 1$?

27. Assume that $f'(x) > 0$, $f"(x) < 0$ for all x, and $f(r) = 0$.

(a) Sketch a possible graph of $y = f(x)$.

(b) Describe the behavior of the sequence of Newton's estimates x_0, x_1,..., x_n, ... when you choose $x_0 > r$. Include a sketch.

(c) Describe the behavior of the sequence if you choose $x_0 < r$. Include a sketch.

28. Let $f(x) = 1/x + 5$

(a) Graph $f(x)$ showing its x-intercepts.

(b) For which x_0 does Newton's Method sequence converge to a solution to $f(x) = 0$?

(c) For which x does Newton Method sequence not converge?

29. Let $f(x) = \frac{1}{x^2} - 5$ and assume the same questions as in the preceding exercise.

30.

(a) Graph $y = x\sin(x)$ for x in $[0, \pi]$.

(b) Using the first and second derivatives, show that it has a unique relative maximum in the interval $[0, \pi]$.

(c) Show that the maximum value of $x\sin(x)$ occurs when $x\cos(x) + \sin(x) = 0$.

(d) Use Newton's method, with $x_0 = \pi/2$, to find an estimate x_1 for a root of $x\cos(x) + \sin(x) = 0$.

(e) Use Newton's method again to find x_2.

31.

(a) Graph $y = x\cos x$ for x in $[0, \pi]$.

(b) Using the first and second derivatives, show that it has a unique relative maximum in the interval $[0, \pi/2]$.

(c) Show that the maximum value of $x\cos x$ occurs when $\cos x - x\sin x = 0$.

(d) Use Newton's method, with $x_0 = \pi/4$, to find an estimate x_1 for a root of $\cos x - x\sin x = 0$.

(e) Use Newton's method again to find x_2.

32. Use Newton's method to estimate the maximum value of $y = 2\sin(x) - x^2$ over the interval $[0, \pi/2]$.

33. Use Newton's method to estimate the maximum value of $y = x^3 + \cos(x)$ over the interval $[0, \pi/2]$.

34. We can show that the error in Newton's method diminishes rapidly (compared to the bisection method). Let x_0 be an estimate of the root r and let x_1 be the second estimate, obtained by Newton's method. Assume $f'(x_0) \neq 0$.
Using the first-order Taylor polynomial with remainder, centered at $a = x_0$, we may write

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(c)}{2}(x - x_1)^2 \qquad (10.4.6)$$

where c is a number between x and x_1. (See page 398 in Section 5.5.)

(a) In (10.4.6), replace x by r and use the definition of x_1 to show that

$$x_1 - r = \frac{f^{(2)}(c)}{2f'(x_0)}(r - x_0)^2,$$

where c is between x_1 and r.

(b) Assume that $x_0 > r$ and that $f'(x)$ and $f''(x)$ are positive for x in $[r, x_0]$. Indicate on a diagram where the numbers x_1, x_2 ... are situated. Then use (a) to discuss how the error, $r - x_n$, behaves as n increases.

35. Let p be a positive number.

(a) Graph $f(x) = 1/x - p$.

(b) For which choices of the initial estimate of a root of f will Newton's Method converge to r?

36. Throughout this section we have assumed we knew the derivative $f'(x)$. However, the derivative may be too complicated, or perhaps you just know the values of $f(x)$ at certain points. When you make an initial guess of a root of f, how would you calculate a plausible "better approximation"? (What could you use instead of the tangent line?)

10.S Chapter Summary

Infinite sequences of numbers a_k, a_{k+1}, ... arise in many contexts. (The initial index, k, can be any non-negative integer.) For instance, they arise when estimating a root of an equation of the form $f(x) = 0$. Any equation, $g(x) = h(x)$ can be transformed to that form, for it is equivalent to $g(x) - h(x) = 0$.

One way to estimate a root of $f(x) = x$ is to pick an estimate, a, of a root and compute $f(a)$, $f(f(a))$, $f(f(f(a)))$, ... If this sequence has a limit, r, then $f(r) = r$.

The bisection method provides estimates of the roots of $f(x) = 0$. One looks for numbers a and b at which $f(x)$ has opposite signs. If f is continuous, it has a root in the interval (a, b). Let m be the midpoint of that interval. Then either m is a root or its sign is opposite the sign of one of $f(a)$ and $f(b)$. Repeat, using either (a, m) or (m, b).

Newton's method for solving $f(x) = 0$ depends on using a tangent to approximate the graph of $f(x)$. It yields the recursion $x_2 = x_1 - f(x_1)/f'(x_1)$. Repeat the process until one has the desired accuracy.

EXERCISES for 10.S 1. Let $a_0 = 0$ and $a_n = a_{n-1} + 2n - 1$ for $n \geq 1$.

(a) Compute a few values of a_n (at least through a_5) and conjecture an explicit formula for a_n.

(b) Show that if your formula is correct for $n = k$, then it is correct for $n = k+1$.

2.

(a) Graph $f(x) = \cos\left(\frac{\pi x}{2}\right)$ for x in $[0, 1]$.

(b) Let a be the unique fixed point of f on $[0, 1]$. Estimate a by looking at your graph in (a).

(c) Use Newton's Method to estimate a to 2 decimal places.

(d) Use the bisection method to estimate a to 2 decimal places.

(e) Does the sequence $\cos\left(\frac{\pi x}{2}\right)$, $\cos\left(\frac{\pi}{2}\cos\left(\frac{\pi x}{2}\right)\right)$, ... converge for every x in $[0, 1]$?

3. In Example 1 in Section 4.1 it was shown that $f(t) = (t^2 - 1)\ln\left(\frac{t}{\pi}\right)$ has one critical number on $[1, \pi]$. Use Newton's Method to estimate this critical number to three decimal digits.

4. In Example 2 in Section 4.1 it was shown that $f(x) = x^3 - 6x^2 + 15x + 3$ has exactly one real root. Use Newton's method to approximate this root to three decimal places.

5.

(a) Graph $y = xe^{-x^2}$.

(b) Estimate the area of the region bounded by $y = xe^{-x^2}$, the line $x + y = 1$, and the x-axis.

You will need Newton's method of estimating a solution of an equation.

6. The spiral $r = \theta$ meets the circle $r = 2\sin(\theta)$ at a point other than the origin. Use Newton's method to estimate the coordinates of that point. (Give both the polar and rectangular coordinates of the point of intersection.)

7. The equation $M = E - e\sin(E)$, known as **Kepler's equation**, occurs in the study of planetary motion. (M involves E, position, and e, the eccentricity of the orbit, a number between 0 and 1.)

(a) Sketch the graph of $M(E) = E - e\sin(E)$ looks like as a function of E when $e = 0.2$.

(b) Show that $M(E) = E - e\sin(E)$ is an increasing function of E for any $0 < e < 1$.

(c) In view of (b), E is a function of M, $E = g(M)$. Use Newton's method to find $g(0.25)$, $g(0.5)$, and $g(1.5)$ if $e = 0.2$. Find all answers to at least three decimal digits.

(d) Repeat (c) with $e = 0.8$.

(A graphing calculator or computer can be used to simplify the calculations.)

Consider the problem of finding a solution to $g(x) = 0$. There are usually several ways to rewrite this equation in the form $f(x) = x$. The challenge is to choose the function f so that the sequence with $a_n = f(a_{n-1})$ converges. Then $L = \lim_{n\to\infty} a_n$ is a solution to $g(x) = 0$. In Exercises 8 to 11 we develop and apply a general result known as the **Fixed Point Theorem**.

8. In this exercise we develop a version of the Fixed Point Theorem that will explain what is happening in Exercises 9 and 11. Basically, if r is a fixed point of f, that is, a number such that $f(r) = r$, then the errors $e_n = r - a_n$ satisfy $r - e_n = f(r - e_{n-1})$.

(a) Fill in the details to show why $r - e_n = f(r - e_{n-1})$.

(b) Replace $f(r - e_{n-1})$ with the linear approximation to f at r and derive the (approximate) result: $e_n \approx f'(r)e_{n-1}$ for all $n \geq 0$.

(c) Show that if $e_n \approx f'(r)e_{n-1}$ for all $n \geq 0$, then $e_n \approx (f'(r))^n e_0$.

(d) Explain why $e_n \to 0$ if $|f'(r)| < 1$ and $\{e_n\}$ diverges if $|f'(r)| > 1$. That is, a_n converges to r if $|f'(r)| < 1$, and $\{a_n\}$ does not converge to r if $|f'(r)| > 1$.

Consider the question of finding a solution to $g(x) = x + \ln(x) = 0$. There are several ways to reformulate this problem as a fixed point problem, that is, to solve an equation of the form $f(x) = x$. Exercises 9 and 10 show that the Fixed Point Theorem can be used to explain why some reformulations are more useful than others for finding a root of $g(x) = 0$.

9.

(a) Let $f_1(x) = -\ln(x)$. Verify that $g(x) = 0$ and $f_1(x) = x$ have the same solution.

(b) Compute $|f_1'(r)|$ where r is close to the solution to $g(x) = 0$. What does this tell you about the sequence with $a_n = f_1(a_{n-1})$?

(c) Let $x_0 = 0.5$ and compute x_1, x_2, x_3, and x_4 using $x_n = f_1(x_{n-1})$. Why can't you compute x_5?

10.

(a) Let $f_2(x) = e^{-x}$. Verify that $g(x) = 0$ and $f_2(x) = x$ have the same solution.

(b) Compute $|f_2'(r)|$ where r is close to the solution to $g(x) = 0$. What does this tell you about the sequence with $a_n = f_2(a_{n-1})$?

(c) Let $x_0 = 0.5$ and compute x_1, x_2, x_3, and x_4 using $x_n = f_2(x_{n-1})$. What happens as $n \to \infty$?

The function $g(x) = x^2 - 2x - 3$ has two roots: $x = 3$ and $x = -1$. In Exercises 11 to 13 we will explore three different ways to use fixed-point iterations to find these roots.

11.

(a) Show that solving $g(x) = 0$ is equivalent to finding a fixed point of $f_1(x) = \sqrt{2x+3}$.

(b) Compute $|f_1'(r)|$, where r is close to either root of $g(x) = 0$. What does this tell you about the sequence $a_n = f_1(a_{n-1})$?

(c) Let $x_0 = 0.5$ and compute x_1, x_2, x_3, and x_4 using $x_n = f_1(x_{n-1})$. What happens $\lim_{n\to\infty} x_n$?

12.

(a) Show that solving $g(x) = 0$ is equivalent to finding a fixed point of $f_2(x) = 3/(x-2)$.

(b) Compute $|f_2'(r)|$, where r is close to either root of $g(x)$. What does this tell you about the sequence $a_n = f_2(a_{n-1})$?

(c) Let $x_0 = 0.5$ and compute x_1, x_2, x_3, and x_4 using $x_n = f_2(x_{n-1})$. What happens $\lim_{n\to\infty} x_n$?

13.

(a) Show that solving $g(x) = 0$ is equivalent to finding a fixed point of $f_3(x) = \frac{1}{2}(x^2 - 3)$.

(b) Compute $|f_3'(r)|$, where r is close to the solutions to $g(x) = 0$. What does this tell you about the sequence $a_n = f_3(a_{n-1})$?

(c) Let $x_0 = 0.5$ and compute x_1, x_2, x_3, and x_4 using $x_n = f_3(x_{n-1})$. What happens $\lim_{n\to\infty} x_n$?

(d) Which of the methods in the these last three exercises is the best way to find the solutions to $g(x) = 0$?

Exercises 14 and 15 will be used in Exercises 16 and 17.

14. Find $\lim_{x\to 0} \frac{\tan(x)-x}{2x-\sin(2x)}$.

15. Find $\lim_{x\to 0} \frac{\tan(x)-x}{x-\sin(x)}$.

16. Let P_n be the perimeter of a regular polygon with n sides that circumscribes a circle of radius 1. Similarly, let p_n be the perimeter of an inscribed regular polygon of n sides. When n is large, which is the better estimate of the perimeter of the circle? To decide, examine the limit of $\frac{P_n-2\pi}{2\pi-p_n}$. (Form an opinion before you calculate.) (See Exercise 14.)

17. Let A_n be the area of a regular polygon with n sides that circumscribes a circle of radius 1. Similarly, let a_n be the area of an inscribed regular polygon of n sides. When n is large, which is the better estimate of the area of the circle? To decide, examine the limit of $\frac{A_n-\pi}{\pi-a_n}$. (Form an opinion before you calculate.) (See Exercise 15.)

18. (Contributed by Frank Saminiego.) Assume that a_i and b_i, $0 \leq i \leq n$, are positive and the ratios a_i/b_i increase as a function of the index i. (That is, $a_0/b_0 < a_1/b_1 < \cdots < a_n/b_n$.) Then it is known that

$$f(x) = \frac{\sum_{i=0}^{n} a_i x^i}{\sum_{i=0}^{n} b_i x^i}$$

is an increasing function for $x > 0$. This fact is used in the statistical theory of reliability.
Verify the assertion for (a) $n = 1$ and (b) $n = 2$. (Show that $f'(x) > 0$.)

19. Let $u(n)$ be the number of ways of tiling a 3 by n rectangle with 1 by 3 dominoes.

(a) Find $u(1)$, $u(2)$, and $u(3)$.

(b) Find a recursive definition of the function u.

(c) Use (b) to find $u(10)$.

20. A tile consists of three 1 by 1 squares arranged to form the letter L. Let $u(n)$ be the number of ways a 2 by n rectangle can be tiled by such tiles.

(a) Find $u(n)$ for $n = 1$, 2, 3, 4, 5, and 6.

(b) Find a recursion for $u(n)$.

(c) Find $u(n)$ for $n = 22$, 23, and 24.

21. Repeat Exercise 20 with $u(n)$ the number of ways a 3 by n rectangle can be tiled by the L-shaped tiles.

22. In the study of the hydrogen atom, one meets the integral

$$\int_0^{\infty} r^n e^{-kr}\, dr$$

Here n is a non-negative integer and k a positive constant. Show that it equals $n!/k^{n+1}$. (First find the value for $n = 0$. Then use integration by parts.) $n!$ is the factorial of n, $n! = 1 \cdot 2 \cdot \cdots \cdot (n-1) \cdot n$

An experiment either succeeds, with probability p, or fails, with probability q ($p+q = 1$). For instance, the probability of getting a five when rolling a six on one die is

$p = 1/6$. If the experiment is repeated n times one would expect near pn successes. For $k = 0, 1, \ldots, n$, the probability of having k successes in n experiments is

$$\frac{n!}{k!(n-k)!}p^k q^{n-k}, \tag{10.S.1}$$

(called the **binomial distribution**). Exercise 23 concerns the case when k is small in comparison to n, showing that (10.S.1) is approximately $k^n e^{-k}/n!$, called the **Poisson distribution**. Exercise 24 obtains an approximation of (10.S.1) when k is "near" pn. This approximation is

$$\frac{1}{\sqrt{2\pi npq}}e^{-\frac{(k-np)^2}{2npq}},$$

which is related to the normal distribution (the famous bell curve).

23. The following limit occurs in the elementary theory of probability:

$$\lim_{N\to\infty} \frac{N!}{n!(N-n)!}\left(\frac{k}{N}\right)^n \left(1-\frac{k}{N}\right)^{N-n},$$

where n is a fixed positive integer and k is a positive constant. Show that the limit is

$$\frac{k^n e^{-k}}{n!}.$$

24. This exercise obtains an approximation of (10.S.1) when k is "near" pn. "Near" means $\lim_{n\to\infty} \frac{k-pn}{n} = 0$. We may write $k = pn + z_k$, where $z_k/n \to 0$ as $n \to \infty$. Note that $k \to \infty$ as $n \to \infty$.

We will use Stirling's approximation to $m!$, namely $\sqrt{2\pi m}(m/e)^m$, developed in Exercise 28 in Section 11.6 on page 987.

(a) Show that (10.S.1) is approximated by

$$\left(\frac{n}{2\pi k(n-k)}\right)^{1/2}\left(\frac{pn}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k}. \tag{10.S.2}$$

(b) Show that

$$\left(\frac{n}{2\pi k(n-k)}\right)^{1/2} \approx \frac{1}{\sqrt{2\pi pqn}}.$$

(c) Show that the other two factors in (10.S.1) equal

$$\frac{1}{\left(1+\frac{z_k}{nq}\right)^{np+z_k}\left(1-\frac{z_k}{nq}\right)^{nq+z_k}}. \tag{10.S.3}$$

(d) Using the approximation $\ln(1+x) = x - x^2/2$, show that the natural log (ln) of the denominator in (10.S.3) is approximately $z_k^2/(2npq)$. (Disregard higher powers of z_k.)

(e) Using (b) and (d), show that (10.S.1) is approximately

$$\frac{1}{\sqrt{2\pi npq}}e^{-\frac{(k-np)^2}{2npq}},$$

Note that in part (e) we meet the function e^{-x^2}, which appears in the formula for the normal distribution. Contrast this with Exercise 23, where e^{-x} appears.

25. Let the mass of a bacteria culture at the end of n intervals of time be C_n. If there is an adequate supply of nutrients, the mass doubles during each interval, that is, $C_{n+1} = 2C_n$. When the population is large it does not reproduce as quickly. In that case, according to the Verhulst model (1848) there is a constant K such that

$$C_{n+1} = \frac{2}{1+\frac{C_n}{K}}C_n.$$

Show that $\lim_{n\to\infty} C_n = K$. (Set $R_n = 1/C_n$.)

26. The recursion $P_{n+1} = re^{\frac{-P_n}{K}}P_n$ was introduced by W. E. Ricker in 1954 in the study of fish populations. P_n denotes the fish population at the n^{th} time interval, while r and K are constants, with r being the maximum reproduction rate.
Examine the recursion when $K = 10,000$, $P_0 = 5,000$ and (a) $r = 20$ and (b) $r = 10$. As you will see, the highly unpredictable sequence $\{P_n\}$ depends dramatically on r. Such sensitivity to r is an early example of "chaos."
References: F. C. Hoppensteadt and C. S. Peskin, *Mathematics in Medicine and the Life Sciences*, Springer, NY 1991 (p. 21)
W. E. Ricker, *Stock and Prerecruitment*, J. Fish Res. Bd., Canada, 11 (1954), pp. 559–623.

27.

Sam: I'm going to prove, using the precise definition, that if $0 < r < 1$, then $\lim_{n\to\infty} r^n = 0$.

Jane: I'll listen.

Sam: I want to show that there is an integer N such that $|r^n - 0| < \epsilon$ if $n > N$, in other words, $r^n < \epsilon$, if n is big enough. To get hold of n, I take logarithms, obtaining $n\ln(r) < \ln(\epsilon)$. Then I'll divide by $\ln(r)$.

Jane: How do you know r has a log?

Sam: Well, $r = e^{\ln(r)}$.

Jane: You mean the equation $r = e^x$ has a solution?

Sam: Sure, that's what a log is all about.

Jane: Since r is less than 1, x would be negative. May I write it as $-p$ where p is positive?

Sam: If you want to, why not?

Jane: So you're saying that r can be written as $(1/e)^p$ for some positive number p. You're assuming that no matter how small r is, there is a positive number p so that $(1/e)^p$ will equal it. Right?

Sam: Right. But why all this fuss?

Jane: To say that $(1/e)^p$ gets as small as you please is just a special case of what you're trying to prove. You're wandering in circles.

Who's right, Jane or Sam? If Sam is right, finish his proof.

Calculus is Everywhere # 13
Hubbert's Peak

In the CIE for Chapter 6, Hubbert combined calculus concepts with counting squares. Later he developed specific functions and used more techniques of calculus in "Oil and Gas Supply Modeling", NBS Special Publication 631, U.S. Department of Commerce, National Bureau of Standards, May, 1982. (NOTE: NBS is now the National Institute of Standards and Technology (NIST).)

In his approach, Q_∞ denotes the total amount of oil reserves a the time oil is first extracted and t, time. The derivative dQ/dt is the rate at which oil is extracted. $Q(t)$ denotes the amount extracted up to time t. Hubbert assumes $Q(0) = 0$ and $(dQ/dt)(0) = 0$. He wants to obtain a formula for $Q(t)$.

"The curve of dQ/dt versus Q between 0 and Q_∞ can be represented by the Maclaurin series

$$\frac{dQ}{dt} = c_0 + c_1 Q + c_2 Q^2 + c_3 Q^3 + \cdots .$$

Since, when $Q = 0$, $dQ/dt = 0$, it follows that $c_0 = 0$.

"Since the curve must return to 0 when $Q = Q_\infty$, the minimum number of terms that permit this, and the simplest form of the equation, becomes the second degree equation

$$\frac{dQ}{dt} = c_1 Q + c_2 Q^2.$$

By letting $a = c_1$ and $b = -c_2$, this can be rewritten as

$$\frac{dQ}{dt} = aQ - bQ^2.$$

"Since when $Q = Q_\infty$, $dQ/dt = 0$,

$$aQ_\infty - bQ_\infty^2 = 0$$

or

$$b = \frac{a}{Q_\infty}$$

and

$$\frac{dQ}{dt} = a\left(Q - \frac{Q^2}{Q_\infty}\right). \tag{C.13.1}$$

"This is the equation of a parabola .... The maximum value occurs when the slope is 0, or when

$$a - \frac{2a}{Q_\infty}Q = 0,$$

or

$$Q = \frac{Q_\infty}{2}.$$

"It is to be emphasized that the curve of dQ/dt versus Q does not have to be a parabola, but that a parabola is the simplest mathematical form that this curve can assume. We may regard the parabolic form as a sort of idealization for all such actual data curves."

He then points out that

$$\frac{dQ/dt}{Q} = a - \frac{a}{Q_\infty}.$$

"This is the equation of a straight line. The plotting of this straight line gives the values for its constraints Q_∞ and a."

Because the rate of production, dQ/dt, and the total amount produced up to time t, namely, $Q(t)$ and observable, the line can be drawn and its intercepts read off the graph. (The two intercepts are $(0, a)$ and $(Q_\infty, 0)$.)

Hubbert then compares this with actual data, which it approximates fairly well.

Equation (C.13.1) can be written as

$$\frac{dQ}{dt} = \frac{a}{Q_\infty} Q\left(Q_\infty - Q\right),$$

which says, "The rate of production is proportional both to the amount already produced and to the reserves $Q_\infty - Q$." This is related to the logistic equation describing bounded growth. (See Exercises 35 to 37 in Section 5.7.)

This approach, which is more formal than the one in CIE 8 at the end of Chapter 6, concludes that as Q approaches Q_∞, the rate of production will decline, approaching 0. This means the Age of Oil will end.

Chapter 11

Series

How is $\sin(\theta)$ computed? One approach might be to draw a right triangle with one angle θ, as in Figure 11.0.1. Then measure the lengths of the opposite side b and the length of the hypotenuse c and calculate b/c ("opposite over hypotenuse"). (Try it,) You are lucky if you get even two decimal places correct. Clearly this method cannot give the many decimal places a calculator displays for $\sin(\theta)$, even if you draw a gigantic triangle.

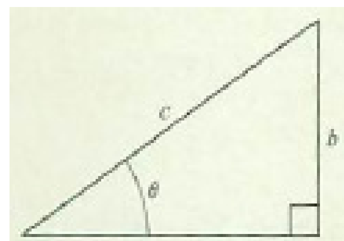


Figure 11.0.1

One way to obtain this accuracy will be described in Chapter 12. The idea is to use polynomials to evaluate important functions like $\sin(x)$, $\arctan(x)$, e^x, and $\ln(x)$ to as many decimal places as we please. For instance, when $|x| \leq 1$, the polynomial

$$x - \frac{x^3}{6} + \frac{x^5}{120}$$

approximates $\sin(x)$ with an error less than 0.0002 (provided angle x is given in radians). This means the estimate will be correct to at least three decimal places for angles less than about 57°.

1 radian $= \frac{180°}{\pi} \approx 57.29578°$

Such an estimate has other uses than simply evaluating a function. Consider the definite integral

$$\int_0^1 \frac{\sin(x)}{x}\, dx.$$

The Fundamental Theorem of Calculus is useless here since $\sin(x)/x$ does not have an elementary antiderivative. But, we can evaluate

$$\int_0^1 \frac{x - \frac{x^3}{6} + \frac{x^5}{120}}{x}\, dx = \int_0^1 \left(1 - \frac{x^2}{6} + \frac{x^4}{120}\right)\, dx.$$

Since the integrand is now a polynomial, the Fundamental Theorem of Calculus can be used to obtain the estimate

$$\left(x - \frac{x^3}{18} + \frac{x^5}{600}\right)\Bigg|_0^1 = 1 - \frac{1}{18} + \frac{1}{600} \approx 0.94611$$

which gives $\int_0^1 \sin(x)/x \, dx$ to three decimal places.

Such approximations and their errors are a central theme in this and the next chapter. Section 11.1 contains an overview of both chapters.

11.1 Informal Introduction to Series

The main goal of this chapter and the next is to show how polynomials can be used to approximate functions that are not polynomials. Table 11.1.1 shows some of the formulas we will obtain.

The larger n is, the better the approximation, as long as we keep x in the appropriate interval.

Function	Approximating Polynomial	Interval
$\frac{1}{1-x}$	$1+x+x^2+x^3+\cdots+x^n$	$\lvert x\rvert < 1$
e^x	$1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots+\frac{x^n}{n!}$	all x
$\ln(1+x)$	$x-\frac{x^2}{2}+\frac{x^3}{3}-\cdots+(-1)^{n-1}\frac{x^n}{n}$	$1 < x \leq 1$
$\sin(x)$	$x-\frac{x^3}{3!}+\frac{x^5}{5!}-\cdots+(-1)^n\frac{x^{2n+1}}{(2n+1)!}$	all x

Table 11.1.1

Example 1 illustrates the use of such polynomials.

EXAMPLE 1 Use the approximations in Table 11.1.1 to estimate $\sqrt{e}$.

$\sqrt{e} = e^{1/2}$

SOLUTION By the first row of the table, for each positive integer n,

$$1+\frac{1}{2}+\frac{\left(\frac{1}{2}\right)^2}{2!}+\frac{\left(\frac{1}{2}\right)^3}{3!}+\cdots+\frac{\left(\frac{1}{2}\right)^n}{n!}$$

is an estimate of $e^{1/2}$. Let us compute some of these estimates, keeping in mind that as n increases we expect the estimates to improve. The sums in the

n	$1+\frac{1}{2}+\frac{\left(\frac{1}{2}\right)^2}{2!}+\frac{\left(\frac{1}{2}\right)^3}{3!}+\cdots+\frac{\left(\frac{1}{2}\right)^n}{n!}$	Decimal Form	Sum
1	$1+\frac{1}{2}$	$1+0.5$	1.5
1	$1+\frac{1}{2}+\frac{\left(\frac{1}{2}\right)^2}{2!}$	$1+0.5+0.125$	1.625
1	$1+\frac{1}{2}+\frac{\left(\frac{1}{2}\right)^2}{2!}+\frac{\left(\frac{1}{2}\right)^3}{3!}$	$1+0.5+0.125+0.02083\ldots$	$1.64583\ldots$
4	$1+\frac{1}{2}+\frac{\left(\frac{1}{2}\right)^2}{2!}+\frac{\left(\frac{1}{2}\right)^3}{3!}+\frac{\left(\frac{1}{2}\right)^4}{4!}$	$1+0.5+0.125+0.02083+0.00260\ldots$	$1.6484375\ldots$

Table 11.1.2

rightmost column form a sequence that converges to $e^{1/2}$. In fact, the estimate with $n = 4$ is correct to three decimal places. ◇

There is little point in making an estimate if we have no idea about the size of its **error** — the difference between an estimate and the number we are estimating. We will focus on two closely related questions.

1. How can we estimate the error?
2. How can we choose n to achieve a prescribed accuracy, say, to 10 decimal places?

Calculus delights in resolving such battles.

Example 1 depicts a battle between two forces. On the one hand, the individual summands are getting very small — shrinking toward 0; so their sums may not get very large. On the other hand, there are more and more summands in each estimate; so their sums might become arbitrarily large.

In Example 1 the first force is stronger, and the sums — no matter how many summands we take — stay less than $\sqrt{e} \approx 1.64872$. But, in Example 2 the sums behave quite differently.

EXAMPLE 2 What happens to sums of the form

$$\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \cdots + \frac{1}{\sqrt{n}} \qquad (11.1.1)$$

as the integer n gets larger and larger? Will they stay less than some fixed number or will they get arbitrarily large, eventually passing 100, then 1,000, and so on?

SOLUTION Table 11.1.3 lists values of (11.1.1) for n up through 5.

n	$\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \cdots + \frac{1}{\sqrt{n}}$	*Decimal Form (7 places)*
1	$\frac{1}{\sqrt{1}}$	1.0000000
2	$\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}}$	1.7071068
3	$\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}}$	2.2844571
4	$\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{4}}$	2.7844571
5	$\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{4}} + \frac{1}{\sqrt{5}}$	3.2316706

Table 11.1.3

As of November 2010, the fastest computer could perform 2.57×10^{15} floating-point computations per second. Source: `http://Top500.org/`.

These computations do not answer the question: What will happen to the sums as n becomes arbitrarily large? In fact, even if we calculated the values of $1/\sqrt{1} + 1/\sqrt{2} + \cdots + 1/\sqrt{n}$ all the way to $n = 1,000,000$, we still would not know the answer. Why? Because we can't be sure what happens to the sums when n is a billion or a quadrillion or larger. Do the sums get arbitrarily large or do they stay below some fixed number? No computer, even the world's fastest supercomputer, can answer that question.

However, an algebraic insight helps us answer the question. Observe that

$$\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \cdots + \frac{1}{\sqrt{n}}$$

has n summands and that the smallest of them is $1/\sqrt{n}$. Therefore (11.1.1) is at least as large as

$$\underbrace{\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}} + \cdots + \frac{1}{\sqrt{n}}}_{n \text{ summands}} = n\left(\frac{1}{\sqrt{n}}\right) = \sqrt{n}.$$

Thus $1/\sqrt{1} + 1/\sqrt{2} + \cdots + 1/\sqrt{n}$ is at least as large as $\sqrt{n}$. (In fact, when $n \geq 2$, the sum is larger than $\sqrt{n}$.)

As n gets larger and larger, $\sqrt{n}$ grows arbitrarily large. For $n = 1{,}000{,}000$, for instance, we have

$$\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \cdots + \frac{1}{\sqrt{1{,}000{,}000}} > \sqrt{1{,}000{,}000} = 1{,}000.$$

So the sums of the form $\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \cdots + \frac{1}{\sqrt{n}}$ also become arbitrarily large. They do *not* stay less than some fixed number. ◇

WARNING *(Traveler's Advisory)* In both Examples 1 and 2, the individual summands form sequences that converge to 0:

$$\lim_{n\to\infty} \frac{\left(\frac{1}{2}\right)^n}{n!} = 0 \qquad \text{and} \qquad \lim_{n\to\infty} \frac{1}{\sqrt{n}} = 0.$$

Yet in the first case, the sums stay less than $\sqrt{e}$, while in the second the sums grow arbitrarily large. This contrast shows that we must be careful when dealing with such sums, especially since they may play a role in approximating important functions.

Summary

THINGS TO COME

In most of this chapter the summands are constants. In Chapter 12 the summands involve a variable.

§11.2 introduces the notion of a "series" as a sequence formed by adding up more and more terms from a sequence of numbers.

§§11.3–11.6 develop methods for determining when these sums converge to a number and, if they do, how big the error is when you use a particular finite sum to estimate that number.

§§12.1 and 12.2 build on Section 5.5 and apply series in various ways. Review Taylor polynomials (5.5) before reading this section.

§§12.3–12.4 shows how a series approximating one function can be used to find a series approximating a related function

§§12.5-12.6 develops complex numbers and uses thems to show that the functions $\sin(x)$ and $\cos(x)$ are intimately related to the exponential function e^x. This relation is used in physics, engineering, and mathematics.

§12.7 introduces series that are the sum of terms of the form $a_n \sin(nx)$ and $b_n \cos(nx)$ for $n = 1, 2, 3, \ldots$.

As you work through Chapters 11 and 12, check back to this outline from time to time. It will help you keep track of what you are doing, and why.

EXERCISES for Section 11.1

1. Estimate $\sqrt[3]{e} = e^{1/3}$ by using the following approximations with $x = \frac{1}{3}$.

(a) $1 + x + \dfrac{x^2}{2!} + \dfrac{x^3}{3!}$

(b) $1 + x + \dfrac{x^2}{2!} + \dfrac{x^3}{3!} + \dfrac{x^4}{4!}$

2. Estimate $1/e = e^{-1}$ using the following approximations with $x = -1$.

(a) $1 + x + \dfrac{x^2}{2!} + \dfrac{x^3}{3!} + \dfrac{x^4}{4!}$

(b) $1 + x + \dfrac{x^2}{2!} + \dfrac{x^3}{3!} + \dfrac{x^4}{4!} + \dfrac{x^5}{5!}$

3. As shown in Section 5.5 the polynomial $x - x^3/6$ is an excellent approximation to $\sin(x)$ (angle measured in radians) for $|x| \leq \frac{1}{2}$. Using a calculator or computer, fill in Table 11.1.4 to seven decimal places.

x	$\sin(x)$	$x - \frac{x^3}{6}$	$\sin(x) - \left(x - \frac{x^3}{6}\right)$
0.1			
0.2			
0.3			
0.4			
0.5			

Table 11.1.4

The results illustrate that for these values of x the estimates are accurate to at least three decimal places.

4. The polynomial $x - x^3/3! + x^5/5!$ is an excellent approximation to $\sin(x)$ (angle in radians) for $|x| \leq 1$. Using a calculator or computer, in (a) and (b) evaluate the expression to at least seven decimal places.

(a) $\sin(1)$,

(b) $x - x^3/3! + x^5/5!$ when $x = 1$.

(c) To how many decimal places do these results agree?

5. Estimate $\int_{1/2}^{1}(e^x-1)/x\,dx$ by approximating e^x by the polynomial

(a) $1+x+\frac{x^2}{2!}$,

(b) $1+x+\frac{x^2}{2!}+\frac{x^3}{3!}$.

(c) The exact value of this definite integral, to seven decimal places, is 0.7477507. To how many decimal places do each of these results agree with the exact value?

6. Estimate $\int_{1/4}^{1/2}\sin(x)/x\,dx$ by approximating $\sin(x)$ by the polynomial

(a) x.

(b) $x-\frac{x^3}{3!}$.

(c) $x-\frac{x^3}{3!}+\frac{x^5}{5!}$.

(d) The exact value of this definite integral, to seven decimal places, is 0.2439738. To how many decimal places do each of these results agree with the exact value?

7.

(a) The polynomial $x-x^2/2+x^3/3-\cdots+(-1)^{n-1}x^n/n$, $|x|\leq 1$, is a good estimate of $\ln(1+x)$ when n is large. So, to estimate $\ln(1.5)$, which is $\ln(1+0.5)$, we use the polynomial with x replaced by $\frac{1}{2}$. Use a calculator or computer to fill in Table 11.1.5.

n	$\frac{1}{2}-\left(\frac{1}{2}\right)^2/2+\left(\frac{1}{2}\right)^3/3-\cdots+(-1)^{n-1}\left(\frac{1}{2}\right)^n/n$	Decimal Form
1		
2		
3		
4		
5		

Table 11.1.5

(b) Use your calculator or a computer to compute $\ln(1.5)$.

(c) What is the error between this approximation and the result for $n=5$ in Table 11.1.5?

8. (See Exercise 7.)

(a) To estimate $\ln(0.5)$, write it as $\ln(1 + (\frac{-1}{2}))$. Fill in Table 11.1.6.

n	$\left(\frac{-1}{2}\right) - \left(\frac{-1}{2}\right)^2/2 + \left(\frac{-1}{2}\right)^3/3 - \cdots + (-1)^{n-1}\left(\frac{-1}{2}\right)^n/n$	Decimal Form
1		
2		
3		
4		
5		

Table 11.1.6

(b) Use your calculator or a computer to compute $\ln(0.5)$.

(c) What is the error between this approximation and the result for $n = 5$ in Table 11.1.6?

9. One way to approximate $\ln(2)$ is to write it as $\ln(1 + 1)$ and use a polynomial in Exercise 7 that approximates $\ln(1 + x)$ with $x = 1$. Another way is to note that $\ln(2) = -\ln(0.5)$ and use the approach of Exercise 8. Using the polynomial approximation of degree 5 ($n = 5$) in both cases, decide which gives the better estimate.

10. What happens to sums of the form

$$\frac{1}{\sqrt[3]{1}} + \frac{1}{\sqrt[3]{2}} + \frac{1}{\sqrt[3]{3}} + \cdots + \frac{1}{\sqrt[3]{n}}$$

as n gets larger? Explore and explain.

11.

(a) Using results from Section 1.4, show that, for $x \neq 1$,

$$1 + x + x^2 + \cdots + x^{n-1} = \frac{1}{1-x} - \frac{x^n}{1-x}. \tag{11.1.2}$$

(b) Now assume that $|x| < 1$. Then x^n approaches 0 as n increases (as was shown in Section 10.1). Thus, for $|x| < 1$ and large n, $1 + x + x^2 + \cdots + x^{n-1}$ is a polynomial approximation for the function $1/(1 - x)$.

(c) Compute $1 + x + x^2 + \cdots + x^{n-1}$ for $n = 6$ and $x = 0.3$. How much does this differ from $1/(1 - 0.3)$?

(d) The same as (c), with $x = -0.9$.

Exercises 12 and 13 use (11.1.2) to derive polynomial approximations to $\ln(1+x)$ and $\arctan(x)$. These two problems both start from the same idea. We begin by expressing (11.1.2) in the form

$$\frac{1}{1-t} = 1 + t + t^2 + t^3 + \cdots + t^{n-1} + \frac{t^n}{1-t} \qquad (t \neq 1).$$

Replace t with $-t$, getting

$$\frac{1}{1+t} = 1 - t + t^2 - t^3 + \cdots + (-1)^{n-1}t^{n-1} + \frac{(-1)^n t^n}{1+t} \qquad (t \neq -1). \qquad (11.1.3)$$

12. This exercise derives the sequence of polynomial approximations to $\ln(1+x)$ listed in Table 11.1.1 on page 919.

(a) Integrate both sides of (11.1.3) over the interval from 0 to x, $x > 0$, to show that

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + \frac{(-1)^{n-1}x^n}{n} + (-1)^n \int_0^x \frac{t^n}{1+t}\,dt.$$

(b) Show that for $0 \leq x \leq 1$, $\int_0^x (t^n/(1+t))\ dt$ approaches 0 as n increases. ($1/(1+t) \leq 1$ for $t \geq 0$.)

13. This exercise obtains a sequence of polynomials that approximate $\arctan(x)$ for $|x| \leq 1$ and shows one way of computing π. The key is that $\frac{d}{dx}\arctan(x) = \frac{1}{1+x^2}$. To begin, replace t by $-t^2$ in (11.1.2) to obtain

$$\frac{1}{1+t^2} = 1 - t^2 + t^4 - t^6 + \cdots + (-1)^{n-1}t^{2n-2} + \frac{(-1)^n t^{2n}}{1+t^2} \qquad (\text{for all } t). \quad (11.1.4)$$

(a) Consider only $0 \leq x \leq 1$. Integrate both sides of (11.1.4) over $[0, x]$ to show that

$$\arctan(x) = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots + (-1)^{n-1}\frac{x^{2n-1}}{2n-1} + (-1)^n \int_0^x \frac{t^{2n}}{1+t^2}\,dt. \qquad (11.1.5)$$

(b) Show that for fixed x, $0 < x < 1$, the integral in (11.1.3) approaches 0 as $n \to \infty$.

(c) Use the polynomial in (a), with $n = 5$ (so its degree is 9) to estimate arctan(1).

(d) Use the result in (c) to estimate π. $(\arctan(1) = \frac{\pi}{4})$

14. In this exercise we will see what happens to sums of the form

$$\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \cdots + \frac{1}{n(n+1)}$$

as n increases. Do these sums get arbitrarily large or do they approach some number?

n	$\frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} + \cdots + \frac{1}{n(n+1)}$	Sum, as fraction	Sum, as decimal
1			
2			
3			
4			
5			

Table 11.1.7

(a) Fill in at least 5 rows of Table 11.1.7. Add more rows if you wish.

(b) On the basis of your computations, what do you think happens to the sums as n increases. (If you don't see a pattern, go up to $n = 10$.)

(c) Justify your opinion in (b).

15.

(a) Use the polynomial in (11.1.5), with $n = 5$, to estimate $\arctan\left(\frac{1}{2}\right)$ in radians. Then translate the answer into degrees.

(b) Use the result in (a) to estimate arctan(2) in radians. (For positive x, what is the relation between $\arctan(1/x)$ and $\arctan(x)$?)

(c) Draw a right triangle with one leg 20 cm long and the other 10 cm; use it and a protractor to estimate arctan(2).

(d) What does your calculator or computer give as an estimate of arctan(2)?

(e) To how many decimal places does the estimate in (b) agree with the value found in (d)? To how many decimal places does the measurement in (c) agree with the value found in (d)?

11.2 Series

The goal of this section is to introduce sequences formed by adding up more and more terms of a given sequence.

Series

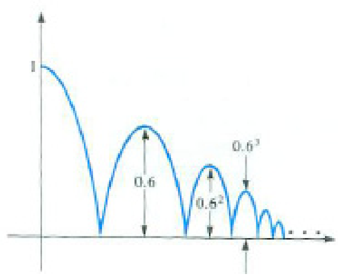


Figure 11.2.1

Similar sums arise in many applications. Exercise 30 is an application to medicine and Exercise 31 presents an example from economics.

Consider a tennis ball that is dropped from a height of 1 meter. It rebounds 0.6 meter. It continues to bounce, and each fall is 60% as high as the previous fall. (See Figure 11.2.1.) What is the total distance the ball falls?

The third fall is $(0.6)^2$ meter, the next is $(0.6)^3$ meter, and so on. In general, the n^{th} time the ball falls, it falls a distance $(0.6)^{n-1}$ meter. While it is clear this geometric sequence converges to zero, we are more interested in the question:

"What happens to the sum $\qquad 1+0.6+(0.6)^2+\cdots+(0.6)^n \qquad$ as $n \to \infty$?"

Example 1 explores this question.

EXAMPLE 1 Given the geometric progression 1, 0.6, $(0.6)^2$, $(0.6)^3$, ..., form a new sequence $\{S_n\}$ as follows:

$$\begin{aligned} S_1 &= 1, \\ S_2 &= 1+0.6, \\ S_3 &= 1+0.6+(0.6)^2, \end{aligned}$$

and, in general,

$$S_n = 1+0.6+(0.6)^2+\cdots+(0.6)^{n-1}.$$

Each S_n is the sum of n terms of the sequence $\{a_n\}$ with $a_n = 0.6^n$ for $n = 0$, 1, 2, ... Does the sequence $\{S_n\}$ converge or diverge? If it converges, what is the limit?

SOLUTION Because S_n is the sum of the first n terms in a geometric sequence whose first term is 1 and whose ratio is 0.6 we have

$$S_n = \frac{1-(0.6)^n}{1-0.6}.$$

Thus

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} \frac{1-(0.6)^n}{1-0.6} = \frac{1}{1-0.6} = 2.5.$$

◇

The rest of this section expands upon the ideas introduced in Example 1.

Let $a_1, a_2, a_3, \ldots, a_n, \ldots$ be a sequence. From this sequence a new sequence $S_1, S_2, S_3, \ldots, S_n, \ldots$ can be formed:

$$\begin{aligned}
S_1 &= a_1 = \sum_{k=1}^{1} a_k, \\
S_2 &= a_1 + a_2 = \sum_{k=1}^{2} a_k, \\
S_3 &= a_1 + a_2 + a_3 = \sum_{k=1}^{3} a_k, \\
&\vdots \\
S_n &= a_1 + a_2 + a_3 + \cdots + a_n = \sum_{k=1}^{\infty} a_k.
\end{aligned}$$

The sequence of sums, $S_1, S_2, S_3, \ldots, S_n, \ldots$, is called the **series** obtained from the sequence $a_1, a_2, a_3, \ldots, a_n, \ldots$. It can also be defined by the recursion, $S_{n+1} = S_n + a_{n+1}$.

Traditionally, $\{S_n\}$ is referred to as "the series whose n^{th} term is a_n." Common notations for the sequence $\{S_n\}$ are $\sum_{k=1}^{\infty} a_k$ and $a_1 + a_2 + a_3 + \cdots + a_k + \cdots$. The sum

$$S_n = a_1 + a_2 + a_3 + \cdots + a_n = \sum_{k=1}^{n} a_k$$

is called a **partial sum** or the $n^{\mathbf{th}}$ **partial sum**. If the sequence of partial sums of a series converges to L, then L is called the **sum** of the series and the series is said to be **convergent**. We write

$$\lim_{n \to \infty} S_n = L.$$

Frequently one writes $L = a_1 + a_2 + \cdots + a_n + \cdots$. Remember, however, that we do *not* add an infinite number of terms; we take the limit of finite sums. A series that is not convergent is called **divergent**.

Only finitely many summands are ever added up.

A Note on Notation Starting with the sequence $a_1, a_2, \ldots, a_n, \ldots$, we form a new sequence, $S_1, S_2, \ldots, S_n, \ldots$, whose terms are the partial sums $S_1 = a_1$, $S_2 = a_1 + a_2$, $\ldots$, $S_n = a_1 + a_2 + \cdots + a_n$. The symbol

$$\sum_{k=1}^{\infty} a_k$$

is short for this sequence $S_1, S_2, \ldots, S_n, \ldots$. If the sequence of partial sums converges to a number L, we also write

$$\sum_{k=1}^{\infty} a_k = L.$$

The symbol $\sum_{k=1}^{\infty} a_k$ has two meanings.

So the symbol $\sum_{k=1}^{\infty} a_k$ stands for two different concepts: a sequence of partial sums and also, if that sequence converges, for its limit. This limit is called the "sum" of the series.

So, in Example 1, we investigated the *series*

$$\sum_{k=1}^{\infty} 0.6^{k-1},$$

namely, the sequence of partial sums 1, $1+0.6$, $1+0.6+0.6^2$, $\dots$, $1+0.6+0.6^2+\cdots+(0.6)^{n-1}$. This sequences converges to 2.5. That permits us to write

$$\sum_{k=1}^{\infty}(0.6)^{k-1} = 2.5,$$

which says, "The series $\sum_{k=1}^{\infty}(0.6)^{k-1}$ converges to the number 2.5." We also say, for the sake of brevity, "Its sum is 2.5."

Just as a sequence need not start with a_1, a series can start with any term, such as a_0 or a_k, and we would write $\sum_{k=0}^{\infty} a_k$ or $\sum_{i=1}^{\infty} a_i$ or $\sum_{j=k}^{\infty} a_j$. Notice that there is nothing special about the index for a series. The most common indices are n, k, j, and i.

Geometric Series

Example 1 concerns the series whose n^{th} term is $(0.6)^{n-1}$:

$$S_n = 1 + 0.6 + 0.6^2 + \cdots + 0.6^{n-1}.$$

Geometric sums with a finite number of terms are discussed in Section 5.5.

It is a special case of a geometric series, which will now be defined.

> **DEFINITION** *(Geometric Series)* Let a and r be real numbers. The series
>
> $$a + ar + ar^2 + \cdots + ar^{n-1} + \cdots$$
>
> is called the 'geometric series with initial term a and ratio r.

The series in Example 1 is a geometric series with initial term 1 and ratio 0.6.

Theorem 11.2.1. *If $-1 < r < 1$, the geometric series*

$$a + ar + ar^2 + \cdots + ar^{n-1} + \cdots \quad \textit{converges to} \quad \frac{a}{(1-r)}.$$

Proof

See Exercise 11 in Section 11.1.

Let S_n be the sum of the first n terms: $S_n = a + ar + \cdots + ar^{n-1}$. The formula for the finite geometric sum is $S_n = \frac{a(1-r^n)}{1-r}$. Since $-1 < r < 1$, $\lim_{n\to\infty} r^n = 0$. Thus

See also Exercise 28.

$$\lim_{n\to\infty} S_n = \frac{a}{1-r},$$

proving the theorem, obtained in Section 1.4. •

The series in Example 1 is a geometric series with first term 1 and ratio $r = 0.6$. It converges and has the sum

$$\frac{1}{1-0.6} = \frac{1}{0.4} = 2.5.$$

The n^{th} Term Test for Divergence

Theorem 11.2.1 says nothing about geometric series in which $r \geq 1$ or $r \leq -1$. The next theorem, which concerns series in general, not just geometric series, will be useful in settling these cases.

Theorem 11.2.2 (n^{th}-Term Test for Divergence.)**.** *If* $\lim_{n\to\infty} a_n \neq 0$, *then the series* $a_1 + a_2 + \cdots + a_n + \cdots$ *diverges. (The same conclusion holds if* $\{a_n\}$ *has no limit.)*

We take an indirect approach.

Proof

Assume that the series $a_1 + a_2 + \cdots + a_n + \cdots$ converges. Since S_n is the sum $a_1 + a_2 + \cdots + a_n$, while S_{n-1} is the sum of the first $n-1$ terms, it follows that $S_n = S_{n-1} + a_n$, or

$$a_n = S_n - S_{n-1}.$$

Because we have assumed the series converges, let $S = \lim_{n\to\infty} S_n$. Then we also have $S = \lim_{n\to\infty} S_{n-1}$, since S_{n-1} runs through the same numbers as S_n. Thus

$$\begin{aligned}
\lim_{n\to\infty} a_n &= \lim_{n\to\infty} (S_n - S_{n-1}) \\
&= \lim_{n\to\infty} S_n - \lim_{n\to\infty} S_{n-1} \\
&= S - S \\
&= 0.
\end{aligned}$$

This proves the theorem. •

If a series converges, its n^{th}-term must approach 0.

The n^{th}-Term Test for Divergence implies that if $a \neq 0$ and $r \geq 1$, the geometric series

$$a + ar + \cdots + ar^{n-1} + \cdots$$

diverges. For instance, if $r = 1$,

$$\lim_{n\to\infty} ar^n = \lim_{n\to\infty} a1^n = a,$$

which is not 0. If $r > 1$, then r^n gets arbitrarily large as n increases; hence $\lim_{n\to\infty} ar^n$ does not exist. Similarly, if $r \le -1$, $\lim_{n\to\infty} ar^n$ does not exist. The above results and Theorem 11.2.1 can be summarized by this statement: The geometric series

$$\sum_{i=1}^{\infty} ar^{i-1} = a + ar + ar^2 + \cdots + ar^{n-1} + \cdots .$$

for $a \neq 0$, converges if and only if $|r| < 1$.

Warning: Even if the n^{th} term approaches 0, the series still can diverge.

The n^{th}-Term Test for Divergence tells us that *if* the series $a_1+a_2+a_3+\cdots$ converges, *then* a_n approaches 0 as $n \to \infty$. The converse of this statement is *not true*. If a_n approaches 0 as $n \to \infty$, it does *not* follow that the series $a_1 + a_2 + a_3 + \cdots$ converges. Be careful to make this distinction.

Recall the series

$$\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{1}} + \cdots + \frac{1}{\sqrt{n}} + \cdots$$

discussed in Example 2 in Section 11.1. Even though its n^{th} term approaches 0 as $n \to \infty$, the sums get arbitrarily large. The n^{th} term approaches 0 so "slowly" that the sums S_n get arbitrarily large.

The harmonic series was so named by the Greeks because of the role of $1/n$ in musical harmony.

Nicole Oresme, 1323–1382, one of the most influential philosophers of the Middle Ages, `http://en.wikipedia.org/wiki/Nicole_Oresme`

In the next example, the n^{th} term approaches 0 much faster than $1/\sqrt{n}$ does. Still, the series diverges. The series in this example is called the **harmonic series**. The argument that it diverges is due to the French mathematician Nicolas of Oresme, who presented it about the year 1360.

EXAMPLE 2 Show that the harmonic series $1/1 + 1/2 + \cdots + 1/n + \cdots$ diverges.

SOLUTION Collect the summands in longer and longer groups. Except for the first two terms, each group contains twice the number of summands as it predecessor:

$$1 + \underbrace{\frac{1}{2}}_{\text{1 term}} + \underbrace{\frac{1}{3} + \frac{1}{4}}_{\text{2 terms}} + \underbrace{\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}}_{\text{4 terms}} + \underbrace{\frac{1}{9} + \frac{1}{10} + \cdots + \frac{1}{16}}_{\text{8 terms}} + \cdots .$$

The sum of the terms in each group is at least $\frac{1}{2}$. For instance,

$$\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} > \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{4}{8} = \frac{1}{2}$$

and

$$\frac{1}{9} + \frac{1}{10} + \cdots + \frac{1}{16} > \frac{1}{16} + \frac{1}{16} + \cdots + \frac{1}{16} = \frac{8}{16} = \frac{1}{2}$$

Since the repeated addition of $\frac{1}{2}$'s produces sums as large as we please, the series diverges. ◇

If the series $a_1 + a_2 + \cdots + a_n + \cdots$ converges, it *follows* that $a_n \to 0$. However, if $a_n \to 0$, it *does not follow* that $a_1 + a_2 + \cdots + a_n + \cdots$ converges. Indeed, there is no general, practical rule for determining whether a series converges or diverges. Fortunately, a few rules suffice to decide on the convergence or divergence of the most common series. They will be presented in this chapter.

An important moral: The n^{th}-term test is only a test for divergence.

Because convergence or divergence of a series is decided by looking at the convergence or divergence of the sequence of partial sums, the basic properties for sequences are also true for series.

Exercise 36 asks for the proof.

Theorem 11.2.3. *A. If $\sum_{i=1}^{\infty} a_i$ is a convergent series with sum L and if c is a number, then $\sum_{i=1}^{\infty} ca_i$ is convergent and has the sum cL.*

B. If $\sum_{k=1}^{\infty} b_i$ is a convergent series with sum M, then $\sum_{k=1}^{\infty}(a_n + b_n)$ is a convergent series with sum $L + M$.

Keep in mind that you can disregard any finite number of terms when deciding whether a series is convergent or divergent. If you delete a finite number of terms from a series and what is left converges, then the series you started with converges. Another way to look at this is to note that a "front end," $a_1 + a_2 + \cdots + a_n$. does not influence convergence or divergence. It is rather a "tail end," $a_{n+1} + a_{n+2} + \cdots$ that matters. The sum of the series is the sum of any tail end plus the sum of the corresponding front end; that is, for any positive integer m,

Front ends do not affect convergence.

$$\sum_{k=1}^{\infty} a_k = \underbrace{\sum_{k=1}^{m} a_k}_{\text{front end}} + \underbrace{\sum_{k=m+1}^{\infty} a_k}_{\text{tail end}} .$$

Suppose that $\sum_{i=1}^{\infty} p_i$ is a series with positive terms and you can show that there is a number B such that every partial sum $S_1 = p_1$, $S_2 = p_1 + p_2$, $\ldots$, $S_n = p_1 + p_2 + \cdots + p_n$, is less than or equal to B. By Theorem 10.1.1 of Section 10.1, they have a limit L, which is less than or equal to B. (See Figure 11.2.2.) This means that $\sum_{k=1}^{\infty} p_i$ is convergent (and its sum is less than or equal to B). This observation will be useful in establishing the convergence of a series of non-negative terms, even though it does not tell us the exact sum of the series.

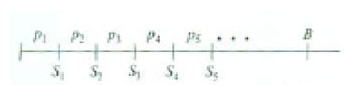


Figure 11.2.2

A similar statement holds for the series $\sum_{k=1}^{\infty} a_i$ in which $a_i \leq 0$ for all n. If there is a number A such that each partial sum is greater than or equal to A, then the series converges and its sum is greater than or equal to A.

Example 3 introduces a series that is representative of many series that arise in the study of $\sin(x)$, $\cos(x)$, and e^x.

EXAMPLE 3 Does the series defined by $\sum_{k=0}^{\infty} \frac{2^k}{k!}$ converge or diverge?

SOLUTION First, note that the first index is $k = 0$, not $k = 1$. This has no bearing on the convergence or divergence of this series (it's part of the front end), but it does affect the value of the series (assuming it converges).

Define $a_k = 2^k/k!$ for $k = 0, 1, 2, \ldots$. The partial sums of the series are $S_n = \sum_{k=0}^{n} a_k$ for $n = 0, 1, 2, \ldots$. From the relation $S_{n+1} = S_n + a_{n+1}$ and the fact that a_{n+1} is positive, we see that $\{S_n\}$ is an increasing sequence.

By the same reasoning used in Section 5.5, we can conclude that for $k > 2$,

$$a_k < \frac{2}{1}\frac{2}{2}\left(\frac{2}{3}\right)^{k-2}$$

This observation that the terms of the series are bounded by the terms of a convergent geometric series is the key to concluding that the partial sums of this series are bounded. For $n \geq 2$:

$$S_n = \sum_{k=0}^{n} a_k = a_0 + a_1 + \sum_{k=2}^{n} a_k < 1 + 2 + \sum_{k=2}^{n} 2\left(\frac{2}{3}\right)^{k-2}.$$

Adding the rest of the terms of the geometric series with first term 2 and ratio 2/3, we conclude that

$$S_n < 1 + 2 + \sum_{k=2}^{\infty} 2\left(\frac{2}{3}\right)^{k-2} = 1 + 2 + \frac{2}{1 - \frac{2}{3}} = 1 + 2 + 6 = 9.$$

Thus, the series $\sum_{k=0}^{\infty} \frac{2^k}{k!}$ converges because the sequence of partial sums for the series is monotone and bounded above (by 9). The actual value of this limit will be found later. It is $e^2 \approx 7.38906$. ◇

The same ideas can be used to prove that $\sum_{K+!}^{\infty} \frac{k^n}{k!}$, for any positive number k, converges.

Summary

Given any sequence $\{a_k\}$ we can form a new sequence $\{S_n\}$,where S_n is the sum of the first n terms of $\{a_k\}$, $S_n = a_1 + a_2 + \cdots + a_n$. The new sequence is called the "series" derived from the original sequence $\{a_k\}$. If the series converges, then a_k must approach 0 as $k \to \infty$. (The converse is not true.) It follows that if a_k does not approach 0 as $k \to \infty$, then the series $a_1 + a_2 + \cdots + a_n + \cdots$ diverges.

If $a_k = ar^{k-1}$, where $|r| < 1$, we obtain the geometric series $\sum_{k=0}^{\infty} ar^k$, which converges to $a/(1 - r)$.

If, for each index, a_k is non-negative and $a_1 + a_2 + \cdots + a_k \leq B$ for some fixed number B for all k, then $\sum_{k=1}^{\infty} a_k$ is convergent and approaches a number no larger than B. This principle was used in this section to show that $\sum_{k=0}^{\infty} \frac{2^k}{k!}$ converges.

EXERCISES for Section 11.2

Exercises 1 to 4 each concern a series $\sum_{k=1}^{\infty} a_k$ and the sequence of its partial sums $\{S_n\}$. (Based on suggestions by James T. Vance Jr.)

1. Suppose you know that $a_n \to 0$ as $n \to \infty$. Which of the following statements are true. (More than one may be true.)

(a) The series definitely converges.

(b) The series definitely diverges.

(c) There is not enough information to decide whether the series diverges or converges.

(d) More information is needed to determine the sum of the series.

(e) $S_n \to 0$ as $n \to \infty$.

(f) $\sum_{k=1}^{\infty} a_k = 0$.

2. Suppose you know that $a_n \to 6$ as $n \to \infty$. Which of the following statements are true. (More than one may be true.)

(a) The series definitely converges.

(b) The series definitely diverges.

(c) There is not enough information to decide whether the series diverges or converges.

(d) More information is needed to determine the sum of the series.

(e) $S_n \to 0$ as $n \to \infty$.

(f) $\displaystyle\sum_{k=1}^{\infty} a_k = 6$.

3. Suppose you know that $S_n \to 3$ as $n \to \infty$. Which of the following statements are true. (More than one may be true.)

(a) The series definitely converges.

(b) The series definitely diverges.

(c) There is not enough information to decide whether the series diverges or converges.

(d) More information is needed to determine the sum of the series.

(e) The sum of the series is 3.

(f) $\displaystyle\sum_{k=1}^{\infty} a_k = 3$.

(g) $\displaystyle\lim_{k\to\infty} a_k = 0$.

4. Suppose you know that $S_n = n/(n+1)$. Which of the following statements are true. (More than one may be true.)

(a) The series definitely converges.

(b) The k^{th} term of the series diverges.

(c) The k^{th} term of the series converges.

(d) The k^{th} term of the series is $1/(k(k+1))$.

(e) The series is a geometric series.

5. This exercise concerns the series $\displaystyle\sum_{k=1}^{\infty} 5(-1/2)^k$.

(a) Express the fourth term of this series as a decimal.

(b) Express the fourth partial sum of this series as a decimal.

(c) Find the limit as $k \to \infty$ of the k^{th} term of the series.

(d) Find the limit as $n \to \infty$ of the n^{th} partial sum of the series.

(e) Does the series converge? If so, what is its sum?

6. This exercise concerns the series $\displaystyle\sum_{k=1}^{\infty} 3(1/10)^k$.

(a) Express the third term of this series as a decimal.

(b) Express the third partial sum of this series as a decimal.

(c) Find the limit as $k \to \infty$ of the k^{th} term of the series.

(d) Find the limit as $n \to \infty$ of the n^{th} partial sum of the series.

(e) Does the series converge? If so, what is its sum?

In Exercises 7 to 14 determine whether the given geometric series converges. If it does, find its sum.

7. $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots + \left(\frac{1}{2}\right)^{k-1} + \cdots$

8. $1 - \frac{1}{3} + \frac{1}{9} - \frac{1}{27} + \cdots + \left(\frac{-1}{3}\right)^{k-1} + \cdots$

9. $\displaystyle\sum_{k=1}^{\infty} 10^{-k}$

10. $\displaystyle\sum_{k=1}^{\infty} 10^{k}$

11. $\displaystyle\sum_{k=1}^{\infty} 5(0.99)^{k}$

12. $\displaystyle\sum_{k=1}^{\infty} 7(-1.01)^{k}$

13. $\displaystyle\sum_{k=1}^{\infty} 4\left(\frac{2}{3}\right)^{k}$

14. $\frac{-3}{2} + \frac{3}{4} - \frac{3}{8} + \cdots + \frac{3}{(-2)^k} + \cdots$

In Exercises 15 to 22 determine whether the given series converge or diverge. Find the sums of the convergent series.

15. $-5 + 5 - 5 + 5 - \cdots + (-1)^k 5 + \cdots$

16. $\sum_{k=1}^{\infty} \frac{1}{(1+(1/k))^k}$

17. $\displaystyle\sum_{k=1}^{\infty} \frac{2}{k}$

18. $\displaystyle\sum_{k=1}^{\infty} \frac{k}{2k+1}$

19. $\displaystyle\sum_{k=1}^{\infty} 6\left(\frac{4}{5}\right)^{k}$

20. $\displaystyle\sum_{k=1}^{\infty} 100\left(\frac{-8}{9}\right)^{k}$

21. $\displaystyle\sum_{k=1}^{\infty} \left(2^{-k} + 3^{-k}\right)$

22. $\displaystyle\sum_{k=1}^{\infty} \left(4^{-k} + k^{-1}\right)$

23. What is the total distance traveled — both up and down — by the ball described in the opening paragraph of this section?

24. A rubber ball, when dropped on concrete, rebounds 90 percent of the distance it falls. If it is dropped from a height of 6 feet, how far does it travel — both up and down — before coming to rest?

25. The repeating decimal

$$3.171717\ldots,$$

where the 17's continue forever, can be viewed as 3 plus a geometric series:

$$3 + \frac{17}{100} + \frac{17}{100^2} + \frac{17}{100^3} + \cdots .$$

Using the formula for the sum of a geometric series, write the decimal as a fraction.

26. (See Exercise 25.) Evaluate the repeating decimal $0.3333\cdots$.

27. (See Exercise 25.) Evaluate the repeating decimal $4.1256256256\ldots$ (with 256 repeating).

28. Show that if $|r| < 1$, the sum of the geometric series $a + ar + ar^2 + \cdots$ differs from S_n by $ar^n/(1-r)$.

29. This is a quote from an economics text: "The present value of the land, if a new crop is planted at time t, $2t$, $3t$, etc., is

$$P = g(t)e^{-rt} + g(t)e^{-2rt} + g(t)e^{-3rt} + \cdots .$$

By the formula for the sum of a geometric series,

$$P = \frac{g(t)e^{-rt}}{1 - e^{-rt}}."$$

Check that the missing step, which simplified the formula for P, was correct.

30. A patient takes A grams of a certain medicine every 6 hours. The amount of each dose active in the body t hours later is Ae^{-kt} grams, where k is a positive constant and time is measured in hours.

(a) Show how immediately after taking the medicine for the n^{th} time, the amount active in the body is

$$S_n = A + Ae^{-6k} + Ae^{-12k} + \cdots + Ae^{-6(n-1)k}.$$

(b) If, as $n \to \infty$, $S_n \to \infty$, the patient would be in danger. Does $S_n \to \infty$? If not, what is $\lim_{n\to\infty} S_n$?

(See also Exercise 117 in the Chapter 5 Summary.)

31. Deficit spending by the federal government inflates the nation's money supply. However, much of the money paid out by the government is spent in turn by those who receive it, thereby producing additional spending. This produces a chain reaction, called by economists the *multiplier effect*. It results in much greater total spending than the government's original expenditure. To be specific, suppose the government spends 1 billion dollars and that the recipients of that expenditure in turn spend 80 percent while retaining 20 percent. Let S_n be the *total* spending generated after n transactions in the chain, 80 percent of receipts being expended at each step.

(a) Show that $S_n = 1 + 0.8 + 0.8^2 + \cdots + 0.8^{n-1}$ billion dollars.

(b) Show that as n increases, the total spending approaches 5 billion dollars. (In this case the multiplier is 5.)

(c) What would the total spending be if 90 percent of receipts is spent at each step instead of 80 percent?

The subprime mortgage foreclosures in 2008 caused a similar ripple effect, threatening a recession.

32. Assume a ball falls $16t^2$ feet in t seconds and bounces upward when it hits the ground. Assume the upward part of a bounce takes as long as the subsequent fall. How long does the ball in Exercise 24 bounce?

Exercises 33 to 35 are related to the following question: A gambler tosses a coin until a head appears. On the average, how many times does she toss it to get a head?

33.

(a) Repeat this experiment 10 times. Each run consists of tossing a coin until a head appears. Average the lengths of the 10 trials.

(b) The probability of a run of length one is $\frac{1}{2}$, since a head must appear on the first toss. The probability of a run of length two is $\left(\frac{1}{2}\right)^2$. The probability of having a head appear for the first time on toss k is $\left(\frac{1}{2}\right)^k$. It is shown in probability theory that the average number of tosses to get a head is $\sum_{k=1}^{\infty} \frac{k}{2^k}$. This is a theoretical average approached as the experiment is repeated many times. Compute $\sum_{k=1}^{8} \frac{k}{2^k}$.

34. Oresme, around the year 1360, summed the series $\sum_{k=1}^{\infty} \frac{k}{2^k}$ by drawing the endless staircase shown in Figure 11.2.3, in which each stair after the first has width 1 and is half as high as the stair immediately to its left.

(a) By looking at the staircase in two ways, show that

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = \frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \cdots .$$

(b) Use (a) to sum $\displaystyle\sum_{k=1}^{\infty} \frac{k}{2^k}$.

(c) Use the same idea to find $\displaystyle\sum_{k=1}^{\infty} kp^k$, when $0 < p < 1$.

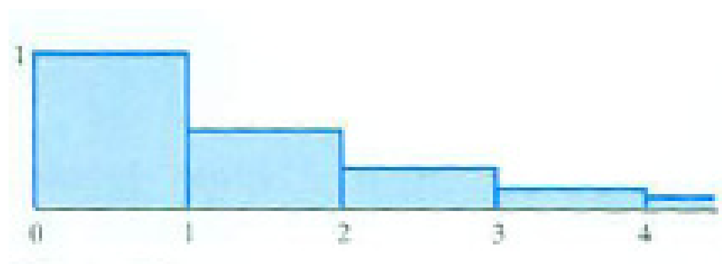


Figure 11.2.3

35.

(a) Using your calculator compute enough partial sums of the series $\displaystyle\sum_{k=1}^{\infty} k3^{-k}$ to offer an opinion as to whether it converges or diverges.

(b) Show that it converges. (The coefficient k is less than 2^k.)

(c) On the basis of (a), what do you think its sum is?

36. Use the precise definition of convergence from Section 10.2 to prove each of the following statements:

(a) If c is a number and $\sum_{k=1}^{\infty} a_k$ is a convergent series with sum L, then $\sum_{k=1}^{\infty} ca_k$ is a convergent series with sum cL.

(b) If $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=1}^{\infty} b_k$ are convergent series with sums L and M, respectively, then $\sum_{k=1}^{\infty}(a_k + b_k)$ is a convergent series with sum $L + M$.

11.3 The Integral Test

In this section we use integrals of the form $\int_a^\infty f(x)\,dx$ to establish convergence or divergence of series whose terms are positive and decreasing. Furthermore, we obtain a way of analyzing the error when we use a partial sum to estimate the sum of the series.

The Integral Test

Let $f(x)$ be a decreasing positive function. We obtain a sequence from $f(x)$ by defining a_n to be $f(n)$. For instance, the sequence $1/1$, $1/2$, $1/3$, $\ldots$, $1/n$, $\ldots$ is obtained from the function $f(x) = 1/x$. It turns out that the convergence (or divergence) of the series $\sum_{k=1}^{\infty} a_k$ is closely connected with the convergence (or divergence) of the improper integral $\int_1^\infty f(x)\,dx$. This connection is expressed in the following theorem:

Theorem 11.3.1 (Integral Test)**.** *Let $f(x)$ be a continuous decreasing function such that $f(x) > 0$ for $x \geq 1$. Let $a_n = f(n)$ for each positive integer n. Then*

A. If $\int_1^\infty f(x)\,dx$ is convergent, then so is the series $\sum_{k=1}^{\infty} a_k$.

B. If $\int_1^\infty f(x)\,dx$ is divergent, then so is the series $\sum_{k=1}^{\infty} a_k$.

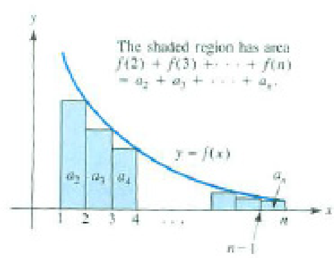


Figure 11.3.1

Proof

Figures 11.3.1 and 11.3.2 are the key to the proof. Note how the rectangles are constructed in each case.

In Figure 11.3.1 the rectangles lie *below* the curve $y = f(x)$. Each rectangle has width 1. Comparing the staircase area with the area under the curve gives the inequality

$$a_2 + a_3 + \cdots + a_n < \int_1^n f(x)\,dx,$$

and therefore

$$a_1 + a_2 + a_3 + \cdots + a_n < a_1 + \int_1^n f(x)\,dx. \qquad (11.3.1)$$

If $\int_1^\infty f(x)\,dx$ is convergent, with value I, then

$$a_1 + a_2 + \cdots + a_n < a_1 + I.$$

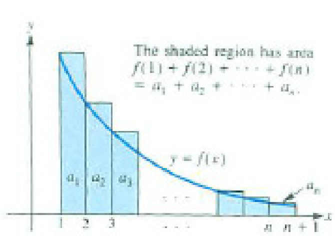


Figure 11.3.2

Since the partial sums of the series $\sum_{k=1}^{\infty} a_k$ are all bounded by the number $a_1 + I$, the series $\sum_{k=1}^{\infty} a_k$ converges and its sum is less than or equal to $a_1 + I$.

Now, Figure 11.3.2 shows that

$$a_1 + a_2 + \cdots + a_n > \int_1^{n+1} f(x)\, dx. \qquad (11.3.2)$$

If follows that if $\int_1^\infty f(x)\, dx$ diverges, then so must the series $\sum_{k=1}^\infty a_k$. •

EXAMPLE 1 Use the integral test to determine the convergence or divergence of

(a) $\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{k} + \cdots = \sum_{k=1}^\infty \frac{1}{k}$

(b) $\frac{1}{1^{1.01}} + \frac{1}{2^{1.01}} + \cdots + \frac{1}{k^{1.01}} + \cdots = \sum_{k=1}^\infty \frac{1}{k^{1.01}}$

SOLUTION

(a) Observe that this is the harmonic series, which was shown in Example 2 in Section 11.2 to diverge. To apply the Integral Test to this series, let $f(x) = 1/x$. This is a decreasing positive function for $x > 0$. Then $a_k = f(k) = 1/k$. We have

$$\int_1^\infty \frac{dx}{x} = \lim_{b\to\infty} \int_1^b \frac{dx}{x} = \lim_{b\to\infty} (\ln(b) - \ln(1)) = \infty$$

Since $\int_1^\infty \frac{dx}{x}$ is divergent, so is the series $\sum_{i=1}^\infty \frac{1}{n}$.

Even though the graphs of $y = \frac{1}{x}$ and $y = \frac{1}{x^{1.01}}$ are near each other, the integrals $\int \frac{dx}{x}$ and $\int \frac{dx}{x^{1.01}}$ behave very differently.

(b) Let $f(x) = 1/x^{1.01}$, which is a decreasing positive function. Then $a_k = f(k) = 1/k^{1.01}$. We have

$$\begin{aligned}\int_1^\infty \frac{dx}{x^{1.01}} &= \lim_{b\to\infty} \int_1^b \frac{dx}{x^{1.01}} = \lim_{b\to\infty} \left.\frac{x^{-1.01+1}}{-1.01+1}\right|_1^b = \lim_{b\to\infty} \left.\frac{x^{-0.01}}{-0.01}\right|_1^b \\ &= \lim_{b\to\infty} \left(\frac{b^{-0.01}}{-0.01} - \frac{1^{-0.01}}{-0.01}\right) = 0 - (-100) = 100.\end{aligned}$$

Since $\int_1^\infty dx/x^{1.01}$ is convergent, so is $\sum_{k=1}^\infty 1/k^{0.01}$. By (11.3.1), its sum is less than $a_1 + 100 = 101$.

◇

The argument in Example 1 extends to a family of series known as p-series.

DEFINITION *([p-series])* For a positive number p, the series

$$\sum_{k=1}^{\infty} \frac{1}{k^p}$$

is called a p-**series**.

For example, when $p = 1$ we obtain the harmonic series $\sum_{k=1}^{\infty} 1/k$ and for $p = 1.01$, the series $\sum_{k=1}^{\infty} 1/k^{1.01}$.

An argument similar to those in Example 1 establishes the following theorem.

Theorem 11.3.2. *If $0 < p \le 1$, the p-series $\sum_{k=1}^{\infty} 1/k^p$ diverges. If $p > 1$, the p-series $\sum_{k=1}^{\infty} 1/k^p$ converges.*

Note that there is a p-series for each positive number p. A negative exponent p would not give a series of interest. For instance, when $p = -1$, we obtain $\sum_{k=1}^{\infty} 1/k^{-1} = \sum_{k=1}^{\infty} k$, which is clearly divergent since its k^{th} term does not approach 0 as $k \to \infty$. (For any negative p, $\lim_{k\to\infty} 1/k^p = \infty$.)

Controlling the Error

Partial sum = front end;
Error = tail end.

When we use a front end of a series (a partial sum) to estimate the sum of the whole series, there will be an error, namely, the sum of the corresponding tail end. For the sum of a front end to be a good estimate of the sum of the whole series, we must be sure that the sum of the corresponding tail end is small. Otherwise, we would be like the carpenter who measures a board as "5 feet long with an error of perhaps as much as 5 feet." That is why we wish to be sure that the sum of the tail end is small.

Let S_n be the sum of the first n terms of a convergent series $\sum_{k=1}^{\infty} a_k$ whose sum is S. The difference

$R_n = \sum_{k=n+1}^{\infty} a_k$

$$R_n = S - S_n = a_{n+1} + a_{n+2} + a_{n+3} + \cdots$$

is called the **remainder** or **error** in using the sum of the first n terms to approximate the sum of the series. That is,

$$\underbrace{a_1 + a_2 + \cdots + a_n}_{\text{partial sum } S_n} + \underbrace{a_{n+1} + a_{n+2} + \cdots}_{\text{tail end } R_n} = \underbrace{a_1 + a_2 + \cdots + a_n + a_{n+1} + a_{n+2} + \cdots}_{\text{sum of series } S}$$

so

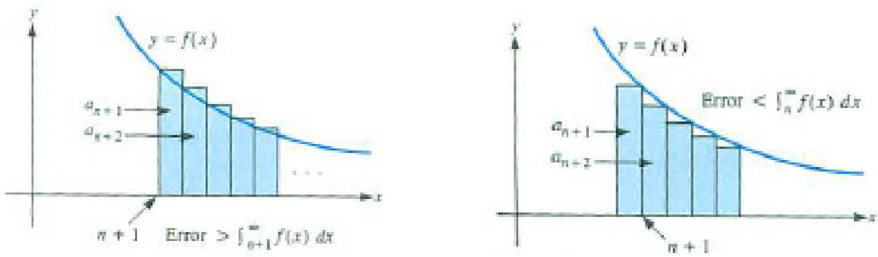
$$S_n + R_n = S.$$

For a series whose terms are positive and decreasing, we can use an improper integral to estimate the error. The reasoning depends again on comparing a staircase of rectangles with the area under a curve.

Recall that $f(x)$ is a continuous decreasing positive function. The error in using $S_n = f(1) + f(2) + \cdots + f(n) = \sum_{i=1}^{n} f(i)$ to approximate $\sum_{i=1}^{\infty} f(i)$ is the sum $\sum_{i=n+1}^{\infty} f(i)$. This sum is the area of the endless staircase of rectangles shown in Figure 11.3.3(a). Comparing the rectangles with the region under the curve $y = f(x)$, we conclude that

$$R_n = a_{n+1} + a_{n+2} + \cdots = f(n+1) + f(n+2) + \cdots > \int\limits_{n+1}^{\infty} f(x)\ dx. \quad (11.3.3)$$

Inequality (11.3.3) gives a *lower* estimate of the error.



(a) (b)

Figure 11.3.3

The staircase in Figure 11.3.3(b), which lies below the curve, gives an *upper* estimate of the error. Inspection of Figure 11.3.3(b) shows that

$$R_n = a_{n+1} + a_{n+2} + \cdots = f(n+1) + f(n+2) + \cdots < \int\limits_{n}^{\infty} f(x)\ dx.$$

Putting these observations together yields the following estimate of the error.

Estimating the error

Theorem 11.3.3 (A bound on the error)**.** *Let $f(x)$ be a continuous decreasing positive function such that $\int_1^{\infty} f(x)\ dx$ is convergent. Then the error R_n in using $f(1) + f(2) + \cdots + f(n)$ to estimate $\sum_{i=1}^{\infty} f(i)$ satisfies the inequality*

$$\int\limits_{n+1}^{\infty} f(x)\ dx \leq R_n \leq \int\limits_{n}^{\infty} f(x)\ dx. \quad (11.3.4)$$

EXAMPLE 2 The first five terms of the series $1/1^2 + 1/2^2 + \cdots + 1/n^2 + \cdots$ are used to estimate the sum of the series.

(a) Put upper and lower bounds on the error in using just those terms.

(b) Use the bounds in (a) to estimate $\sum_{k=1}^{\infty} 1/k^2$.

SOLUTION First, observe that the series with terms $a_k = 1/k^2$ is the p-series with $p = 2$. Since $p > 1$, this series converges. Also, the function $f(x) = 1/x^2$ is continuous, decreasing, and positive for $x \geq 1$.

(a) By inequality (11.3.4) of Theorem 11.3.3, the error R_5 satisfies the inequalities

$$\int_6^{\infty} \frac{dx}{x^2} < R_5 < \int_5^{\infty} \frac{dx}{x^2}.$$

Now, $$\int_5^{\infty} \frac{dx}{x^2} = \left.\frac{-1}{x}\right|_5^{\infty} = 0 - \left(\frac{-1}{5}\right) = \frac{1}{5}.$$

Similarly, $$\int_6^{\infty} \frac{dx}{x^2} = \frac{1}{6}.$$

Thus $$\frac{1}{6} < R_5 < \frac{1}{5}.$$

(b) The sum of the first five terms of the series is

$$S_5 = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} \approx 1.463611.$$

Keep more digits than you need until all calculations have been done. Then, "round down" lower bounds and "round up" upper bounds.

Since the sum of the remaining terms (the "tail end") is between $\frac{1}{6}$ and $\frac{1}{5}$, the sum of the series is between 1.463611+0.166666 and 1.463611+0.2, hence between 1.6302 and 1.6636. (In the 17th century Euler proved that this sum is $\pi^2/6 \approx 1.644934068$.

◇

Estimating a Partial Sum S_n

We still restrict our attention to series that satisfy the hypotheses of the integral test in Theorem 11.3.1. That is, there is a continuous, positive, and decreasing function $f(x)$ such that $f(n) = a_n$.

Just as we can use an (improper) integral to estimate the sum of a tail end of such a series, we can also use a (definite) integral to estimate a partial sum $S_n = a_1 + a_2 + \cdots + a_n$.

In the course of proving Theorem 11.3.1, we obtained equations (11.3.1) and (11.3.2). Taken together, they give us the inequalities

$$\int_1^{n+1} f(x)\,dx < a_1 + a_2 + \cdots + a_n < a_1 + \int_1^n f(x)\,dx. \qquad (11.3.5)$$

If we can evaluate $\int_1^{n+1} f(x)\,dx$ and $\int_1^n f(x)\,dx$ by the Fundamental Theorem of Calculus, we may use (11.3.5) to put upper and lower bounds on $S_n = \sum_{k=1}^n a_k$. These estimates are valid whether the series $\sum_{k=1}^\infty a_k$ converges or diverges.

EXAMPLE 3 Use (11.3.5) to estimate the sum of the first million terms of the harmonic series.
SOLUTION By (11.3.5)

$$\int_1^{1{,}000{,}001} \frac{dx}{x} < \sum_{k=1}^{1{,}000{,}000} \frac{1}{k} < 1 + \int_1^{1{,}000{,}000} \frac{dx}{x}.$$

hence
$$\ln(1,000,001) < \sum_{k=1}^{1{,}000{,}000} \frac{1}{k} < 1 + \ln(1,000,000).$$

Evaluating the logarithm with a calculator, we conclude that

$$13.8155 < \sum_{i=1}^{1{,}000{,}000} \frac{1}{i} < 14.8156.$$

$\diamond$

Summary

We developed a test for convergence or divergence for series whose terms a_k are of the form $f(k)$ for a continuous, positive, decreasing function $f(x)$. The series converges if $\int_1^\infty f(x)\,dx$ converges, and diverges if $\int_1^\infty f(x)\,dx$ diverges.

We also used integrals to analyze the error in using a partial sum S_n of such a series as an estimate of the sum of the series. (Rather than memorizing the formulas, just draw the appropriate staircase diagrams.)

We assumed $f(x)$ is decreasing for $x \geq 1$. Actually, Theorem 11.3.1 holds if we assume that $f(x)$ is decreasing from some point on, that is, there is some number a such that $f(x)$ is decreasing for $x \geq a$. (The argument for this type of integral involves similar staircase diagrams.)

EXERCISES for Section 11.3

Use the integral test in Exercises 1 to 8 to determine whether each series diverges or converges.

1. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^{1.1}}$

2. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^{0.9}}$

3. $\displaystyle\sum_{k=1}^{\infty} \frac{k}{k^2+1}$

4. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^2+1}$

5. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k\ln(k)}$

6. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k+1,000}$

7. $\displaystyle\sum_{k=1}^{\infty} \frac{\ln(k)}{k}$

8. $\displaystyle\sum_{k=1}^{\infty} \frac{k^3}{e^k}$

Use Theorem 11.3.2 in Exercises 9 to 12 to determine whether each series diverges or converges.

9. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{\sqrt[3]{k}}$

10. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^3}$

11. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{\sqrt{k}}$

12. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^{0.999}}$

13.

(a) Prove that if $p > 1$, the p-series converges.

(b) Give two numbers between which its sum lies.

14.

(a) If you used S_{100} to estimate $\sum_{k=1}^{\infty} 1/k^2$, what could you say about the error R_{100}?

(b) How large should you choose k to be sure that the error R_k is less than 0.0001?

15.

(a) If you used S_{1000} to estimate $\sum_{k=1}^{\infty} 1/k^3$, what could you say about the error R_{1000}?

(b) How large should you choose k to be sure that the error R_k is less than 0.0001?

16.

(a) How many terms of the series $\sum_{k=1}^{\infty} 1/k^4$ should you use to be sure that the remainder is less than 0.0001?

(b) Estimate $\sum_{k=1}^{\infty} 1/k^4$ to three decimal places.

17. Repeat Exercise 16 for the series $\sum_{k=1}^{\infty} 1/k^5$.

In each of Exercises 18 to 21 (a) compute the sum of the first four terms of the series to four decimal places, (b) give upper and lower bound on the error R_4, (c) combine (a) and (b) to estimate the sum of the series.

18. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^3}$

19. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^4}$

20. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^2+1}$

21. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^2+k}$

22. Prove that if $p \leq 1$, the p-series diverges.

23. What does the integral test say about the geometric series $\sum_{k=1}^{\infty} p^k$, when $0 < p < 1$?

24. Let $f(x)$ be a positive continuous function that is decreasing for $x \geq a$. Let $a_k = f(k)$. Show in detail (with appropriate diagrams and exposition) why $\int_a^\infty f(x)\, dx$ and $\sum_{k=1}^\infty a_k$ both converge or both diverge. Use your own words. Don't just mimic the book's treatment of the case $a = 1$.

25. (See Exercise 24.) Show that $\sum_{k=1}^\infty k^{10}e^{-k}$ converges.

26. Show that for $n \geq 2$,

$$2\sqrt{n+1} - 2 < \sum_{k=1}^n \frac{1}{\sqrt{k}} < 2\sqrt{n} - 1.$$

27.

(a) By comparing the sum with integrals, show that

$$\ln\left(\frac{201}{100}\right) < \frac{1}{100} + \frac{1}{101} + \frac{1}{102} + \cdots + \frac{1}{200} < \ln\left(\frac{200}{99}\right).$$

(b) Find $\lim_{n\to\infty}\left(\frac{1}{n} + \frac{1}{n+1} + \cdots + \frac{1}{2n}\right)$.

28. In Example 1 we showed that the p-series for $p = 1$ diverges but the p-series for $p = 1.01$ converges. This contrast occurs even though the corresponding terms of the two series seem to resembe each other so closely. (For instance, $1/7^{1.01} \approx 0.140104$, $1/7^1 \approx 0.142857$.) What happens to the ratio $(1/k^{1.01})/(1/k)$ as $k \to \infty$?

In Exercises 29 and 30 concern products, rather than sums, of numbers.

29. Let $\{a_n\}$ be a sequence of positive numbers. Denote the product $(1+a_1)(1+a_2)\cdots(1+a_n)$ by $\prod_{k=1}^n(1+a_k)$.

(a) Show that $\sum_{k=1}^\infty a_k \leq \prod_{k=1}^n(1+a_k)$.

(b) Show that if $\lim_{k\to\infty}\prod_{k=1}^n(1+a_k)$ exists, then $\sum_{k=1}^\infty a_k$ is convergent.

30. (This continues Exercise 29.)

(a) Show that $1 + a_k \leq e^{a_k}$. (Show that $1 + x \leq e^x$ for $x > 0$.)

(b) Show that if the series $\sum_{k=1}^\infty a_k$ is convergent, then $\lim_{n\to\infty}\prod_{k=1}^n(1+a_k)$ exists.

31. Here is an argument that there is an infinite number of primes. Assume that there is only a finite number of primes, p_1, p_2, ..., p_m.

(a) Show that
$$\frac{1}{1-1/p_k} = 1 + \frac{1}{p_k} + \frac{1}{p_k^2} + \frac{1}{p_k^3} + \cdots .$$

(b) Show then that
$$\frac{1}{1-1/p_1}\frac{1}{1-1/p_2}\cdots\frac{1}{1-1/p_m} = \sum_{k=1}^{\infty}\frac{1}{k}.$$

Assume the series can be multiplied term-by-term.

(c) From (b) obtain a contradiction.

11.4 The Comparison Tests

So far in this chapter three tests for the convergence (or divergence) of a series have been presented. The first concerned a special type of series, a geometric series. The second, the n^{th}-term test for divergence, asserts that if the n^{th} term of a series does *not* approach 0, the series diverges. The third, the integral test, applies to certain series of positive terms. In this section two further tests are developed; the comparison and limit-comparison tests. We still consider only tests for series with positive terms.

Comparison Tests

The first test is similar to the comparison test for improper integrals in Section 7.8.

Comparison Tests

Theorem 11.4.1 (Comparison Tests for Convergence and Divergence)**.**

(a) If $0 \leq p_k \leq c_k$ for each k and $\sum_{k=1}^{\infty} c_k$ converges, so does $\sum_{k=1}^{\infty} p_k$.

(b) If $0 \leq d_k \leq p_k$ for each k and $\sum_{k=1}^{\infty} d_k$ diverges, so does $\sum_{k=1}^{\infty} p_k$.

Proof

(a) Let the sum of the series $c_1 + c_2 + \cdots$ be C. Let S_n denote the partial sum $p_1 + p_2 + \cdots + p_n$. Then, for each n,

$$S_n = p_1 + p_2 + \cdots + p_n \leq c_1 + c_2 + \cdots + c_n \leq C.$$

Since the p_n's are non-negative,

$$S_1 \leq S_2 \leq \cdots \leq S_n \leq \cdots .$$

$S_1 \leq S_2 \leq \cdots \leq S_n \leq \cdots \leq C$

Since each S_n is less than or equal to C, Theorem 10.1.1 of Section 10.1 assures us that the sequence $\{S_n\}$ converges to a number L (less than or equal to C). In other words, the series $p_1 + p_2 + \cdots$ converges (and its sum is less than or equal to the sum $c_1 + c_2 + \cdots$).

(b) The divergence test follows immediately from the convergence test. If the series $p_1 + p_2 + \cdots$ converged, so would the series $d_1 + d_2 + \cdots$, which is assumed to diverge.

•

Figure 11.4.1 presents the two comparison tests in Theorem 11.4.1 in terms of endless staircases.

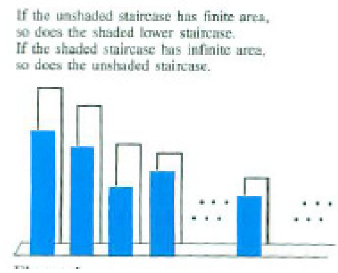


Figure 11.4.1

In order to apply the comparison test to a series of positive terms you have to compare it to a series whose convergence or divergence you already know. What series can you use for comparison? You know the p-series converges for $p > 1$ and diverges for $p \leq 1$. Also a geometric series $\sum_{k=1}^{\infty} r^k$ with positive terms converges for $0 \leq r < 1$ but diverges for $r \geq 1$. Moreover, when we multiply one of theses series by a non-zero constant, we don't affect its convergence or divergence.

EXAMPLE 1 Does the series

$$\sum_{k=1}^{\infty} \frac{k+1}{k+2}\frac{1}{k^2} = \frac{2}{3}\cdot\frac{1}{1^2} + \frac{3}{4}\cdot\frac{1}{2^2} + \frac{4}{5}\cdot\frac{1}{3^2} + \cdots$$

converge or diverge?
SOLUTION The coefficients $\frac{2}{3}$, $\frac{3}{4}$, and $\frac{4}{5}$, ... approach 1 as $k \to \infty$, so they are a minor influence. The series resembles the series

$$\frac{1}{1^2} + \frac{1}{2^2} + \cdots + \frac{1}{k^2} + \cdots ,$$

which was shown by the integral test to be convergent. Since the fraction $(k+1)/(k+2)$ is less than 1,

$$\frac{k+1}{k+2}\frac{1}{k^2} < \frac{1}{k^2}.$$

Thus, by the comparison test for convergence, the series

$$\frac{2}{3}\cdot\frac{1}{1^2} + \frac{3}{4}\cdot\frac{1}{2^2} + \frac{4}{5}\cdot\frac{1}{3^2} + \cdots$$

also converges. However, the test does not tell us the sum of the series. ◇

EXAMPLE 2 Does the series

$$\sum_{k=1}^{\infty} \frac{k+1}{k+2}\cdot\frac{1}{k} = \frac{2}{3}\cdot\frac{1}{1} + \frac{3}{4}\cdot\frac{1}{2} + \cdots + \frac{k+1}{k+2}\cdot\frac{1}{k} + \cdots$$

converge or diverge?
SOLUTION Again the coefficient $(k+1)/(k+2)$ is a minor influence. We suspect that $1/k$ is the main influence and that the series diverges.

Unfortunately, the terms in this series are *less* than the terms of the harmonic series $\sum_{k=1}^{\infty} \frac{1}{k}$. So the divergence test does not directly apply. However, $(k+1)/(k+2)$ is greater than $1/2$. Now, the series

$$\frac{1}{2} \cdot \frac{1}{1} + \frac{1}{2} \cdot \frac{1}{2} + \cdots + \frac{1}{2} \cdot \frac{1}{k} + \cdots$$

is also divergent, since it's just a multiple of a divergent series. The divergence part of the comparison test applies: the series

$$\sum_{k=1}^{\infty} \frac{k+1}{k+2} \frac{1}{k}$$

is, term by term, larger than the terms of the divergent series

$$\sum_{k=1}^{\infty} \frac{1}{2} \frac{1}{k}.$$

Hence, $\sum_{k=1}^{\infty} \frac{k+1}{k+2} \frac{1}{k}$ is divergent. ◇

Limit-Comparison Tests

There is a variation of the comparison test that produces a much quicker solution of Example 2. It is the **limit-comparison test**.

Limit-Comparison Tests

Theorem 11.4.2 (Limit-Comparison Tests for Convergence and Divergence)**.** *Let $\sum_{k=1}^{\infty} p_k$ be a series of positive terms to be tested for convergence or divergence.*

A. *Let $\sum_{k=1}^{\infty} c_k$ be a convergent series of positive terms. If $\lim_{k\to\infty} \frac{p_k}{c_k}$ exists, then $\sum_{k=1}^{\infty} c_k$ also converges.*

B. *Let $\sum_{k=1}^{\infty} d_k$ be a divergent series of positive terms. If $\lim_{k\to\infty} \frac{p_k}{d_k}$ exists and is not 0 or if the limit is infinite, then $\sum_{k=1}^{\infty} p_k$ also diverges.*

Proof

We shall prove part (a). Let $a = \lim_{k\to\infty} \frac{p_k}{c_k}$. Since as $k \to \infty$, $p_k/c_k \to a$, there must be an integer N such that, for all $n \geq N$, p_k/c_k remains less than, say, $a+1$. Thus

$$p_k < (a+1)c_k \qquad \text{for all } n \geq N.$$

Now the series

$$(a+1)c_N + (a+1)c_{N+1} + \cdots + (a+1)c_k + \cdots ,$$

being $a+1$ times the tail end of a convergent series, is itself convergent. By the comparison test,

$$p_N + p_{N+1} + \cdots + p_k + \cdots$$

is convergent. Hence $p_1 + p_2 + \cdots + p_k + \cdots$ is convergent.

Part (B) can be proved in a similar manner. •

Note that in part B of the Limit-Comparison Test nothing is said about the case $\lim_{k\to\infty} p_k/d_k = 0$. In this circumstance the series $\sum_{k=1}^{\infty} p_k$ can either converge or diverge. For instance, take $\sum_{k=1}^{\infty} d_k$ to be the divergent series $\sum_{k=1}^{\infty} \frac{1}{\sqrt{n}}$. The series $\sum_{k=1}^{\infty} \frac{1}{k^2}$ is convergent and $\lim_{k\to\infty} \frac{1/k^2}{1/\sqrt{k}} = 0$. Contrarily, the harmonic series $\sum_{k=1}^{\infty} \frac{1}{k}$ is divergent and again $\lim_{k\to\infty} \frac{1/k}{1/\sqrt{k}} = 0$.

The next example shows how convenient the limit-comparison test is. Contrast the solution in Example 3 with that in Example 2.

EXAMPLE 3 Does the series

$$\sum_{k=1}^{\infty} \frac{k+1}{k+2} \cdot \frac{1}{k} = \frac{2}{3} \cdot \frac{1}{1} + \frac{3}{4} \cdot \frac{1}{2} + \cdots + \frac{k+1}{k+2} \cdot \frac{1}{k} + \cdots$$

converge or diverge?

SOLUTION As with Example 2, we expect this series to behave like the harmonic series. For this reason we examine the ratio between corresponding terms:

$$\lim_{k\to\infty} \frac{\frac{k+1}{k+2} \cdot \frac{1}{k}}{\frac{1}{k}} = \lim_{k\to\infty} \frac{k+1}{k+2} = 1.$$

Since the limit is not 0, and the harmonic series diverges, the Limit-Comparison Test tells us that $\displaystyle\sum_{k=1}^{\infty} \frac{k+1}{k+2} \cdot \frac{1}{k}$ diverges. ◇

EXAMPLE 4 Does

$$\sum_{k=1}^{\infty} \frac{(1+1/k)^k \left(1+(-1/2)^k\right)}{2^k}$$

converge or diverge?

SOLUTION Note that as $k\to\infty$, $(1+1/k)^k \to e$ and $1+(-1/2)^k \to 1$. The major influence is the 2^k in the denominator. So use the Limit-Comparison Test. The given series resembles the convergent geometric series with first term $\frac{1}{2}$ and ratio also $\frac{1}{2}$: $\frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^k} + \cdots$. Then

See Section 2.2.

$$\lim_{k\to\infty} \frac{\frac{\left(1+\frac{1}{k}\right)^k \left(1+\left(\frac{-1}{2}\right)^k\right)}{2^k}}{\frac{1}{2^k}} = \lim_{k\to\infty} \left(1+\frac{1}{k}\right)^k \left(1+\left(\frac{-1}{2}\right)^k\right) = e \cdot 1 = e.$$

Since $\sum_{k\to\infty} 2^{-k}$ is convergent, so is the given series. ◇

EXAMPLE 5 Does $\sum_{k=1}^{\infty} k^3 3^{-k}$ converge or diverge?
SOLUTION The typical term $k^3 3^{-k}$ is dominated by the exponential factor, $1/3^k$. For this reason we suspect that the series $\sum_{k=1}^{\infty} k^3 3^{-k}$ might also converge. We try the Limit-Comparison Test, obtaining

$$\lim_{k\to\infty} \frac{\frac{k^3}{3^k}}{\frac{1}{3^k}} = \lim_{k\to\infty} k^3 = \infty.$$

Since the limit is not finite, the test gives no information. So we start over and look at $k^3/3^k$ a little closer.

See also Section 5.6.

The numerator k^3 approaches ∞ much more slowly than 3^k, so we still suspect that $\sum_{k=1}^{\infty} k^3/3^k$ converges. Now, k^3 approaches ∞ *more slowly than any exponential* b^k with $b > 1$. For example, for large k, k^3 is less than $(1.5)^k$. This means that for large k

$$\frac{k^3}{3^k} < \frac{(1.5)^k}{3^k} = (0.5)^k.$$

The geometric series $\sum_{k=1}^{\infty}(0.5)^k$ converges. Since $k^3/3^k < (0.5)^k$ for all but a finite number of values of k, the Comparison Test tells us that $\sum_{k=1}^{\infty} k^3/3^k$ converges. ◇

Summary

We developed two tests for convergence or divergence of a series with positive terms, $\sum_{k=1}^{\infty} p_k$. If, for each k, p_k is less than the corresponding term of a convergent series, then $\sum_{k=1}^{\infty} p_k$ converges. If p_k is larger than the corresponding term of a divergent series of positive terms, then $\sum_{k=1}^{\infty} p_k$ diverges. This Comparison Test is the basis for the Limit-Comparison Test, which is often easier to apply. This test depends only on the limit of the ratio of p_k to the corresponding term of a series of positive terms known to converge or diverge.

EXERCISES for Section 11.4

Use the comparison test in Exercises 1 to 4 to determine whether each series converges or diverges.

1. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^2+3}$

2. $\displaystyle\sum_{k=1}^{\infty} \frac{k+2}{(k+1)\sqrt{k}}$

3. $\displaystyle\sum_{k=1}^{\infty} \frac{\sin^2(k)}{k^2}$

4. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k2^k}$

Use the limit-comparison test in Exercises 5 to 8 to determine whether each series converges or diverges.

5. $\displaystyle\sum_{k=1}^{\infty} \frac{5k+1}{(k+2)k^2}$

6. $\displaystyle\sum_{k=1}^{\infty} \frac{2^k+k}{3^k}$

7. $\displaystyle\sum_{k=1}^{\infty} \frac{k+1}{(5k+2)\sqrt{k}}$

8. $\displaystyle\sum_{k=1}^{\infty} \frac{(1+1/k)^k}{k^2}$

In Exercises 9 to 28 use any test discussed so far in this chapter to determine whether each series converges or diverges.

9. $\displaystyle\sum_{k=1}^{\infty} \frac{k^2k}{3^k}$

10. $\displaystyle\sum_{k=1}^{\infty} \frac{2^k}{k^2}$

11. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^k}$

12. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k!}$

13. $\displaystyle\sum_{k=1}^{\infty} \frac{4k+1}{(2k+3)k^2}$

14. $\displaystyle\sum_{k=1}^{\infty} \frac{k^2(2^k+1)}{3^k+1}$

15. $\displaystyle\sum_{k=1}^{\infty} \frac{1+\cos(k)}{k^2}$

16. $\displaystyle\sum_{k=1}^{\infty} \frac{\ln(k)}{k}$

17. $\displaystyle\sum_{k=1}^{\infty} \frac{ln(k)}{k^2}$

18. $\displaystyle\sum_{k=1}^{\infty} \frac{5^k}{k^k}$

19. $\displaystyle\sum_{k=1}^{\infty} \frac{2^k}{k!}$

20. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{\sqrt{k}\ln(k)}$

21. $\displaystyle\sum_{k=1}^{\infty} \frac{e^{2k}}{\pi^k}$

22. $\displaystyle\sum_{k=1}^{\infty} \frac{k^2e^k}{\pi^k}$

23. $\displaystyle\sum_{k=1}^{\infty} \frac{3k+1}{2k+10}$

24. $\displaystyle\sum_{k=1}^{\infty} \frac{4}{2k^2-k}$

25. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{\ln(k)}$

26. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{\sin(1/k)}$

27. $\displaystyle\sum_{k=1}^{\infty} \left(\frac{k+1}{k+3}\right)^k$

28. $\displaystyle\sum_{k=1}^{\infty} \left(\frac{k}{2k-1}\right)^k$

In Exercises 29 to 34, assume that $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=1}^{\infty} b_k$ are series with positive terms. What, if anything, can we conclude about the convergence or divergence of $\sum_{k=1}^{\infty} a_k$ if:

29. If $\sum_{k=1}^{\infty} b_k$ is divergent and $\lim_{k\to\infty} \frac{a_k}{b_k} = 0$?

30. If $\sum_{k=1}^{\infty} b_k$ is convergent and $\lim_{k\to\infty} \frac{a_k}{b_k} = \infty$?

31. If $\sum_{k=1}^{\infty} b_k$ is convergent and $3b_k \le a_k \le 5b_k$?

32. If $\sum_{k=1}^{\infty} b_k$ is divergent and $3b_k \le a_k \le 5b_k$?

33. If $\sum_{k=1}^{\infty} b_k$ is convergent and $a_k < b_k^2$?

34. If $\sum_{k=1}^{\infty} b_k$ is divergent and $b_k \to 0$ as $k \to \infty$, and $a_k < b_k^2$?

35. For which values of the positive number x does the series $\displaystyle\sum_{k=1}^{\infty} \frac{x^k}{k2^k}$ converge? diverge?

36. For which values of the positive exponent m does the series $\sum_{k=1}^{\infty} \frac{1}{k^m \ln(k)}$ converge? diverge?

37. Prove part B of the Limit-Comparison Test for Convergence and Divergence.

38. For which constants p does $\sum_{k=1}^{\infty} k^p e^{-k}$ converge?

39.

(a) Show that $\sum_{k=1}^{\infty} 1/(1+2^k)$ converges.

(b) Show that the sum of the series in (a) is between 0.64 and 0.77. (Use the first three terms and control the sum of the rest of the series by comparing it to the sum of a geometric series.)

40.

(a) Show that $\sum_{k=n+1}^{\infty} 1/k!$ is less than the sum of the geometric series whose first term is $1/(n+1)!$ and whose ratio is $1/(n+2)$.

(b) Use (a) with $n = 4$ to show that

$$1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} < \sum_{k=0}^{\infty} \frac{1}{k!} < 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} \cdot \frac{1}{1 - \frac{1}{6}}.$$

(c) From (b) deduce that

$$2.71 < \sum_{k=0}^{\infty} \frac{1}{k!} < 2.72.$$

(d) Find a value of n such that $\sum_{k=n+1}^{\infty} 1/k! < 0.0005$.

(e) Use (d) to estimate $\sum_{k=0}^{\infty} 1/k!$ to three decimal places.

41. Prove the following result, which is used in the statistical theory of stochastic processes: Let $\{a_n\}$ and $\{c_n\}$ be two sequences of non-negative numbers such that

$\sum_{k=1}^{\infty} a_k c_k$ converges and $\lim_{n\to\infty} c_n = 0$. Then $\sum_{k=1}^{\infty} a_k c_k^2$ converges.

42. Find a specific number B, expressed as a decimal, such that

$$\sum_{k=1}^{\infty} \frac{\ln(k)}{k^2} < B.$$

43. Find a specific number B, expressed as a decimal, such that

$$\sum_{k=1}^{\infty} \frac{k+2}{k+1} \cdot \frac{1}{n^3} < B.$$

44. Estimate $\sum_{k=1}^{\infty} \frac{1}{k2^k}$ to three decimal places.

45. Let $\sum_{k=1}^{\infty} a_k$ be a convergent series with only positive terms. Must $\sum_{k=1}^{\infty} (a_k)^2$ also converge?

46. Let $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=1}^{\infty} b_k$ be convergent series with only positive terms. Must $\sum_{k=1}^{\infty} a_k b_k$ converge?

11.5 Ratio Tests

The next test is suggested by the test for the convergence of a geometric series. In a geometric series the ratio between consecutive terms is constant. The "Ratio Test" concerns series when this ratio is "almost constant".

The Ratio Test

Theorem 11.5.1 (Ratio Test)**.** *Let $p_1 + p_2 + \cdots + p_n + \cdots$ be a series of positive terms. Assume $\lim_{k\to\infty} p_{k+1}/p_k$ exists and call it r.*

Ratio Test

(a) If r is less than 1, the series converges.

(b) If r is greater than 1 or r is infinite, the series diverges.

(c) If r is equal to 1 or r does not exist, no conclusion can be drawn (the series may converge or may diverge).

Proof

The idea behind the Ratio Test is to compare the original series to a geometric series. Here is how that works.

(a) Assume $r = \lim_{k\to\infty} \frac{p_{k+1}}{p_k} < 1$. Select a number s such that $r < s < 1$. Then there is an integer N such that for all $k \geq N$,

$$\begin{aligned} \frac{p_{k+1}}{p_k} &< s \\ \text{and, therefore,} \qquad p_{k+1} &< sp_k. \end{aligned}$$

Using this inequality, we deduce that

$$\begin{aligned} p_{N+1} &< sp_N \\ p_{N+2} < sp_{N+1} &< s(sp_N) = s^2 p_N \\ p_{N+3} < sp_{N+2} &< s(s^2 p_N) = s^3 p_N, \end{aligned}$$

and so on.

Thus the terms of the series

$$p_N + p_{N+1} + p_{N+2} + \cdots$$

are less than the corresponding terms of the geometric series

$$p_N + sp_N + s^2 p_N + \cdots$$

(except for the first term, p_N, which equals the first term of the geometric series). Since $s < 1$, the latter series converges. By the comparison test, $p_N + p_{N+1} + p_{N+2} + \cdots$ converges. Adding in the front end,

$$p_1 + p_2 + \cdots + p_{N-1},$$

still results in a convergent series.

(b) If $r > 1$ or is infinite, then for all k from some point on p_{k+1} is larger than p_k. Thus the n^{th} term of the series $p_1 + p_2 + \cdots$ cannot approach 0. By the n^{th}-term test for divergence the series diverges.

•

No information if r is 1 or does not exist.

When $r = 1$ or r does not exist, anything can happen; the series may diverge or it may converge. (Exercise 21 illustrates these possibilities.) In these cases, one must look to other tests to determine whether the series diverges or converges.

The Ratio Test is a natural test to try if the k^{th} term of a series involves powers of a fixed number, or factorials, as the next two examples show.

EXAMPLE 1 Show that the series $p + 2p^2 + 3p^3 + \cdots + kp^k + \cdots$ converges for any fixed number p for which $0 < p < 1$.

SOLUTION Let a_k denote the k^{th} term of the series. Then

$$a_k = kp^k \qquad \text{and} \qquad a_{k+1} = (k+1)p^{k+1}.$$

The ratio between consecutive terms is

$$\frac{a_{k+1}}{a_k} = \frac{(k+1)p^{k+1}}{kp^k} = \frac{k+1}{k}p.$$

The sum of this series is found in Exercise 34.

Thus

$$r = \lim_{k\to\infty} \frac{a_{k+1}}{a_k} = p < 1,$$

and the series converges. $\diamond$

EXAMPLE 2 Determine the positive values of x for which the series

$$\frac{1}{0!} + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^k}{k!} \cdots$$

converges and for which values of x it diverges. (Each choice of x determines a specific series with constant terms.)

SOLUTION If we start the series with $k = 0$, then the n^{th} term, a_k is $x^k/k!$. Thus

$$a_{k+1} = \frac{x^{k+1}}{(k+1)!},$$

and therefore

$$\frac{a_{k+1}}{a_k} = \frac{\frac{x^{k+1}}{(k+1)!}}{\frac{x^k}{k!}} = x\frac{k!}{(k+1)!} = \frac{x}{k+1}.$$

Since x is fixed,

$$r = \lim_{k\to\infty} \frac{a_{k+1}}{a_k} = \lim_{k\to\infty} \frac{x}{k+1} = 0.$$

By the Ratio Test, the series converges for all positive x. ◇

In the next section, it will be shown that this series converges for all negative values of x, too.

The next example uses the Ratio Test to establish divergence.

EXAMPLE 3 Show that the series $2/1 + 2^2/2 + \cdots + 2^k/k + \cdots$ diverges.
SOLUTION In this case, $a_k = 2^k/k$ and

$$\frac{a_{k+1}}{a_k} = \frac{\frac{2^{k+1}}{k+1}}{\frac{2^k}{k}} = \frac{2^{k+1}}{k+1}\frac{k}{2^k} = 2\frac{k}{k+1}.$$

Thus

$$r = \lim_{k\to\infty} \frac{a_{k+1}}{a_k} = 2,$$

which is larger than 1. By the Ratio Test, this series diverges. ◇

The series is like a geometric series with ratio 2.

It is not really necessary to call on the powerful Ratio Test to establish the divergence of the series in Example 3. Since $\lim_{k\to\infty} 2^k/k = \infty$, its k^{th} term gets arbitrarily large; by the k^{th}-term test, the series diverges. (Comparison with the harmonic series also demonstrates divergence.)

The Root Test

The next test, closely related to the Ratio Test, is of use when the k^{th} term contains only k^{th} powers, such as k^k or 3^k. It is not useful if factorials such as $k!$ are present.

Root Test

Theorem 11.5.2 (Root Test)**.** *Let $\sum_{k=1}^{\infty} p_k$ be a series of positive terms. Assume $\lim_{k\to\infty} \sqrt[k]{p_k}$ exists and call it r. Then*

A. If r is less than 1, the series converges.

B. If r is greater than 1 or r is infinite, the series diverges.

C. If r is equal to 1 or r does not exist, no conclusion can be drawn (the series may converge or may diverge).

The proof of the Root Test is outlined in Exercises 22 and 23.

EXAMPLE 4 Use the Root Test to determine whether $\sum_{k=1}^{\infty} 3^k/k^{k/2}$ converges or diverges.

SOLUTION We have

$$r = \lim_{k\to\infty} \sqrt[k]{\frac{3^k}{k^{k/2}}} = \lim_{k\to\infty} \frac{3}{\sqrt{k}} = 0.$$

By the Root Test, the series converges. $\diamond$

Summary

We developed two tests for convergence or divergence of a series $\sum_{k=1}^{\infty} p_k$ with positive terms, both motivated by geometric series. In the Ratio Test, we examine $\lim_{k\to\infty} p_{k+1}/p_k$ and in the Root Test, $\lim_{k\to\infty} \sqrt[k]{p_k}$. The Ratio Test is convenient to use when the terms involve powers and factorials. The Root Test is convenient when only powers appear.

EXERCISES for Section 11.5

In Exercises 1 to 6 apply the Ratio Test to decide whether the series converges or diverges. If that test gives no information, use another test to decide.

1. $\displaystyle\sum_{k=1}^{\infty} \frac{k^2}{3^k}$

2. $\displaystyle\sum_{k=1}^{\infty} \frac{(k+1)^2}{k2^k}$

3. $\displaystyle\sum_{k=1}^{\infty} \frac{k\ln(k)}{3^k}$

4. $\displaystyle\sum_{k=1}^{\infty} \frac{k!}{3^k}$

5. $\displaystyle\sum_{k=1}^{\infty} \frac{(2k+1)(2^k+1)}{3^k+1}$

6. $\displaystyle\sum_{k=1}^{\infty} \frac{k!}{k^k}$

In Exercises 7 and 8 use the Root Test to determine whether the series converge or diverge.

7. $\displaystyle\sum_{k=1}^{\infty} \frac{k^k}{3^{k^2}}$

8. $\displaystyle\sum_{k=1}^{\infty} \frac{(1+1/k)^k(2k+1)^k}{(3k+1)^k}$

Each series found in Exercises 9 to 14 converges. Use any legal means to find a number B in decimal form that is larger than the sum of the series.

9. $\displaystyle\sum_{k=1}^{\infty} \frac{k^2}{2^k}$

10. $\displaystyle\sum_{k=1}^{\infty} \frac{k}{3^k}$

11. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^3}$

12. $\displaystyle\sum_{k=1}^{\infty} \frac{\sin^2(k)}{k^2}$

13. $\displaystyle\sum_{k=1}^{\infty} \frac{\ln(k)}{k^2}$

14. $\displaystyle\sum_{k=1}^{\infty} \frac{\left(1+\frac{2}{k}\right)^k}{1.1^k}$

Each series in Exercises 15 to 18 diverges. Use any legal means to find a number m such that the m^{th} partial sum of the series exceeds 1,000.

15. $\displaystyle\sum_{k=1}^{\infty} \frac{\ln(k)}{k}$

16. $\displaystyle\sum_{k=1}^{\infty} \frac{k}{k^2+1}$

17. $\displaystyle\sum_{k=1}^{\infty} (1.01)^k$

18. $\displaystyle\sum_{k=1}^{\infty} \frac{(k+2)^2}{k+1} \cdot \frac{1}{\sqrt{k}}$

19. Use the result of Example 2 to show that, for $x > 0$, $\lim_{k\to\infty} x^k/k! = 0$. This was established directly in Section 11.2.

20. Solve Example 3 using the Root Test.

21. This exercise shows that the Ratio Test gives no information if $\displaystyle\lim_{k\to\infty} \frac{p_{k+1}}{p_k} = 1$.

(a) Show that for $p_k = 1/k$, $\sum_{k=1}^{\infty} p_k$ diverges and $\lim_{k\to\infty} \frac{p_{k+1}}{p_k} = 1$.

(b) Show that for $p_k = 1/k^2$, $\sum_{k=1}^{\infty} p_k$ converges and $\lim_{k\to\infty} \frac{p_{k+1}}{p_k} = 1$.

22. This exercise shows that the Root Test gives no information if $\displaystyle\lim_{k\to\infty} \sqrt[k]{p_k} = 1$.

(a) Show that for $p_k = 1/k$, $\sum_{k=1}^{\infty} p_k$ diverges and $\lim_{k\to\infty} \sqrt[k]{p_k} = 1$.

(b) Show that for $p_k = 1/k^2$, $\sum_{k=1}^{\infty} p_k$ converges and $\lim_{k\to\infty} \sqrt[k]{p_k} = 1$.

23. (Proof of the Root Test, Theorem 11.5.2.)

(a) Assume that $r = \lim_{k\to\infty} \sqrt[k]{p_k} < 1$. Pick any s with $r < s < 1$, and then pick N such that $\sqrt[k]{p_k} < s$ for all $k > N$. Show that $p_k < s^k$ for all $k > N$ and compare a tail end of $\sum_{k=1}^{\infty} p_k$ to a geometric series.

(b) Assume that $r = \lim_{k\to\infty} \sqrt[k]{p_k} > 1$. Pick any s with $1 < s < r$, and then pick N such that $\sqrt[k]{p_k} > s$ for all $k > N$. Show that $p_k > s^k$ for all $k > N$. From this conclude that $\sum_{k=1}^{\infty} p_k$ diverges.

Skill Drill

In Exercises 24 to 26 a, b, and c are constants. In each case verify the following derivative formulas.

24. $\dfrac{d}{dx}\left(x\sin(ax)\right) = \sin(ax) + ax\cos(ax)$

25. $\dfrac{d}{dx}\left(\ln\left|ax^2+bx+c\right|\right) = \dfrac{2ax+b}{ax^2+bx+c}$

26. $\dfrac{d}{dx}\left(x\arctan(ax) - \dfrac{1}{2a}\ln\left(1+a^2x^2\right)\right) = \arctan(ax)$

In Exercises 27 to 32 a, b, c, and n are constants and n is positive. Use integration techniques to obtain each of the following reduction formulas.

27. $\displaystyle\int x^n \sin(ax)\ dx = -\frac{1}{a}\cos(ax) + \frac{n}{a}\int x^{n-1}\cos(ax)\ dx$

28. $\displaystyle\int x^n \cos(ax)\ dx = \frac{1}{a}\cos(ax) - \frac{n}{a}\int x^{n-1}\sin(ax)\ dx$

29. $\displaystyle\int \frac{dx}{x^2\sqrt{ax+b}} = \frac{-\sqrt{ax+b}}{bx} - \frac{a}{2b}\int \frac{dx}{x\sqrt{ax+b}}$

30. $\displaystyle\int \frac{dx}{(ax^2+c)^{n+1}} = \frac{1}{2nc}\frac{x}{(ax^2+c)^n} + \frac{2n-3}{2nc}\int \frac{dx}{(ax^2+c)^n}$

31. $\displaystyle\int \frac{dx}{(ax^2+bx+c)^{n+1}} = \frac{2ax+b}{n(4ac-b^2)(ax^2+bx+c)^n} + \frac{2(2n-1)a}{n(4ac-b^2)}\int \frac{dx}{(ax^2+bx+c)^n}$

32. $\displaystyle\int (\ln(ax))^2\ dx = x^2\left((\ln(ax))^2 - 2\ln(ax) + 2\right)$

11.6 Tests for Series with Both Positive and Negative Terms

The tests for convergence or divergence in Sections 11.3 to 11.5 concern series whose terms are positive. This section examines series that have both positive and negative terms. Two tests for the convergence of such a series are presented. The alternating-series test applies to series whose terms alternate in sign (+, -, +, -, . . .) and decrease in absolute value. In the absolute-convergence test, the signs may vary in any way.

Alternating Series

DEFINITION *(Alternating Series)* If $p_1, p_2, \ldots, p_n, \ldots$ is a sequence of positive numbers, then the series

$$\sum_{k=1}^{\infty}(-1)^{k+1}p_k = p_1 - p_2 + p_3 - p_4 + \cdots + (-1)^{k+1}p_k + \cdots$$

and the series

$$\sum_{k=1}^{\infty}(-1)^{k}p_k = -p_1 + p_2 - p_3 + p_4 - \cdots + (-1)^{k}p_k + \cdots$$

are called **alternating series**.

For instance,

$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots + (-1)^{k+1}\frac{1}{2k-1} + \cdots$$

and

$$1 - 1 + 1 - 1 + \cdots + (-1)^k + \cdots$$

are alternating series.

By the n^{th}-term test, the second series diverges. The following theorem implies that the first series converges.

Alternating-Series Test

Theorem 11.6.1. *(Alternating-Series Test) If $p_1, p_2, \ldots, p_k, \ldots$ is a decreasing sequence of positive numbers such that $\lim_{k\to\infty} p_k = 0$, then the series whose k^{th} term is $(-1)^{k+1}p_k$,*

$$\sum_{k=1}^{\infty}(-1)^{k+1}p_k = p_1 - p_2 + p_3 - \cdots + (-1)^{k+1}p_k + \cdots ,$$

converges.

Proof

We will prove the theorem in the special case when $p_k = 1/k$, that is, the **alternating harmonic series**

$$\sum_{k=1}^{\infty}(-1)^{k+1}\frac{1}{k} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots + (-1)^{k+1}\frac{1}{k} + \cdots .$$

The argument easily generalizes to prove the general theorem. (See Exercise 33.)

Consider first the partial sums of an *even* number of terms, S_2, S_4, S_6, $\ldots$. For clarity, group the summands in pairs:

$$\begin{array}{lcllcl} S_2 & = & (1-\frac{1}{2}) & & & \\ S_4 & = & (1-\frac{1}{2}) + (\frac{1}{3}-\frac{1}{4}) & & = & S_2 + (\frac{1}{3}-\frac{1}{4}) \\ S_6 & = & (1-\frac{1}{2}) + (\frac{1}{3}-\frac{1}{4}) + (\frac{1}{5}-\frac{1}{6}) & & = & S_4 + (\frac{1}{5}-\frac{1}{6}) \\ \vdots & & & & & \end{array}$$

Since $\frac{1}{3}$ is larger than $\frac{1}{4}$, the difference $\frac{1}{3}-\frac{1}{4}$ is positive. Therefore, S_4, which equals $S_2 + (\frac{1}{3}-\frac{1}{4})$, is larger than S_2. Similarly, $S_6 > S_4$. More generally:

$$S_2 < S_4 < S_6 < S_8 < \cdots .$$

The sequence of even partial sums, $\{S_{2n}\}$ is increasing. (See Figure 11.6.1.)

Figure 11.6.1

Next, it will be shown that S_{2n} is less than 1, the first term of the sequence. First of all,

$$S_2 = 1 - \frac{1}{2} < 1.$$

Next, consider S_4:

$$\begin{array}{lcll} S_4 & = & 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} & \\ & = & 1 - (\frac{1}{2} - \frac{1}{3}) - \frac{1}{4} & \\ & < & 1 - (\frac{1}{2} - \frac{1}{3}) & \text{because } \frac{1}{4} \text{ is positive} \\ & < & 1 & \text{because } \frac{1}{2} - \frac{1}{3} \text{ is positive.} \end{array}$$

Similarly,

$$\begin{array}{lcll} S_6 & = & 1 - (\frac{1}{2} - \frac{1}{3}) - (\frac{1}{4} - \frac{1}{5}) - \frac{1}{6} & \\ & < & 1 - (\frac{1}{2} - \frac{1}{3}) - (\frac{1}{4} - \frac{1}{5}) & \text{because } \frac{1}{6} \text{ is positive} \\ & < & 1 - (\frac{1}{2} - \frac{1}{3}) & \text{because } \frac{1}{4} - \frac{1}{5} \text{ is positive} \\ & < & 1 & \text{because } \frac{1}{2} - \frac{1}{3} \text{ is positive.} \end{array}$$

In general then,

$$S_{2n} < 1 \qquad \text{for all } n.$$

The sequence

$$S_2,\ S_4,\ S_6,\ \ldots$$

is therefore increasing and yet bounded by the number 1, as indicated in Figure 11.6.2. By Theorem 10.1.1 of Section 10.1, $\lim_{n\to\infty} S_{2n}$ exists. Call this limit S, which is less than or equal to 1. (See Figure 11.6.2.)

All that remains is to show that the *odd* partial sums

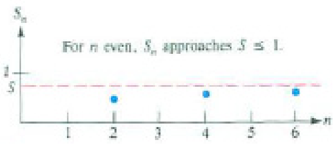


Figure 11.6.2

$$S_1, S_3, S_5, \ldots$$

also converge to S.

Note that

$$\begin{array}{rclcl} S_3 & = & 1 - \frac{1}{2} + \frac{1}{3} & = & S_2 + \frac{1}{3} \\ S_5 & = & 1 - \frac{1}{3} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} & = & S_4 + \frac{1}{5}. \end{array}$$

In general,

$$S_{2k+1} = S_{2k} + \frac{1}{2k+1}.$$

Thus

$$\lim_{k\to\infty} S_{2k+1} = \lim_{k\to\infty}\left(S_{2k} + \frac{1}{2k+1}\right) = \lim_{k\to\infty} S_{2k} + \lim_{k\to\infty}\frac{1}{2k+1} = S + 0 = S.$$

Since the sequence of even partial sums, $S_2,\ S_4,\ S_6,\ \ldots,\ S_{2k},\ \ldots$, and the sequence of odd partial sums, $S_1,\ S_3,\ S_5,\ \ldots,\ S_{2k+1},\ \ldots$, both have the same limit, S, it follows that

$$\lim_{k\to\infty} S_k = S.$$

Thus the alternating harmonic series

$$\sum_{k=1}^{\infty}(-1)^k\frac{1}{k} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots$$

See Exercise 29.

converges. In Chapter 12 it will be shown that this sum is ln(2).

A similar argument applies to any alternating series whose k^{th} term approaches 0 and whose terms decrease in absolute value. •

Decreasing alternating series

An alternating series, such as the alternating harmonic series, whose terms decrease in absolute value as k increases will be called a **decreasing alternating series**. Theorem 11.6.1 shows that a decreasing alternating series whose

k^{th} term approaches zero as $k \to \infty$ converges.

EXAMPLE 1 Estimate the sum S of the alternating harmonic series.
SOLUTION These are the first five partial sums:

$$\begin{array}{lcll} S_1 &=& 1 &= 1.00000 \\ S_2 &=& 1-\frac{1}{2} &= 0.50000 \\ S_3 &=& 1-\frac{1}{2}+\frac{1}{3} \approx 0.5+0.33333 &= 0.83333 \\ S_4 &=& S_3-\frac{1}{4} \approx 0.83333-0.25 &= 0.58333 \\ S_5 &=& S_4+\frac{1}{5} \approx 0.58333+0.2 &= 0.78333 \end{array}$$

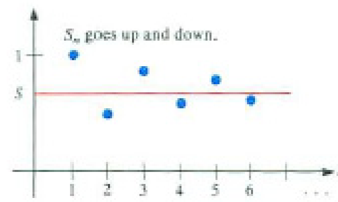


Figure 11.6.3

Figure 11.6.3 is a graph of S_n as a function of n. The odd partial sums S_1, S_3, ... approach S from above. The even partial sums S_2, S_4, ... approach S from below. For instance,

$$S_4 < S < S_5$$

gives the information that $0.58333 < S < 0.8334$. (See Figure 11.6.4.) ◇

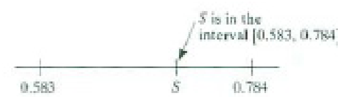


Figure 11.6.4

As Figure 11.6.3 suggests, any partial sum of a series satisfying the hypothesis of the alternating-series test differs from the sum of the series by less than the absolute value of the first omitted term. That is, if S_n is the sum of the first n terms of the series and S is the sum of the series, then the error

$$R_n = S - S_n$$

has absolute value at most p_{n+1}, which is the absolute value of the first omitted term. Moreover, S is between S_n and S_{n+1} for every n.

The error in estimating the sum of a decreasing alternating series.

EXAMPLE 2 Does the series

$$\frac{3}{1!} - \frac{3^2}{2!} + \frac{3^3}{3!} - \frac{3^4}{4!} + \frac{3^5}{5!} - \cdots + (-1)^{k+1}\frac{3^k}{k!} + \cdots$$

converge or diverge?
SOLUTION This is an alternating series. By Example 2 of Section 11.2, its k^{th} term approaches 0. Let us see whether the absolute values of the terms decrease in size, term-by-term. The first few absolute values are

$$\begin{array}{lcl} \frac{3}{1!} &=& 3 \\ \frac{3^2}{2!} &=& \frac{9}{2} = 4.5 \\ \frac{3^3}{3!} &=& \frac{27}{6} = 4.5 \\ \frac{3^4}{4!} &=& \frac{81}{24} = 3.375. \end{array}$$

At first the terms increase, but then they decrease.

At first, they increase. However, the fourth term is less than the third. Let us show that the rest of the terms decrease in size. For instance,

$$\begin{aligned} \frac{3^5}{5!} &= \frac{3}{4}\frac{3^4}{4!} < \frac{3^4}{4!}, \\ \text{and, for } n \geq 3, \qquad \frac{3^{k+1}}{(k+1)!} &= \frac{3}{k+1}\frac{3^k}{k!} < \frac{3^k}{k!}. \end{aligned}$$

By the alternating-series test, the tail end that begins

$$\frac{3^3}{3!} - \frac{3^4}{4!} + \frac{3^5}{5!} - \frac{3^6}{6!} - \cdots$$

converges. Call its sum S. If the front end

$$\frac{3}{1!} - \frac{3^2}{2!}$$

is added on, we obtain the original series, which therefore converges and has the sum

$$\frac{3}{1!} - \frac{3^2}{2!} + S.$$

$\diamond$

As Example 2 illustrates, the alternating-series test works as long as the k^{th} term approaches 0 and the terms decrease in size from some point on.

It may seem that any alternating series whose k^{th} term approaches 0 converges. *This is not the case*, as shown by this series:

$$\frac{2}{1} - \frac{1}{1} + \frac{2}{2} - \frac{1}{2} + \frac{2}{3} - \frac{1}{4} + \cdots, \tag{11.6.1}$$

whose terms alternate $2/k$ and $-1/k$.

Let S_n be the sum of the first n terms of (11.6.1). Then

$$\begin{array}{lllll} S_2 &=& \frac{2}{1} - \frac{1}{1} &=& \frac{1}{1}, \\ S_4 &=& \left(\frac{2}{1} - \frac{1}{1}\right) + \left(\frac{2}{2} - \frac{1}{2}\right) &=& \frac{1}{1} + \frac{1}{2}, \\ S_6 &=& \left(\frac{2}{1} - \frac{1}{1}\right) + \left(\frac{2}{2} - \frac{1}{2}\right) + \left(\frac{2}{3} - \frac{1}{3}\right) &=& \frac{1}{1} + \frac{1}{2} + \frac{1}{3}, \end{array}$$

and, more generally,

$$S_{2n} = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}.$$

Recall that the harmonic series diverges.

Since S_{2n} gets arbitrarily large as $n \to \infty$, the series (11.6.1) diverges.

Absolute Convergence

Consider the series

$$a_1 + a_2 + \cdots + a_n \cdots ,$$

whose terms may be positive, negative, or zero. It is reasonable to expect it to behave at least as "nicely" as the corresponding series with non-negative terms

$$|a_1| + |a_2| + \cdots + |a_n| + \cdots ,$$

since by making all the terms positive we give the series more chance to diverge. This is similar to the case with improper integrals in Section 7.8, where it was shown that if $\int_a^\infty |f(x)|\, dx$ converges, then so does $\int_a^\infty f(x)\, dx$. The next theorem (and its proof) is similar to the Absolute-Convergence Test for Improper Integrals in Section 7.8. (Re-read it. It's on page 662.)

Theorem 11.6.2. *(Absolute-Convergence Test) If the series $\sum_{k=1}^\infty |a_k|$ converges, then so does the series $\sum_{k=1}^\infty a_k$. Furthermore, if $\sum_{k=1}^\infty |a_k| = S$, then $\sum_{k=1}^\infty a_k$ is between $-S$ and S.*

Absolute-Convergence Test

Proof

We introduce two series in order to record the behavior of the positive and negative terms in $\sum_{k=1}^\infty a_k$ separately. Let

$$b_k = \begin{cases} a_k & \text{if } a_k \text{ is positive} \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad c_k = \begin{cases} a_k & \text{if } a_k \text{ is negative} \\ 0 & \text{otherwise} \end{cases} .$$

Note that $a_k = b_k + c_k$. To establish the convergence of $\sum_{k=1}^\infty a_k$ we show that both $\sum_{k=1}^\infty b_k$ and $\sum_{k=1}^\infty c_k$ converge. First of all, since b_k is non-negative and $b_k \le |a_k|$, the series of positive terms, $\sum_{k=1}^\infty b_k$, converges by the comparison test. In fact, it converges to a number $P \le S$.

Since c_k is non-positive, and $c_k \ge -|a_k|$, the series of negative terms, $\sum_{k=1}^\infty c_k$, converges to a number $N \ge -S$. Thus $\sum_{k=1}^\infty a_k = \sum_{k=1}^\infty (b_k + c_k)$ converges to $P + N$, which is between $-S$ and S. •

EXAMPLE 3 Examine the series

$$\frac{\cos(x)}{1^2} + \frac{\cos(2x)}{2^2} + \frac{\cos(3x)}{3^2} + \cdots + \frac{\cos(kx)}{k^2} + \cdots \tag{11.6.2}$$

for convergence or divergence.

SOLUTION The number x is fixed. The numbers $\cos(kx)$ may be positive, negative, or zero, in an irregular manner. However, for all k, $|\cos(kx)| \le 1$.

The series

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{k^2}$$

In Section 12.7 it is shown that for $0 \leq x \leq 2\pi$, series (11.6.2) sums to $\frac{1}{12}(3x^2 - 6\pi x + 2\pi^2)$.

is the p-series with $p = 2$, which converges (by the integral test). Since $\left|\frac{\cos(kx)}{k^2}\right| \leq \frac{1}{k^2}$, the series

$$\frac{|\cos(x)|}{1^2} + \frac{|\cos(2x)|}{2^2} + \frac{|\cos(3x)|}{3^2} + \cdots + \frac{|\cos(kx)|}{k^2} + \cdots \tag{11.6.3}$$

converges by the comparison test. Theorem 11.6.2 then tells us that (11.6.2) converges. ◇

WARNING *(Converse of Theorem 11.6.2 is false)* If $\sum_{k\to\infty} a_k$ converges, then $\sum_{k\to\infty} |a_k|$ may converge or diverge. The standard counterexample to the converse of Theorem 11.6.2 is the alternating harmonic series, $\frac{1}{1} - \frac{1}{2} + \frac{1}{3} - \cdots$. This series converges, as was shown by the alternating-series test (Theorem 11.6.1). But, when all of the terms are replaced by their absolute values, the resulting series is the harmonic series, $\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots$, which diverges (it is a p-series with $p = 1$).

The following definitions are frequently used in describing these various cases of convergence or divergence.

DEFINITION *(Absolute Convergence)* A series $a_1 + a_2 + \cdots$ is said to **converge absolutely** if the series $|a_1| + |a_2| + \cdots$ converges.

Theorem 11.6.2 can then be stated simply: "If a series converges absolutely, then it converges."

DEFINITION *(Conditional Convergence)* A series $a_1 + a_2 + \cdots$ is said to **converge conditionally** if it converges but does *not* converge absolutely.

$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$ converges conditionally.

For instance, the alternating harmonic series $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$ is conditionally convergent.

Absolute-Limit-Comparison Test

When you combine the limit-comparison test for positive series with the absolute-convergence test, you obtain a single test, described in Theorem 11.6.3.

Absolute-Limit-Comparison Test

Theorem 11.6.3. *(Absolute-Limit-Comparison Test) Let $\sum_{k=1}^{\infty} a_k$ be a series whose terms may be negative or positive. Let $\sum_{k=1}^{\infty} c_k$ be a convergent series of positive terms. If*

$$\lim_{k\to\infty} \left|\frac{a_k}{c_k}\right|$$

exists, then $\sum_{k=1}^{\infty} a_k$ is absolutely convergent, hence convergent.

Proof

Note that $|a_k/c_k| = |a_k|/c_k$, since c_k is positive. The limit-comparison test tells us that $\sum_{k=1}^{\infty} |a_k|$ converges. Then the absolute-convergence test assures us that $\sum_{k=1}^{\infty} a_k$ converges. •

One advantage of the absolute-convergence test over the limit-comparison test is that we don't have to follow it by the absolute-convergence test. Another is that we don't have to worry about the arithmetic of negative numbers.

EXAMPLE 4 Show that

$$\frac{3}{2}\left(\frac{1}{2}\right) - \frac{5}{2}\left(\frac{1}{2}\right)^2 + \frac{7}{3}\left(\frac{1}{2}\right)^3 - \cdots + (-1)^{k+1}\frac{2k+1}{k}\left(\frac{1}{2}\right)^k + \cdots \qquad (11.6.4)$$

converges.

SOLUTION Consider the series with positive terms

$$\frac{3}{2}\left(\frac{1}{2}\right) + \frac{5}{2}\left(\frac{1}{2}\right)^2 + \frac{7}{3}\left(\frac{1}{2}\right)^3 + \cdots + \frac{2k+1}{k}\left(\frac{1}{2}\right)^k + \cdots .$$

The fact that $(2k+1)/k \to 2$ as $k \to \infty$ suggests use of the limit-comparison test, comparing the second series to the convergent geometric series $\sum_{k=1}^{\infty}(1/2)^k$. We have

$$\lim_{k\to\infty} \frac{\frac{2k+1}{k}\left(\frac{1}{2}\right)^k}{\left(\frac{1}{2}\right)^k} = 2.$$

Thus $\sum_{k=1}^{\infty}((2k+1)/k)(1/2)^k$ converges. Consequently, the first series (11.6.4), with both positive and negative terms, converges absolutely. Thus it converges. ◇

Absolute-Ratio Test

The ratio test of Section 11.5 also has an analog that applies to series with both negative and positive terms.

Theorem 11.6.4 (Absolute-Ratio Test)**.** *Let $\sum_{k=1}^{\infty} a_k$ be a series such that* Absolute-Ratio Test

$$\lim_{k\to\infty}\left|\frac{a_{k+1}}{a_k}\right| = r < 1.$$

Then $\sum_{k=1}^{\infty} a_k$ converges. If $r > 1$ or if $\lim_{k\to\infty}|a_{k+1}/a_k| = \infty$, then $\sum_{k=1}^{\infty} a_k$ diverges. If $r = 1$, then the Absolute-Ratio Test gives no information.

Proof

Take the case $r < 1$. By the Ratio Test, $\sum_{k=1}^{\infty} |a_k|$ converges. Since $\sum_{k=1}^{\infty} |a_k|$ converges, it follows that $\sum_{k=1}^{\infty} a_k$ converges also.

The case $r > 1$ is treated in Exercise 34.

The case $r = \infty$ can be treated as follows. If $\lim_{k\to\infty} |a_{k=1}/a_k| = \infty$, the ratio $|a_{k+1}|/|a_k|$ gets arbitrarily large as $k \to \infty$. So from some point on the positive numbers $|a_k|$ increase. By the k^{th}-Term Test for Divergence, $\sum_{k=1}^{\infty} a_k$ is divergent. •

The Absolute-Ratio Test avoids work with minus signs.

Theorem 11.6.4 establishes the convergence of the series in Example 4 as follows. Let $a_k = (-1)^{k+1}\frac{(2k+1)}{k2^k}$. Then

$$\left|\frac{a_{k+1}}{a_k}\right| = \left|\frac{(-1)^{k+2}\frac{(2k+3)}{(k+1)2^{k+1}}}{(-1)^{k+1}\frac{(2k+1)}{k2^k}}\right| = \frac{2k+3}{2k+1}\cdot\frac{k}{k+1}\cdot\frac{1}{2},$$

which approaches $r = \frac{1}{2}$ as $k \to \infty$. Thus $\sum_{k=1}^{\infty} a_k$ converges (in fact, absolutely).

Rearrangements

$1 + 13 + 15 + 27 = 13 + 27 + 15 + 1$

The sum of a finite collection of numbers does not depend on the order in which they are added. A series that converges absolutely is similar: no matter how the terms of an absolutely convergent series are rearranged, the new series converges and has the same sum as the original series. It might be expected that any convergent series has this property, but this is not the case. For instance, the alternating harmonic series

$$\frac{1}{1} - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots \tag{11.6.5}$$

does not. To show this, rearrange the terms so that two positive terms alternate with one negative term, as follows:

$$\frac{1}{1} + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \cdots. \tag{11.6.6}$$

Rearranging the terms in a conditionally convergent series is dangerous.

The positive summands in (11.6.6) have much more influence than the negative summands. In the battle between the positives and the negatives, the positives will win by a bigger margin in (11.6.6) than in (11.6.5). In fact, the sum of (11.6.6) is $\frac{3}{2}\ln(2)$, while Exercise 28 shows that the sum of (11.6.5) is $\ln(2)$.

Conditionally convergent series are so sensitive that they can be made to sum to any number that you choose. To be precise, Riemann proved: if $\sum_{k=1}^{\infty} a_k$ is a conditionally convergent series and s is any real number, then there is a rearrangement of the a_ks whose sum is s. This is proved in Exercise 40.

Summary

Earlier in this chapter we described ways to test for the convergence or divergence of series whose terms are all positive. This section describes several tests for series that may be a mix of positive and negative terms.

- If the signs alternate and the absolute value of the terms decreases and approach 0, the series converges. [Alternating-Series Test]
- If the series converges when "all the terms are made positive," then it converges. [Absolute-Convergence Test]
- This Absolute-Convergence Test in combination with the Limit-Comparison Test gives us a single test, called the Absolute-Limit-Comparison Test.
- The Absolute-Convergence Test in combination with the Ratio Test gives us the Absolute-Ratio Test. (This will be the most important test in Chapter 12.)

EXERCISES for Section 11.6

Exercises 1 to 8 concern alternating series. Determine which series converge and which diverge. Explain your reasoning.

1. $\frac{1}{2} - \frac{2}{3} + \frac{3}{4} - \frac{4}{5} + \cdots + (-1)^{k+1}\frac{k}{k+1} + \cdots$

2. $-\frac{1}{1+\frac{1}{2}} + \frac{1}{1+\frac{1}{4}} - \frac{1}{1+\frac{1}{8}} + \cdots + (-1)^k\frac{1}{1+2^{-k}} + \cdots$

3. $\frac{1}{\sqrt{1}} - \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} - \frac{1}{\sqrt{4}} + \cdots + (-1)^{k+1}\frac{1}{\sqrt{k}} + \cdots$

4. $\frac{5}{1!} - \frac{5^2}{2!} + \frac{5^3}{3!} - \frac{5^4}{4!} + \cdots + (-1)^{k+1}\frac{5^k}{k!} + \cdots$

5. $\frac{3}{\sqrt{1}} - \frac{2}{\sqrt{1}} + \frac{3}{\sqrt{2}} - \frac{2}{\sqrt{2}} + \frac{3}{\sqrt{3}} - \frac{2}{\sqrt{3}} + \cdots$

6. $\sqrt{1} - \sqrt{2} + \sqrt{3} - \sqrt{4} + \cdots + (-1)^{k+1}\sqrt{k} + \cdots$

7. $\frac{1}{3} - \frac{2}{5} + \frac{3}{7} - \frac{4}{9} + \frac{5}{11} - \cdots + (-1)^{k+1}\frac{k}{2k+1} + \cdots$

8. $\frac{1}{1^2} - \frac{1}{2^2} + \frac{1}{3^2} - \frac{1}{4^2} + \cdots + (-1)^{k+1}\frac{1}{k^2} + \cdots$

9. Consider the alternating harmonic series

$$\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}.$$

(a) Compute S_5 and S_6 to five decimal places.

(b) Is the estimate S_5 smaller or larger than the sum of the series?

(c) Use (a) and (b) to find two numbers between which the sum of the series must lie.

10. Consider the series $\sum_{k=1}^{\infty}(-1)^{k+1}\frac{2^{-k}}{k}$.

(a) Estimate the sum of the series using S_6.

(b) Estimate the error R_6.

11. Does the series

$$\frac{2}{1} - \frac{3}{2} + \frac{4}{3} - \frac{5}{4} + \cdots + (-1)^{k+1}\left(\frac{n+1}{n}\right) + \cdots.$$

converge or diverge?

In Exercises 12 to 26 determine which series diverge, converge absolutely, or converge conditionally. Explain your answers.

12. $\sum_{k=1}^{\infty} \frac{(-1)^k}{\sqrt[3]{k^2}}$

13. $\sum_{k=1}^{\infty} \ln\left(\frac{1}{k}\right)$

14. $\sum_{k=1}^{\infty} \frac{(-1)^k}{k\ln(k)}$

15. $\sum_{k=1}^{\infty} \frac{\sin(k)}{k^{1.01}}$

16. $\sum_{k=1}^{\infty} \left(1-\cos\left(\frac{\pi}{k}\right)\right)$

17. $\sum_{k=1}^{\infty} (-1)^k \cos\left(\frac{\pi}{k^2}\right)$

18. $\sum_{k=1}^{\infty} \frac{(-2)^k}{k!}$

19. $\frac{1}{1^2}+\frac{1}{2^2}-\frac{1}{3^2}-\frac{1}{4^2}+\frac{1}{5^2}+\frac{1}{6^2}-\cdots$ There are two +'s alternating with two −'s.

20. $\sum_{k=1}^{\infty} \frac{(-3)^k(1+k^2)}{k!}$

21. $\sum_{k=1}^{\infty} \frac{\cos(k\pi)}{2k+1}$

22. $\sum_{k=1}^{\infty} \frac{(-1)^k(k+5)}{k^2}$

23. $\sum_{k=1}^{\infty} \frac{(-9)^k}{10^k+k}$

24. $\sum_{k=1}^{\infty} \frac{(-1)^k}{\sqrt[3]{k}}$

25. $\sum_{k=1}^{\infty} \frac{(-1.01)^k}{k!}$

26. $\sum_{k=1}^{\infty} \frac{(-\pi)^{2k+1}}{(2k+1)!}$

27. For which values of x does $\sum_{k=1}^{\infty} \frac{x^k}{k!}$ converge?

28. The series $\sum_{k=1}^{\infty}(-1)^{k+1}2^{-k}$ is both a geometric series and a decreasing alternating series whose k^{th} term approaches 0.

(a) Compute S_6 to three decimal places.

(b) Using the fact that the series is a decreasing alternating series, put a bound on R_6.

(c) Using the fact that the series is a geometric series, compute R_6 exactly.

29.

(a) How many terms of the series $\sum_{k=1}^{\infty} \frac{\sin(k)}{k^2}$ must you take to be sure the error is less than 0.005? Explain.

(b) Estimate $\sum_{k=1}^{\infty} \frac{\sin(k)}{k^2}$ to two decimal places.

30. Estimate $\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots$ to two decimal places. Explain your reasoning.

31.

(a) Show $\sum_{k=1}^{\infty} \frac{2^k}{k!}$ converges.

(b) Estimate the sum of the series in (a) to two decimal places.

32. Let $P(x)$ and $Q(x)$ be two polynomials of degree at least one. Assume that for $n \geq 1$, $Q(n) \neq 0$. What relation must there be between the degrees of $P(x)$ and $Q(x)$ if

(a) $\frac{P(k)}{Q(k)} \to 0$ as $k \to \infty$?

(b) $\sum_{k=1}^{\infty} \frac{P(k)}{Q(k)}$ converges absolutely?

(c) $\displaystyle\sum_{k=1}^{\infty}(-1)^k\frac{P(k)}{Q(k)}$ converges absolutely?

33. The Alternating-Series Test was proved only for the alternating harmonic series. Prove it in general. (The only difference is that the k^{th} term is $(-1)^{k+1}p_k$ instead of $(-1)^{k+1}/k$.)

34. This exercise treats the second half of the absolute-ratio test.

(a) Show that if

$$r=\lim_{k\to\infty}\left|\frac{a_{k+1}}{a_k}\right|>1,$$

then $|a_k|\to\infty$ as $k\to\infty$. (First show that there is a number s, $s>1$, such that for some integer N, $|a_{k+1}|>s|a_k|$ for all $k\geq N$.)

(b) From (a) deduce that a_k does not approach 0 as $k\to\infty$.

35. For which values of x does the series $\displaystyle\sum_{k=1}^{\infty}\frac{kx^k}{2k+1}$ diverge? converge conditionally? converge absolutely? Record your conclusions in a diagram on the x-axis.

36. Repeat Exercise 35 for the series *(a)* $\displaystyle\sum_{k=1}^{\infty}\frac{x^k}{k!}$ and *(b)* $\displaystyle\sum_{k=1}^{\infty}\frac{x^k}{k^2}$.

37. Is this argument okay? Add the alternating harmonic series to half of itself:

$$\begin{array}{ccccccccccccccc}
1 & -\frac{1}{2} & +\frac{1}{3} & -\frac{1}{4} & +\frac{1}{5} & -\frac{1}{6} & +\frac{1}{7} & -\frac{1}{8} & +\frac{1}{9} & -\frac{1}{10} & -\frac{1}{11} & +\frac{1}{12} & +\cdots & = & S \\
 & \frac{1}{2} & & -\frac{1}{4} & & +\frac{1}{6} & & -\frac{1}{8} & & +\frac{1}{10} & & -\frac{1}{12} & +\cdots & = & \frac{1}{2}S \\
\hline
1 & & +\frac{1}{3} & -\frac{1}{2} & +\frac{1}{5} & & +\frac{1}{7} & -\frac{1}{4} & +\frac{1}{9} & & -\frac{1}{11} & & +\cdots & = & \frac{3}{2}S
\end{array}$$

Rearranging the last line produces the alternating harmonic series, whose sum is S. Thus $S=\frac{3}{2}S$, from which it follows that $S=0$.

38.

Sam: I have a neat proof that absolute convergence implies convergence. First of all,

$$a_n=a_n+|a_n|-|a_n|.$$

Jane: True, but why do that?

Sam: Don't interrupt me. Just wait. Now $a_n+|a_n|$ is 0 if a_n is negative and it's $2|a_n|$ if a_n is positive. Right?

Jane: If you say so.

Sam: Just think.

Jane: Yes, I agree.

Sam: So $0 \leq a_n + |a_n| \leq 2|a_n|$. Right? So $\sum(a_n + |a_n|)$ converges.

Jane: Yes.

Sam: You can fill in the rest, yes?

Jane: Oh, neat.

Sam: Yeh, mathematicians really like this proof.

Is the proof correct? (Explain your answer.) Which proof do you prefer, this one or the one on page 973?

39. If $\displaystyle\sum_{k=1}^{\infty} a_k$ converges and $a_k > 0$ for all k, what, if anything, can we say about the convergence or divergence of (a) $\displaystyle\sum_{k=1}^{\infty} \sin(a_k)$ and (b) $\displaystyle\sum_{k=1}^{\infty} \cos(a_k)$?

40. Prove that if $\sum_{k=1}^{\infty} a_k$ is a conditionally convergent series and s is any real number, then there is a rearrangement of $\sum_{k=1}^{\infty} a_k$ whose sum is s. (A conditionally convergent series must have an endless supply of both positive and negative numbers. Moreover, the series of positive terms and the series of negative terms, separately, diverge. Use these facts to explain how to construct a rearrangement that converges to s.)

41. In the proof of the Absolute-Convergence Theorem, why does $\sum_{k=1}^{\infty} c_k$ converge and have a sum greater than or equal to $-S$?

42. The Absolute-Convergence Test asserts that $\sum_{k=1}^{\infty} a_k$ is between $-S$ and S. Why is that?

11.S Chapter Summary

This chapter concerns sequences formed by adding a finite number of terms from another sequence: $S_n = a_1 + a_2 + \cdots + a_n$. Two questions motivate the sections:

- Does $\lim_{n\to\infty} S_n$ exist?
- If the limit exists, how do we estimate it?

If the limit exists, it is denoted $\displaystyle\sum_{k=1}^{\infty} a_k$, though we never add an infinite number of summands.

Some of the tests for convergence or divergence apply only to series whose terms are positive (or all are negative): the Integral Test, the Comparison Tests, and the Ratio Tests.

For series whose terms a_k may be both positive and negative, the key is that if $\sum_{k=1}^{\infty} |a_k|$ converges so must $\sum_{k=1}^{\infty} a_k$. Moreover, if $\sum_{k=1}^{\infty} |a_k| = L$, then $-L \leq \sum_{k=1}^{\infty} a_k \leq L$.

If the series alternates, $a_1 - a_2 + a_3 - a_4 + \cdots$ and $a_k \to 0$ monotonically, then $\sum_{k=1}^{\infty} a_k$ converges.

The Integral Test, the Comparison Tests, and the formula for the sum of a geometric series also provide ways to estimate the error in using a particular S_n to approximate the sum of the series.

EXERCISES for 11.S

1. Explain in your own words.

(a) Why the Comparison Test for convergence works.

(b) Why the Ratio Test for convergence works.

(c) Why the Alternating-Series Test works.

(d) Why the Absolute-Convergence Test works.

2. How many terms of the series $\sum_{k=1}^{\infty}(-1)^{n+1}(1/n^2)$ should be used to estimate its sum to three-decimal place accuracy?

3. For which type of series does each of these tests imply convergence:

(a) Alternating-Series Test

(b) Integral Test

(c) Comparison Test

(d) Absolute-Convergence Test

(e) Absolute-Ratio Test.

4. Assume that $|a_k| \leq 1/2^n$ for $n \geq 1$.

(a) Must $\sum_{k=1}^{\infty} |a_k|$ converge? If so, what can you say about its sum?

(b) Must $\sum_{k=1}^{\infty} a_k$ converge? If so, what can you say about its sum?

Sometimes convergence or divergence of a series can be established by more than one of the tests developed in this chapter. In Exercises 5 to 10 determine the convergence or divergence of the given series by as many tests that can be applied in each case.

5. $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2}$

6. $\displaystyle\sum_{i=1}^{\infty} \frac{(-1)^i}{3^i}$

7. $\displaystyle\sum_{k=1}^{\infty} \frac{\sqrt{k}}{k^2+1}$

8. $\displaystyle\sum_{k=1}^{\infty} \frac{\sqrt{k}}{k^2-2}$

9. $\displaystyle\sum_{i=1}^{\infty} \left(\frac{3+1/n}{2+1/n}\right)^n$

10. $\displaystyle\sum_{n=1}^{\infty} \left(\frac{2}{3+1/n}\right)^n$

11. What is the Comparison Test and how can it be used to estimate the error when using part of a series to approximate the sum of the series?

12. What do the three expressions "convergent," "conditionally convergent," and "absolutely convergent" mean?

13. What tests could be used to to test a series for convergence if you know that $\lim_{k\to\infty} a_{k+1}/a_k = -1/3$? Explain.

14. Assume that $\lim_{k\to\infty} a_k = 2$. For what values of s does $\sum_{k=1}^{\infty} a_k s^k$ converge?

15. For what values of p does $\sum_{k=1}^{\infty} 1/k^p$ converge?

16. If $\lim_{k\to\infty} a_{k+1}/a_k = 1$, what can we conclude about the series $\sum_{k\to\infty} a_k$?

17. For what values of q does $\sum_{k=1}^{\infty}(-1)^k k^q$ (a) converge? (b) converge absolutely?

18. If $\sum_{k=0}^{\infty} a_k$ is convergent, does it follow that

(a) $\lim_{n\to\infty} a_k = 0$?

(b) $\lim_{n\to\infty}(a_k + a_{k+1}) = 0$?

(c) $\lim_{n\to\infty} \sum_{k=n}^{2n} a_k = 0$?

(d) $\lim_{n\to\infty} \sum_{k=n}^{\infty} a_k = 0$?

Compare with Exercise 5 in Chapter 7.

19. Let $\sum_{k=0}^{\infty} a_k$ be a conditionally convergent series. It is made up of a subsequence of non-negative terms and a subsequence of negative terms.

(a) Could both of these subsequences be convergent?

(b) Could exactly one of theme be convergent?

(c) Could neither be convergent?

20. In an energy problem one meets the integral

$$\int_0^{\pi/2} \frac{\sin x}{e^x - 1}\, dx.$$

Note that the integrand is not defined at $x = 0$. Is that a big obstacle? Is this integral convergent or divergent? Do not try to evaluate the integral.

21. Give an example of a convergent series of positive terms $\{a_k\}$ such that $\lim_{n\to\infty} \frac{a_{n+1}}{a_n}$ does not exist but $\lim_{n\to\infty} \frac{a_{n+1}}{a_n}$ is not ∞.

22. Assume that f is continuous on $[0,\infty)$ and has period one, that is, $f(x) = f(x+1)$ for all x in $[0,\infty)$. Assume also that $\int_0^{\infty} e^{-x} f(x)\, dx$ is convergent. Show that

$$\int_0^{\infty} e^{-x} f(x)\, dx = \frac{e}{e-1} \int_0^{1} e^{-x} f(x)\, dx.$$

In Exercises 23 to 28 a short formula for estimating $n!$ is obtained.

23. Let f have the properties that for $x \geq 1$, $f(x) \geq 0$, $f'(x) > 0$, and $f''(x) < 0$. Let a_n be the area of the region below the graph of $y = f(x)$ and above the line segment that joins $(n, f(n))$ with $(n+1, f(n+1))$.

(a) Draw a large-scale version of Figure 11.S.1. The individual regions of area a_1, a_2, a_3, and a_4 should be clear and not too narrow.

(b) Using geometry, show that the series $a_1 + a_2 + a_3 + \cdots$ converges and has a sum no larger than the area of the triangle with vertices $(1, f(1))$, $(2, f(2))$, $(1, f(2))$.

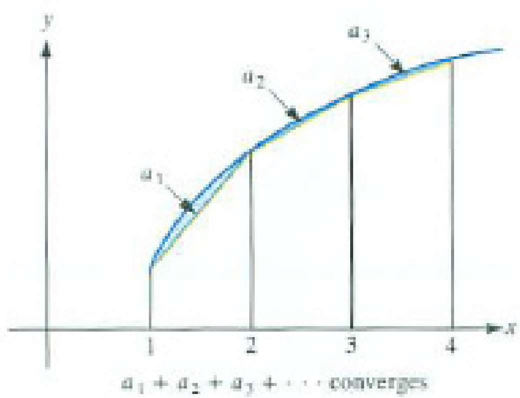


Figure 11.S.1

24. Let $y = \ln(x)$.

(a) Using Exercise 23, show that as $n \to \infty$,

$$\int_1^n \ln(x)\, dx - \left(\frac{\ln(1)+\ln(2)}{2} + \frac{\ln(2)+\ln(3)}{2} + \cdots + \frac{\ln(n-1)+\ln(n)}{2}\right)$$

has a limit; denote this limit by C.

(b) Show that (a) is equivalent to the assertion

$$\lim_{n\to\infty} \left(n\ln(n) - n + 1 - \ln(n!) + \ln(\sqrt{n})\right) = C.$$

25. From Exercise 24(b), deduce that there is a constant k such that

$$\lim_{n\to\infty} \frac{n!}{k(n/e)^n\sqrt(n)} = 1.$$

Exercises 26 and 27 are related. Review Example 8 of Section 8.3 first.

26. Let $I_n = \int_0^{\pi/2} \sin^n(\theta)\, d\theta$, where n is is a nonnegative integer.

(a) Evaluate I_0 and I_p.

(b) Show that

$$I_{2n} = \frac{2n-1}{2n}\frac{2n-3}{2n-2}\cdots\frac{3}{4}\frac{1}{2}\frac{\pi}{2} \qquad \text{and} \qquad I_{2n+1} = \frac{2n}{2n+1}\frac{2n-2}{2n-1}\cdots\frac{4}{5}\frac{2}{3}.$$

(c) Show that

$$\frac{I_7}{I_6} = \frac{6}{7}\frac{6}{5}\frac{4}{5}\frac{4}{3}\frac{2}{3}\frac{2}{1}\frac{2}{\pi}.$$

(d) Show that

$$\frac{I_{2n+1}}{I_{2n}} = \frac{2n}{2n+1}\frac{2n}{2n-1}\frac{2n-2}{2n-1}\cdots\frac{2}{3}\frac{2}{1}\frac{2}{\pi}.$$

(e) Show that

$$\frac{2n}{2n+1}I_{2n} < \frac{2n}{2n+1}I_{2n-1} = I_{2n+1} < I_{2n},$$

and thus $\lim_{n\to\infty} \frac{I_{2n+1}}{I_{2n}} = 1$.

(f) From (d) and (e), deduce that

$$\lim_{n\to\infty} \frac{2\cdot 2}{1\cdot 3}\frac{4\cdot 4}{3\cdot 5}\frac{6\cdot 6}{5\cdot 7}\cdots\frac{(2n)(2n)}{(2n-1)(2n+1)} = \frac{\pi}{2}.$$

This is **Wallis's formula**, usually written in shorthand as

$$\frac{2\cdot 2}{1\cdot 3}\frac{4\cdot 4}{3\cdot 5}\frac{6\cdot 6}{5\cdot 7}\cdots = \frac{\pi}{2}$$

27.

(a) Show that $2\cdot 4\cdot 6\cdot 8\cdots 2n = 2^n n!$.

(b) Show that $1\cdot 3\cdot 5\cdot 7\cdots(2n-1) = \frac{(2n)!}{2^n n!}$.

(c) From Exercise 26 deduce that

$$\lim_{n\to\infty} \frac{(n!)^2 4^n}{(2n)!\sqrt{2n+1}} = \sqrt{\frac{\pi}{2}}.$$

28.

(a) Using Exercise 27(c), show that k in Exercise 25 equals $\sqrt{2\pi}$. Thus a good estimate of $n!$ is provided by the formula

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

This is known as **Stirling's formula**.

(b) Using the factorial key on a calculator, compute (20)!. Then compute the ratio $\sqrt{2\pi n}(n/e)^n/n!$ for $n = 20$.

29. Let $\{a_k\}$ and $\{b_k\}$ be sequences of positive terms. Assume that for all k

$$\frac{a_{k+1}}{a_k} \leq \frac{b_{k+1}}{b_k}.$$

(a) Prove that if $\sum_{k=1}^{\infty} b_k$ converges, so does $\sum_{k=1}^{\infty} a_k$. (Rewrite the inequality as $a_{k+1}/b_{k+1} \leq a_k/b_k$,)

(b) Use the result in (a) to prove that if $\lim_{k\to\infty} a_{k+1}/a_k = r < 1$, then $\sum_{k=1}^{\infty} a_k$ converges.

Calculus is Everywhere # 14

$E = mc^2$

The equation $E = mc^2$ relates the energy associated with an object to its mass and the speed of light. Where does it come from?

Newton's second law of motion reads: "Force is the rate at which the momentum of an object changes." The momentum of an object of mass m and velocity v is the product mv. Denoting the force by F, we have

$$F = \frac{d}{dt}(mv).$$

If the mass is constant, this reduces to the familiar "force equals mass times acceleration." But what if the mass m is not constant? What if the mass of an object changes as its velocity changes?

According to Einstein's Special Theory of Relativity, announced in 1905, the mass does change, in a manner described by the equation:

$$m = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}}. \tag{C.14.1}$$

Here m_0 is the mass at rest, v is the velocity, and c is the velocity of light. If v is not zero, m is larger than m_0. When v is small (compared to the velocity of light) then m is only slightly larger than m_0. However, as v approaches the velocity of light, the mass becomes arbitrarily large.

For a satellite circling the Earth at 17,000 miles per hour, v/c is less than 1/2500.

Consider moving an object, initially at rest, in a straight line. If the velocity at time t is $v(t)$, then the displacement is $x(t) = \int_0^t v(s)\ ds$. Assuming the object is initially at rest $v(0) = 0$, the work done by a varying force F in moving the object during the time interval $[0, T]$ is

$$\begin{aligned}
\int_0^T F(t)v(t)\ dt &= \int_0^T (mv)'v\ dt \\
&= (mv)v|_0^T - \int_0^T mv(v')\ dt \\
&= m(v(T))^2 - \int_0^T \frac{m_0 v v'}{\sqrt{1-\frac{v^2}{c^2}}}\ dt \\
&= m(v(T))^2 - \left(-c^2 m_0 \sqrt{1-\frac{v^2}{c^2}}\right)\Bigg|_0^T \\
&= m(v(T))^2 - \left(-c^2 m_0 \sqrt{1-\frac{(v(T))^2}{c^2}} + c^2 m_0 \sqrt{1-\frac{0^2}{c^2}}\right) \\
&= m(v(T))^2 + c^2 m_0 \sqrt{1-\frac{(v(T))^2}{c^2}} - m_0 c^2 \\
&= m(v(T))^2 + mc^2\left(1-\frac{(v(T))^2}{c^2}\right) - m_0 c^2 \\
&= m(v(T))^2 + mc^2 - m(v(T))^2 - m_0 c^2 \\
&= mc^2 - m_0 c^2. \qquad \text{(C.14.2)}
\end{aligned}$$

We can interpret this as saying that the "total energy associated with the object" increases from m_0c^2 to mc^2. The energy of the object at rest is then m_0c^2, called its **rest energy**.

That is the mathematics behind the equation $E = mc^2$. It suggests that mass may be turned into energy, as Einstein predicted. For instance, in a nuclear reactor some of the mass of the uranium is indeed turned into energy in the fission process. Also, the mass of the sun decreases as it emits radiant energy.

What about the equation that states kinetic energy is half the product of the mass and the square of the velocity? Exercise 2 uses (C.14.2) when v is small (compared to c), to show the total increase in energy is approximated by the familiar kinetic energy

$$mc^2 - m_0c^2 \approx \frac{1}{2}m_0v^2. \qquad \text{(C.14.3)}$$

EXERCISES

1. Provide a brief explanation for each step in the derivation of (C.14.2).

2. This exercise requires the use of the Binomial Theorem, which is not presented until Section 12.1. See also, Exercise 29 in Section 12.1. Assume v is small, compared to c. Use the first two terms of the binomial series for $\left(1 - x^2\right)^{-1/2}$, with $x = v^2/c^2$, to derive (C.14.3). That is, show that

$$mc^2 - m_0c^2 \approx \frac{1}{2}m_0v^2.$$

Chapter 12

Applications of Series

The preceding chapter developed several tests for determining the convergence or divergence of an infinite series. This chapter uses infinite series to approximate functions, such as e^x, evaluate integrals, and find limits in the indeterminate form 'zero-over-zero." After a section devoted to complex numbers, we will use them to expose the close link between trigonometric and exponential functions.

12.1 Taylor Series

Section 5.5 introduced the n^{th}-order Taylor polynomial of a function f centered at a as the polynomial P_n that agrees with f and its first n derivatives at $x = a$:

$$\begin{aligned} P_n(x;a) &= f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n \\ &= \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k \end{aligned}$$

The sequence of Taylor polynomials $P_0(x;a)$, $P_1(x;a)$, ..., $P_n(x;a)$, ... can now be viewed as the sequence of partial sums of the infinite series

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k.$$

This series is called the **Taylor series at a associated with the function f**. When $a = 0$, the series is also called the **Maclaurin series associated with f**.

A partial sum of a Taylor series is a Taylor polynomial; a partial sum of a Maclaurin series is a Maclaurin polynomial.

EXAMPLE 1 Show that the limit of the Maclaurin series associated with e^x is e^x,
SOLUTION By Section 5.5 the series is $\sum_{k=0}^{\infty} \frac{x^n}{n!}$. We want to show that the series converges to e^x. The absolute ratio test shows that the series converges, but it does not tell us that its limit is e^x.

Also by Section 5.5, the difference between $f(x)$ and its Maclaurin polynomial up through the power x^n has the form

$$\frac{f^{(n+1)}(c_n)}{(n+1)!}x^{n+1} \tag{12.1.1}$$

for some number c_n between 0 and x. In the case $f(x) = e^x$, we have $f^{(n+1)}(x) = e^x$, hence $f^{(n+1)}(c_n) = e^{c_n}$. Thus the error (12.1.1) equals

$$\frac{e^{c_n}x^{n+1}}{(n+1)!}.$$

For $x > 0$, we know $c_n < x$ so $c^{c_n} < e^x$; for $x < 0$, $c_n < 0$, so $e^{c_n} < 1$. In either case e^{c_n} is less than a fixed number, which we call M. Thus $e^{c_n} < M$ for all n. Keeping in mind that x is fixed, we see that

$$\lim_{n\to\infty} \frac{|e^{c_n}x^{n+1}|}{(n+1)!} \leq M\frac{|x|^{n+1}}{(n+1)!}. \tag{12.1.2}$$

It was shown in Section 11.2 that $\lim_{n\to\infty} k^n/n!$ is 0 for any fixed number k. Thus (12.1.2) approaches 0 as $n \to \infty$, which means that the sum of the series is e^x. We have:

For all x

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

This provides a way to estimate e^x using only addition, multiplication, and division. In particular, when $x = 1$, it gives a series representation of e:

$$e = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} + \cdots .$$

Euler used this formula to evaluate e to 23 decimal places (without the aid of a calculator). ◇

EXAMPLE 2 Use the Maclaurin series in Example 1 to estimate $\sqrt{e} = e^{1/2}$ with an error of at most 0.001.

SOLUTION The error in using the front end $\sum_{k=0}^{n}(1/2)^k/k!$ has the form

$$e^{c_n}\frac{\left(\frac{1}{2}\right)^{n+1}}{(n+1)!}$$

where c_n is between 0 and 1/2. Then $e^{c_n} < e^{1/2}$, which is less than 2, because $2^2 > 3$. So we want to find n large enough so that

$$\frac{2\left(\frac{1}{2}\right)^{n+1}}{(n+1)!} < 0.001.$$

To find such a number n, we experiment, making a little table, with 4-decimal place accuracy We stop at $n = 4$ with an error less than 0.001. Rounded to

n	1	2	3	4
$2\left(\frac{1}{2}\right)^{n+1}/(n+1)!$	0.2500	0.0417	0.0026	0.0005

Table 12.1.1

five decimal places, the estimate for $\sqrt{e}$ is

$$1 + \frac{1}{2} + \frac{\left(\frac{1}{2}\right)^2}{2!} + \frac{\left(\frac{1}{2}\right)^3}{3!} + \frac{\left(\frac{1}{2}\right)^4}{4!} \approx 1.64843,$$

which is close to what a calculator shows: 1.6487. ◇

In Section 5.5 the Maclaurin polynomial associated with $\sin(x)$ was computed. Using that result, we conclude that the Maclaurin series associated with $\sin(x)$ is

$$\sum_{k=0}^{\infty}(-1)^k\frac{x^{2k+1}}{(2k+1)!} = x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\cdots+(-1)^n\frac{x^{2n+1}}{(2n+1)!}+\cdots. \quad (12.1.3)$$

The next Example shows that its sum is $\sin(x)$.

EXAMPLE 3 Show that $\sum_{k=0}^{\infty}(-1)^k\frac{x^{2k+1}}{(2k+1)!} = \sin(x)$.
SOLUTION To show that the series converges to $\sin(x)$ we must show that the difference between $\sin(x)$ and $\sum_{k=0}^{n}(-1)^k x^{2k+1}/(2k+1)!$ approaches 0 as $n\to\infty$.

To do this we again make use of Lagrange's formula, which involves the higher derivatves of $\sin(x)$, which are $\pm\sin(x)$ and $\pm\cos(x)$. In any case, if $f(x)=\sin(x)$, $\left|f^{(n)}(x)\right|\le 1$. Thus we have

$$\left|\frac{f^{(n+1)}(c_n)x^{n+1}}{(n+1)!}\right| \le \frac{|x|^{n+1}}{(n+1)!}.$$

Because the expression $|x|^{n+1}/(n+1)!$ approaches 0 as $n\to\infty$, no matter how large x is, the difference between the Maclaurin polynomials and $\sin(x)$ approaches 0 as their degree is chosen larger. We conclude that the Taylor series (12.1.3) converges to $\sin(x)$ for all x. ◇

Therefore, we may write

For all x

$$\sin(x) = x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\cdots+(-1)^n\frac{x^{2n+1}}{(2n+1)!}+\cdots = \sum_{k=0}^{\infty}(-1)^k\frac{x^{2k+1}}{(2k+1)!}.$$

In a similar manner, we have

For all x

$$\cos(x) = 1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\cdots+(-1)^n\frac{x^{2n}}{(2n)!}+\cdots = \sum_{k=0}^{\infty}\frac{(-1)^k x^{2k}}{(2k)!}.$$

Taylor Series in Powers of $x - a$

Just as there are Taylor polynomials "around 0," there are such polynomials around any number, a. The Taylor series around a associated with $f(x)$ involves powers of $x - a$ instead of powers of x $(= x - 0)$:

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{n!}(x-a)^k.$$

This series may or may not converge; if it converges, it may not converge to $f(x)$.

EXAMPLE 4 Find the Taylor series associated with $1/x$ in powers of $x-1$.
SOLUTION Here $f(x) = 1/x$. This table shows a few of the higher derivatives evaluated at 1. In general,

n	1	2	3	4	5
$f^{(n)}(x)$	$-1/x^2$	$2/x^3$	$\frac{-3\cdot 2}{x^4}$	$\frac{4\cdot 3\cdot 2}{x^5}$	$\frac{-5\cdot 4\cdot 3\cdot 2}{x^6}$
$f^{(n)}(1)$	-1	2	$-3!$	$4!$	$-5!$

Table 12.1.2

$$f^{(n)}(1) = (-1)^n n!.$$

Thus the typical term in the Taylor series around 1 is

$$\frac{(-1)^n n!(x-1)^n}{n!} = (-1)^n (x-1)^n.$$

The series begins

$$1 - (1-x) + (1-x)^2 - (1-x)^3 + \cdots .$$

By the n^{th} term test, the series does not converge if $|x-1| > 1$, that is, $x > 2$ or $x < 0$.

If $x = 0$, the series becomes $\sum_{k=0}^{\infty}(-1)^n(-1)^n = \sum_{k=0}^{\infty} 1$, which, by the n^{th} term test, does not converge. Similarly, it does not converge when $x = 2$. To deal with x in $(0, 2)$ we use the absolute-ratio test,examining

$$\lim_{n\to\infty} \left| \frac{(-1)^{n+1}(x-1)^{n+1}}{(-1)^n (x-1)^n} \right| = \lim_{n\to\infty} |x-1| = |x-1|.$$

So, if $|x-1| < 1$, the series converges. But, does it converge to $1/x$?

The Lagrange formula for the remainder is

$$\frac{f^{(n+1)}(c_n)(x-1)^{n+1}}{(n+1)!} = \frac{(n+1)!}{(c_n)^{n+1}} \frac{(x-1)^{n+1}}{(n+1)!} = \frac{(x-1)^{n+1}}{(c_n)^{n+1}} \qquad (12.1.4)$$

where c_n is between 1 and x. So we need to show that $|(x-1)/c_n|^{n+1} \to 0$ as $n \to \infty$. There is trouble.

For instance, if x is near 0, $|x-1|$ is near 1 and c_n may be near 0, for we know only that c_n is between x and 1. Perhaps the ratio $|(x-1)/c_n|$ is a large number.

However, if x is in $(1,2)$ then we have $c_n > 1$ while $|x-1| < 1$, so

$$0 < \frac{|x-1|}{c_n} < |x-1|.$$

Thus the remainder approaches 0 as $n \to \infty$. So we see that for x in $(1,2)$, $1-(x-1)+(x-1)^2-(1-x)^3-\cdots = 1/x$. The Lagrange formula justifies the same conclusion for x in $(-1/2, 1)$, but doesn't help for x in $(0, 1/2]$, as Exercise 32 shows.

However, $1-(x-1)+(x-1)^2-(1-x)^3-\cdots$ is a geometric series with first term 1 and ratio $r = -(x-1)$. It converges to

$$\frac{1}{1-r} = \frac{1}{1-(1-(-(x-1)))} = \frac{1}{1-x-1} = \frac{1}{x}.$$

This argument covers all x in $(0,2)$ at once. $\diamond$

The General Binomial Theorem

The binomial theorem will be reviewed in an appendix. $\left(\begin{array}{c} r \\ k \end{array}\right) = \frac{r!}{k!(r-k)!}$

If r is 0 or a positive integer, $(1+x)^r$, when multiplied out, is a polynomial of degree r. Its Maclaurin series has only a finite number of nonzero terms, the one of highest degree being x^r. The formula

$$(1+x)^r = \sum_{k=0}^{r} \frac{r!}{k!(r-k)!} x^k$$

To remember it, recall that the coefficient of x^k has k factors in both the numerator and denominator. The factors in the numerator start from r and decrease by one. The factors in the denominator start from 1 and increase by 1.

is known as the **binomial theorem**. It can also be written as

$$(1+x)^r = \sum_{k=0}^{r} \frac{r(r-1)\cdots(r-(k-1))}{1\cdot 2\cdots k} x^k.$$

Example 5 generalizes the binomial theorem to arbitrary exponents r.

EXAMPLE 5 Find the Maclaurin series associated with $f(x) = (1+x)^r$, when r is not 0 or a positive integer and determine its radius of convergence.
SOLUTION The following table will help in computing $f^{(k)}(0)$:

k	$f^{(k)}(x)$	$f^{(k)}(0)$
0	$(1+x)^r$	1
1	$r(1+x)^{r-1}$	r
2	$r(r-1)(1+x)^{r-2}$	$r(r-1)$
3	$r(r-1)(r-2)(1+x)^{r-3}$	$r(r-1)(r-2)$
$\vdots$	$\vdots$	$\vdots$
k	$r(r-1)\cdots(r-k+1)(1+x)^{r-k}$	$r(r-1)(r-2)\cdots(r-k+1)$

Table 12.1.3

Consequently, the Maclaurin series associated with $(1+x)^r$ is

$$1 + rx + \frac{r(r-1)}{1\cdot 2}x^2 + \frac{r(r-1)(r-2)}{1\cdot 2\cdot 3}x^3 + \cdots. \qquad (12.1.5)$$

Note that the series has an infinite number of non-zero terms (it does not stop) if r is not a positive integer or 0.

For $x = 0$, the series clearly converges. So consider $x \neq 0$. The presence of x^k, which can be positive or negative, and of $k!$ in the denominator of the general term suggests using the absolute-ratio test. Let a_k be the term in the Maclaurin series for $(1+x)^r$ that contains the power x^k. Then

$$a_k = \frac{r(r-1)(r-2)\cdots(r-k+1)}{1\cdot 2\cdot 3\cdots k}x^k,$$

and

$$a_{k+1} = \frac{r(r-1)(r-2)\cdots(r-k+1)(r-k)}{1\cdot 2\cdot 3\cdots k(k+1)}x^{k+1}.$$

Thus

$$\left|\frac{a_{k+1}}{a_k}\right| = \left|\frac{\frac{r(r-1)(r-2)\cdots(r-k+1)(r-k)}{1\cdot 2\cdot 3\cdots k(k+1)}x^k}{\frac{r(r-1)(r-2)\cdots(r-k+1)}{1\cdot 2\cdot 3\cdots k}x^{k+1}}\right| = \left|\frac{r-k}{k+1}x\right|.$$

Since r is fixed,

$$\lim_{k\to\infty}\left|\frac{a_{k+1}}{a_k}\right| = |x|.$$

By the absolute-ratio test, series (12.1.5) converges when $|x| < 1$ and diverges when $|x| > 1$. ◇

In Example 5 it was shown that for $|x| < 1$ the Maclaurin series associated with $(1+x)^r$ converges to something, but does it converge to $(1+x)^r$?

Let us check the case $r = -1$. When $r = -1$, series (12.1.5) becomes

$$1 + (-1)x + \frac{(-1)(-2)}{1 \cdot 2}x^2 + \frac{(-1)(-2)(-3)}{1 \cdot 2 \cdot 3}x^3 + \cdots,$$

or

$$1 - x + x^2 - x^3 + \cdots.$$

This series is a geometric series with first term 1 and ratio $-x$. It therefore converges for $|x| < 1$. Moreover, it does represent the function $(1+x)^r = (1+x)^{-1}$. (See Exercises 34 to 37 in Section 12.4.)

It is true that for $|x| < 1$ series (12.1.5) does converge to $(1+x)^r$. The fact that $(1+x)^r$ is represented by the series (12.1.5) is known as the **general binomial theorem** or, simply, the **binomial theorem**. Series (12.1.5) is called the **binomial expansion** of $(1+x)^r$.

Summary

The Taylor series associated with a function is the series whose partial sums are its n^{th}-order Taylor polynomials. This series represents the original function only for inputs such that the remainder of the n^{th}-order Taylor polynomial approaches zero as $n \to \infty$: $\lim_{n\to\infty} R_n(x,a) = 0$. The Lagrange form of the remainder, Theorem 5.5.1 from Section 5.5, helps to show that the remainder converges to zero, though, as Example 3 illustrates, in some cases it may not be strong enough to do that.

Function	Series	Interval of Convergence
e^x	$\sum_{k=0}^{\infty} \frac{x^k}{k!}$	all x: $(-\infty, \infty)$
$\sin(x)$	$\sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!}$	all x: $(-\infty, \infty)$
$\cos(x)$	$\sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!}$	all x: $(-\infty, \infty)$
$(1+x)^r$	$\sum_{k=0}^{\infty} \frac{r(r-1)(r-2)\cdots(r-k+1)}{1\cdot 2\cdot 3\cdots k}x^k$	$\|x\| < 1$

Table 12.1.4

EXERCISES for Section 12.1

1. State without using any mathematical symbols the formula for the terms of a Taylor series of a function around a number that may not be zero and Lagrange's formula for the remainder.

2. State without using any mathematical symbols the formula for the terms of a Maclaurin series of a function and Lagrange's formula for the remainder.

In Exercises 3 to 9 compute the Maclaurin series associated with the given function

3. $1/(1+x)$

4. $1/(1-x)$

5. $\ln(1+x)$

6. $\ln(1-x)$

7. $\sin(x)$

8. e^{-x}

9. $\sqrt{1+x}$

10. Let $f(x) = e^x$. Show that $\lim_{n\to\infty} R_n(x; 0) = 0$ for any negative number x. This completes the proof that the exponential function is represented by its Maclaurin series for all numbers x (see Example 2).

11. Show that the Maclaurin series associated with $\sin(x)$ represents $\sin(x)$ for all x.

12. Show that the Maclaurin series associated with e^{-x} represents e^{-x} for all x.

13.

(a) Why will there be no terms of even degree in the Maclaurin series for $\arctan(x)$? (That is, all terms of the form x^{2k} have coefficient zero.)

(b) Obtain the first two non-zero terms of the Maclaurin series for $\arctan(x)$.

In Section 12.4 we use a shortcut to find the entire series.

14.

(a) Use the Lagrange formula to show that the Maclaurin series associated with $1/(1+x)$ represents $1/(1+x)$ for all $-1/2 < x < 1$. (Examine $R_n(x; 0)$.)

(b) Use the fact that it is a geometric series to show that the representation holds for $-1 < x < 1$.

15. Show that the Taylor series in powers of $x - a$ for e^x represents e^x for all x.

16. Show that the Taylor series in powers of $x - a$ for $\cos(x)$ represents $\cos(x)$ for all x.

17.

(a) Write out the first four terms of the binomial expansion of

$$(1 + x)^{-2} = 1/(1 + x)^2.$$

(b) What is the coefficient of the general term x^n?

18. Write out the first four terms of the binomial expansion of $(1+x)^{1/2} = \sqrt{1 + x}$.

19. What is the typical term in the Maclaurin series associated with $(1 - x)^r$? (Exploit the binomial expansion of $(1 + x)^r$; don't start from scratch.)

20. Suppose one uses the Maclaurin series for e^x to find e^{100}.

(a) What are the first four terms?

(b) Does the series converge to e^{100}?

(c) If your answer to (b) is "yes" how many terms would you use to estimate e^{100} with an error less than 0.005?

(d) Which terms in the series are largest?

21.

(a) Use the Maclaurin series for e^x to estimate $\sqrt[3]{e}$ to three decimal places.

(b) Compare your answer in (a) to the value of $\sqrt[3]{e}$ returned by your calculator.

22. Find the Maclaurin series associated with $1/(1 - x)^2$.

23. This problem examines three ways to estimate the error in using a front-end of $\sum_{k=0}^{\infty} \frac{(-1)^k}{k!}$ to estimate e^{-1}.

(a) Use the Lagrange formula to obtain an estimate of the error in using the front-end up through $(-1)^m/m!$ to estimate e^{-1}

(b) Estimate the error by noticing the series is alternating and the terms decrease in absolute value

(c) Estimate the error by comparing $\sum_{k=m+1}^{\infty}\left|\frac{(-1)^k}{k!}\right|$ to a geometric series, which is easy to sum.

(d) Which of the three methods provides the smallest (best) estimate of the error?

24.

(a) Use the Taylor series around $\pi/4$ to estimate $\cos(50°)$ to two decimal places. (That is, with an error less than 0.005.) Approximate π by 3.1416 and $\sqrt{2}$ by 1.4142.

(b) Check your calculation by calculating $\cos(50°)$ with your calculator.

25. Do there exist any polynomials $p(x)$ such that $\sin(x) = p(x)$ for all x in the interval $[1, 1.0001]$? Explain.

26. Do there exist any polynomials $p(x)$ such that $\ln(x) = p(x)$ for all x in the interval $[1, 1.0001]$? Explain.

27. Let f be a function that has derivatives of all orders for all x. Assume that $\left|f^{(n)}(x)\right| \leq n$ for all 100^n. Show why $f(x)$ is represented by its Maclaurin series for all x.

28.

(a) From the Maclaurin series for $\cos(x)$ obtain the Maclaurin series for $\sin^2(x)$. (Use a trigonometric identity.)

(b) From (a), and another trigonometric identity, obtain the Maclaurin series for $\cos^2(x)$.

29. In the CIE on Einstein's theory of relativity at the end of Chapter 11 (see pages 989—990) it is shown that the total increase in energy is $mc^2 - m_0c^2$. Use the first two terms of the binomial series for $\left(1-x^2\right)^{-1/2}$, with $x = v^2/c^2$, to derive (C.14.3). That is, show that

$$mc^2 - m_0c^2 \approx \frac{1}{2}m_0v^2.$$

Exercises 30 and 31 present a non-zero function whose Maclaurin series has the value 0 for all x, and therefore does not represent the function. This function is so "flat" at the origin that all its derivatives are zero there.

30. The following steps show that $\lim_{x\to 0} \dfrac{e^{1/x^2}}{x^n} = 0$ for all positive numbers n:

(a) Why does it suffice to consider only $x > 0$?

(b) Let $v = 1/x^2$ and translate the limit to

$$\lim_{v\to\infty} v^{n/2}e^{-v}.$$

(c) This limit is similar to a limit treated in Section 5.6. Show that it equals 0.

(d) Show that $\lim_{n\to\infty} \dfrac{p(x)e^{-1/x^2}}{x^n} = 0$ for any polynomial $p(x)$.

31. Let $f(x) = e^{-1/x^2}$ if $x \neq 0$ and $f(0) = 0$.

(a) Show f is continuous at 0.

(b) Show f is differentiable at 0.

(c) Show that $f'(0) = 0$.

(d) Show that $f''(0) = 0$.

(e) Explain why $f^{(n)}(0) = 0$ for all integers $n \geq 0$.

(f) What is the Maclaurin series associated with f?

(g) Why does the example use e^{-1/x^2} instead of the simpler $e^{-1/x}$?

32. Explain why it is not possible to use the Lagrange formula to show that the Taylor series in powers of $(x-1)$ associated with $1/x$ converges to $1/x$ for x in $(0, 1/2)$.

12.2 Two Applications of Taylor Series

If a Taylor series associated with a function $f(x)$ represents the function, then any front end (or Taylor polynomial) approximates $f(x)$. This can be used to evaluate some indeterminate limits and to estimate some definite integrals.

Using a Taylor Series to Find a Limit

The next example shows how series can be used to evaluate the limit of a quotient that is an indeterminate form.

EXAMPLE 1 Find $\lim_{x\to 0} \frac{\sin(x^2)}{\sqrt{1+3x^2}-1}$.

SOLUTION The Maclaurin series for $\sin(x^2)$ and $\sqrt{1+3x^2}$, we have

$$\lim_{x\to 0} \frac{\sin(x^2)}{\sqrt{1+3x^2}-1} = \lim_{x\to 0} \frac{x^2 - \frac{x^6}{6} + \cdots}{1+\frac{1}{2}\left(3x^2\right) - \frac{1}{8}\left(3x^2\right)^2 + \cdots - 1}$$

$$= \frac{x^2 - \frac{1}{6}x^6 + \cdots}{\frac{3}{2}x^2 - \frac{9}{8}x^6 + \cdots} = \frac{1}{\frac{3}{2}} = \frac{2}{3}.$$

◇

In Example 1 we needed only enough terms of each series to know the smallest power of x that appears in the numerator and in the denominator. The next example illustrates this.

EXAMPLE 2 Find $\lim_{x\to 0} \frac{\sqrt{1+x}-\sqrt{1+2x}}{\sqrt{1+2x}-\sqrt{1+4x}}$.
SOLUTION This limit could be bound by l'Hôpital's method. However, it is faster to use Taylor series.

For a number r, and $|x|<1$, the binomial theorem asserts that

$$(1+x)^r = 1 + rx + \cdots .$$

Thus the limit is

$$\begin{aligned}
&\lim_{x\to 0} \frac{\left(1+\frac{1}{2}x+\cdots\right)-\left(1+\frac{1}{2}(2x)+\cdots\right)}{\left(1+\frac{1}{2}(2x)+\cdots\right)-\left(1+\frac{1}{2}(4x)+\cdots\right)} \\
&= \lim_{x\to 0} \frac{\left(x\left(\frac{1}{2}+\cdots\right)\right)-\left(x\left(\frac{1}{2}(2)+\cdots\right)\right)}{\left(x\left(\frac{1}{2}(2)+\cdots\right)\right)-\left(x\left(\frac{1}{2}(4)+\cdots\right)\right)} \\
&= \lim_{x\to 0} \frac{\left(\frac{1}{2}+\cdots\right)-\left(\frac{2}{2}+\cdots\right)}{\left(\frac{1}{2}+\cdots\right)-\left(\frac{2}{4}+\cdots\right)} \\
&= \frac{\frac{1}{2}-\frac{2}{2}}{\frac{2}{2}-\frac{4}{2}} = \frac{1}{2}
\end{aligned}$$

◇

Using a Taylor Series to Estimate an Integral

The integral describes the "bell curve."

In statistics, the integral $\int_{-\infty}^{b}(1/\sqrt{2\pi})e^{-x^2/2}\,dx$ is of major importance. Since $e^{-x^2/2}$ does not have an elementary antiderivative, the integral must be estimated by other means. Tables of values of this function can be found in almost any mathematical handbook.

The next example shows how to estimate $\int_a^b f(x)\,dx$ when $f(x)$ is represented by a Taylor series.

EXAMPLE 3 Use the Maclaurin series for e^x to estimate $\int_0^1 e^{-x^2}\,dx$.
SOLUTION The first step is to obtain the Maclaurin series for the integrand: e^{-x^2}. Because

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

we can replace x with $-x^2$ to obtain

$$e^{-x^2} = 1 - x^2 + \frac{x^4}{2!} - \frac{x^6}{3!} + \cdots . \qquad (12.2.1)$$

For $0 \le |x| \le 1$, (12.2.1) is a convergent alternating series. Every partial sum that ends with a negative term is smaller than e^{-x^2}; every partial sum that ends with a positive term is larger than e^{-x^2}. For example,

$$1 - x^2 + \frac{x^4}{2!} - \frac{x^6}{3!} < e^{-x^2} < 1 - x^2 + \frac{x^4}{2!} - \frac{x^6}{3!} + \frac{x^8}{4!}.$$

Hence

$$\int_0^1 \left(1 - x^2 + \tfrac{x^4}{2!} - \tfrac{x^6}{3!}\right)\,dx \;<\; \int_0^1 e^{-x^2}\,dx \;<\; \int_0^1 \left(1 - x^2 + \tfrac{x^4}{2!} - \tfrac{x^6}{3!} + \tfrac{x^8}{4!}\right)\,dx,$$

or

$$1 - \tfrac{1}{3} + \tfrac{1}{5\cdot 2!} - \tfrac{1}{7\cdot 3!} \;<\; \int_0^1 e^{-x^2}\,dx \;<\; 1 - \tfrac{1}{3} + \tfrac{1}{5\cdot 2!} - \tfrac{1}{7\cdot 3!} + \tfrac{1}{9\cdot 4!}.$$

From this it follows that $0.742 < \int_0^1 e^{-x^2}\,dx < 0.748.$ ◇

Summary

The Taylor series associated with a function can be used to evaluate some indeterminate limits and to estimate definite integrals. In many cases there is no need to write the formula for all the terms, for usually only a few at the front end are needed.

EXERCISES for Section 12.2

In Exercises 1 to 4 use Taylor series to find the limits.

1. $\lim_{x\to 0} \frac{\sqrt{1+x}-1}{\sqrt{1+3x}-1}$.

2. $\lim_{x\to 0} \frac{\sin(4x)}{\sqrt{1+3x}-1}$.

3. $\lim_{x\to 0} \frac{e^{x^2}-1}{\sin(x^2)}$.

4. $\lim_{x\to 0} \frac{\cos(x)-1+\frac{x^2}{2}}{\sin(x^4)}$.

In Exercises 5 to 11 find the limits two ways. First use a Taylor series and then again using l'Hôpital's rule.

5. $\lim_{x\to 0} \frac{\cos(x)e^{2x^2}-1}{x\sin(x)}$.

6. $\lim_{x\to 0} \frac{\sqrt{1+3x}(e^x-1)x}{1-\cos(2x)}$.

7. $\lim_{x\to 0} \frac{\cos(x)-\sqrt{1+x}}{\cos(2x)-\sqrt[3]{1+2x}}$.

8. $\lim_{x\to 0} \frac{\ln(1+3x)}{\sin(2x)}$.

9. $\lim_{x\to 0} \frac{e^{x^2}-1}{e^{3x^2}-1}$.

10. $\lim_{x\to 0} \frac{\left(\sin(x^2)+e^{x^3}-1\right)\sqrt[3]{5+x}}{\sqrt{1+5x^2}-1}$.

11. $\lim_{x\to 4} \frac{(8-2x)e^{x^2}}{\sqrt[3]{4-x}}$. First write $4-x$ as $4(1-x/4)$ and factor 4 out of the radical. See Exercises Exercise 34 to 37 for more on the binomial theorem for $(a+b)^r$.

12.

(a) Write out the first four terms of the binomial series for $(1+x)^{-2}$

(b) What is the general form?

13.

(a) Find the limit in Example 2 by l'Hôpital's rule.

(b) Find the limit in Example 1 by l'Hôpital's rule.

14.

(a) Show $\int_0^1 (e^x - 1)/x \, dx$ is finite, even though the integrand is not defined at 0.

(b) Show that $1 + \frac{1}{2 \cdot 2!} + \frac{1}{3 \cdot 3!} + \frac{1}{4 \cdot 4!} + \frac{1}{5 \cdot 5!}$ is an estimate of the integral in (a).

(c) The error in using the sum in (b) is $\frac{1}{6 \cdot 6!} + \frac{1}{7 \cdot 7!} + \frac{1}{8 \cdot 8!} + \frac{1}{9 \cdot 9!} + \cdots$. Show that this is less than $\frac{1}{6 \cdot 6!}\left(1 + \frac{1}{7}\left(\frac{1}{7}\right) + \frac{1}{7}\left(\frac{1}{7}\right)^2 + \frac{1}{7}\left(\frac{1}{7}\right)^3 + \cdots\right)$.

(d) From (c) deduce that the error is less than 0.000237.

15.

(a) Show that for x in $[0, 2]$

$$x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} \leq e^x - 1 \leq x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \frac{e^2 x^{n+1}}{(n+1)!}.$$

(b) Use (a) to find $\int_0^2 \frac{e^x - 1}{x} \, dx$ to three decimal places.

16. Find $\int_0^1 \frac{1 - \cos(x)}{x} \, dx$ to three decimal places, using an approach like that in Exercise 15.

17. Estimate $\int_0^\infty e^{-5x^2} \, dx$ following these steps:

(a) Find a number b such that

$$\int_b^\infty e^{-5x^2} \, dx < 0.0005.$$

(Use the fact that $e^{-5x^2} < e^{-5x}$ for $x > 1$.)

(b) Let b be the number you found in (a). Estimate $\int_0^b e^{-5x^2} \, dx$ with an error of less than 0.0005. (Use the Maclaurin series for e^{-5x^2}.)

(c) Combine (a) and (b) to get a two decimal place estimate of $\int_0^\infty e^{-5x^2} \, dx$.

18. Estimate $\int_0^\infty \frac{\cos(x^6/100)-1}{x^6}\,dx$, following these steps:

(a) Find a number b such that

$$\left|\int_b^\infty \frac{\cos(x^6/100)-1}{x^6}\,dx\right| < 0.001.$$

(Use the fact that $|\cos(x)| \leq 1$.)

(b) Let b be the number you found in (a). Estimate

$$\int_0^b \frac{\cos(x^6/100)-1}{x^6}\,dx,$$

with an error less that 0.001. (Use the Maclaurin series for $\cos(x)$.)

(c) Combine (a) and (b) to get a two decimal place estimate for

$$\int_0^\infty \frac{\cos(x^6/100)-1}{x^6}\,dx.$$

19. Evaluate $\int_0^1 \frac{dx}{1+x^2}$ by

(a) the Fundamental Theorem of Calculus (approximate π to 3 decimal places),

(b) Simpson's method (six sections),

(c) trapezoid method (six sections),

(d) using the first six non-zero terms of the series $1 - x^2 + x^4 - \cdots$ for $1/(1+x^2)$.

20. If $|a/b| < 1$, use the binomial theorem to expand $(a+b)^r$ as the sum of terms of the form ca^pb^q.

21. If $|b/a| < 1$, use the fundamental theorem to expand $(a+b)^r$ as the sum of terms of the form ca^pb^q. (As a check, the series starts with b^r.)

22. Write out the first four (4) terms of the series for $(8+x)^{1/3}$ if (a) $x > 8$, and (b) $x < 8$. See Exercises 20 and 21.)

23.

Sam: I was playing with the binomial theorem.

Jane: Is that possible?

Sam: I looked at $(3+5)^{1/3}$, which I know is two. But I can write it as $5^{1/3}\left(1+\frac{3}{5}\right)^{1/3}$ and get

$$5^{1/3}\left(1+\frac{1}{3}\frac{3}{5}+\frac{1}{2!}\frac{1}{3}\frac{-2}{3}\left(\frac{3}{5}\right)^2+\cdots\right)$$

so

$$2=5^{1/3}+\frac{1}{3}5^{-2/3}(3)-\frac{1}{9}5^{-5/3}3^2+\cdots.$$

Jane: That's a fancy way to estimate 2.

Sam: But I can write $(3+5)^{1/3}$ as $3^{1/3}\left(1+\frac{5}{3}\right)^{1/3}$ and get

$$2=3^{1/3}+\frac{1}{3}3^{-2/3}(5)-\frac{1}{9}3^{-5/3}5^2+\cdots$$

Jane: Another nutty way to estimate 2.

Sam: My point is that they can't both be right.

Can they both be right?

24. Repeat Exercise 19 for $\int_0^1 \frac{dx}{1+x^3}$.

25. In R. P. Feynman, *Lectures on Physics*, Addison-Wesley, Reading, MA 1963, this statement appears in Section 15.8 of Volume 1:

> An approximate formula to express the increase of mass, for the case when the velocity is small, can be found by expanding $m_0/\sqrt{1-v^2/c^2}=m_0(1-v^2/c^2)^{-1/2}$ in a power series, using the binomial theorem. We get
>
> $$m_0\left(1-\frac{v^2}{c^2}\right)^{-1/2}=m_0\left(1+\frac{1}{2}\frac{v^2}{c^2}+\frac{3}{8}\frac{v^4}{c^4}+\cdots\right).$$
>
> We see clearly from the formula that the series converges rapidly when v is small and the terms after the first two or three are negligible.

Check the expansion and justify the equation.

26. A fluid mechanics text has the following argument in a discussion of flow through a nozzle:

The pressure p equals

$$\left(1+\frac{\gamma-1}{2}M^2\right)^{\gamma/(1-\gamma)}.$$

By the binomial theorem and the fact that $v^2 = M^2\gamma RT$:

$$p = 1 - \frac{1}{2}\frac{v^2}{RT} + \frac{\gamma(2\gamma-1)}{8}M^4 + \cdots.$$

Fill in the steps. γ is the specific heat, which is about 1.4, and M is a Mach number, which is in the range 1 to 2.

27.

(a) The ellipse $x^2/a^2 + y^2/b^2 = 1$ for $a \le b$ has the parameterization

$$x = a\cos(t), \qquad y = b\sin(t).$$

Show that the arc length of one quadrant of an ellipse is

$$b\int_0^{\pi/2} \sqrt{1-\left(1-\left(\frac{a}{b}\right)^2\right)\sin(t)^2}\; dt.$$

The integrand does not have an elementary antiderivative.

(b) If $a < b$, the integral in (a) has the form $b\int_0^{\pi/2}\sqrt{1-k^2\sin(t)^2}\; dt$, where $0 < k < 1$. The "elliptic integral"

$$E = b\int_0^{\pi/2}\sqrt{1-k^2\sin(t)^2}\; dt$$

is tabulated in mathematical handbooks for many values of k in $[0, 1]$. Using the binomial theorem and the formula for $\int_0^{\pi/2}\sin^n(\theta)\; d\theta$ (Formula 74 in the table of integrals), obtain the first four non-zero terms of E as a series in powers of k^2.

12.3 Power Series and Their Interval of Convergence

Our use of Taylor polynomials to approximate a function led us to consider series of the form

$$\sum_{k=0}^{\infty} b_k(x-a)^k = b_0 + b_1(x-a) + b_2(x-a)^2 + \cdots + b_k(x-a)^k + \cdots .$$

Such a series is called a **power series** in $x - a$. If $a = 0$, we obtain a series in powers of x:

$$\sum_{k=0}^{\infty} b_k x^k = b_0 + b_1 x + b_2 x^2 + \cdots + b_k x^k + \cdots .$$

We will now look at some properties of power series and see that they behave very much like polynomials.

The Radius of Convergence of a Power Series

The power series $b_0 + b_1x + b_2x^2 + \cdots$ certainly converges when $x = 0$. It may or may not converge for other choices of x. However, as Theorem 12.1.3 will show, if the series converges at a certain value c, it converges at any number x whose absolute value is less than $|c|$, that is, throughout the interval $(-|c|, |c|)$. Since the proof of Theorem 12.3.1 uses the comparison test and the absolute-convergence test, it offers a nice review of important concepts from Chapter 11.

For each fixed choice of x, a power series becomes a series with constant terms.

Theorem 12.3.1. *Let c be a nonzero number such that $\sum_{k=0}^{\infty} b_k c^k$ converges. Then, if $|x| < |c|$, $\sum_{k=0}^{\infty} b_k x^k$ converges. In fact, it converges absolutely.*

The proof is at the end of this section.

By Theorem 12.3.1, the set of numbers x such that $\sum_{k=0}^{\infty} b_k x^k$ converges has no holes. In other words, it is one unbroken piece, which includes the number 0. Moreover, if r is in the set, so is the entire open interval $(-|r|, |r|)$.

The x's for which the series converges form an interval with 0 at its midpoint.

There are two possibilities. In the first case, there are arbitrarily large r's such that the series converges for x in $(-r, r)$. This means that the series converges for all x. In the second case, there is an upper bound on the numbers r such that the series converges for x in $(-r, r)$. It is shown in advanced calculus that there is then a smallest upper bound on the r's; call it R.

See Figure 12.3.1.

Note that convergence or divergence at R and $-R$ is not mentioned.

Consequently, either

1. $b_0 + b_1 x + b_2 x^2 + \cdots$ converges for all x

or

2. there is a positive number R such that $b_0 + b_1 x + b_2 x^2 + \cdots$ converges for all x such that $|x| < R$ but diverges for $|x| > R$.

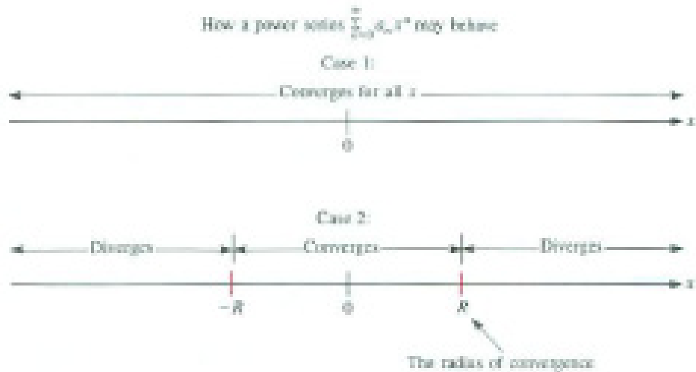


Figure 12.3.1

In the second case, R is called the **radius of convergence** of the series. In the first case, the radius of convergence is said to be infinite, $R = \infty$. For the geometric series $1 + x + x^2 + \cdots + x^k + \cdots$, $R = 1$, since the series converges when $|x| < 1$ and diverges when $|x| > 1$. (It also diverges when $x = 1$ and $x = -1$.) A power series with radius of convergence R may or may not converge at R and at $-R$. These observations are summarized as Theorem 12.3.2.

Theorem 12.3.2. *Radius of Convergence Let R be the radius of convergence for the power series* $\sum_{k=0}^{\infty} b_k x^k$. *If* $R = 0$, *the series converges only for* $x = 0$. *If* R *is a positive number, the series converges for* $|x| < R$ *and diverges for* $|x| > R$. *If* R *is* ∞, *the series converges for all* x.

EXAMPLE 1 Find the radius of convergence, R, for $x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + \frac{(-1)^{k+1} x^k}{k} + \cdots$.

SOLUTION Because of the presence of x^k and the fact that x may be negative, use the absolute-ratio test. The absolute value of the ratio of successive terms is

$$\left| \frac{\frac{(-1)^{k+2} x^{k+1}}{k+1}}{\frac{(-1)^{k+1} x^k}{k}} \right| = \frac{k}{k+1} |x|.$$

As $k \to \infty$, $k/(k+1) \to 1$. Thus,

$$\lim_{k\to\infty} \frac{k}{k+1}|x| \quad = \quad |x|.$$

Consequently, by the absolute-ratio test, if $|x| < 1$ the series converges. If $|x| > 1$, it diverges.

The radius of convergence is $R = 1$. It remains to see what happens at the endpoints, 1 and -1.

The absolute-ratio test takes care of $|x| < 1$ and $|x| > 1$.

For $x = 1$, we obtain the alternating harmonic series:

Checking convergence at $x = 1$

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots .$$

This series converges, by the alternating-series test.

What about $x = -1$? The series becomes

Checking convergence at $x = -1$

$$(-1) - \frac{(-1)^2}{2} + \frac{(-1)^3}{3} - \frac{(-1)^4}{4} + \cdots + \frac{(-1)^{k+1}(-1)^k}{k} + \cdots$$

or

$$-1 - \frac{1}{2} - \frac{1}{3} - \frac{1}{4} - \cdots - \frac{1}{k} + \cdots ,$$

which, being the negative of the harmonic series, diverges.

All told, this series converges only for $-1 < x \le 1$. Figure 12.3.2 records what we found.

Figure 12.3.2

◇

Earlier we saw that $\sum_{k=0}^{\infty} x^n/n!$ has radius of convergence $R = \infty$. The next example represents the opposite extreme, $R = 0$.

EXAMPLE 2 Find the radius of convergence of the series

$$\sum_{k=1}^{\infty} k^k x^k = 1x + 2^2x^2 + 3^3x^3 + \cdots + k^k x^k + \cdots .$$

Every power series converges for at least one value of x.

SOLUTION The series converges for $x = 0$.

If $x \neq 0$, consider the k^{th} term $k^k x^k$, which can be written as $(kx)^k$. As $k \to \infty$, $|kx| \to \infty$. By the n^{th} term test, this series diverges. In short, the series converges only when $x = 0$. The radius of convergence is $R = 0$. $\diamond$

The Radius of Convergence of $\sum_{k=0}^{\infty} b_k(x-a)^k$

Just as a power series in x has an associated radius of convergence, so does a power series in $x - a$. To see this, consider any such power series,

$$\sum_{k=0}^{\infty} b_k(x-a)^k = b_0 + b_1(x-a) + b_2(x-a)^2 + \cdots . \tag{12.3.1}$$

Let $u = x - a$. Then series (12.3.1) becomes

$$\sum_{k=0}^{\infty} b_k u^k = b_0 + b_1 u + b_2 u^2 + \cdots . \tag{12.3.2}$$

R may be zero, positive, or infinite.

Series (12.3.2) has a certain radius of convergence R. That is, (12.3.2) converges for $|u| < R$ and diverges for $|u| > R$. Consequently (12.3.1) converges for $|x-a| < R$ and diverges for $|x-a| > R$. The number R is called the radius of convergence of the series (12.3.1).

Figure 12.3.3

As Figure 12.3.3 suggests, the series $\sum_{k=0}^{\infty} b_k(x-a)^k$ converges in an interval $(a-R, a+R)$, whose midpoint is a. The question marks in Figure 12.3.3 indicate that the series may converge or may diverge at the ends of the interval, $a - R$ and $a + R$. These cases must be decided separately.

These observations are summarized in the following theorem.

Theorem 12.3.3. *Let R be the radius of convergence for the power series $\sum_{k=0}^{\infty} b_k(x-a)^k$. If $R = 0$, the series converges only for $x = a$. If R is a positive real number, the series converges for $|x-a| < R$ and diverges for $|x-a| > R$. If $R = \infty$, the series converges for all numbers x.*

EXAMPLE 3 Find all values of x for which

$$\sum_{k=1}^{\infty}(-1)^{k-1}\frac{(x-1)^k}{k} = (x-1)-\frac{(x-1)^2}{2}+\frac{(x-1)^3}{3}-\frac{(x-1)^4}{4}+\cdots \quad (12.3.3)$$

converges.

SOLUTION Note that this is Example 1 with x replaced by $x-1$. Thus $x-1$ plays the role that x played in Example 1. Consequently, series (12.3.3) converges for $-1 < x-1 \le 1$, that is, for $0 < x \le 2$, and diverges for all other values of x. Its radius of convergence is $R = 1$. The set of values where the series converges is an interval $(0,2]$.

Figure 12.3.4

The convergence of (12.3.3) is recorded in Figure 12.3.4. ◇

Proof of Theorem 12.3.1

Proof (of Theorem 12.3.1)

Since $\sum_{k=0}^{\infty} b_k c^k$ converges, the k^{th} term $a_k c^k$ approaches 0 as $k \to \infty$. Thus there is an integer N such that for $k \ge N$, $|b_k c^k| \le 1$, say. From here on, consider only $k \ge N$. Now,

$$b_k x^k = b_k c^k \left(\frac{x}{c}\right)^k.$$

Since $$\left|b_k x^k\right| = \left|b_k c^k\right| \left|\frac{x}{c}\right|^k,$$

it follows that for $k \geq N$,

$$\left|b_k x^k\right| \leq \left|\frac{x}{c}\right|^k \qquad (\text{since } |b_k c^k| \leq 1 \text{ for } k \geq N).$$

The series $\sum_{k=0}^{\infty} \left|\frac{x}{c}\right|^k$ is a geometric series with the ratio $|x/c| < 1$. Hence it converges.

Since $|b_k x^k| \leq \left|\frac{x}{c}\right|^k$ for $k \geq N$, the series $\sum_{k=0}^{\infty} \left|b_k x^k\right|$ converges (by the comparison test). Thus $\sum_{k=N}^{\infty} b_k x^k$ converges (in fact, absolutely). Putting in the front end, $\sum_{k=0}^{N-1} b_k x^k$, we conclude that the series $\sum_{k=0}^{\infty} b_k x^k$ converges absolutely if $|x| < |c|$. •

You may wonder why it's called "radius of convergence," when no circles seem to be involved. Sections 12.5 and 12.6, which use complex numbers, explain why.

Summary

Motivated by Taylor series, we investigated series of the form $\sum_{k=0}^{\infty} b_k x^k$ and, more generally, $\sum_{k=0}^{\infty} b_k (x-a)^k$. Associated with each such series is a radius of convergence R. (If the series converges for all x, we take R to be infinite.) If $\sum_{k=0}^{\infty} b_k x^k$ has radius of convergence R, then it converges (absolutely) for all x in $(-R, R)$, but diverges for all x such that $|x| > R$. Similarly, if $\sum_{k=0}^{\infty} b_k (x-a)^k$ has radius of convergence R, it converges for all x such that x is in $(a-R, a+R)$ but diverges for $|x-a| > R$. Convergence or divergence at the endpoints of the interval of convergence must be checked separately.

EXERCISES for Section 12.3

In Exercises 1 to 12 draw the appropriate diagrams (like Figure 12.3.4) showing where the series converge or diverge. Explain your work.

1. $\sum_{k=1}^{\infty} \frac{x^k}{k^2}$

2. $\sum_{k=1}^{\infty} \frac{x^k}{\sqrt{k}}$

3. $\sum_{k=0}^{\infty} \frac{x^k}{3^k}$

4. $\sum_{k=1}^{\infty} k^2 e^{-k} x^k$

5. $\sum_{k=0}^{\infty} \frac{2k^2+1}{k^2-5} x^k$

6. $\sum_{k=1}^{\infty} \frac{x^k}{k}$

7. $\sum_{k=0}^{\infty} \frac{x^k}{(2k)!}$

8. $\sum_{k=0}^{\infty} \frac{2^k x^k}{k!}$

9. $\sum_{k=0}^{\infty} \frac{x^k}{(2k+1)!}$

10. $\sum_{k=0}^{\infty} k! x^k$

11. $\sum_{k=1}^{\infty} \frac{(-1)^{k+1} x^k}{k}$

12. $\sum_{k=1}^{\infty} \frac{2^k x^k}{n}$

13. Assume that $\sum_{k=0}^{\infty} b_k x^k$ converges for $x = 9$ and diverges when $x = -12$. What, if anything, can be said about

(a) convergence when $x = 7$?

(b) absolute convergence when $x = -7$?

(c) absolute convergence when $x = 9$?

(d) convergence when $x = -9$?

(e) divergence when $x = 10$?

(f) divergence when $x = -15$?

(g) divergence when $x = 15$?

14. Assume that $\sum_{k=0}^{\infty} b_k x^k$ converges for $x = -5$ and diverges when $x = 8$. What, if anything, can be said about

(a) convergence when $x = 4$?

(b) absolute convergence when $x = 4$?

(c) convergence when $x = 7$?

(d) absolute convergence when $x = -5$?

(e) convergence when $x = -9$?

(f) convergence when $x = -9$?

15. If $\sum_{k=0}^{\infty} b_k x^k$ converges whenever x is positive, must it converge whenever x is negative?

16. If $\sum_{k=0}^{\infty} b_k 6^k$ converges, what can be said abou the convergence of

(a) $\sum_{k=0}^{\infty} b_k (-6)^k$?

(b) $\sum_{k=0}^{\infty} b_k 5^k$?

(c) $\sum_{k=0}^{\infty} b_k (-5)^k$?

In Exercises 17 to 28 draw the appropriate diagrams showing where the series converge and diverge.

17. $\sum_{k=0}^{\infty} \frac{(x-2)^k}{k!}$

18. $\sum_{k=0}^{\infty} \frac{(x-1)^k}{k3^k}$

19. $\sum_{k=0}^{\infty} \frac{(x-1)^k}{k+3}$

20. $\sum_{k=0}^{\infty} \frac{(x-4)^k}{2k+1}$

21. $\sum_{k=0}^{\infty} \frac{k(x-2)^k}{2k+3}$

22. $\sum_{k=0}^{\infty} \frac{(x-5)^k}{k\ln(k)}$

23. $\sum_{k=0}^{\infty} \frac{(x+3)^k}{5^k}$

24. $\sum_{k=0}^{\infty} k(x+1)^k$

25. $\sum_{k=0}^{\infty} \frac{(x-5)^k}{k^2}$

26. $\sum_{k=0}^{\infty} (-1)^k \frac{(x+4)^k}{k+2}$

27. $\sum_{k=0}^{\infty} k!(x-1)^k$

28. $\sum_{k=0}^{\infty} \frac{k^2+1}{k^3+1}(x+2)^k$

In Exercises 29 to 34 write out the first five (non-zero) terms of the binomial expansion of the given functions.

29. $(1+x)^{1/2}$

30. $(1+x)^{1/3}$

31. $(1+x)^{3/2}$

32. $(1+x)^{-2}$

33. $(1+x)^{-3}$

34. $(1+x)^{-4}$

35.

(a) If a power series $\sum_{k=0}^{\infty} b_k x^k$ diverges when $x = 3$, at which x must it diverge?

(b) If a power series $\sum_{k=0}^{\infty} b_k (x+5)^k$ diverges when $x = -3$, at which x must it diverge?

36. If $\sum_{k=0}^{\infty} b_k (x-3)^k$ converges for $x = 7$, at what other values of x must the series necessarily converge?

37. Find the radius of convergence of $\sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!}$.

38. If $\sum_{k=0}^{\infty} b_k x^k$ has a radius of convergence 3 and $\sum_{k=0}^{\infty} c_k x^k$ has a radius of convergence 5, what can be said about the radius of convergence of $\sum_{k=0}^{\infty}(b_k + c_k)x^k$?

39.

(a) Use the first four nonzero terms of the Maclaurin series for $\sqrt{1+x^3}$ to estimate $\int_0^1 \sqrt{1+x^3}\,dx$. (This integral cannot be evaluated by the Fundamental Theorem of Calculus.)

(b) Evaluate the integral in (a) to three decimal places by Simpson's method.

40.

(a) Write the first four terms of the Maclaurin series associated with $f(x) = (1+x)^{-3}$.

(b) Find a formula for the general term in the Maclaurin series associated with $f(x)$.

(c) Replace x by $-x$ in your answer to (b) to obtain the first four nonzero terms in the Maclaurin series for $(1-x)^{-3}$.

41. Find the radius of convergence for the Maclaurin series associated with

(a) e^x

(b) $\sin(x)$

(c) $\cos(x)$

(d) $\ln(1+x)$

(e) $\arctan(x)$

(f) $(1+x)^{1/3}$

(g) $(1+2x)^{3/5}$

12.4 Manipulating Power Series

Where they converge, power series behave like polynomials. We can differentiate or integrate them term-by-term. We can add, subtract, multiply, and divide them. While most of the discussion will be on power series in x, the same ideas apply to power series in $(x-a)$. Proofs can be found in any advanced calculus text.

Differentiating a Power Series

In Section 3.3 we showed that you can differentiate the sum of a finite number of functions by adding their derivatives. Theorem 12.4.1 generalizes this to power series in x.

See the Sum and Difference Rules in Section 3.3

Theorem 12.4.1 (Differentiating a power series)**.** *Assume $R > 0$ and that $\sum_{k=0}^{\infty} b_k x^k$ converges to $f(x)$ for $|x| < R$. Then for $|x| < R$, f is differentiable, $\sum_{k=1}^{\infty} k b_k x^{k-1}$ converges to $f'(x)$, and*

$$f'(x) = b_1 + 2b_2x^2 + 3b_3x^3 + \cdots .$$

This theorem is *not* covered by the fact that the derivative of the sum of a *finite* number of functions is the sum of their derivatives.

Because f is differentiable it is continuous. Thus the limit as x approaches 0 of $\displaystyle\sum_{k=0}^{\infty} b_k x^k$ is b_0, the value of the series when $x=0$. This property was used without justification in Example 2 in Section 12.2.

EXAMPLE 1 Obtain a power series for the function $1/(1-x)^2$ from the power series for $1/(1-x)$.

SOLUTION From the formula for the sum of a geometric series, we know that

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots \qquad \text{for } |x| < 1.$$

According to Theorem 12.4.1, differentiating both sides of this equation produces a valid equation, namely

$$\frac{1}{(1-x)^2} = 0 + 1 + 2x + 3x^2 + \cdots \qquad \text{for } |x| < 1.$$

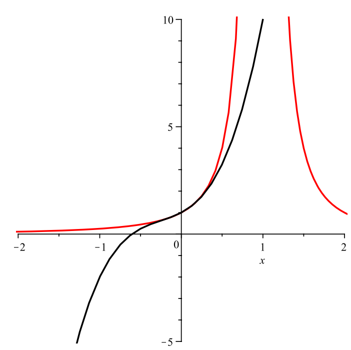


Figure 12.4.1

This can be expressed in summation notation. The geometric series is $\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$. When we differentiate both sides of this equation, we obtain $\frac{1}{(1-x)^2} = \sum_{k=1}^{\infty} k x^{k-1}$. (See Figure 12.4.1.)

Note that the Maclaurin series for $1/(1-x)^2$ can also be written as $\sum_{k=0}^{\infty}(k+1)x^k$.

Theorem 12.4.1 does not say anything about convergence of the Maclaurin series for $1/(1-x)^2$ at the endpoints of the interval of convergence. When

$x = 1$ the series is $\sum_{k=1}^{\infty} k$, which diverges (because the terms of this series do not approach 0). This is not surprising, because the derivative (and, in fact, the original function) are not defined when $x = 1$. When $x = -1$, $\frac{1}{(1-x)^2} = \frac{1}{4}$, so the derivative of the function is well-defined. But, when the series for the derivative is evaluated at $x = -1$ we get the series $\sum_{k=0}^{\infty}(-1)^{k-1}k$. As when $x = 1$, the terms of this series do not converge to zero and the series diverges. $\diamond$

Suppose that $f(x)$ has a power-series representation $b_0 + b_1x + b_2x^2 + \cdots$; Theorem 12.4.2 enables us to find the coefficients b_0, b_1, b_2, $\ldots$.

Theorem 12.4.2 (Formula for b_k)**.** *Let R be a positive number and suppose that $f(x)$ is represented by the power series $\sum_{k=0}^{\infty} b_kx^k$ for $|x| < R$; that is,*

$$f(x) = b_0 + b_1x + b_2x^2 + \cdots + b_kx^k + \cdots \qquad \text{for } |x| < R.$$

Then

$$b_k = \frac{f^{(k)}(0)}{k!}. \tag{12.4.1}$$

The proof is practically the same as the derivation of the formulas for the coefficients of Taylor polynomials in Section 5.5. It consists of repeated differentiation and evaluation of the higher derivatives at 0.

Theorem 12.4.2 also tells us that there can be at most one series of the form $\sum_{k=0}^{\infty} b_kx^k$ that represents $f(x)$, for the coefficients b_k are completely determined by $f(x)$ and its derivatives. That series must be the Maclaurin series we obtained in Section 12.1. For instance, the series $1+x+x^2+x^3+\cdots$, which sums to $1/(1-x)$ for $|x| < 1$ must be the Maclaurin series for $1/(1-x)$.

Integrating a Power Series

Just as we may differentiate a power series term by term, we can integrate it term by term.

Theorem 12.4.3. *(Integrating a power series) Assume that $R > 0$ and*

$$f(x) = b_0 + b_1x + b_2x^2 + \cdots + b_kx^k + \cdots \qquad \text{for } |x| < R.$$

Then

$$b_0x + b_1\frac{x^2}{2} + b_2\frac{x^3}{3} + \cdots + b_k\frac{x^{k+1}}{k+1} + \cdots$$

converges for $|x| < R$, and

$$\int_0^x f(t)\ dt = b_0x + b_1\frac{x^2}{2} + b_2\frac{x^3}{3} + \cdots + b_k\frac{x^{k+1}}{k+1} + \cdots.$$

> **WARNING** *(Choosing Variables of Integration)* Note that t is used as the variable of integration. This is done to avoid writing $\int_0^x f(x)\,dx$, an expression in which x describes both the interval $[0,x]$ and the independent variable of the integrand.

The next example shows the power of Theorem 12.4.3.

EXAMPLE 2 Integrate the power series for $1/(1+x)$ to obtain the power series in x for $\ln(1+x)$.

SOLUTION Start with the geometric series $1/(1-x) = 1 + x + x^2 + \cdots$ for $|x| < 1$. Replace x by $-x$ and obtain

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + x^4 - \cdots \qquad \text{for } |x| < 1.$$

By Theorem 12.4.3, $\int_0^x \frac{dt}{1+t} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{5} - \cdots$ for $|x| < 1$.

Now,

$$\begin{aligned} \int_0^x \frac{dt}{1+t} &= \ln(1+t)\big|_0^x \\ &= \ln(1+x) - \ln(1+0) \\ &= \ln(1+x). \end{aligned}$$

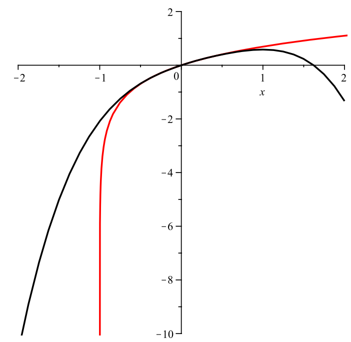


Figure 12.4.2

Thus

Power series for $\ln(1+x)$

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{5} - \cdots \text{ for } |x| < 1.$$

◇

The power series for $\ln(1+x)$ can also be found using Theorem 12.4.2 on page 1024 but this requires calculating the derivatives of $\ln(1+x)$ and evaluating them at $x = 0$.

The derivation in Example 2 is more straightforward, and it gives the radius of convergence without additional work.

The Algebra of Power Series

In addition to differentiating and integrating power series, we may also add, subtract, multiply, and divide them just like polynomials, as Theorem 12.4.4 asserts.

Theorem 12.4.4. *The algebra of power series. Assume that*

$$f(x) = \sum_{k=0}^{\infty} b_k x^k = b_0 + b_1 x + b_2 x^2 + \cdots \quad \textit{for } |x| < R_1$$

and

$$g(x) = \sum_{k=0}^{\infty} c_k x^k = c_0 + c_1 x + c_2 x^2 + \cdots \quad \textit{for } |x| < R_2.$$

Let R be the smaller of R_1 and R_2. Then, for $|x| < R$,

$$\begin{aligned}
f(x)+g(x) &= \sum_{k=0}^{\infty}(b_k+c_k)x^k = (b_0+c_0)+(b_1+c_1)x+(b_2+c_2)x^2+\cdots \\
f(x)-g(x) &= \sum_{k=0}^{\infty}(b_k-c_k)x^k = (b_0-c_0)+(b_1-c_1)x+(b_2-c_2)x^2+\cdots \\
f(x)g(x) &= (b_0c_0)+(b_0c_1+b_1c_0)x+(b_0c_2+b_1c_1+b_2c_0)x^2+\cdots \\
f(x)/g(x) &\ \textit{is}\ \ \textit{obtainable by long division, provided } g(x)\neq 0 \textit{ for all } |x|<R.
\end{aligned}$$

This says "multiply two power series the way you multiply polynomials — term by term: start with the constant terms and work up."

EXAMPLE 3 Find the first four terms of the Maclaurin series for $e^x/(1-x)$.

SOLUTION There are at least three ways to approach this problem. The direct approach is to use Theorem 12.4.2; this requires finding the first three derivatives of $e^x/(1-x)$ evaluated at $x=0$. A second idea is to divide the power series for e^x by $1-x$. The third idea is to multiply the power series for e^x and the power series for $1/(1-x)$.

See Exercise 6

As multiplication is generally easier to carry out than division, that is the option we choose. The power series for e^x is $1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots$ (radius of convergence is ∞) and the power series for $1/(1-x)$ is $1+x+x^2+x^3+\cdots$ (radius of convergence is 1):

$$\begin{aligned}
e^x\frac{1}{1-x} &= \left(1+x+\frac{x^2}{2!}+\frac{x^3}{3}+\cdots\right)\left(1+x+x^2+x^3+\cdots\right) \\
&= (1\cdot 1)+(1\cdot 1+1\cdot 1)x+\left(1\cdot 1+1\cdot 1+\frac{1}{2!}\cdots\right)x^2 \\
&\quad +\left(1\cdot 1+1\cdot 1+\frac{1}{2!}\cdots+\frac{1}{3!}\cdot 1\right)x^3+\cdots \\
&= 1+2x+\frac{5}{2}x^2+\frac{8}{3}x^3+\cdots.
\end{aligned}$$

According to Theorem 12.4.2, the power series for $e^x/(1-x)$, whose first four terms we just found, has radius of convergence $R=1$. ◇

EXAMPLE 4 Find the first four nonzero terms of the Maclaurin series associated with $e^x/\cos(x)$.

SOLUTION We attack this problem with Theorem 12.4.4. The Maclaurin series associated with $e^x/\cos(x)$ is the quotient of the Maclaurin series associated with e^x and $\cos(x)$. Long division shows us that

$$\frac{e^x}{\cos(x)} = 1+x+x^2+\frac{2x^3}{3}+\cdots$$

Even though the power series for e^x and $\cos(x)$ both have infinite radius of convergence, the fact that $\cos(\pi/2) = 0$ reduces the radius of convergence to $\pi/2$.

What happens when $|x| = \pi/2$?

We could have found the front-end of the Maclaurin series using Theorem 12.4.2, but this approach does not give any information about the radius of convergence of this power series. ◇

Power Series Around a

Power series in $x - a$

The various theorems and methods of this section were stated for power series in $x = x - 0$. Analogous theorems hold for power series in $x - a$. Such series may be differentiated and integrated term by term inside the interval in which they converge. For instance, Theorem 12.4.2 generalizes:

Theorem 12.4.5 (Formula for b_k)**.** *Let R be a positive number and suppose that $f(x)$ is represented by the power series $\sum_{k=0}^{\infty} b_k(x-a)^k$ for $|x-a| < R$; that is,*

$$f(x) = b_0 + b_1(x-a) + b_2(x-a)^2 + \cdots + b_k(x-a)^k + \cdots \qquad \text{for } |x-a| < R.$$

Then

$$b_k = \frac{f^{(k)}(a)}{k!}.$$

The proof of Theorem 12.4.5 is similar to that of Theorem 12.4.2.

Endpoints

Each theorem in this section includes information on the radius of convergence of a power series obtained from another power series. Convergence at the endpoints is never mentioned; it must be checked separately in every case.

In Example 1 we found the power series in x for $1/(1-x)^2$ is

$$1 + 2x + 3x^2 + \cdots = \sum_{k=1}^{\infty} kx^{k-1} \tag{12.4.2}$$

for $|x| < 1$. When $x = 1$ this series becomes $\sum_{k=1}^{\infty} k$, and, when $x = -1$ it is $\sum_{k=1}^{\infty} k(-1)^{k-1}$. Each of these series diverges because its terms do not approach 0 as $k \to \infty$. Thus, (12.4.2) converges only on the open interval $(-1, 1)$.

In Example 2 the power series for $\ln(1+x)$ was found to be

$$x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}x^k}{k} \tag{12.4.3}$$

again for $|x| < 1$.

When $x = 1$ the series becomes $\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}$. This is the alternating harmonic series, which converges to $\ln(2)$, as Exercise 29 shows. When $x = -1$ the series becomes $\sum_{k=1}^{\infty} \frac{-1}{k}$ which diverges because it is the negative of the harmonic series. This means the interval of convergence for (12.4.3) is $(-1, 1]$.

Some series converge at both endpoints. You can never tell what will happen until you check each endpoint.

How Some Calculators Find e^x

The power series in x for e^x is

$$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^k}{k!} + \cdots .$$

For $x = 10$, this would give

$$e^{10} = 1 + 10 + \frac{10^2}{2!} + \frac{10^3}{3!} + \cdots + \frac{10^k}{k!} + \cdots .$$

Although the terms eventually become very small, the first few terms are quite large. (For instance, the fifth term, $10^4/4!$, is about 417.) So when x is large, the series for e^x provides a time-consuming procedure for calculating e^x.
Some calculators use the following method instead.
The values of e^x at certain inputs are built into the memory:

$$\begin{aligned} e^1 &\approx 2.718281828459 \\ e^{10} &\approx 22,026.46579 \\ e^{100} &\approx 2.6881171 \times 10^{43} \\ e^{0.1} &\approx 1.1051709181 \\ e^{0.01} &\approx 1.0100501671 \\ e^{0.001} &\approx 1.0010005002. \end{aligned}$$

To find $e^{315.425}$, say, the calculator makes use of the identities $e^{x+y} = e^x e^y$ and $(e^x)^y = e^{xy}$ and computes

$$\left(e^{100}\right)^3 \left(e^{10}\right)^1 \left(e^1\right)^5 \left(e^{0.1}\right)^4 \left(e^{0.01}\right)^2 \left(e^{0.001}\right)^5 \approx 9.71263198 \times 10^{136}.$$

This result is accurate to six decimal places.

Summary

We showed how to operate with power series to obtain new power series — by differentiation, integration, or an algebraic operation, such as multiplying or dividing two series. For instance, from the geometric series for $1/(1+x)$, you can obtain the series for $\ln(1+x)$ by integration, or the series for $-1/(1+x)^2$ by differentiation.

In many cases the radius of convergence for a derived power series can be determined directly from the radius of convergence of the original series and the operation performed. However, convergence at the endpoints must be checked for each series.

EXERCISES for Section 12.4

1. Differentiate the Maclaurin series for $\sin(x)$ to obtain the Maclaurin series for $\cos(x)$.

2. Differentiate the Maclaurin series for e^x to show that $D(e^x) = e^x$.

3. Prove Theorem 12.4.2 by carrying out the necessary differentiations.

4.

(a) Show that, for $|t| < 1$, $1/(1 + t^2) = 1 - t^2 + t^4 - t^6 + \cdots$.

(b) Use Theorem 12.4.3 to show that, for $|x| < 1$, $\arctan(x) = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots$.

(c) Give the formula for the k^{th} term of the series in (b).

(d) How many terms of the series in (b) are needed to approximate $\arctan(1/2)$ to three decimal places?

(e) Use the formula in (b) to estimate $\arctan(1/2)$ to three decimal places.

Exercise 22 shows that the series in (b) converges to $\arctan(x)$ also when $x = -1$ and $x = 1$.

5.

(a) Using Theorem 12.4.3, show that for $|x| < 1$,

$$\int_0^x \frac{dt}{1 + t^3} = x - \frac{x^4}{4} + \frac{x^7}{7} - \frac{x^{10}}{10} + \cdots.$$

(b) Use (a) to express $\int_0^{0.7} dt/(1 + t^3)$ as a series whose terms are numbers.

(c) How many terms of the series in (a) are needed to estimate $\int_0^{0.7} dt/(1 + t^3)$ to three decimal places?

(d) Use (b) to evaluate $\int_0^{0.7} dt/(1 + t^3)$ to three decimal places.

(e) Describe how you would evaluate $\int_0^{0.7} dt/(1 + t^3)$ using the fundamental theorem of calculus. (*Do not carry out the details.*)

(f) Use a computer algebra system to find the exact value of $\int_0^{0.7} dt/(1 + t^3)$.

6.

(a) Find the first four nonzero terms of the Maclaurin series for $e^x/(1-x^2)$ by division of series. (Keep the first five terms of e^x.)

(b) Find the first four nonzero terms of the Maclaurin series for $e^x/(1-x^2)$ by using the formula for them in terms of derivatives.

7.

(a) Find the first three nonzero terms of the Maclaurin series for $\tan(x)$ by dividing the series for $\sin(x)$ by the series for $\cos(x)$.

(b) Find the first two nonzero terms of the Maclaurin series for $\tan(x)$ by using the formula for the k^{th} term, $b_k = f^{(k)}(0)/k!$.

8.

(a) Find the first four nonzero terms of the Maclaurin series for $(1-\cos(x))/(1-x^2)$ by division of series.

(b) Find the first four nonzero terms of the Maclaurin series for $(1-\cos(x))/(1-x^2)$ by multiplication of series.

In Exercises 9 and 10, obtain the first three nonzero terms in the Maclaruin series for the indicated functions by algebraic operations with known series. Also, state the radius of convergence.

9. $e^x \sin(x)$

10. $\frac{x}{\cos(x)}$

In Exercises 11 to 16 use power series to determine the limits.

11. $\lim_{x\to 0} \frac{(1-\cos(x))^3}{x^6}$

12. $\lim_{x\to 0} \frac{\sin(3x)}{\sin(2x)}$

13. $\lim_{x\to 0} \frac{\sin^2(x^3)e^x}{(1-\cos(x^2))^3}$

14. $\lim_{x\to 0} \left(\frac{1}{\sin(x)} - \frac{1}{\ln(1+x)}\right)$

15. $\lim_{x\to 0} \frac{(e^x-1)^2(\cos(3x))^2}{\sin(x^2)}$

16. $\lim_{x\to 0} \frac{\sin(x)(1-\cos(x))}{e^{x^3}-1}$

17. Estimate $\int_0^{1/2} \sqrt{x}e^{-x}\, dx$ to four decimal places.

18. Let $f(x) = \sum_{k=0}^{\infty} k^2 x^k$.

(a) What is the domain of f?

(b) Find $f^{(100)}(0)$.

19. Let $f(x) = \arctan(x)$. Making use of the Maclaurin series for $\arctan(x)$, find

(a) $f^{(100)}(0)$

(b) $f^{(101)}(0)$.

20. Since $e^x e^y = e^{x+y}$, the product of the Maclaurin series for e^x and e^y should be the Maclaurin series for e^{x+y}. Check that for terms up to degree 3 in the series for e^{x+y} this is the case.

21.

(a) Give a numerical series whose sum is $\int_0^1 \sqrt{x}\sin(x)\, dx$.

(b) How many terms of the series in (a) are needed to approximate this integral to four decimal places?

(c) Use (a) to evaluate the integral to four decimal places.

22. The Taylor series for $\arctan(x)$ is $\displaystyle\sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} x^{2k+1}$. While the interval of convergence of this power series is easily found to be $[-1, 1]$, Theorem 12.4.3 tells us only that this series converges to $\arctan(x)$ on the open interval $(-1, 1)$.

(a) Show that, when $x = 1$, the series sums to $\arctan(1)$. (Look at the Lagrange Form for the Remainder.)

(b) Repeat (a), using $x = -1$.

(c) Because $\arctan(1) = \pi/4$, the Maclaurin series for $\arctan(1)$ provides one way to obtain approximations to π. Approximate π using the first 5 non-zero terms in the Maclaurin series for $\arctan(1)$.

(d) Estimate the error in the approximation to π found in (c).

(e) How many terms in the Maclaurin series are needed to obtain an approximate value of π accurate to 2 decimal places? 4 decimal places? 12 decimal places?

23.

(a) From the Maclaurin series for $\cos(x)$, obtain the Maclaurin series for $\cos(2x)$.

(b) Exploiting the identity $\sin^2(x) = \frac{1}{2}(1-\cos(2x))$, obtain the Maclaurin series for $\sin^2(x)/x^2$.

(c) Estimate $\int_0^1 (\sin(x)/x)^2 \, dx$ using the first three nonzero terms of the series in (b).

(d) Find a bound on the error in the estimate in (c).

24. Let $\sum_{k=0}^{\infty} b_k x^k$ and $\sum_{k=0}^{\infty} c_k x^k$ converge for $|x| < 1$. If they converge to the same limit for each x in $(-1,1)$ must $b_k = c_k$ for every $k = 0, 1, 2, \ldots$?

25. This exercise outlines a way to compute logarithms of numbers larger than 1.

(a) Show that every number $y > 1$ can be written in the form $(1+x)/(1-x)$ for some x in $(0,1)$.

(b) When $y = 3$, find x.

(c) Show that if $y = (1+x)/(1-x)$, then $\ln(y) = 2(x + x^3/3 + \cdots + x^{2n+1}/(2n+1) + \ldots)$.

(d) Use (b) and (c) to estimate $\ln(3)$ to two decimal places. (To control the error, compare a tail end of the series to an appropriate geometric series.)

(e) Is the error in (d) less than the first omitted term?

26. Sam has an idea: "I have a more direct way of estimating $\ln(y)$ for $y > 1$. I just use the identity $\ln(y) = -\ln(1/y)$. Because $1/y$ is in $(0,1)$ I can write it as $1-x$, and x is still in (0,1). In short, $\ln(y) = -\ln(1/y) = -\ln(1-x) = x + x^2/2 + x^3/3 + \ldots$. It's even an easier formula. And it's better because it doesn't have that coefficient 2 in front."

(a) Is Sam's formula correct?

(b) Use his method to estimate $\ln(3)$ to two decimal places.

(c) Which is better, Sam's method or the one in Exercise 25?

27. Use the method of Exercise 25 to estimate $\ln(5)$ to two decimal places. Include a description of your procedure.

28. Here are five ways to compute $\ln(2)$. Which seems to be the most efficient? least efficient? Explain.

(a) The series for $\ln(1+x)$ when $x = 1$.

(b) The series for $\ln(1+x)$ when $x = \frac{-1}{2}$. This gives $\ln\left(\frac{1}{2}\right) = -\ln(2)$.

(c) The series for $\ln((1+x)/(1-x))$ when $x = \frac{1}{3}$.

(d) Simpson's method applied to the integral $\int_1^2 dx/x$.

(e) The root of $e^x = 2$. (Use Newton's method.)

(f) The sum $\frac{1}{n} + \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n}$

29. In the discussion of endpoints for the Maclaurin series for $\ln(1+x)$, we showed that the series converges for $x = 1$, but we did not show that its sum is $\ln(2)$. To show that it does equal $\ln(2)$, integrate both sides of the following equation over $[0, 1]$:

$$\frac{1 + (-x)^{n+1}}{1+x} = 1 - x + x^2 - x^3 + \cdots + (-1)^n x^n.$$

(Separate the left-hand side into two separate integrals. Then, take the limit as $n \to \infty$.)

30.

(a) Compute the product of the Maclaurin series of degree 5 for e^x and e^y.

(b) How does the result compare with the first few terms of teh Maclaurin series for e^{x+y}?

31.

(a) For which x does $\sum_{k=0}^{\infty} k^2 x^k$ converge?

(b) Starting with the Maclaurin series for $x^2/(1-x)$, sum the series in (a).

(c) Does your formula seem to give the correct answer when $x = \frac{1}{3}$?

32. This exercise uses power series to give a new perspective on l'Hôpital's rule. Assume that f and g can be represented by power series in some open interval containing 0:

$$f(x) = \sum_{k=0}^{\infty} b_k x^k \quad \text{and} \quad g(x) = \sum_{k=0}^{\infty} c_k x^k.$$

Assume that $f(0) = 0$, $g(0) = 0$, and $g'(0) \neq 0$. Under these assumptions explain why

$$\lim_{x \to 0} \frac{f(x)}{g(x)} = \lim_{x \to 0} \frac{f'(x)}{g'(x)}.$$

33. If R. P. Feynman, *Lectures on Physics*, Addison-Wesley, Reading, MA, 1963, appears this remark:

> Thus the average velocity is
>
> $$\langle E \rangle = \frac{\hbar\omega(0 + x + 2x^2 + 3x^3 + \cdots)}{1 + x + x^2 + \cdots}.$$
>
> Now the two sums which appear here we shall leave for the reader to play with and have some fun with. When we are all finished summing and substituting for x in the sum, we should get — if we make no mistakes in the sum —
>
> $$\langle E \rangle = \frac{\hbar\omega}{e^{\hbar\omega/kT} - 1}.$$
>
> This, then, was the first quantum-mechanical formula ever known, or ever discussed, and it was the beautiful culmination of decades of puzzlement.

Have the aforementioned fun, given that $x = e^{-\hbar\omega/kT}$.

Exercises 34 to 37 outline a proof that the Maclaurin series associated with $(1+x)^r$ converges to $(1+x)^r$ for $|x| < 1$. This justifies the assertion that $(1+x)^r = \sum_{k=0}^{\infty} \binom{n}{k} x^k$ for $|x| < 1$. The notation $\binom{n}{k}$ stands for $\frac{n!}{k!(n-k)!}$.

34. Show that

$$k\binom{r}{k} + (k+1)\binom{r}{k+1} = r\binom{r}{k}.$$

(This is needed in Exercise 35.) (First, rewrite the equation as $(k+1)\binom{r}{k+1} = (r-k)\binom{r}{k}$.)

35. Let $f(x) = \sum_{k=0}^{\infty} \binom{r}{k} x^k$.

(a) Find the interval of convergence for $f(x)$.

(b) Show that $(1+x)f'(x) = rf(x)$. (First, write out the first four terms to see the pattern.)

36. Using the result from Exercise 35, show that the derivative of $f(x)/(1+x)^r$ is 0.

37. Show that $f(x)/(1+x)^r = 1$, which implies that $\sum_{k=0}^{\infty} \binom{r}{k} x^k = (1+x)^r$. What is the interval of convergence

38. Newton obtained the Maclaurin series for $\arcsin(x)$ with the aid of the binomial series for $\sqrt{1-x^2}$, as follows.
Consider the circle $x^2 + y^2 = 1$ and the point $Q = (x, y)$ on it, as shown in Figure 12.4.3. Then $\theta = \arcsin(x) = \angle QOR$.

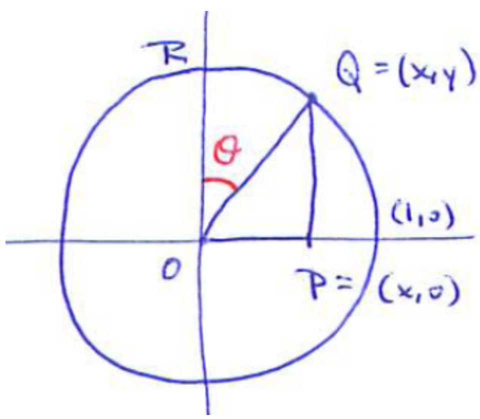


Figure 12.4.3

(a) Then

$$\begin{aligned} \frac{\theta}{2} &= \text{area}OQR = \text{area}OPQR - \text{area}OPQ \\ &= \int_0^x \sqrt{1-t^2}\, dt - \frac{1}{2}x\sqrt{1-x^2}. \end{aligned}$$

Use this equation to obtain Newton's result:

$$\theta = x + \frac{1}{6}x^3 + \frac{3}{40}x^5 + \frac{5}{112}x^7 + \cdots \tag{12.4.4}$$

(b) Use the fact that $\theta = \arcsin(x) = \int_0^x \frac{dt}{\sqrt{1-t^2}}$ to derive (12.4.4).

12.5 Complex Numbers

The number line of real numbers coincides with the x-axis of the xy coordinate system. With its addition, subtraction, multiplication, and division, it is a small part of a number system that occupies the plane, and which obeys the usual rules of arithmetic. This section describes that system, known as the **complex numbers**. One of the important properties of the complex numbers is that any nonconstant polynomial has a root; in particular, the equation $x^2 = -1$ has two solutions.

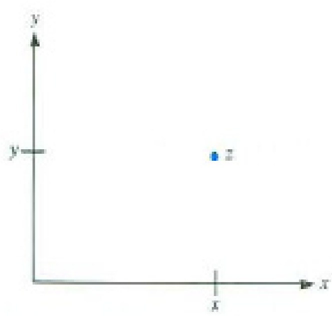


Figure 12.5.1

The Complex Numbers

By a complex number z we mean an expression of the form $x + iy$ or $x + yi$, where x and y are real numbers and i is a symbol with the property that $i^2 = -1$. This expression will be identified with the point (x, y) in the xy plane, as in Figure 12.5.1. Every point in the xy plane may therefore be thought of as a complex number.

To add or multiply two complex numbers, follow the usual rules of arithmetic of real numbers, with one new proviso:

Whenever you see i^2, replace it by -1.

For instance, to add the complex numbers $3 + 2i$ and $-4 + 5i$, just collect like terms:

$$(3 + 2i) + (-4 + 5i) = (3 - 4) + (2i + 5i) = -1 + 7i.$$

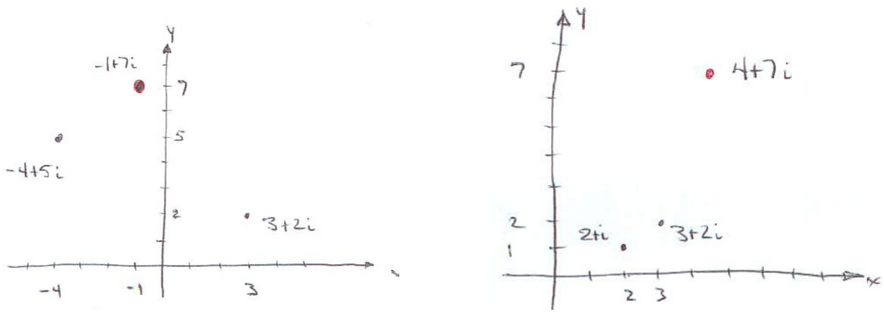
(See Figure 12.5.2(a).) Addition does not make use of the fact that $i^2 = -1$. However, multiplication does, as Example 1 shows.

EXAMPLE 1 Compute the product $(2 + i)(3 + 2i)$.
SOLUTION We can multiply the complex numbers just as we would multiply binomials: we have

$$(2+i)(3+2i) = 2\cdot 3+2\cdot 2i+i\cdot 3+i\cdot 2i = 6+4i+3i+2i^2 = 6+4i+3i-2 = 4+7i.$$

Figure 12.5.2(b) shows the complex numbers $2 + i$, $3 + 2i$, and their product $4 + 7i$. ◇

Note that $(-i)(-i) = i^2 = -1$. Both i and $-i$ are square roots of -1. The symbol $\sqrt{-1}$ traditionally denotes i rather than $-i$.



(a) (b)

Figure 12.5.2

A complex number that lies on the y-axis is called **imaginary**. Every complex number z is the sum of a real number and an imaginary number, $z = x + iy$. The number x is called the **real part of** z, and y is called the **imaginary part**. One often writes "$\operatorname{Re} z = x$" and "$\operatorname{Im} z = y$."

Real numbers are on the x-axis, imaginary on the y-axis.

We have seen how to add and multiply complex numbers. Subtraction is straightforward. For instance,

$$(3 + 2i) - (4 - i) = (3 - 4) + (2i - (-i)) = -1 + 3i.$$

Division of complex numbers requires rationalizing the denominator. This involves the **conjugate** of a complex number. The conjugate of the complex number $z = x + yi$ is the complex number $x - yi$, which is denoted $\bar{z}$. Note that

conjugate of z

$$\begin{aligned} z\bar{z} &= (x + yi)(x - yi) = x^2 + y^2 \\ z + \bar{z} &= (x + yi) + (x - yi) = 2x \\ \text{and} \qquad z - \bar{z} &= (x + yi) - (x - yi) = 2yi. \end{aligned}$$

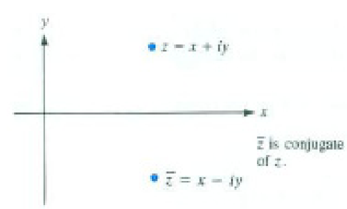


Figure 12.5.3

Thus, $z\bar{z}$ and $z + \bar{z}$ are real, and $z - \bar{z}$ is imaginary. Figure 12.5.3 shows that $\bar{z}$ is the mirror image of z reflected across the x-axis. To "rationalize the denominator" means to find an equivalent fraction with a real-valued denominator. If the fraction is w/z, the denominator can be rationalized by multiplying by $\bar{z}/\bar{z}$.

rationalizing the denominator

EXAMPLE 2 Compute the quotient $\frac{1+5i}{3+2i}$.
SOLUTION To rationalize the denominator, we multiply by $\frac{3-2i}{3-2i}$:

$$\frac{1 + 5i}{3 + 2i} = \frac{1 + 5i}{3 + 2i} \cdot \frac{3 - 2i}{3 - 2i} = \frac{3 - 2i + 15i + 10}{9 - 6i + 6i + 4i^2} = \frac{13 + 13i}{13} = 1 + i.$$

◇

Now All Polynomials Have Roots

Every polynomial has a root in the complex numbers.

The complex numbers provide the equation $x^2 + 1 = 0$ with two solutions, i and $-i$. This illustrates an important property of complex numbers: If $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$ is any polynomial of degree $n \geq 1$, with real or complex coefficients, then there is a complex number z such that $f(z) = 0$. This fact, known as the **Fundamental Theorem of Algebra**, is illustrated in Example 3. Its proof requires advanced mathematics.

EXAMPLE 3 Solve the quadratic equation $z^2 - 4z + 5 = 0$.
SOLUTION By the quadratic formula, the solutions are

$$\begin{aligned} z &= \frac{-(-4) \pm \sqrt{(-4)^2 - 4 \cdot 1 \cdot 5}}{2 \cdot 1} \\ &= \frac{4 \pm \sqrt{-4}}{2} = \frac{4 \pm 2i}{2} = 2 \pm i. \end{aligned}$$

The two solutions are $2 + i$ and $2 - i$.

These solutions can be checked by substitution in the original equation. For instance,

$$\begin{aligned} (2+i)^2 - 4(2+i) + 5 &= (4 + 4i + i^2) - 8 - 4i + 5 \\ &= 4 + 4i - 1 - 8 - 4i + 5 = 0 + 0i = 0. \end{aligned}$$

Yes, it checks. The solution $2 - i$ can be checked similarly. ◇

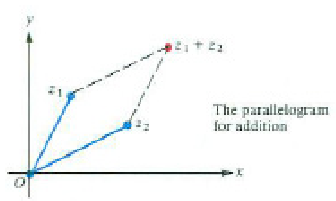


Figure 12.5.4

The sum of the complex numbers z_1 and z_2 is the fourth vertex (opposite O) in the parallelogram determined by the origin O and the points z_1 and z_2, as shown in Figure 12.5.4.

The Geometry of the Product

Amplitude is a synonym for magnitude.

The symbols $|z|$ and $\arg(z)$

The geometric relation between z_1, z_2 and their product $z_1 z_2$ is easily described in terms of the magnitude and argument of a complex number. Each complex number z other than the origin is at a (positive) distance r from the origin and has a polar angle θ relative to the positive x-axis. The distance r is called the **magnitude of** z, and θ is called the **argument of** z. A complex number has an infinity of arguments differing from each other by an integer multiple of 2π. The complex number 0, which lies at the origin, has magnitude 0 and any polar angle as argument. In short, we may think of magnitude and argument as polar coordinates r and θ of z, with the restriction that r is nonnegative. The magnitude of z is denoted $|z|$. The symbol $\arg(z)$ denotes any of the arguments of z, it being understood that if θ is an argument of z, then so is $\theta + 2n\pi$ for any integer n.

EXAMPLE 4

(a) Draw all complex numbers with magnitude 3.

(b) Draw the complex number z of magnitude 3 and argument $\pi/6$.

SOLUTION

(a) The complex numbers of magnitude 3 form a circle of radius 3 with center at 0. (See Figure 12.5.5.)

(b) The complex number of magnitude 3 and argument $\pi/6$ is shown (in red) in Figure 12.5.5.

◇

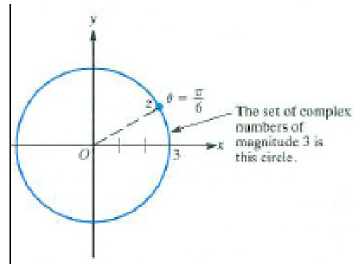


Figure 12.5.5

Note that $|x+iy| = \sqrt{x^2+y^2}$, by the Pythagorean theorem. Each complex number $z = x+iy$ other than 0 can be written as the product of a positive real number and a complex number of magnitude 1. To show this, let $z = x+iy$ have magnitude r and argument θ. Recalling the relation between polar and rectangular coordinates, we conclude that

$$z \quad = \quad r\cos(\theta) + r\sin(\theta)i = r(\cos(\theta) + i\sin(\theta)).$$

The number r is a positive real number. The magnitude of the number $\cos(\theta) + i\sin(\theta)$ is $\sqrt{\cos(\theta)^2 + \sin(\theta)^2} = 1$. Figure 12.5.6 shows the numbers r and $\cos(\theta) + i\sin(\theta)$, whose product is z. (The expression $\cos(\theta) + i\sin(\theta)$ appears so frequently when working with complex numbers that the shorthand notation $\text{cis}(\theta)$ is used, that is, $\text{cis}(\theta) = \cos(\theta) + i\sin(\theta)$. While this is convenient, you have to be careful not to confuse "cis" with "cos.")

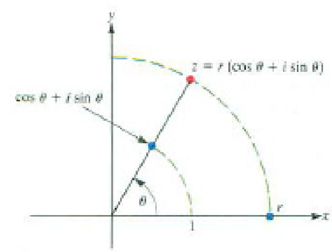


Figure 12.5.6 ARTIST: Draw the point for (b) in red.

The next theorem describes *how to multiply two complex numbers if they are given in polar form,* that is, in terms of their magnitudes and arguments.

Theorem 12.5.1. *Assume that z_1 has magnitude r_1 and argument θ_1 and that z_2 has magnitude r_2 and argument θ_2. Then the product z_1z_2 has magnitude r_1r_2 and argument $\theta_1+\theta_2$.*

Proof

$$\begin{aligned} z_1z_2 &= r_1(\cos(\theta_1) + i\sin(\theta_1))r_2(\cos(\theta_2) + i\sin(\theta_2)) \\ &= r_1r_2(\cos(\theta_1) + i\sin(\theta_1))(\cos(\theta_2) + i\sin(\theta_2)) \\ &= r_1r_2(\cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2) + i(\sin(\theta_1)\cos(\theta_2) + \cos(\theta_1)\sin(\theta_2)) \\ &= r_1r_2(\cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2)) \end{aligned}$$

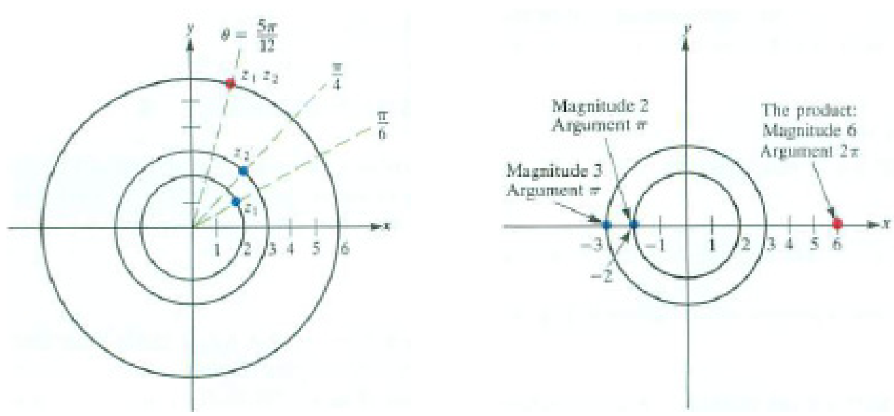
The last step uses the identities for $\cos(u+v)$ and $\sin(u+v)$.

Thus, the magnitude of z_1z_2 is r_1r_2 and the argument of z_1z_2 is $\theta_1 + \theta_2$. This proves the theorem. •

In practical terms, this theorem says:

> "To multiply two complex numbers, add their arguments and multiply their magnitudes."

EXAMPLE 5 Find z_1z_2 for z_1 and z_2 in Figure 12.5.7(a).
SOLUTION z_1 has magnitude 2 and argument $\pi/6$; z_2 has magnitude 3 and argument $\pi/4$. Thus, z_1z_2 has magnitude $2 \cdot 3 = 6$ and argument $\pi/6 + \pi/4 = 5\pi/12$. (See Figure 12.5.7 ◇



(a) (b)

Figure 12.5.7

EXAMPLE 6 Using the geometric description of multiplication, find the product of the real numbers -2 and -3.
SOLUTION The number -2 has magnitude 2 and argument π. The number -3 has magnitude 3 and argument π. Therefore $(-2) \cdot (-3)$ has magnitude $2 \cdot 3 = 6$ and argument $\pi + \pi = 2\pi$. The complex number with magnitude 6 and argument 2π is just our old friend, the real number 6. Thus $(-2) \cdot (-3) = 6$, in agreement with the statement "the product of two negative numbers is positive." (See Figure 12.5.7(b).) ◇

Division of Complex Numbers

Division of complex numbers given in polar form is similar, except that the magnitudes are divided and the arguments are subtracted:

See Exercise 29.

$$\frac{r_1(\cos(\theta_1) + i\sin(\theta_1))}{r_2(\cos(\theta_2) + i\sin(\theta_2))} = \frac{r_1}{r_2}\left(\cos(\theta_1 - \theta_2) + i\sin(\theta_1 - \theta_2)\right).$$

EXAMPLE 7 Let $z_1 = 6(\cos(\pi/2) + i\sin(\pi/2))$ and $z_2 = 3(\cos(\pi/6) + i\sin(\pi/6))$. Find (a) $z_1 z_2$ and (b) z_1/z_2.
SOLUTION See Figure 12.5.8

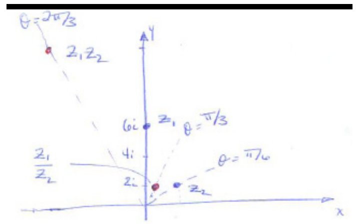


Figure 12.5.8

(a)

$$\begin{aligned} z_1 z_2 &= 6\cdot 3\left(\cos\left(\frac{\pi}{2} + \frac{\pi}{6}\right) + i\sin\left(\frac{\pi}{2} + \frac{\pi}{6}\right)\right) \\ &= 18\left(\cos\left(\frac{2\pi}{3}\right) + i\sin\left(\frac{2\pi}{3}\right)\right) \\ &= 18\left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i\right) = -9 + 9\sqrt{3}i. \end{aligned}$$

(b)

$$\begin{aligned} \frac{z_1}{z_2} &= \frac{6}{3}\left(\cos\left(\frac{\pi}{2} - \frac{\pi}{6}\right) + i\sin\left(\frac{\pi}{2} - \frac{\pi}{6}\right)\right) = 2\left(\cos\left(\frac{\pi}{3}\right) + i\sin\left(\frac{\pi}{3}\right)\right) \\ &= 2\left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right) = 1 + \sqrt{3}i \end{aligned}$$

◇

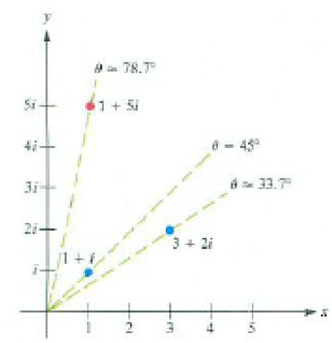


Figure 12.5.9

EXAMPLE 8 Compute the product $(1+i)(3+2i)$ arithmetically and check the answer in terms of magnitudes and arguments.
SOLUTION

$$(1+i)(3+2i) = 3 + 2i + 3i + 2i^2 = 3 + 2i + 3i - 2 = 1 + 5i.$$

To check this calculation, we first verify that $|1+5i| = |1+i||3+2i|$. We have

$$\begin{aligned} |1+5i| &= \sqrt{1^2+5^2} = \sqrt{26}, \\ |1+i| &= \sqrt{1^2+1^2} = \sqrt{2}, \\ |3+2i| &= \sqrt{3^2+2^2} = \sqrt{13}. \end{aligned}$$

Since $\sqrt{26} = \sqrt{2}\sqrt{13}$, the magnitude of $1+5i$ is the product of the magnitudes of $1+i$ and $3+2i$.

$\arg(x+iy) = \arctan(y/x)$ for $x+iy$ in the first or fourth quadrants.

Next, consider the arguments. First, $\arg(1+5i) = \arctan(5) \approx 1.3734$. Similarly, $\arg(1+i) = \arctan(1) \approx 0.7854$ and $\arg(3+2i) = \arctan(2/3) \approx 0.5880$. Since $0.7854+0.5880 = 1.3734$, it indeed appears that the argument of $1+5i$ is the sum of the arguments of $1+i$ and $3+2i$. (See also Figure 12.5.9.) ◇

Powers of z

When the polar coordinates of z are known, it is easy to compute the powers z^2, z^3, z^4, ... Let z have magnitude r and argument θ. Then $z^2 = z \cdot z$ has magnitude $r \cdot r = r^2$ and argument $\theta + \theta = 2\theta$. So, to square a complex number, just square its magnitude and double its argument (angle).

How to compute z^n

More generally, to compute z^n for any positive integer n, find $|z|^n$ and multiply the argument of z by n. In short, we have **DeMoivre's Law**:

$$\boxed{(r(\cos(\theta) + i\sin(\theta)))^n . = r^n\left(\cos(n\theta) + i\sin(n\theta)\right)}$$

Example 9 illustrates the geometric view of computing powers.

EXAMPLE 9 Let z have magnitude 1 and argument $2\pi/5$. Compute and sketch z, z^2, z^3, z^4, z^5, and z^6.

SOLUTION Since $|z| = 1$, it follows that $|z^2| = |z|^2 = 1^2 = 1$. Similarly, for all positive integers n, $|z^n| = 1$; that is, z^n is a point on the unit circle with center at the origin, O. All that remains is to examine the arguments of z^2, z^3, etc..

The argument of z^2 is twice the argument of z: $2(2\pi/5) = 4\pi/5$. Similarly, $\arg(z^3) = 6\pi/5$, $\arg(z^4) = 8\pi/5$, $\arg(z^5) = 10\pi/5 = 2\pi$, and $\arg(z^6) = 12\pi/5$. Observe that $z^5 = 1$, since it has magnitude 1 and argument 2π. Similarly, $z^6 = z$, since both z and z^6 have magnitude 1 and their arguments differ by an integer multiple of 2π. (Or, algebraically, $z^6 = z^{5+1} = z^5 \cdot z = 1 \cdot z = z$.) Figure 12.5.10 shows that the powers of z form the vertices of a regular pentagon. ◇

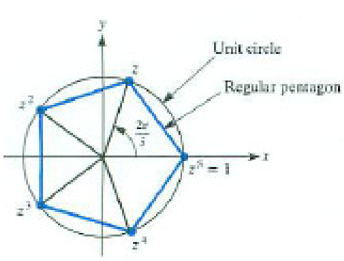


Figure 12.5.10

The equation $x^5 = 1$ has only one real root, namely, $x = 1$. However, it has five complex roots. For instance, the number z shown in Figure 12.5.10 is a solution of $x^5 = 1$ since $z^5 = 1$. Another root is z^2, since $\left(z^2\right)^5 = z^{10} = \left(z^5\right)^2 = 1^2 = 1$. Similarly, z^3 and z^4 are roots of $x^5 = 1$. There is a total of five roots: 1, z, z^2, z^3, and z^4.

The powers of i.

The powers of i will be needed in the next section. They are $i^2 = -1$, $i^3 = i^2 \cdot i = (-1)i = -i$, $i^4 = i^3 \cdot i = (-i)i = -i^2 = 1$, $i^5 = i^4 \cdot i = i$, and so on. They repeat in blocks of four: for any integer n, $i^{n+4} = i^n$.

It is often useful to express a complex number $z = x + iy$ in polar form. Recall that $|z| = \sqrt{x^2 + y^2}$. To find θ, it is best to sketch z in order to see in which quadrant it lies. Although $\arctan(\theta) = y/x$ we cannot say that $\theta = \arctan(y/x)$, since $\arctan(u)$ lies between $-\pi/2$ and $\pi/2$ for any real number u. However, the angle of z may lie in the second- or third-quadrant

For instance, to put $z = -2 - 2i$ in polar form, first sketch z, as in Figure 12.5.11. We have $|z| = \sqrt{(-2)^2 + (-2)^2} = \sqrt{8}$ and $\arg(z) = 5\pi/4$. Thus

$$z = \sqrt{8}\left(\cos\left(\frac{5\pi}{4}\right) + i\sin\left(\frac{5\pi}{4}\right)\right).$$

Note that $\arctan(-2/(-2))$ is $\pi/4$ which is *not* an argument of z.

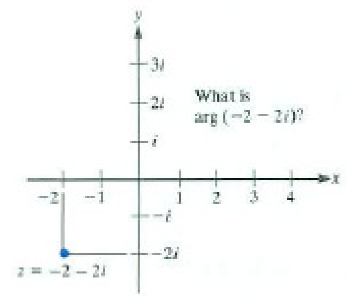


Figure 12.5.11

Roots of z

Each complex number z, other than 0, has exactly n n^{th} roots for each positive integer n. These can be found by expressing z in polar coordinates. If $z = r(\cos(\theta) + i\sin(\theta))$, that is, has magnitude r and argument θ, then one n^{th} root of z is

$$r^{1/n}\left(\cos\left(\frac{\theta}{n}\right) + i\sin\left(\frac{\theta}{n}\right)\right).$$

To check that this is an n^{th} root of z, just raise it to the n^{th} power.

To find the other n^{th} roots of z, change the argument z from θ to $\theta + 2k\pi$, where $k = 1, 2, \ldots, n-1$. Then

$$r^{1/n}\left(\cos\left(\frac{\theta + 2k\pi}{n}\right) + i\sin\left(\frac{\theta + 2k\pi}{n}\right)\right)$$

is also an n^{th} root of z. (Why?)

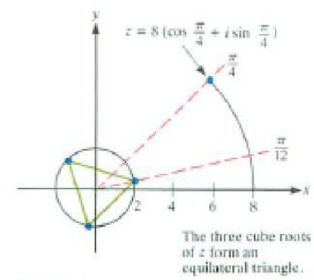


Figure 12.5.12

The n roots of the equation $z^n = a$ are the vertices of a regular polygon with n sides.

For instance, let $z = 8(\cos(\pi/4) + i\sin(\pi/4))$. Then the three cube roots of z all have magnitude $8^{1/3} = 2$. Their arguments are

$$\frac{\pi/4}{3} = \frac{\pi}{12}, \qquad \frac{\pi/4 + 2\pi}{3} = \frac{\pi}{12} + \frac{2\pi}{3}, \qquad \frac{\pi/4 + 4\pi}{3} = \frac{\pi}{12} + \frac{4\pi}{3}.$$

These three roots are shown in Figure 12.5.12, along with z.

Summary

The real numbers, with which we all grew up, are just a small part of the complex numbers, which fill up the xy plane. We add complex numbers by a "parallelogram law." To multiply them "we multiply their magnitudes and add their angles." Using the complex numbers we can see that "negative real time negative real is positive," since $180° + 180° = 360°$, which describes the positive x-axis. We also saw how to raise a complex number to a power and how to take its roots. We can now view points in the xy plane as "numbers." However, mathematicians have shown that we cannot treat points in three-dimensional space as "numbers" that satisfy the usual rules of addition and multiplication.

EXERCISES for Section 12.5

In Execises 1 to 4 express the given complex numbers in the form $x + iy$.

1.

(a) $(2 + 3i) + (5 - 2i)$

(b) $(2 + 3i)(2 - 3i)$

(c) $\frac{1}{2-i}$

(d) $\frac{3+2i}{4-i}$

2.

(a) $(2 + 3i)^2$

(b) $\frac{4}{3-i}$

(c) $(1 + i)(3 - i)$

(d) $\frac{1+5i}{2-3i}$

3.

(a) $(1 + 3i)^2$

(b) $(1 + i)(1 - i)$

(c) i^{-3}

(d) $\frac{4+\sqrt{2}i}{2+i}$

4.

(a) $(1 + i)^3$

(b) $\frac{i}{1-i}$

(c) $(3 + i)^{-1}$

(d) $(5 + 2i)(5 - 2i)$

In Exercises 5 to 8 express the number in polar form $r(\cos(\theta) + i\sin(\theta))$ with θ in $[0, 2\pi)$.

5. $\sqrt{3} + i$

6. $\sqrt{3} - i$

7. $\sqrt{2} + \sqrt{2}i$

8. $-4 + 4i$

In Exercises 9 to 12 express the number in both polar and rectangular forms.

9. $(-1 + i)^{10}$

10. $(\sqrt{3} + i)^4$

11. $(2 + 2i)^8$

12. $1 - \sqrt{3}i)^7$

13. Rationalize the denominator in each fraction. That is, express the fraction as an equivalent fraction whose denominator does not have a square root or i.

(a) $\frac{1}{1+\sqrt{2}}$

(b) $\frac{1}{2-i}$

(c) $\frac{2-\sqrt{3}}{\sqrt{3}+2}$

(d) $\frac{3+2i}{i-3}$

14. For each equation, (i) find all solutions, (ii) plot all solutions in the complex plane, and (iii) check that the solutions satisfy the equations.

(a) $x^2 + x + 1 = 0$

(b) $x^2 - 3x + 5 = 0$

(c) $2x^2 + x + 1 = 0$

(d) $3x^2 + 4x + 5 = 0$

15.

(a) Use the quadratic formula to find all solutions of the equation $x^2 + x + 1 = 0$.

(b) Plot the solutions in (a).

(c) Check that the solutions in (a) satisfy $x^2 + x + 1 = 0$.

16. Let z_1 have magnitude 2 and argument $\pi/6$, and let z_2 have magnitude 3 and argument $\pi/3$.

(a) Plot z_1 and z_2.

(b) Find $z_1 z_2$ using the polar form.

(c) Write z_1 and z_2 in the rectangular form $x + yi$.

(d) With the aid of (c) compute $z_1 z_2$.

(e) Check that (b) and (d) give the same point.

17. Let z_1 have magnitude 2 and argument $\pi/4$, and let z_2 have magnitude 3 and argument $3\pi/4$.

(a) Plot z_1 and z_2.

(b) Find $z_1 z_2$ using the polar form.

(c) Write z_1 and z_2 in the form $x + yi$.

(d) With the aid of (c) compute $z_1 z_2$.

(e) Check that (b) and (d) give the same point.

18. The complex number z has argument $\pi/3$ and magnitude 1. Find and plot (a) z^2, (b) z^3, (c) z^4, and (d) $1/\overline{z}$.

19. Find and plot (a) i^3, (b) i^4, (c) i^5, and (d) i^{73}.

20. Let z have magnitude 2 and argument $\pi/6$.

(a) What are the magnitude and argument of z^2, z^3, and z^4?

(b) Sketch z, z^2, z^3, and z^4.

(c) What are the magnitude and argument of z^n?

21. Let z have magnitude 0.9 and argument $\pi/4$.

(a) Find and plot z^2, z^3, z^4, z^5, and z^6.

(b) What happens to z^n as $n \to \infty$?

22. Find and plot all solutions of the equation $z^5 = 32\left(\cos(\pi/4) + i\sin(\pi/4)\right)$.

23. Find and plot all solutions of the equation $z^4 = 8 + 8\sqrt{3}i$. (First draw $8 + 8\sqrt{3}i$.)

24. Let z have magnitude r and argument θ. Let w have magnitude $1/r$ and argument $-\theta$. Show that $zw = 1$. w is called the **reciprocal of** z, and denoted z^{-1} or $1/z$.

25. Find z^{-1} if $z = 4 + 4i$. See Exercise 24.

26.

(a) By substitution, verify that $2+3i$ is a solution of the equation $x^2 - 4x + 13 = 0$.

(b) Use the quadratic formula to find all solutions of the equation $x^2 - 4x + 13 = 0$.

27. Write in polar form

(a) $5 + 5i$,

(b) $-\frac{1}{2} - \frac{\sqrt{3}}{2}i$,

(c) $-\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i$,

(d) $3 + 4i$, and

(e) $1/\overline{(3 + 4i)}$.

28. Write in rectangular form as simply as possible:

(a) $3\left(\cos\left(\frac{3\pi}{4}\right) + i\sin\left(\frac{3\pi}{4}\right)\right)$,

(b) $2\left(\cos\left(\frac{\pi}{6}\right) + i\sin\left(\frac{\pi}{6}\right)\right)$,

(c) $10\left(\cos(\pi) + i\sin(\pi)\right)$,

(d) $\frac{1}{5}\left(\cos(22^\circ) + i\sin(22^\circ)\right)$ (Express the answer to at least three decimal places.)

29. Let z_1 have magnitude r_1 and argument θ_1, and let z_2 have magnitude r_2 and argument θ_2.

(a) Explain why the magnitude of z_1/z_2 is r_1/r_2.

(b) Explain why the argument of z_1/z_2 is $\theta_1 - \theta_2$.

30. Compute

$$\frac{\cos\left(\frac{5\pi}{4}\right) + i\sin\left(\frac{5\pi}{4}\right)}{\cos\left(\frac{3\pi}{4}\right) + i\sin\left(\frac{3\pi}{4}\right)}$$

by two ways: (a) by the result of Exercise 29, (b) by rationalizing the denominator.

31. Compute

(a) $(2+3i)(1+i)$

(b) $\frac{2+3i}{1+i}$

(c) $(7-3i)(\overline{7-3i})$

(d) $3(\cos(42^\circ) + i\sin(42^\circ)) \cdot 5(\cos(168^\circ) + i\sin(168^\circ))$

(e) $\frac{\sqrt{8}(\cos(147^\circ)+i\sin(147^\circ)}{\sqrt{2}(\cos(57^\circ)+i\sin(57^\circ))}$

(f) $1/(3-i)$

(g) $((\cos(52^\circ) + i\sin(52^\circ))^{-1}$

(h) $\left(\cos\left(\frac{\pi}{6}\right) + i\sin\left(\frac{\pi}{6}\right)\right)^{12}$

32. Compute

(a) $(4+3i)(4-3i)$

(b) $\frac{3+5i}{-2+i}$

(c) $\frac{1}{2+i}$

(d) $\left(\cos(\left(\frac{\pi}{12}\right) + i\sin(\left(\frac{\pi}{12}\right)\right)^{20}$

(e) $(r(\cos(\theta) + i\sin(\theta))^{-1}$

(f) $\mathrm{Re}\left((r(\cos(\theta) + i\sin(\theta)))^{10}\right)$

(g) $\frac{3\left(\cos\left(\frac{\pi}{6}\right)+i\sin\left(\frac{\pi}{6}\right)\right)}{5-12i}$

33. Find and plot all solutions of $z^3 = i$.

34. Sketch all complex numbers z such that (a) $z^6 = 1$, (b) $z^6 = 64$, (c) $z^6 = -1$.

35.

(a) Why is the symbol $\sqrt{-4}$ ambiguous?

(b) Draw all solutions of $z^2 = -4$.

36. If z_k has argument θ_k and magnitude r_k, $k = 1, 2$, write each of the following in the form $r(\cos(\theta) + i\sin(\theta))$.

(a) z_1^2

(b) $1/z_1$

(c) $\overline{z_1}$

(d) $z_1 z_2$

(e) z_1/z_2

(f) $1/\overline{z_1}$

37. Draw the six sixth roots of

(a) 1

(b) 64

(c) i

(d) -1

(e) $\frac{-1}{2} + \frac{\sqrt{3}}{2}i$

38. Using the fact that

$$(\cos(\theta) + i\sin(\theta))^n = \cos(n\theta) + i\sin(n\theta)$$

find formulas for $\cos(3\theta)$ and $\sin(3\theta)$ in terms of $\cos(\theta)$ and $\sin(\theta)$.

39.

(a) If $|z_1| = 1$ and $|z_2| = 1$, how large can $|z_1 + z_2|$ be? (Draw some pictures.)

(b) If $|z_1| = 1$ and $|z_2| = 1$, what can be said about $|z_1 z_2|$?

40. Show that (a) $\overline{z_1 z_2} = \overline{z_1}\,\overline{z_2}$, (b) $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$.

41. Let $z = \frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}}$.

(a) Compute z^2 algebraically.

(b) Compute z^2 by putting z into polar form.

(c) Sketch the numbers z, z^2, z^3, z^4, and z^5.

42. Let a, b, and c be complex numbers such that $a \neq 0$ and $b^2 - 4ac \neq 0$. Show that $ax^2 + bx + c = 0$ has two distinct roots.

43. Find and plot the roots of $x^2 + ix + 3 - i = 0$.

44. Compute the roots of the following equations and plot them relative to the same axes:

(a) $x^2 - 3x + 2 = 0$

(b) $x^2 - 3x + 2.25 = 0$

(c) $x^2 - 3x + 2.5 = 0$

(d) $x^2 - 3x + 1.5 = 0$

45. The complex number $z = t + i$ (t a real number) lies on the line $y = 1$.

(a) Plot z^2 for $t = 0$, 1, -1, and for at least two other values of t.

(b) Find the equation of the curve on which z^2 lies.

46. The complex number $z = t + i/t$ (t a positive real number) lies on the parabola $y = 1/x$.

(a) Sketch the curve on which z^2 lies.

(b) Find the equation in rectangular coordinates, u and v, for the curve in (b).

47. Let z be the typical complex number on the curve C whose polar equation is $r = \cos(\theta)$.

(a) Sketch z^2 for at least four choices of z on C.

(b) Find the equation in polar coordinates for the curve swept out by z^2 for z on C.

(c) With the aid of the equation in (b), sketch the curve.

48. Let C be the same curve as in the preceding exercise.

(a) Sketch $1/\overline{z}$ for at least four choices of z on C.

(b) Find the equation in polar and in rectangular coordinates for the curve swept out by $1/\overline{z}$ for points on C.

49.

(a) Draw the curve on which $z = t + ti$ lies.

(b) Draw the curve on which z^2 lies.

(c) Give the equations in rectangular coordinates for both curves.

50. In parts (a) through (d) plot z, $\overline{z}$, and $1/z$ on the same set of axes.

(a) Let $z = 1 + \sqrt{3}i$.

(b) Let $z = (1 + i)/\sqrt{2}$.

(c) Let $z = 3$.

(d) Let $z = 2i$.

(e) For an arbitrary complex number z, give a verbal explanation (no equations and no graphs) of the relationships among z, $\overline{z}$, and $1/z$.

51. For which complex numbers z is $\overline{z} = 1/z$?

52. Let z be a point on the line $x + y = 1$.

(a) Plot z^2 for at least 5 points on the line.

(b) Find the equation for the curve in rectangular coordinates u and v.

(c) What type of curve is the curve in (b)?

(See Exercise 50.)

53. Let $z = \frac{1}{2} + \frac{i}{2}$.

(a) Sketch the numbers z^n for $n = 1$, 2, 3, 4, and 5.

(b) What happens to z^n as $n \to \infty$?

54. Let $z = 1 + i$.

(a) Sketch the numbers $z^n/n!$ for $n = 1$, 2, 3, 4, and 5.

(b) What happens to $z^n/n!$ as $n \to \infty$?

55.

(a) Graph $r = \cos(\theta)$ in polar coordinates.

(b) Pick five points on the curve in (a). Viewing each as a complex number z, plot z^2.

(c) As z runs through the curve in (a), what curve does z^2 sweep out? (Give its polar equation.)

56. The partial-fraction representation of a rational function is much simpler when we have complex numbers available. No second-degree polynomial $ax^2 + bx + c$ is needed. This exercise indicates why this is the case.
Let z_1 and z_2 be the roots of $ax^2 + bx + c = 0$, $a \neq 0$.

(a) Using the quadratic formula (or by other means), show that $z_1 + z_2 = -b/a$ and $z_1 z_2 = c/a$.

(b) From (a) deduce that

$$ax^2 + bx + c = a(x - z_1)(x - z_2).$$

(c) With the aid of (b) show that

$$\frac{1}{ax^2 + bx + c} = \frac{1}{a(z_1 - z_2)} \left(\frac{1}{x - z_1} - \frac{1}{x - z_2} \right).$$

Part (c) shows that the theory of partial fractions, described in Section 8.4, becomes much simpler when complex numbers are allowed as the coefficients of the polynomials. Only partial fractions of the form $k/(ax + b)^n$ are needed.

57. Let $f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4$, where each coefficient is real.

(a) Show that if c is a root of $f(x) = 0$, then so is $\overline{c}$.

(b) Show that if c is a root of f and is not real, then $(x - c)(x - \overline{c})$ divides $f(x)$.

(c) Using the fundamental theorem of algebra, show that any fourth-degree polynomial with real coefficients can be expressed as the product of polynomials of degree at most 2 with real coefficients.

Exercise 58 is related to Exercise 90 on page 780. (See also Exercises 6 and 7 on page 1084.)

58. Let a point $\mathcal{O}$ be a distance $a \neq 1$ from the center of a unit circle.

(a) Show that the average value of the (natural) logarithm of the distance from $\mathcal{O}$ to points on the circumference is

$$\frac{1}{2\pi} \int_0^{2\pi} \frac{1}{2} \ln(1 + a^2 - 2a\cos(\theta))\, d\theta.$$

(b) Spend at least three minutes, but at most five minutes, trying to evaluate the integral in (a).

12.6 The Relation Between Exponential and Trigonometric Functions

With the aid of complex numbers Leonard Euler discovered in 1743 that the trigonometric functions can be expressed in terms of the exponential function e^z, where z is complex. This section retraces his discovery. In particular, it will be shown that

Expressing $\sin(x)$ and $\cos(x)$ in terms of the exponential function.

$$e^{i\theta} = \cos(\theta) + i\sin(\theta), \qquad \cos(\theta) = \frac{e^{i\theta}+e^{-i\theta}}{2}, \quad \text{and} \quad \sin(\theta) = \frac{e^{i\theta}-e^{-i\theta}}{2i}.$$

Complex Series

In order to relate the exponential function to the trigonometric functions, we will use infinite series such as $\sum_{k=0}^{\infty} z_k$, where the z_k's are complex numbers. Such a series is said to converge to S if its n^{th} partial sum S_n approaches S in the sense that $|S - S_n| \to 0$ as $n \to \infty$. It is shown in Exercise 37 that if $\sum_{k=0}^{\infty} |z_k|$ (a series with real-valued terms) converges, so does $\sum_{k=0}^{\infty} z_k$, and the series $\sum_{k=0}^{\infty} z_k$ is said to *converge absolutely*. If a series converges absolutely, we may rearrange the terms in any order without changing the sum.

Here, $|\cdot|$ refers to the magnitude of a complex number.

Let $z_k = x_k + iy_k$, where x_k and y_k are real. If $\sum_{k=0}^{\infty} z_k$ converges, so do $\sum_{k=0}^{\infty} x_k$ and $\sum_{k=0}^{\infty} y_k$. If $\sum_{k=0}^{\infty} z_k = S = a + bi$, then $\sum_{k=0}^{\infty} x_k = a$ and $\sum_{k=0}^{\infty} y_k = b$. $\sum_{k=0}^{\infty} x_k$ is called the *real part* of $\sum_{k=0}^{\infty} z_k$ and $\sum_{k=0}^{\infty} y_k$ is the *imaginary part* of $\sum_{k=0}^{\infty} z_k$.

$\mathrm{Re}\left(\sum_{k=0}^{\infty} z_k\right) = \sum_{k=0}^{\infty} x_k$
$\mathrm{Im}\left(\sum_{k=0}^{\infty} z_k\right) = \sum_{k=0}^{\infty} y_k$

EXAMPLE 1 Determine for which complex numbers z, $\sum_{k=0}^{\infty} z^k/k!$ converges.

SOLUTION We will examine absolute convergence, that is, the convergence of $\sum_{k=0}^{\infty} |z^k|/k!$. This series has real terms. In fact, it is the Maclaurin series for $e^{|z|}$, which converges for all real numbers. Since $\sum_{k=0}^{\infty} z^n/n!$ converges absolutely for all z, it converges for all z. ◇

$|z|$ is a real number

Defining e^z

The Maclaurin series for e^x when x is real suggests the following definition:

> **DEFINITION** *(e^z for complex z.)* Let z be a complex number. Define e^z to be the sum of the convergent series $\sum_{k=0}^{\infty} z^k/k!$.

In some treatments of exponentials e^z is defined as a power series and e is defined as the value of the series when $z = 1$.

Observe that when z happens to be real, $z = x$, e^z is our familiar real-valued exponential function, e^x. It can be shown by multiplying the series for e^{z_1} and e^{z_2} that $e^{z_1+z_2} = e^{z_1}e^{z_2}$ in accordance with the basic law of exponents.

When the expression for z is complicated, we sometimes write e^z as $\exp(z)$. For example, in exp notation the law of exponents becomes $\exp(z_1 + z_2) = (\exp(z_1))(\exp(z_2))$.

Euler's Formula: The Link between $e^{i\theta}$, $\cos(\theta)$, and $\sin(\theta)$

The following theorem of Euler provides the key link between the exponential function e^z and the trigonometric functions $\cos(\theta)$ and $\sin(\theta)$.

Euler's Formula

Theorem 12.6.1 (Euler's Formula). *Let θ be a real number. Then*

$$e^{i\theta} = \cos(\theta) + i\sin(\theta).$$

Proof

By definition of e^z for any complex number,

$$\begin{aligned} e^{i\theta} &= 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \cdots. \\ &= 1 + i\theta + \frac{i^2\theta^2}{2!} + \frac{i^3\theta^3}{3!} + \frac{i^4\theta^4}{4!} + \cdots. \\ &= 1 + i\theta - \frac{\theta^2}{2!} - \frac{i\theta^3}{3!} + \frac{\theta^4}{4!} + \cdots. \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots\right) + i\left(\theta - \frac{\theta^3}{3!} + \cdots\right) \quad \text{(rearranging)} \\ &= \cos(\theta) + i\sin(\theta). \end{aligned}$$

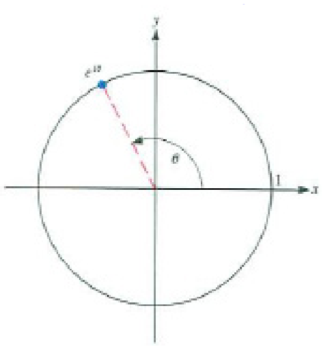


Figure 12.6.1

Figure 12.6.1 shows $e^{i\theta}$, which lies on the standard unit circle. •

Theorem 12.6.1 asserts, for instance, that

$$e^{i\pi} = \cos(\pi) + i\sin(\pi) = -1 + i \cdot 0 = -1.$$

The equation $e^{i\pi} = -1$ is remarkable in that it links e (the fundamental number in calculus), π (the fundamental number in trigonometry), i (the fundamental complex number), and the negative number -1. The history of that short equation would recall the struggles of hundreds of mathematicians to create the number system that we now take for granted. It is as important in mathematics as $F = ma$ or $E = mc^2$ in physics.

There is an old saying: "God created the complex numbers; anything less is the work of man."

With the aid of Theorem 12.6.1, both $\cos(\theta)$ and $\sin(\theta)$ may be expressed in terms of the exponential function.

Theorem 12.6.2. *Let θ be a real number. Then*

$$\boxed{\cos(\theta) = \frac{e^{i\theta}+e^{-i\theta}}{2} \quad \text{and} \quad \sin(\theta) = \frac{e^{i\theta}-e^{-i\theta}}{2i}}$$

Figure 12.6.2 The license plate of mathematician Martin Davis, whose e-mail signature is "eipye, add one, get zero."

Proof
We begin with Euler's formula (Theorem 12.6.1),

$$e^{i\theta} = \cos(\theta) + i\sin(\theta). \tag{12.6.1}$$

Replacing θ by $-\theta$ in (12.6.1), we obtain

$$e^{-i\theta} = \cos(\theta) - i\sin(\theta). \tag{12.6.2}$$

The sum of (12.6.1) and (12.6.2) yields

$$e^{i\theta} + e^{-i\theta} = 2\cos(\theta),$$

hence

$$\cos(\theta) = \frac{e^{i\theta}+e^{-i\theta}}{2}.$$

Subtraction of (12.6.2) from (12.6.1) yields

$$e^{i\theta} - e^{-i\theta} = 2i\sin(\theta),$$

hence

$$\sin(\theta) = \frac{e^{i\theta}-e^{-i\theta}}{2i}.$$

This establishes the two equations in this theorem. •

The hyperbolic functions $\cosh(x)$ and $\sinh(x)$ were defined in terms of the exponential function by

$$\cosh(x) = \frac{e^x+e^{-x}}{2} \quad \text{and} \quad \sinh(x) = \frac{e^x-e^{-x}}{2}.$$

sinh and cosh were defined in Section 4.1, see Exercises 49 to 52 on page 301.

Theorem 12.6.2 shows the trigonometric functions could be similarly defined in terms of the exponential function — if complex numbers were available. This means one could bypass right triangles and unit circles when defining $\sin(\theta)$ and $\cos(\theta)$.

Just as Maxwell discovered the connection between light and electricity, Euler discovered the connection between the exponential and trigonometric functions.

Indeed, from the complex numbers and e^z we could even obtain the derivative formulas for $\sin(\theta)$ and $\cos(\theta)$. For instance,

$$\frac{d}{d\theta}\sin(\theta) = \left(\frac{e^{i\theta}-e^{-i\theta}}{2i}\right)' = \frac{ie^{i\theta}+ie^{-i\theta}}{2i} = \frac{e^{i\theta}+e^{-i\theta}}{2} = \cos(\theta).$$

(That the familiar rules for differentiation extend to complex-valued functions is justified in a course in complex variables.)

Sketching e^z

Magnitude and argument of e^{x+iy}

If $z = x + iy$, the evaluation of e^z can be carried out as follows:

$$e^z = e^{x+iy} = e^x e^{iy} = e^x(\cos(y) + i\sin(y)).$$

The magnitude of e^{x+iy} is e^x and the argument of e^{x+iy} is y.

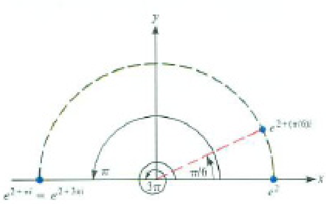


Figure 12.6.3

EXAMPLE 2 Compute and sketch (a) $e^{2+(\pi/6)i}$, (b) $e^{2+\pi i}$, and (c) $e^{2+3\pi i}$.
SOLUTION (a) $e^{2+(\pi/6)i}$ has magnitude e^2 and argument $\pi/6$. (b) $e^{2+\pi i}$ has magnitude e^2 and argument π; it equals $-e^2$. (c) $e^{2+3\pi i}$ has magnitude e^2 and argument 3π, so is the same number as the number in (b). The results are sketched in Figure 12.6.3. ◇

The next example illustrates a typical computation in alternating currents. Electrical engineers frequently use j as the symbol for i (so they can use i to represent current).

EXAMPLE 3 Find the real part of $100e^{j(\pi/6)}e^{j\omega t}$. Here t refers to time, ω is a real constant related to frequency, and j is the mathematician's i.
SOLUTION

$$\begin{aligned} 100e^{j(\pi/6)}e^{j\omega t} &= 100e^{j(\pi/6)+j\omega t} \\ &= 100e^{j(\pi/6+\omega t)} \\ &= 100\left(\cos\left(\frac{\pi}{6}+\omega t\right) + i\sin\left(\frac{\pi}{6}+\omega t\right)\right). \end{aligned}$$

Thus

$$\operatorname{Re}\left(100e^{j(\pi/6)}e^{j\omega t}\right) = 100\cos\left(\frac{\pi}{6}+\omega t\right).$$

◇

It is sometimes convenient to think of $\cos(\theta)$ as $\operatorname{Re}\left(e^{i\theta}\right)$. The next example exploits this point of view.

EXAMPLE 4 Evaluate $\sum_{k=0}^{\infty} \frac{\cos(k\theta)}{2^k}$.
SOLUTION Recall that $e^{ik\theta} = \cos(k\theta) + i\sin(k\theta)$. Hence $\cos(k\theta) = \operatorname{Re}\left(e^{ik\theta}\right)$, and we have

Recall the definition of e^z.

$$\sum_{k=0}^{\infty} \frac{\cos(k\theta)}{2^k} = \sum_{k=0}^{\infty} \operatorname{Re}\left(\frac{e^{ik\theta}}{2^k}\right) = \operatorname{Re}\left(\sum_{k=0}^{\infty} \frac{e^{ik\theta}}{2^k}\right) \qquad (12.6.3)$$

To simplify the complex-valued expression inside the parentheses, notice that

$$\frac{e^{ik\theta}}{2^k} = \left(\frac{e^{i\theta}}{2}\right)^k.$$

Now, because $\left|e^{i\theta}/2\right| = 1/2 < 1$, this "geometric" series converges with sum

$$\frac{1}{1-\left(\frac{e^{i\theta}}{2}\right)} = \frac{2}{2-\cos(\theta)-i\sin(\theta)} = \frac{2(2-\cos(\theta)+i\sin(\theta))}{(2-\cos(\theta))^2+(\sin(\theta))^2}. \qquad (12.6.4)$$

Inserting (12.6.4) as the sum of the series in (12.6.3) gives

$$\sum_{k=0}^{\infty}\frac{\cos(k\theta)}{2^k} = \mathrm{Re}\left(\sum_{k=0}^{\infty}\frac{e^{ik\theta}}{2^k}\right) = \mathrm{Re}\left(\frac{2(2-\cos(\theta)+i\sin(\theta))}{5-4\cos(\theta)}\right) = \frac{2(2-\cos(\theta))}{5-4\cos(\theta)}.$$

◇

Summary

Using power series, we obtained the fundamental relation $e^{i\theta} = \cos(\theta)+i\sin(\theta)$ and showed that $\cos(\theta)$ and $\sin(\theta)$ can be expressed in terms of the exponential function. Since $\ln(x)$ is the inverse of e^x, it too is obtained from the exponential function. We may define even x^n, $x > 0$, in terms of the exponential function as $e^{n\ln(x)}$. Similarly, a^x, $a > 0$, can be defined as $e^{x\ln(a)}$. These observations suggest that the most fundamental function in calculus is e^x, where x is real or complex.

EXERCISES for Section 12.6

In Exercises 1 to 6 sketch the numbers given and state their real and imaginary parts.

1. $e^{5\pi i/4}$

2. $5e^{\pi i/4}$

3. $2e^{\pi i/4} + 3e^{\pi i/6}$

4. e^{2+3i}

5. $e^{\pi i/6}e^{3\pi i/4}$

6. $2e^{\pi i} \cdot 3e^{-\pi i/3}$

In Exercises 7 to 10 express the given numbers in the form $re^{i\theta}$ for a positive real number r and argument θ, where $-\pi < \theta \leq \pi$.

7. $\frac{e^2}{\sqrt{2}} - \frac{e^2}{\sqrt{2}}i$

8. $3\left(\cos\left(\frac{\pi}{4}\right) + i\sin\left(\frac{\pi}{4}\right)\right)$

9. $5\left(\cos\left(\frac{\pi}{6}\right) + i\sin\left(\frac{\pi}{6}\right)\right) \cdot 3\left(\cos\left(\frac{\pi}{2}\right) + i\sin\left(\frac{\pi}{2}\right)\right)$

10. $7\left(\cos\left(\frac{7\pi}{3}\right) + i\sin\left(\frac{7\pi}{3}\right)\right)$

In Exercises 11 to 14 plot $\exp(z)$ for the given values of z:

11. $z = 2$

12. $\pi i/2$

13. $2 - \pi i/3$

14. $-1 + 17\pi i/6$

In Exercises 15 to 18 plot the given complex numbers:

15. $\exp(\pi i/4 + 3\pi i)$

16. $\exp(1 + 9\pi i/4$

17. $\exp(2 - \pi i/3)$

18. $\exp(-1 + 17\pi i/6)$

19. Let $z = e^{a+bi}$. Find (a) $|z|$, (b) $\bar{z}$, (c) z^{-1}, (d) $\mathrm{Re}(z)$, (e) $\mathrm{Im}(z)$, and (f) $\arg(z)$. In (f), assume a and b are positive.

20. How far is $\exp(x + iy)$ from the origin?

21. How far is $\exp(x + iy)$ from the x-axis? From the y-axis?

22. For which values of a and b is $\lim_{n\to\infty}\left(e^{a+ib}\right)^n = 0$?

23. Find all complex numbers z such that $e^z = 1$.

24. Find all complex numbers z such that $e^z = -1$. Because there is no real number r such that $e^r = -1$, the number -1 has no real logarithm. However, it has an infinite number of complex logarithms, all situated on the y-axis.

25.

(a) Find $\left|e^{3+4i}\right|$.

(b) Plot the complex number e^{3+4i}.

26.

(a) Plot all complex numbers of the form e^{x+4i}, x real.

(b) Plot all complex numbers of the form e^{3+yi}, y real.

27. If z lies on the line $y = 1$, where does $\exp(z)$ lie?

28. If z lies on the line $x = 1$, where does $\exp(z)$ lie?

29. In Claude Garrod's *Twentieth Century Physics*, Faculty Publishing, Davis, Calif., p. 107, there is the remark: "Using the fact that

$$\left(e^{-i\omega_0 t}\right)^* \left(e^{-i\omega_0 t}\right) = 1,$$

we can easily evaluate the probability density for these standard waves." Justify this equation. In this text, z^* denotes the conjugate of z and ω_0 is real.

30. Use the fact that $1 + \cos(\theta) + \cos(2\theta) + \cdots + \cos((n-1)\theta)$ is the real part of $1 + e^{\theta i} + e^{2\theta i} + \cdots + e^{(n-1)\theta i}$ to find a short formula for that trigonometric sum.

31. Find all z such that $e^z = 3 + 4i$.

32. Assuming that $e^{z_1+z_2} = e^{z_1}e^{z_2}$ for complex numbers z_1 and z_2, obtain the trigonometric identities for $\cos(A + B)$ and $\sin(A + B)$.

33. Evaluate

$$\sum_{k=0}^{\infty} \frac{\cos(k\theta)}{k!}.$$

First, show that the series converges (absolutely).

34. Evaluate
$$\sum_{k=0}^{\infty} \frac{\sin(k\theta)}{k!}.$$
First, show that the series converges.

35. Evaluate
$$\sum_{k=1}^{\infty} \frac{\sin(k\theta)}{k}.$$
First, show that the series converges.

36. Evaluate
$$\sum_{k=1}^{\infty} \frac{\cos(k\theta)}{k}.$$
First, show that the series converges (absolutely).

37. This problem shows that if $\sum_{k=0}^{\infty} |z_k|$ converges, so does $\sum_{k=0}^{\infty} z_k$. Let $z_k = x_k + iy_k$ and assume that $\sum_{k=0}^{\infty} |z_k|$ converges.

(a) Show that $\sum_{k=0}^{\infty} |x_k|$ and $\sum_{k=0}^{\infty} |x_{k+1}|$ both converge. ($|a| \leq \sqrt{a^2 + b^2}$)

(b) Show that $\sum_{k=0}^{\infty} x_k$ and $\sum_{k=0}^{\infty} y_k$ both converge.

(c) Show that $\sum_{k=0}^{\infty} (x_k + iy_k)$ converges.

38. Let $f(z)$ be a polynomial with real coefficients.

(a) Show that if $f(a) = 0$, then $f(\overline{a}) = 0$. (This shows that roots of f occur in conjugate pairs.)

(b) Show that $\overline{e^z} = e^{\overline{z}}$.

(c) Show that $\overline{\sin(z)} = \sin(\overline{z})$.

39. When z is real, $|\sin(z)| \leq 1$ and $|\cos(z)| \leq 1$. Do these inequalities hold for all complex z?

40. Does the equation $\cos^2(z) + \sin^2(z) = 1$ hold for complex z?

41. Let
$$z = \frac{1+i}{\sqrt{2}}.$$

(a) Plot z, $z^2/2!$, $z^3/3!$, and $z^4/4!$.

(b) Plot $1 + z + z^2/2! + z^3/3! + z^4/4!$, which is an estimate for $\exp\left((1+i)/\sqrt{2}\right)$.

(c) Plot $\exp\left((1+i)/\sqrt{2}\right)$ on the xy plane.

42. An integral table lists $\int xe^{ax}\,dx = e^{ax}\left(ax-1\right)/a^2$. At first glance, finding $\int xe^{ax}\cos(bx)\,dx$ may appear to be a much harder problem. However, by noticing that $\cos(bx) = \mathrm{Re}\left(e^{ibx}\right)$, we can reduce it to a simpler problem. Following this approach, find $\int xe^{ax}\cos(bx)\,dx$. (The formula for $\int xe^{ax}\,dx$ holds when a is complex.)

43. In Section 4.1 we define $\cosh(x) = \left(e^x + e^{-x}\right)/2$ and $\sinh(x) = \left(e^x - e^{-x}\right)/2$. We can use the same definitions when x is complex. In view of Theorem 12.6.2, let us define sine and cosine for complex z by $\sin(z) = \left(e^{iz} - e^{-iz}\right)/(2i)$ and $\cos(z) = \left(e^{iz} + e^{-iz}\right)/2$. Establish the following links between the hyperbolic and trigonometric functions:

(a) $\cosh(z) = \cos(iz)$

(b) $\sinh(z) = -i\sin(iz)$

44. Show that

(a) $\sin(z) = i\sinh(iz)$.

(b) $\cos(z) = \cosh(iz)$.

(c) $\cosh(z)^2 - \sinh(z)^2 = 1$

45. Sam is at it again: "I don't need power series to define e^z. I just write z as $x + iy$ and define e^{x+iy} to be $e^x(\cos(y) + i\sin(y))$. That's all there is to it. If I call this function $E(z)$, then it's easy to check that $E(z_1 + z_2) = E(z_1)E(z_2)$. Moreover, if z is real, then $y = 0$ and $E(z) = e^x$, agreeing with our familiar $\exp(x)$."

(a) Is Sam right?

(b) Does his $E(z)$ obey the basic law of exponents, as he claims?

(c) Jane asks him, "But where did you get the idea for that definition? It seems to float in out of thin air." What is Sam's answer?

46.

Sam: I can show that $e^{i\theta} = \cos(\theta) + i\sin(\theta)$ without using Taylor series.

Jane: That would be nice.

Sam: I differentiate the quotient $e^{i\theta}/(\cos(\theta)+i\sin(\theta))$ and get 0. So it's a constant. Then it's easy to show the constant is 1. That does it.

Jane: But you used that the derivative of e^z is e^z.

Sam: I did, but that follows from the definition of e^z as $\sum_{k=0}^{\infty} z^k/k!$, the only power series needed.

Jane: You may be right, but once again why did you think of $\cos(\theta) + i\sin(\theta)$?

Check Sam's calculations. Is his reasoning correct?

47. For which z is

(a) $e^z = e^{-z}$,

(b) $e^{iz} = e^{-ix}$

(c) $\sin(z) = 0$.

48. Let z be a complex number and θ a real number. What is the geometric relationship between z and $e^{i\theta}z$? Experiment, conjecture, and explain.

12.7 Fourier Series

In Section 12.1 we used sums of terms of the form ax^n, where n is a non-negative integer and a is a number, to represent a function. This required a function to have derivatives of all orders. Now, instead, we will use sums of terms of the form $a\cos(kx)$ and $b\sin(kx)$, where a, b, and k are numbers. This method applies to a much broader class of functions, even, for instance, the absolute value function, $f(x) = |x|$, which is not differentiable at 0, and some functions that are not even continuous. The technique, called **Fourier Series**, is used in such varied fields as heat conduction, electric circuits, the theory of sound and mechanical vibrations.

At first glance, the use of sine and cosine, which are periodic functions, may seem a surprising choice. However, if you think in terms of sound, it is quite plausible. Every tuning fork produces a pure pitch at a specific frequency. With a collection of such devices, each at a different pitch, struck simultaneously, you can approximate the sound made by a band or an orchestra. Each tuning fork corresponds to $\sin(kt)$ or $\cos(kt)$, where t is time. The one set at concert A vibrates at the rate of 440 cycles per second, that is, 440 Hertz (440 Hz). In this case the acoustic wave is expressed as $\sin(400\,(2\pi t))$, for, as t increases by $1/400$ second, the argument $400\,(2\pi t)$ increases by 2π, enabling the function to complete one cycle.

To listen to several tuning forks, go to `http://www.onlinetuningfork.com/`.

Periodic Functions

The function $\cos(x)$ (and $\sin(x)$) has period 2π, that is, $\cos(x+2\pi) = \cos(x)$. Changing the input by 2π does not change the output. It follows that $\cos(x-2\pi) = \cos(x)$, $\cos(x+4\pi) = \cos(x)$. Moreover, for any integer n, $\cos(x)$ has $n(2\pi)$ as a period. A function's **natural period**, also called the **period** is its shortest period. When we say "$\cos(x)$ has period 2π" we are stating that the natural period of $\cos(x)$ is 2π.

EXAMPLE 1 Find the period of (a) $\cos(3\pi x)$, (b) $\cos(k\pi x/L)$, where k is a positive integer and L is a positive number.

SOLUTION In each case we ask, "How much must x change in order for the argument (the input) to change by 2π?"

(a) For $3\pi x$ to change by 2π, we solve the equation $3\pi x = 2\pi$, obtaining $x = 2/3$. Thus $\cos(3\pi x)$ has period $2/3$.

(b) For $\cos(k\pi x/L)$ the reasoning used in (a) leads us to conclude the period is $2L/k$.

Note that in (b) the larger L is, the longer the period. Also, the larger k is, the shorter the period. For each k, $2L$ is among its periods. ◇

Fourier Series for Functions with Period 2π

We first treat the familiar case of functions that have period 2π. Then we consider the general case, where the period is $2L$, for any positive number L.

The formulas for a_k and b_k are known as "**Euler's formulas**." Euler published them in 1777, but Fourier was unaware of them.

Let $f(x)$ have period 2π. Its values are determined by its values on any interval of length 2π. We choose the interval $(-\pi, \pi]$ rather than $[0, 2\pi)$ to simplify some computations that we will encounter momentarily.

Let $f(x)$ be a function of period 2π. The Fourier Series associated with this function is

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos(kx) + b_k \sin(kx)\right) \qquad (12.7.1)$$

Note that the formula for a_k includes the case for a_0.

where

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(kx)\ dx \qquad k = 0, 1, 2, \dots \qquad (12.7.2)$$

$$b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(kx)\ dx \qquad k = 1, 2, \dots. \qquad (12.7.3)$$

(This assumes the integrals in (12.7.2) and (12.7.3) exist.)

After we compute two Fourier series, we will show why the coefficients are given by the integrals in (12.7.2) and (12.7.3).

Constant term is $a_0/2$

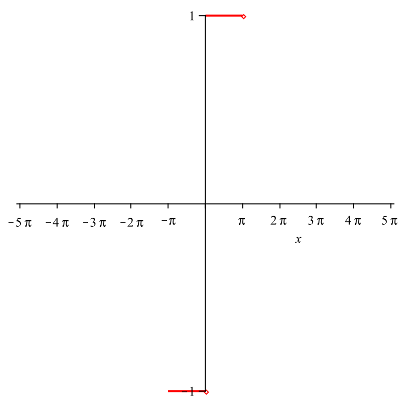
The numbers a_k and b_k are called the **Fourier coefficients** for $f(x)$. The formula for a_0 reduces to $a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\ dx$. This means that the constant term $a_0/2$ is the average value of the function $f(x)$ over one period. Note that the formula for a_k in (12.7.2) also holds for $k = 0$ because the constant term in (12.7.1) is $a_0/2$. (The 2 was included in (12.7.1) so (12.7.2) would hold when $k = 0$.)

EXAMPLE 2 Find the Fourier series associated with the function defined by

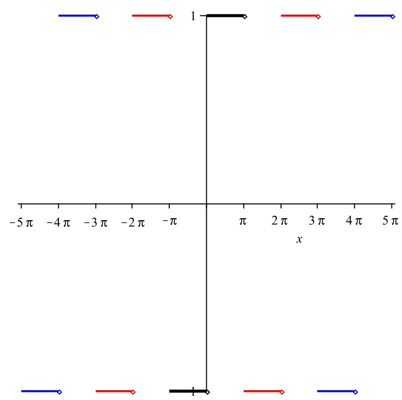
$$f(x) = \begin{cases} -1 & -\pi < x \le 0 \\ 1 & 0 < x \le \pi. \end{cases}$$

Because $f(x)$ is (almost) an odd function, we expect only sines to appear in its Fourier series.

To make $f(x)$ have period 2π, just repeat the graph on every interval of the form $[-\pi + 2n\pi, \pi + 2n\pi)$. The graph of $f(x)$ is shown in Figure 12.7.1(a) and the extension of $f(x)$ is shown in Figure 12.7.1(b).



(a)



(b)

Figure 12.7.1

SOLUTION

$$\begin{aligned}
a_0 &= \tfrac{1}{\pi}\int_{-\pi}^{\pi} f(x)\ dx \\
&= \tfrac{1}{\pi}\int_{-\pi}^{0} f(x)\ dx + \tfrac{1}{\pi}\int_{0}^{\pi} f(x)\ dx \\
&= \tfrac{1}{\pi}\int_{-\pi}^{0} -1\ dx + \tfrac{1}{\pi}\int_{0}^{\pi} 1\ dx \\
&= \tfrac{1}{\pi}(-\pi) + \tfrac{1}{\pi}(\pi) = 0.
\end{aligned}$$

Similarly, for $k \geq 1$,

$$\begin{aligned}
a_k &= \tfrac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(kx)\ dx \\
&= \tfrac{1}{\pi}\int_{-\pi}^{0} f(x)\cos(kx)\ dx + \tfrac{1}{\pi}\int_{0}^{\pi} f(x)\cos(kx)\ dx \\
&= \tfrac{1}{\pi}\int_{-\pi}^{0}(-\cos(kx))\ dx + \tfrac{1}{\pi}\int_{0}^{\pi}\cos(kx)\ dx \\
&= \tfrac{1}{\pi}\ \frac{-\sin(kx)}{k}\bigg|_{-\pi}^{0} + \tfrac{1}{\pi}\ \frac{\sin(kx)}{k}\bigg|_{0}^{\pi} = 0 + 0 = 0
\end{aligned}$$

and

$$\begin{aligned}
b_k &= \tfrac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(kx)\ dx \\
&= \tfrac{1}{\pi}\int_{-\pi}^{0} f(x)\sin(kx)\ dx + \tfrac{1}{\pi}\int_{0}^{\pi} f(x)\sin(kx)\ dx \\
&= \tfrac{1}{\pi}\int_{-\pi}^{0}(-\sin(kx))\ dx + \tfrac{1}{\pi}\int_{0}^{\pi}\sin(kx)\ dx \\
&= \tfrac{1}{\pi}\ \frac{\cos(kx)}{k}\bigg|_{-\pi}^{0} + \tfrac{1}{\pi}\ \frac{-\cos(kx)}{k}\bigg|_{0}^{\pi} = \tfrac{1}{\pi}\left(\frac{1-\cos(-k\pi)}{k}\right) + \tfrac{1}{\pi}\left(\frac{-\cos(k\pi)+1}{k}\right)
\end{aligned}$$

Because $\cos(-k\pi) = \cos(k\pi)$, we have

$$b_k = \tfrac{1}{k\pi}\left((1-\cos(k\pi)) + (1-\cos(k\pi))\right) = \frac{2(1-\cos(k\pi))}{k\pi}.$$

When k is even, $1-\cos(k\pi) = 1-1 = 0$. And, when k is odd, $1-\cos(k\pi) = 1-(-1) = 2$. Thus

$$b_k = \begin{cases} 0 & \text{when } k \text{ is even} \\ \frac{4}{k\pi} & \text{when } k \text{ is odd.} \end{cases}$$

The Fourier Series (12.7.1) in this case has only terms involving $\sin(kx)$

with k odd. It is

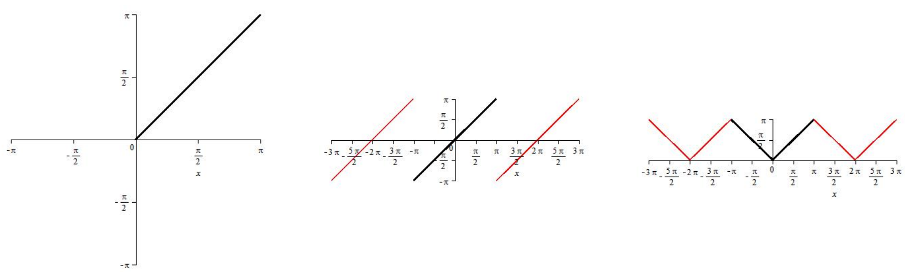
$$\frac{4}{\pi}\sin(x) + \frac{4}{3\pi}\sin(3x) + \frac{4}{5\pi}\sin(5x) + \dots .$$

In particular, when $x = \pi/2$, $f(x) = 1$ and we have

$$\begin{array}{lrcl} & 1 & = & \frac{4}{\pi}\sin\left(\frac{(\pi}{2}\right) + \frac{4}{3\pi}\sin\left(\frac{3\pi}{2}\right) + \frac{4}{5\pi}\sin\left(\frac{5\pi}{2}\right) + \dots \\ & 1 & = & \frac{4}{\pi} - \frac{4}{3\pi} + \frac{4}{5\pi} - \dots \\ \text{Thus} & \frac{\pi}{4} & = & 1 - \frac{1}{3} + \frac{1}{5} - \dots . \end{array}$$

This result was obtained previously in Exercise 22 in Section 12.4 with the aid of the Maclaurin series for $\arctan(x)$. $\diamond$

The fact that the function $f(x)$ in Example 2 is defined on a full period is quite convenient. In many applications the function is given only on one half of the period. For example, $f(x) = x$ for $0 \le x < \pi$ (see Figure 12.7.2(a)). Because $f(x)$ is not periodic, the first step is to replace $f(x)$ with a function $g(x)$ that has period 2π and coincides with $f(x)$ on its domain, that is, on $[0, \pi)$. Two possible periodic extensions of $f(x)$ are shown in Figure 12.7.2(b) and (c). Both have period 2π; one is odd, the other even.



(a) (b) (c)

Figure 12.7.2

EXAMPLE 3 Find the Fourier series of the triangular wave with period 2π shown in Figure 12.7.2(c).

SOLUTION Let $T(x)$ denote the triangular wave. To compute the Fourier series of $T(x)$ we need to know the definition of $T(x)$ on an interval with length 2π.

$T(x) = |x|$ for x in $[-\pi, \pi)$

$$T(x) = \begin{cases} x & \text{for } 0 \le x \le \pi \\ -x & \text{for } -\pi\mathcal{L}\,[x] < 0 \end{cases} .$$

If $T(x) = T(-x)$, then $\int_{-\pi}^{\pi} T(x)\,dx = 2\int_0^{\pi} T(x)\,dx$.

Because $T(x)$ is an even function, $b_k = 0$ for $k = 1, 2, \dots$. Then

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} T(x)\,dx = \frac{2}{\pi}\int_0^{\pi\pi} x\,dx = \frac{1}{\pi}\,x^2\Big|_0^{\pi} = \pi.$$

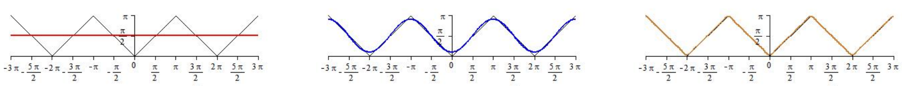
The coefficients of the cosine terms are

$$\begin{aligned}
a_k &= \frac{1}{\pi}\int_{-\pi}^{\pi} T(x)\cos(kx)\ dx = \frac{2}{\pi}\int_{0}^{\pi} x\cos(kx)\ dx && \text{because } T(x)\cos(kx) \text{ is even}\\
&= \frac{2}{\pi}\left(\frac{x}{k}\sin(kx)\Big|_0^{\pi} - \frac{1}{k}\int_0^{\pi}\sin(kx)\ dx\right) && \text{integrate by parts}\\
&= \frac{2}{\pi}\left(0 + \frac{1}{k^2}\cos(kx)\Big|_0^{\pi}\right) && \sin(k\pi)=0 \text{ for all integers } k\\
&= \frac{2}{k^2\pi}(\cos(k\pi)-1) = \frac{2((-1)^k-1)}{k^2}
\end{aligned}$$

When k is an even integer, $a_k = 2((-1)^k - 1)/(k^2\pi) = 0$. And, when k is an odd integer, $a_k = 2((-1)^k - 1)/(k\pi)^2 = -4/(k^2\pi)$.

Then, the Fourier series for the triangular wave is

$$T(x) = \frac{\pi}{2} - \frac{4}{\pi}\left(\cos(x) + \frac{1}{9}\cos(3x) + \frac{1}{25}\cos(5x) + \dots\right). \qquad (12.7.4)$$



(a) (b) (c)

Figure 12.7.3

◇

Figure 12.7.3 shows the partial Fourier sums for the triangular wave with 1, 2, and 5 terms. In an advanced calculus course it is proved that the partial sums converge to the function for every real number. As is easy to check, replacing x by 0 in (12.7.4) shows the sum of the reciprocals of the squares of all the positive odd integers is $\pi^2/8$.

The Origins of the Formulas for a_k and b_k

We will derive the formulas for the Fourier coefficients in the special case when the period is 2π. Exercises 12 and 13 outline the similar argument for the general case when the period is $2L$.

The keys are the following three integrals:

$$\int_{-\pi}^{\pi} \sin(kx)\sin(mx)\,dx = \begin{cases} \pi & \text{if } m = k,\ k = 1, 2, \ldots \\ 0 & \text{if } m \neq k,\ k = 1, 2, \ldots \end{cases}$$

$$\int_{-\pi}^{\pi} \cos(kx)\cos(mx)\,dx = \begin{cases} \pi & \text{if } m = k,\ k = 1, 2, \ldots \\ 0 & \text{if } m \neq k,\ k = 1, 2, \ldots \end{cases}$$

$$\int_{-\pi}^{\pi} \sin(kx)\cos(mx)\,dx = 0 \quad \text{for any } m = 1, 2, \ldots \text{ and any } k = 1, 2, \ldots.$$

The third one is immediate, for the integrand, being the product of an odd function and an even function, is an odd function. The other two depend on trigonometric identities, and were developed in Exercises 17 to 19 inSection 8.5.

The formula for a_m, $m = 1, 2, \ldots$, is found by multiplying $f(x)$ by $\cos(mx)$ and integrating term-by-term over one period of length 2π:

We are assuming it's legal to switch the order, integrate term-by-term, then sum: "$\int_{-\pi}^{\pi}\sum_{k=1}^{\infty} = \sum_{k=1}^{\infty}\int_{-\pi}^{\pi}$".

$$\begin{aligned}
&\int_{-\pi}^{\pi} f(x)\cos(mx)\,dx \\
&= \int_{-\pi}^{\pi} \left(\frac{a_0}{2} + \sum_{k=1}^{\infty}\left(a_k\cos(kx) + b_k\sin(kx)\right)\right)\cos(mx)\,dx \\
&= \frac{a_0}{2}\int_{-\pi}^{\pi}\cos(mx)\,dx \\
&\quad + \sum_{k=1}^{\infty}\left(a_k\int_{-\pi}^{\pi}\cos(kx)\cos(mx)\,dx + b_k\int_{-\pi}^{\pi}\sin(kx)\cos(mx)\,dx\right).
\end{aligned}$$

Each integral in this last expression is zero — except the coefficient of a_m. This gives the equation

$$\int_{-\pi}^{\pi} f(x)\cos(mx)\,dx = a_m\int_{-\pi}^{\pi}\left(\cos(kx)\right)^2\,dx = a_m\pi.$$

Solving for a_m, we find that

$$a_m = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(mx)\,dx.$$

The derivation of the formulas for a_0 and for b_k are similar. (See Exercises 12 and 13.)

Remarks on the Underlying Theory

Just as a Taylor series associated with a function may not represent the function, the Fourier series associated with a function may not represent it, even if the function is continuous. However, there are several theorems that assure us that for many functions met in applications the series does converge to the function. First, a couple of definitions.

Recall that the right-hand limit of $f(x)$ at a is defined as the limit of $f(x)$ as x approaches a through values larger than a, and is denoted $\lim_{x\to a^+} f(x)$. Similarly, the left-hand limit, denoted $\lim_{x\to a^-} f(x)$, is defined as the limit of $f(x)$ as x approaches a through values smaller than a. If both these limits exist at a and are different, we say that the function has a "jump discontinuity at a."

Theorem 12.7.1. *Let $f(x)$ have period $2L$. Assume that in the interval $[-L, L)$ (a) $f(x)$ is differentiable exept at a finite number of points, where there are jump discontinuities, and (b) at L the right-hand limit of $f(x)$ exists and at $-L$ the left-hand limit of $f(x)$ exists. Then,*

I. *if the function is continuous at a, its associated Fourier series converges to $f(a)$.*

II. *if $f(x)$ has a jump discontinuity at a, then the series converges to the average of the left- and right-hand limits at a.*

III. *at the endpoints, L and $-L$, the Fourier series converges to the average of $\lim_{x\to -L^+} f(x)$ and $\lim_{x\to L^-} f(x)$.*

Note that there is no mention of the existence of any second-order, or higher-order derivatives.

The name Joseph Fourier (1768—1830) is attached to trigonometric series because he explored and applied them in his classic Analytic Theory of Heat, published in 1822. He came upon the formulas for the coefficients by an indirect route, starting with the Maclaurin series for $\sin(x)$ and $\cos(x)$. For the details, see Morris Kline's *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, New York, 1972 (especially pages 671– 675), but see further references in its index. In the nineteenth and twentieth centuries mathematicians developed a variety of conditions that implied the series converges to the function. The most recent is due to Lennart Carleson (1928–) in 1966, which settled a famous conjecture.

Summary

While Taylor Series are useful for dealing with a function that is very smooth (having derivatives of all orders), Fourier series can represent a function that is not even continuous. While the coefficients in Taylor series are expressed in terms of derivatives, those in Fourier series are expressed in terms of integrals. Even non-periodic functions can be represented by Fourier series. For instance, to deal with x^2 on, say, $[0, 100)$ just extend its domain to the whole x-axis by defining a function of period 100 that agrees with x^2 on $[0, 100)$.

EXERCISES for Section 12.7

The following table of integrals will be helpful in evaluating some of the integrals in these exercises.

$$\begin{array}{rcl}
\int x\sin(ax)\ dx &=& \frac{1}{a^2}\sin(ax) - \frac{x}{a}\cos(ax) + C \\
\int x\cos(ax)\ dx &=& \frac{1}{a^2}\cos(ax) + \frac{x}{a}\sin(ax) + C \\
\int x^2\sin(ax)\ dx &=& \frac{2}{a^3}\cos(ax) + \frac{2x}{a^2}\sin(ax) - \frac{x^2}{a}\cos(ax) + C \\
\int x^2\cos(ax)\ dx &=& \frac{-2}{a^3}\sin(ax) + \frac{2x}{a^2}\cos(ax) + \frac{x^2}{a}\sin(ax) + C \\
\int \sin(x)\sin(ax)\ dx &=& \frac{1}{2(a-1)}\sin((a-1)x) - \frac{1}{2(a+1)}\sin((a+1)x) + C \\
\int \sin(x)\cos(ax)\ dx &=& \frac{1}{2(a-1)}\cos((a-1)x) - \frac{1}{2(a+1)}\cos((a+1)x) + C \\
\int \cos(x)\sin(ax)\ dx &=& \frac{-1}{2(a-1)}\cos((a-1)x) - \frac{1}{2(a+1)}\cos((a+1)x) + C \\
\int \cos(x)\cos(ax)\ dx &=& \frac{1}{2(a-1)}\sin((a-1)x) + \frac{1}{2(a+1)}\sin((a+1)x) + C \\
\int e^x\sin(ax)\ dx &=& \frac{1}{1+a^2}e^x\sin(ax) - \frac{a}{1+a^2}e^x\cos(ax) + C \\
\int e^x\cos(ax)\ dx &=& \frac{a}{1+a^2}e^x\sin(ax) + \frac{1}{1+a^2}e^x\cos(ax) + C
\end{array}$$

In Exercises 1 to 8 give the period of the function

1. $\tan(x)$

2. $2/\cos^2(x)$

3. $\sin(3x)$

4. $\sin(2\pi x)$

5. $\sin(x/5)$

6. $\cos(2\pi x/5)$

7. $\sin(\pi x/3)$

8. $\sin(x/3)$

9. Let $f(x) = x^2$ for x in $[-\pi, \pi)$ and have period 2π.

(a) Find $f(\pi)$, $f(2\pi)$, $f(3\pi)$, $f(-\pi)$, $f(-2\pi)$, and $f(-3\pi)$.

(b) Graph $f(x)$ for x in $[-4\pi, 4\pi]$.

(c) Why will the Fourier series for $f(x)$ have no sine terms?

(d) Find the Fourier series for $f(x)$.

10. Let $f(x) = -x^2$ for x in $[-\pi, 0)$ and x^2 for x in $[0, \pi)$ and have period 2π.

(a) Find $f(\pi)$, $f(2\pi)$, $f(-\pi)$, and $f(-2\pi)$.

(b) Graph $f(x)$ for x in $[-4\pi, 4\pi]$.

(c) Show that f is "almost" an odd function. For what x is $f(x) \neq -f(x)$?

(d) Show that the Fourier series of $f(x)$ is

$$2\frac{\pi^2-4}{\pi}\sin(x)-\pi\sin(2x)+2\frac{9\pi^2-4}{27\pi}\sin(3x)-\frac{\pi}{2}\sin(4x)+2\frac{25\pi^2-4}{125\pi}\sin(5x)-\frac{\pi}{3}\sin(6x)$$

(e) Why are there no cosine terms in the series?

11. Let $f(x)=x$ for x in $[-\pi,\pi)$ and have period 2π. This function is known as a sawtooth function.

(a) Find $f(\pi)$, $f(2\pi)$, $f(-\pi)$, and $f(-2\pi)$.

(b) Graph $f(x)$ for x in $[-4\pi,4\pi]$.

(c) Show that the Fourier series of $f(x)$ is

$$2\sin(x)-\sin(2x)+\frac{2}{3}\sin(3x)-\frac{1}{2}\sin(4x)+\cdots.$$

(d) What does the series converge to at the discontinuities of $f(x)$?

Exercises 12 and 13 complete the derivation of the Fourier series associated with a function with period 2π. That is, of (12.7.1) with coefficients given by (12.7.2) and (12.7.3).

12. Derive (12.7.2).

13. Derive (12.7.3).

Exercises 14 to 16 develop the formulas for the Fourier Series for a function with period $2L$ (instead of 2π).

14. Show that

$$\int_{-L}^{L}\sin\left(\frac{k\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)\,dx=\begin{cases}L & \text{if } m=k,\ k=1,\,2,\,\ldots\\ 0 & \text{if } m\neq k,\ k=1,\,2,\,\ldots\end{cases}.$$

(Use the trigonometric identity $\sin(u)\sin(v)=\frac{1}{2}\left(\cos(u-v)-\cos(u+v)\right)$.)

15. Show that

$$\int_{-L}^{L}\cos\left(\frac{k\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)\,dx=\begin{cases}L & \text{if } m=k,\ k=1,\,2,\,\ldots\\ 0 & \text{if } m\neq k,\ k=1,\,2,\,\ldots\end{cases}.$$

(Use the trigonometric identity $\cos(u)\cos(v) = \frac{1}{2}\left(\cos(u-v) + \cos(u+v)\right)$.)

16. Show that

$$\int_{-L}^{L} \sin\left(\frac{k\pi x}{L}\right) \cos\left(\frac{m\pi x}{L}\right)\, dx = 0.$$

(While you could use a trigonometric identity, this exercise can be completed at a glance.)

Exercises 17 to 20 explore the fact that any integer multiple of the natural period of a function is also a period of the function.

17. Find the Fourier series of $f(x) = \sin(x)$, viewed as a function of period 2π.

18. Find the Fourier series of $f(x) = \sin(x)$, viewed as a function of period 4π.

19. Find the Fourier series of $f(x) = \cos(2x)$, viewed as a function of period π.

20. Find the Fourier series of $f(x) = \cos(2x)$, viewed as a function of period 4π.

In Exercises 21 to 30, (a) sketch at least two periods of the function, (b) compute the Fourier series of the indicated function, and (c) indicate any points where the function and its Fourier series do not agree. In each case assume the function is periodic.

21. $f(x) = x^2$, $-1 \leq x < 1$ (period 2)

22. $f(x) = x^2$, $-2 \leq x < 2$ (period 4)

23. $f(x) = \begin{cases} 0 & \text{for } -1 \leq x < 0 \\ 1 & \text{for } 0 \leq x < 1 \end{cases}$ (period 2)

24. $f(x) = \begin{cases} 1 & \text{for } -1 \leq x < 0 \\ 0 & \text{for } 0 \leq x < 1 \end{cases}$ (period 2)

25. $f(x) = \begin{cases} 0 & \text{for } -1 \leq x < 0 \\ x & \text{for } 0 \leq x < 1 \end{cases}$ (period 2)

26. $f(x) = \begin{cases} 1 & \text{for } -1 \leq x < 0 \\ x & \text{for } 0 \leq x < 1 \end{cases}$ (period 2)

27. $f(x) = \begin{cases} 0 & \text{for } -\pi \leq x < 0 \\ \sin(x) & \text{for } 0 \leq x < \pi \end{cases}$ (period 2π)

28. $f(x) = \begin{cases} 1 & \text{for } -\pi \leq x < 0 \\ \cos(x) & \text{for } 0 \leq x < \pi \end{cases}$ (period 2π)

29. $f(x) = \begin{cases} 0 & \text{for } -2\pi \leq x < 0 \\ \sin(x) & \text{for } 0 \leq x < 2\pi \end{cases}$ (period 4)

30. $f(x) = \begin{cases} 1 & \text{for } -2\pi \leq x < 0 \\ \cos(x) & \text{for } 0 \leq x < 2\pi \end{cases}$ (period 4)

In Exercises 31 to 36, (a) extend the given function to be an odd periodic function with period $2L$, (b) compute the Fourier series of the function found in (a), (c) graph at least two periods of the first three non-zero terms of the Fourier series found in (b).

31. $f(x) = 1$, $0 \le x < 1$ $(L = 1)$

32. $f(x) = x$, $0 \le x < 1$ $(L = 1)$

33. $f()x) = x^2$, $0 \le x < 1$ $(L = 1)$

34. $f(x) = |x - 1|$, $0 \le x < 2$ $(L = 2)$

35. $f(x) = \sin(x)$, $0 \le x < \pi$ $(L = \pi)$

36. $f(x) = \cos(x)$, $0 \le x < \pi$ $(L = \pi)$

In Exercises 37 to 42, (a) extend the given function to be an even periodic function with period $2L$, (b) compute the Fourier series of the function found in (a), (c) graph at least two periods of the function corresponding to the Fourier series found in (b).

37. $f(x)$ from Exercise 31

38. $f(x)$ from Exercise 32

39. $f(x)$ from Exercise 33

40. $f(x)$ from Exercise 34

41. $f(x)$ from Exercise 35

42. $f(x)$ from Exercise 36

43. Show that any function, $f(x)$, can be written as the sum of an even function (f_{even}) and an odd function (f_{odd}). (Write $f(x) = f_{even}(x) + f_{odd}(x)$. Use the properties of f_{even} and f_{odd} to express $f(-x)$ in terms of $f_{even}(x)$ and $f_{odd}(x)$.)

44. Write each of the following functions as the sum of an even function and an odd function.

(a) $f(x) = x^2 + 2x$

(b) $f(x) = x^3 - 2x$

(c) $f(x) = x^3 + 3x^2 - 2x + 1$

(d) $f(x) = \sin(4x) - 3x^3$

(e) $f(x) = |x| \sin(x)$

(f) $f(x) = |x| \cos(x)$

(g) $f(x) = (\sin(x) + 1)^3$

(h) $f(x) = (\cos(x) + 1)^3$

45. Let $f(x) = x$ for x in $[-1, 1)$ and have period 2. This function is known as a sawtooth function.

(a) Find $f(1)$, $f(2)$, $f(-1)$, and $f(-2)$.

(b) Graph $f(x)$ for x in $[-4, 4]$.

(c) Find the Fourier series of $f(x)$.

(d) Why are there no sine terms in the Fourier series?

(e) What is the average value of $f(x)$ over any interval of length 2π?

(f) What does the series converge to at the jump discontinuities?

(g) How does this Fourier series compare with the one in Exercise 11?

46. In Section 11.6, Example 3, it is claimed that the series

$$\frac{\cos(x)}{1^2} + \frac{\cos(2x)}{2^2} + \frac{\cos(3x)}{3^2} + \cdots + \frac{\cos(kx)}{k^2} + \cdots$$

converges to $\frac{1}{12}(3x^2 - 6\pi x + 2\pi^2)$ for $0 \le x \le 2\pi$. Use Fourier series to verify this claim. (Note that this is a closed interval. What happens at the endpoints?)

47. Let $f(x)$ be a periodic function with period $2L$.

(a) Show that $\int_0^{2L} f(x)\ dx = \int_{-L}^{L} f(x)\ dx$.

(b) Show that $\int_{-2L}^{0} f(x)\ dx = \int_{-L}^{L} f(x)\ dx$.

(c) Show that $\int_a^{a+2L} f(x)\ dx = \int_{-L}^{L} f(x)\ dx$ for any number a.

Just as the complex numbers helped expose a close tie between the exponential and trigonometric functions, they also reveal a relation between power series and Fourier series. Exercise 48 helps to make this connection.

48. A Taylor series $\sum_{k=0}^{\infty} a_k z^k$ does not look like a Fourier series. However, when a_k is written as $b_k + ic_k$ and z is expressed as $r(\cos(\theta) + i\sin(\theta))$, where r is constant, the connection becomes clear. To check that this is so, write the series in the form $A + Bi$ where A and B are real. What two Fourier series appear as the real and imaginary parts arise from these manipulations?

12.S Chapter Summary

The Taylor polynomials first encountered in Section 5.5 suggested the power series associated with a function that has derivatives of all orders at a, namely

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k \qquad (12.S.1)$$

which certainly converges when x is a. It may even converge for other values of x, but not necessarily to $f(x)$. For the common functions e^x, $\sin(x)$, and $\cos(x)$ the corresponding power series does converge to the function for all values of x.

The error in using a front end up through the power $(x-a)^n$ to estimate $f(x)$ is given by Lagrange's formula,

$$f(x)-\sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1} \qquad \text{for some } c \text{ between } x \text{ and } a. \qquad (12.S.2)$$

For some functions, such as $\tan(x)$, it is not easy to find the k^{th} derivative. So, we should be glad that e^x, $\sin(x)$, and $\cos(x)$ have such convenient higher derivatives.

Replace x by $-x^2$.

One can obtain a few terms of the Maclaurin series for $\tan(x)$ by dividing the series for $\sin(x)$ by the series for $\cos(x)$. The series for $1/(1+x^2)$ is easily found by massaging the sum of the geometric series $1/(1-x) = 1+x+x^2+\dots$. Integration of that series yields painlessly the Maclaurin series for $\arctan(x)$.

Each power series $\sum_{k=0}^{\infty} a_k(x-a)^k$ has a radius of convergence, R. For $|x-a|<R$, the series converges absolutely and for $|x-a|>R$ the series does not converge. If it converges for all x, then $R=\infty$. For $|x-a|<R$, one may safely differentiate and integrate a series, producing new series.

Estimating an integrand $f(x)$ by the front end of a power series, we can then estimate $\int_a^b f(x)\,dx$. Also, power series are of use in finding indeterminate limits of the type zero-over-zero. that is, $\lim_{x\to 0} \frac{f(x)}{g(x)}$, where both $\lim_{x\to 0} f(x) = 0$ and $\lim_{x\to 0} g(x) = 0$.

Maclaurin series, combined with complex numbers, exposed a fundamental relation between exponential and trigonometric functions:

$$e^{i\theta} = \cos(\theta) + i\sin(\theta).$$

Other important truths, not covered in this chapter, are revealed with the aid of complex numbers. For instance, if we allow complex coefficients, every polynomial can be written as the product of first-degree polynomials, thus simplifying the partial fractions of Section 8.4. Complex numbers can also help us find the radius of convergence. For instance, what is the radius of

Function	Maclaurin Series	R	How Found?
e^x	$\sum_{k=0}^{\infty} \frac{x^k}{k!}$	∞	Taylor's Theorem
$\sin(x)$	$\sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!}$	∞	Taylor's Theorem
$\cos(x)$	$\sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!}$	∞	Taylor's Theorem
$\frac{1}{1-x}$	$\sum_{k=0}^{\infty} x^k$	1	Geometric Series
$\ln(1+x)$	$\sum_{k=1}^{\infty} (-1)^{k+1} \frac{x^k}{k}$	1	Integrate Geometric Series
$\arctan(x)$	$\sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k+1}$	1	Integrate Geometric Series
$\arcsin(x)$	$x + \frac{1}{2}\frac{x^3}{3} + \frac{1\cdot 3}{2\cdot 4}\frac{x^5}{5}$		
	$+\frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\frac{x^7}{7} + \cdots$	1	Integrate Geometric Series
$(1+x)^r$	$1 + rx + \frac{r(r-1)}{2!}x^2$		
	$+\frac{r(r-1)(r-2)}{3!}x^3 + \cdots$	1	Taylor's Theorem
$\frac{1}{(1-x)^2}$	$\sum_{k=0}^{\infty} kx^{k-1}$	1	Differentiate Geometric Series

Table 12.S.1

convergence of the Taylor series in powers of $x-3$ associated with $1/(1+x^2)$? Answer: it is the distance from the point $(3,0)$ to the nearest complex number at which $1/(1+x^2)$ "blows up," that is, when $1+x^2=0$. This occurs when x is i or $-i$, both of which, by the Pythagorean Theorem, are at a distance $\sqrt{1^2+3^2} = \sqrt{10}$ from $(3,0)$. So, $R = \sqrt{10}$.

The final section introduced Fourier series. In contrast to Taylor series, its coefficients are given by integrals, rather than by derivatives. Consequently, Fourier series apply to a larger class of functions. However, this method applies directly only to periodic functions. In the case of a non-periodic function, one restricts the domain to an interval $(-L, L)$ and extends the function to have period $2L$

SHERMAN: Moved from Chapter 1 Summary - find optimal location in Chapter 12.

Isaac Newton, Area, Logarithms, and Geometric Series

The area under the curve $y = 1/(1+x)$ above the interval $[0, c]$, when c is positive, is approximated by $c - c^2/2 + \cdots \pm c^n/n$. When c is negative, the area above the interval $[c, 0]$ is approximated by $-(c + c^2/2 + c^3/3 + \cdots + c^n/n)$.

When he was a student, Isaac Newton calculated the area under the curve $y = 1/(1+x)$ and above the intervals $[-0.1, 0]$ ($c = -0.1$) and $[0, 0.1]$ ($c = 0.1$) to 53 decimal places. See Figure 12.S.1. In Chapter 12 we will see that the first area equals $(0.1) - \frac{(0.1)^2}{2} + \frac{(0.1)^3}{3} - \frac{(0.1)^4}{4} + \frac{(0.1)^5}{5} - \cdots$. The $\cdots$ at the end mean that the more terms you include in the sum, the closer you get to the exact area. The area above $[0.9, 1]$ is $(0.1) + \frac{(0.1)^2}{2} + \frac{(0.1)^3}{3} + \frac{(0.1)^4}{4} + \frac{(0.1)^5}{5} + \cdots$.

When you examine the manuscript you can follow Isaac's orderly calculations, done with a quill pen, not with a calculator or any other computational aid. (Notice the evidence in the manuscript that he found — and corrected — an error in the value of $(0.1)^{23}/23$.)

In Chapter 6 you will learn that the two areas are $\ln(1+c)$ and $-\ln(1-c)$, respectively. (See Exercises 29 and 30 in Section 6.5.) The connection between the geometric series and logarithms will become clear in Chapter 12. (See Exercise 2 in Section 12.7.)

EXERCISES for 12.S

1. What are the polar coordinates of e^{x+iy}?

Exercise 2 provides additional detail for the historical discussion (see page 1082) about Newton's calculation of the area under a hyperbola to more than 50 decimal places. (See also Exercises 29 and 30 in Section 6.5.)

2. Let c be a positive constant.

(a) Show that the area under the curve $y = 1/(1+x)$ above the interval $[0, c]$ is $-\sum_{k=1}^{\infty} \frac{(-c)^k}{k}$.

(b) Show that the area under the curve $y = 1/(1+x)$ above the interval $[-c, 0]$ is $\sum_{k=1}^{\infty} \frac{c^k}{k}$.

3. As pointed out by Frank Samaniego, the following problem arises in statistics. Let a_1, a_2, a_3, $\ldots$ approach a. Show that $\lim_{n\to\infty} (1 + a_n/n)^n$ is e^a. (Show that its natural natural logarithm approaches a.)

4. The integral $\int_0^{2\pi} \frac{1-\cos(x))}{x}\, dx$ occurs in the theory of antennas.

(a) Show that it is *not* an improper integral.

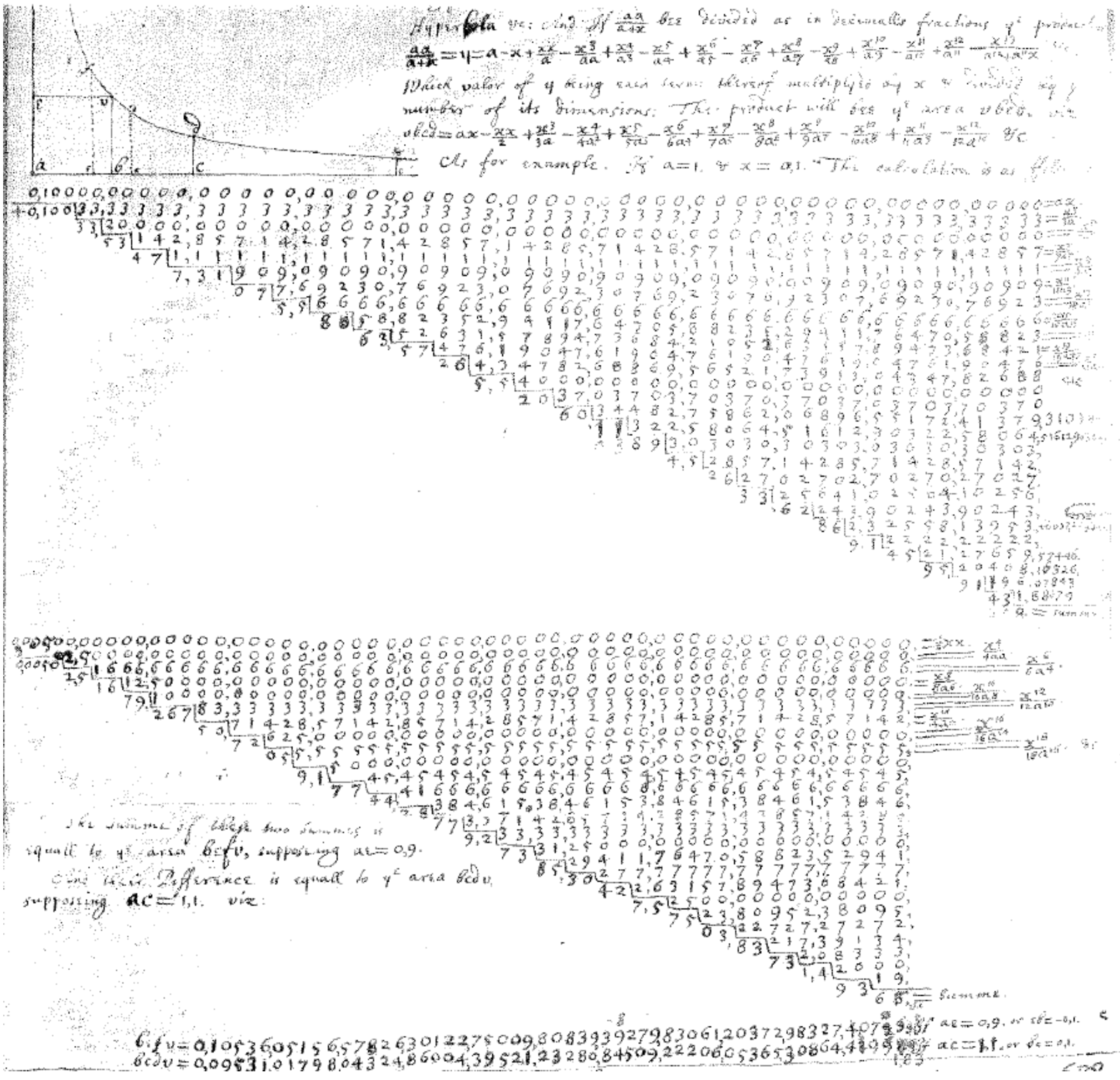


Figure 12.S.1 Excerpt from Isaac Newton's student notebook showing his calculation of the areas under the graph of $y = 1/(1 + x)$ above the intervals $[-0.1, 0.0]$ and $[0.0, 0.1]$. This manuscript is provided courtesy of the Cambridge University Library.

(b) Show that there is a continuous function whose domain is $[0, 2\pi]$ that coincides with the integrand when x is not 0.

(c) The integrand does not have an elementary antiderivative. How many terms are needed in the Maclaurin series for the integrand to obtain an approximation to the integral that is accurate to 3 decimal places?

Exercises 5 to 7 use complex numbers to find the average value of the logarithm of a certain function. Exercise 5 is related to Exercise 90 on page 780 and to Exercise 58 on page 1056.

5. Let a point $\mathcal{O}$ be a distance $a \neq 1$ from the center of a unit circle.

(a) Show that the average value of the (natural) logarithm of the distance from $\mathcal{Q}$ to points on the circumference is

$$\frac{1}{2\pi} \int_0^{2\pi} \frac{1}{2} \ln(1 + a^2 - 2a\cos(\theta))\ d\theta.$$

(b) Spend at least three minutes, but at most 5 minutes, trying to evaluate the integral in (a).

6. This algebraic exercise is needed in Exercise 7. Let $z_0, z_1, \ldots, z_{n-1}$ be the n n^{th} roots of 1. Then it is shown in an algebra course that

$$(z - z_0)(z - z_1)(z - z_2) \cdots (z - z_{n-1}) = z^n - 1.$$

Check that this equation holds when n is (a) 2, (b) 3, (c) 4.

7. Let $z_0, z_1, \ldots, z_{n-1}$ be the n n^{th} roots of 1.

(a) Why is $\displaystyle\frac{1}{n}\sum_{i=0}^{n-1} \ln|a - z_i|$ an estimate of the average distance?

(b) Show that the average in (a) equals

$$\frac{1}{n}\ln|a^n - 1|. \tag{12.S.3}$$

(c) If $0 < a < 1$, show that the limit of (12.S.3) as $n \to \infty$ is 0.

(d) The case when $a = 1$ is not covered by parts (c) and (d). In this case, choose Q to be a point on the unit circle whose polar angle is not a rational multiple of π. (So no z_i coincides with Q.) Then argue as in parts (c) or (e).

(e) If $a > 1$, show that the limit of (12.S.3) as $n \to \infty$ is $\ln(a)$.

(f) Use the results in (c) and (d) to evaluate the integral in Exercise 5(a).

8. Find

(a) $\lim_{x\to\infty} \frac{xe^x}{e^{x^2}}$

(b) $\lim_{x\to 0} \frac{x(e^{\sqrt{x}}-1)}{e^{x^2}-1}$

9. Does $\sum_{n=1}^{\infty} \left(1 - \cos\left(\frac{1}{n}\right)\right)$ converge or diverge? Explain.

10. Assume that $f(x)$ has a continuous fourth derivative. Let M_4 be the maximum of $\left|f^{(4)}(x)\right|$ for x in $[-1, 1]$. Show that

$$\left| \int_{-1}^{1} f(x)\, dx - f\left(\frac{1}{\sqrt{3}}\right) - f\left(\frac{-1}{\sqrt{3}}\right) \right| \le \frac{7M_4}{270}.$$

(Use the representation $f(x) = f(0) + f'(0)x + f''(0)x^2/2 + f^{(3)}(0)x^3/6 + f^{(4)}(c)x^4/24$, where c depends on x.)

11. Justify this statement, found in a biological monograph:

Expanding the equation

$$a \cdot \ln(x+p) + b \cdot \ln(y+q) = M,$$

we obtain

$$a\left(\ln(p) + \frac{x}{p} - \frac{x^2}{2p^2} + \frac{x^3}{3p^3} - \cdots\right) + b\left(\ln(q) + \frac{y}{q} - \frac{y^2}{2q^2} + \frac{y^3}{3q^3} - \cdots\right) = M.$$

12. Estimate $\int_1^3 e^{-x^2}\, dx$ using a Taylor series at $x = 2$ associated with e^{-x^2}.

13. Explain why both $\cos(x)$ and $\sin(x)$ can be expressed in terms of the exponential function e^z.

14. State some of the advantages of complex numbers over real numbers.

15. Why is the "radius of convergence" called "the radius of convergence" rather than the "interval of convergence."

16. Starting with $1 + x + x^2 + \cdots + x^n + \cdots = \frac{1}{1-x}$ obtain the Maclaurin series for

(a) $1/(1-x)^2$

(b) $1/(1+x)$

(c) $\frac{1}{1+x^2}$

(d) $\ln(1+x)$

(e) $\arctan(x)$

17. Find the radius of convergence for each series in Exercise 16

18. Show that each series in Exercise 16 converges to the given function.

19. Sam says, "According to their book, if I multiply the Maclaurin series for e^x by the one for e^{-x} I should get the Maclaurin series for $e^x e^{-x}$, which is just 1. I don't believe that the product could be that simple." Multiply enough terms of the two series to calm Sam down.

20.

(a) Graph the circle $r = \sqrt{2}\cos(\theta)$.

(b) Show that the function $f(z) = z^2$ maps the circle in (a) into the cardioid $r = 1 + \cos(\theta)$.

21. Suppose f is a function with the property that $f^{(n)}(x)$ is "small" in the sense that $|f^{(n)}(x)| \leq |(x+100)^n|$ for all x. Show that the Maclaurin series represents $f(x)$ for all x.

Exercises 22 and 23 treat the complex logarithms of a complex number. They show that $z = \ln(w)$ is not single-valued.

22. Let w be a nonzero complex number. Show that there are an infinite number of complex numbers z such that $e^z = w$. (Use Euler's formula.)

23. (See Exercise 22.) When $e^z = w$, we write $z = \ln(w)$ although $\ln(w)$ is not a uniquely defined number. If b is a nonzero complex number and q is a complex number, define b^q to be $e^{q\ln(b)}$. Since $\ln(b)$ is not unique, b^q is usually not unique. List all possible values of (a) $(-1)^i$, (b) $10^{1/2}$, (c) 10^3,

Calculus is Everywhere # 15

Sparse Traffic

Customers arriving at a checkout counter, cars traveling on a one-way road, raindrops falling on a street and cosmic rays entering the atmosphere all illustrate one mathematical idea — the theory of sparse traffic involving independent events. We will develop the mathematics, which is the basis of the study of waiting time – whether customers at the checkout counter or telephone calls at a switchboard.

First we sketch a bit of probability theory.

Some Probability Theory

The probability that an event occurs is measured by a number p, which can be anywhere from 0 up to 1; $p = 1$ implies the event will certainly occur with negligible exceptions and $p = 0$ that it will not occur with negligible exceptions. The probability that a penny turns up heads is $p = 1/2$ and that a die turns up 2 is $p = 1/6$. (The phrase "certainly occurs with negligible exceptions" means, roughly, that the times the event does not occur are so rare that we may disregard them. Similarly, the phrase "certainly will not occur with negligible exceptions" means, roughly, that the times the event does not occur are so rare that we may disregard them.)

The probability that two events that are independent of each other both occur is the product of their probabilities. For instance, the probability of getting heads when tossing a penny and a 2 when tossing the die is $p = \left(\frac{1}{2}\right)\left(\frac{1}{6}\right) = \frac{1}{12}$.

The probability that exactly one of several mutually exclusive events occurs is the sum of their probabilities. For instance, the probability of getting a 2 or a 3 with a die is $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

With that thumbnail introduction, we will analyze sparse traffic on a one-way road. We will assume that the cars enter the traffic independently of each other and travel at the same speed. Finally, to simplify matters, we assume each car is a point.

The Model

To construct our model we introduce the functions $P_0, P_1, P_2, \ldots, P_n, \ldots$ where $P_n(x)$ shall be the probability that any interval of length x contains exactly n

cars (independently of the location of the interval). Thus $P_0(x)$ is the probability that an interval of length x is empty. We shall assume that

$$P_0(x) + P_1(x) + \cdots + P_n(x) + \cdots = 1 \qquad \text{for any } x.$$

We also shall assume that $P_0(0) = 1$ ("the probability is 1 that a given point contains no cars").

For our model we make the following two major assumptions:

(a) The probability that exactly one car is in any fixed short section of the road is approximately proportional to the length of the section. That is, there is some positive number k such that

Δ, pronounced del-t$_ə$, is the Greek letter corresponding to the Latin "D".

$$\lim_{\Delta x \to 0} \frac{P_1(\Delta x)}{\Delta x} = k.$$

(b) The probability that there is more than one car in any fixed short section of the road is neglible, even when compared to the length of the section. That is,

$$\lim_{\Delta x \to 0} \frac{P_2(\Delta x) + P_3(\Delta x) + P_4(\Delta x) + \cdots}{\Delta x} = 0. \qquad \text{(C.15.1)}$$

We shall now put assumptions (a) and (b) into more useful forms. If we let

$$\epsilon = \frac{P_1(\Delta x)}{\Delta x} - k \qquad \text{(C.15.2)}$$

The Greek letter ϵ, pronounced ep-s$_ə$-l$_ə$n, corresponds to the Latin letter "e"

. where ϵ depends on Δx, assumption (a) tells us that $\lim_{\Delta x \to 0} \epsilon = 0$. Thus, solving (C.15.2) for $P_1(\Delta x)$, we see that assumption (a) can be phrased as

$$P_1(\Delta x) = k\Delta x + \epsilon \Delta x \qquad \text{(C.15.3)}$$

where $\epsilon \to 0$ as $\Delta x \to 0$.

Since $P_0(\Delta x) + P_1(\Delta x) + \cdots + P_n(\Delta x) + \cdots = 1$, assumption (b) may be expressed as

$$\lim_{\Delta x \to 0} \frac{1 - P_0(\Delta x) - P_1(\Delta x)}{\Delta x} = 0. \qquad \text{(C.15.4)}$$

In light of assumption (a), equation (C.15.4) is equivalent to

$$\lim_{\Delta x \to 0} \frac{1 - P_0(\Delta x)}{\Delta x} = k. \qquad \text{(C.15.5)}$$

In the manner in which we obtained (C.15.3), we may deduce that

$$1 - P_0(\Delta x) = k\Delta x + \delta \Delta x,$$

where $\delta \to 0$ as $\Delta x \to 0$. Thus

$$P_0(\Delta x) = 1 - k\Delta x - \delta \Delta x, \tag{C.15.6}$$

where $\delta \to 0$ as $\Delta x \to 0$. *On the basis of (a) and (b), expressed in* (C.15.3) *and* (C.15.6)*, we shall obtain an explicit formula for each* P_n.

Let us determine P_0 first. Observe that a section of length $x+\Delta x$ is vacant if its left-hand part of length x is vacant and its right-hand part of length Δx is also vacant. Since the cars move independently of each other, the probability that the whole interval of length $x + \Delta x$ being empty is the product of the probabilities that the two smaller intervals of lengths x and Δx are both empty. (See Figure C.15.1.) Thus we have

Figure C.15.1 No cars in a section of length $x+\Delta x$.

$$P_0(x + \Delta x) = P_0(x)P_0(\Delta x). \tag{C.15.7}$$

Recalling (C.15.6), we write (C.15.7) as

$$P_0(x + \Delta x) = P_0(x)(1 - k\Delta x - \delta\Delta x)$$

which a little algebra transforms to

$$\frac{P_0(x + \Delta x) - P_0(x)}{\Delta x} = -(k + \delta)P_0(x). \tag{C.15.8}$$

Taking limits on both sides of (C.15.8) as $\Delta x \to 0$, we obtain

$$P_0'(x) = -kP_0(x). \tag{C.15.9}$$

(Recall that $\delta \to 0$ as $\Delta x \to 0$.) From (C.15.9) it follows that there is a constant A such that $P_0(x) = Ae^{-kx}$. Since $1 = P_0(0) = Ae^{-k0} = A$, we conclude that $A = 1$, hence

$$P_0(t) = e^{-kx}.$$

This explicit formula for P_0 is reasonable; e^{-kx} is a decreasing function of x, so that the larger an interval, the less likely that it is empty.

Now let us determine P_1. To do so, we examine $P_1(x + \Delta x)$ and relate it to $P_0(x)$, $P_0(\Delta x)$, $P_1(x)$, and $P_1(\Delta x)$, with the goal of finding an equation involving the derivative of P_1.

Again, imagine an interval of length $x + \Delta x$ cut into two intervals, the left-hand subinterval of length x and the right-hand subinterval of length Δx. Then there is precisely one car in the whole interval if *either* there is exactly one car in the left-hand interval and none in the right-hand subinterval *or* there is none in the left-hand subinterval and exactly one in the right-hand subinterval. (See Figure C.15.2.) Thus we have

$$P_1(x + \Delta x) = P_1(x)P_0(\Delta x) + P_0(x)P_1(\Delta x). \tag{C.15.10}$$

(a) (b)

Figure C.15.2 The two ways to have exactly one car in an interval of length $x + \Delta x$.

In view of (C.15.3) and (C.15.6), we may write (C.15.10) as

$$P_1(x + \Delta x) = P_1(x)(1 - k\Delta x - \delta\Delta x) + P_0(x)(k\Delta x + \epsilon\Delta x)$$

which a little algebra changes to

$$\frac{P_1(x + \Delta x) - P_1(x)}{\Delta x} = -(k + \delta)P_1(x) + (k + \epsilon)P_0(x). \qquad \text{(C.15.11)}$$

Letting $\Delta x \to 0$ in (C.15.11) and remembering that $\delta \to 0$ and $\epsilon \to 0$ as $\Delta x \to 0$, we obtain $P_1'(x) = -kP_1(x) + kP_0(x)$; recalling that $P_0(x) = e^{-kx}$, we deduce that

$$P_1'(x) = -kP_1(x) + ke^{-kx}. \qquad \text{(C.15.12)}$$

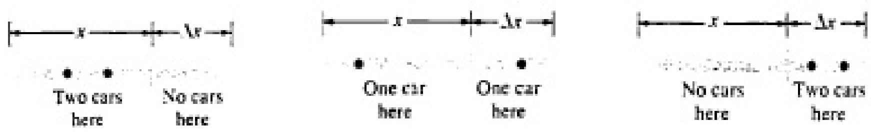
From (C.15.12) we shall obtain an explicit formula for $P_1(x)$. Since $P_0(x)$ involves e^{-kx} and so does (C.15.12), it is reasonable to guess that $P_1(x)$ involves e^{-kx}. Therefore let us express $P_1(x)$ as $g(x)e^{-kx}$ and determine the form of $g(x)$. (Since we have the identity $P_1(x) = \left(P_1(x)e^{kx}\right)e^{-kx}$, we know that $g(x)$ exists.)

According to (C.15.12) we have $\left(g(x)e^{-kx}\right)' = -kg(x)e^{-kx} + ke^{-kx}$; hence

$$g'(x)e^{-kx} + g(x)(ke^{-kx}) = -kg(x)e^{-kx} + ke^{-kx}$$

from which it follows that $g'(x) = k$. Hence $g(x) = kx + c_1$, where c_1 is some constant: $P_1(x) = (kx + c_1)e^{-kx}$. Since $P_1(0) = 0$, we have $P_1(0) = (k \cdot 0 + c_1)e^{-k \cdot 0} = c_1$ and hence $c_1 = 0$. Thus we have shown that

$$P_1(x) = kxe^{-kx} \qquad \text{(C.15.13)}$$



(a) (b) (c)

Figure C.15.3 The three ways to have exactly two cars in an interval of length $x + \Delta x$.

and P_1 is completely determined.

To obtain P_2 we argue as we did in obtaining P_1. Instead of (C.15.10) we have

$$P_2(x + \Delta x) = P_2(x)P_0(\Delta x) + P_1(x)P_1(\Delta x) + P_0(x)P_2(\Delta x) \qquad \text{(C.15.14)}$$

an equation that records the three ways in which two cars in a section of length $x + \Delta x$ can be situated in a section of length x and a section of length Δx. (See Figure C.15.3.)

Similar reasoning shows that

See Exercise 8.

$$P_2(x) = \frac{k^2x^2}{2}. \qquad \text{(C.15.15)}$$

Then, applying the same reasoning inductively leads to

$$P_n(x) = \frac{(kx)^n}{n!}e^{-kx}. \qquad \text{(C.15.16)}$$

We have obtained in (C.15.16) the formulas on which the rest of our analysis will be based. Note that these formulas refer to a road section of any length, though the assumptions (a) and (b) refer only to short sections. What has enabled us to go from the "microscopic" to the "macroscopic" is the additional assumption that the traffic in any one section is independent of the traffic in any other section. The formulas (C.15.16) are known as the **Poisson formulas**.

The Meaning of k

The constant k was defined in terms of arbitrarily short intervals, at the "microscopic level". How would we compute k in terms of observable data, at the "macroscopic level"? It turns out that k records the traffic density: the average number of cars in an interval of length x is kx.

The average number of cars in a section of length x is defined as $\sum_{n=0}^{\infty} nP_n(x)$. This weights each possible number of events (n) with it's likelihood of occurring ($P_n(x)$). This average is

$$\sum_{n=0}^{\infty} nP_n(x) = \sum_{n=1}^{\infty} n\frac{(kx)^n e^{-kx}}{n!} = kxe^{-kx}\sum_{n=1}^{\infty}\frac{(kx)^{n-1}}{(n-1)!} = kxe^{-kx}e^{kx} = kx.$$

Thus the expected number of cars in a section is proportional to the length of the section. This shows that the k appearing in assumption (a) is the measure of traffic density, the number of cars per unit length of road.

To estimate k, in the case of traffic for instance, divide the number of cars in a long section of the road by the length of that section.

EXAMPLE 1 (Traffic at a checkout counter.) Customers arrive at a checkout counter at the rate of 15 per hour. What is the probability that exactly five customers will arrive in any given 20-minute period?

SOLUTION We may assume that the probability of exactly one customer coming in a short interval of time is roughly proportional to the duration of that interval. Also, there is only a negligible probability that more than one customer may arrive in a brief interval of time. Therefore conditions (a) and (b) hold, if we replace "length of section" by "length of time". Without further ado, we conclude that the probability of exactly n customers arriving in a period of x minutes is given by (C.15.16). Moreover, the "customer density" is one per 4 minutes; hence $k = 1/4$, and thus the probability that exactly five customers arrive during a 20-minute period, $P_5(20)$, is

$$\left(\frac{1}{4}\cdot 20\right)^5 \frac{e^{-(1/4)\cdot 20}}{5!} = \frac{5^5 e^{-5}}{120} \approx 0.17547.$$

◇

Modeling of the type within this section is of use in predicting the length of waiting lines (or times) or the waiting time to cross. This is part of the theory of queues. See, for instance, Exercises 2 and 3. (See also Exercise 65 in the Summary Exercises in Chapter 4.)

EXERCISES

1.

(a) Why would you expect that $P_0(a+b) = P_0(a) \cdot P_0(b)$ for any a and b?

(b) Verify that $P_0(x) = e^{-kx}$ satisfies the equation in (a).

2. A cloud chamber registers an average of four cosmic rays per second.

(a) What is the probability that no cosmic rays are registered for 6 seconds?

(b) What is the probability that exactly two are registered in the next 4 seconds?

3. Telephone calls during the busy hour arrive at a rate of three calls per minute. What is the probability that none arrives in a period of (a) 30 seconds, (b) 1 minute, (c) 3 minutes?

4. In a large continually operating factory there are, on the average, two accidents per hour. Let $P_n(x)$ denote the probability that there are exactly n accidents in an interval of time of length x hours.

(a) Why is it reasonable to assume that there is a constant k such that $P_0(x)$, $P_1(x)$, ... satisfy 1 and 2 on page 1088?

(b) Assuming that these conditions are satisfied, show that $P_n(x) = (kx)^n e^{-kx}/n!$.

(c) Why must $k = 2$?

(d) Compute $P_0(1)$, $P_1(1)$, $P_2(1)$, $P_3(1)$, and $P_4(1)$.

5. A typesetter makes an average of one mistake per page. Let $P_n(x)$ be the probability that a section of x pages (x need not be an integer) has exactly n errors.

(a) Why would you expect $P_n(x) = x^n e^{-x}/n!$?

(b) Approximately how many pages would be error-free in a 300-page book?

6. In a light rainfall you notice that on one square foot of pavement there are an average of 3 raindrops. Let $P_n(x)$ be the probability that there are n raindrops on an area of x square feet.

(a) Check that assumptions 1 and 2 are likely to hold.

(b) Find the probability that an area of 5 square feet has exactly two raindrops.

(c) What is the most likely number of raindrops to find on an area of one square foot?

7. Write x^2 in the form $g(x)e^{-kx}$.

8. Show that $P_2(x) = \frac{k^2x^2}{2}e^{-kx}$.

9.

(a) Why would you expect $P_3(a+b) = P_0(a)P_3(b) + P_1(a)P_2(b) + P_2(a)P_1(b) + P_3(a)P_0(b)$?

(b) Do functions defined in (C.15.16) satisfy the equation in (a)?

10.

(a) Why would you expect $\lim_{n\to\infty} P_n(x) = 0$?

(b) Show that the functions defined in (C.15.16) have the limit in (a).

11. We obtained $P_0(x) = e^{-kx}$ and $P_1(x) = kxe^{-kx}$. Verify that $\lim_{\Delta x\to 0} P_1(\Delta x)/\Delta x = k$, and $\lim_{\Delta x\to 0} P_0(\Delta x)/\Delta x = 1-k$. Hence show that $\lim_{\Delta x\to 0}(P_2(\Delta x)+P_3(\Delta x)+\cdots+)/\Delta x = 0$, and that assumptions 1 and 2 on page 1088 are indeed satisfied.

12.

(a) Obtain assumption 1 from equation (C.15.3).

(b) Obtain equation (C.15.3) from assumption 2.

(c) Obtain assumption 2 from equation (C.15.6).

13.

(a) What length of road is most likely to contain exactly one car? That is, what x maximizes $P_1(x)$?

(b) What length of road is most likely to contain three cars?

14. For any $x \geq 0$, $\sum_{n=0}^{\infty} P_n(x)$ should equal 1 because it is certain that some number of cars is in a given section of length x (maybe 0 cars). Check that $\sum_{n=0}^{\infty} P_n(x) = 1$. This provides a probabilistic argument that $e^u = \sum_{n=0}^{\infty} u^n/n!$ for $n \geq 0$.

Chapter 13

Introduction to Differential Equations

This chapter is a brief introduction to one of the major applications of calculus, differential equations. A **differential equation** is simply an equation that involves derivatives of an unknown function. The goal is usually to find the unknown function or, at least, to determine some of its properties.

As Section 13.1 reminds us, we have already met such equations, for instance the equation describing natural growth and decay, $\frac{dP}{dt} = kP$. Section 13.2 shows how to solve certain differential equations involving only the function and its first derivative. The next two sections are concerned with solving certain equations that involve a function and its first and second derivatives. For most differential equations it is not possible to find explicit solutions. In these cases, one settles for approximate solutions. One method for doing this is described in Section 13.5.

13.1 Introduction and Review: Separable Equations and Direction Fields

A differential equation is an equation that involves the derivatives of a function. We have already met three such equations.

Section 5.7 concerned the differential equation that describes natural growth or decay, namely

$$\frac{dP}{dt} = kP(t). \tag{13.1.1}$$

Section 3.6 has the equation for an antiderivative $F(x)$ of a function $f(x)$,

$$\frac{dF}{dx} = f(x). \tag{13.1.2}$$

The study of motion with constant acceleration in Section 3.7 is based on solving the equation

$$\frac{d^2y}{dt^2} = a, \qquad \text{where } a \text{ is a constant.} \tag{13.1.3}$$

In this section we present the terminology of differential equations (DE), a way to visualize solutions to DEs, and a method for solving DEs of a special form.

Terminology

A **solution** of a differential equation is any function that satisfies the equation. **Solving** a differential equation means finding *all* its solutions. For instance, the solutions of (13.1.1) are $P(t) = Ce^{kt}$ for any constant C. The solutions of (13.1.3) are $y(t) = \frac{a}{2}t^2 + vt + s$, where v is the velocity and s is the position when $t = 0$.

In algebra the solutions of equations are numbers, and usually there are only a finite number of solutions. For example, the equation $x^2 - 3x + 2 = 0$ has two solutions, $x = 1$ and $x = 2$. A differential equation behaves quite differently. Its solutions are functions (not numbers). Moreover, there is usually an infinite number of solutions. (Consider, for instance, the solutions of the two equations (13.1.1) and (13.1.3).)

The **order** of a differential equation is the order of the highest-order derivative that appears in the equation. Both (13.1.1) and (13.1.2) have order one; (13.1.3) has order two. Most differential equations met in applications have orders one or two. However, there are cases of higher-order, for instance the equation used in modeling the bending of a beam, such as a diving board:

$$\frac{d^4y}{dx^4} - \frac{d^2y}{dx^2} = -W. \tag{13.1.4}$$

The unknown function $y = y(x)$ is the deflection of the beam at a distance x along the beam and W is the weight on the beam, assumed to be uniformly distributed. In Exercises 28 and 29 in Section 13.4 we will see that the general solution turns out to be

$$y(x) = ae^x + be^{-x} + c + dx + \frac{W}{2}x^2, \tag{13.1.5}$$

where a, b, c, and d are constants determined by properties of the particular beam. (See also Exercise 37.)

The number of constants that appear in the general solution usually equals the order of the differential equation. For a second-order DE the **initial conditions** would be $y(t_0) = y_0$ (initial position) and $y'(t_0) = v_0$ (initial velocity). This is like determining the orbit of a comet from its position and velocity at a given time. For instance, in Example 4 of Section 3.7, the initial conditions are $y(0) = 96$ and $y'(0) = 64$.

A differential equation together with a set of initial conditions is called an **initial value problem**.

In some problems the constraints are given at two (or more) different times. For example, in a second-order DE the position at two times might be known, $y(t_1)$ and $y(t_2)$. These are called **boundary conditions**. (This is like determining an orbit of a comet from its positions at two different times.)

A differential equation together with a set of boundary conditions is called a **boundary value problem**.

Slope Fields and Differential Equations

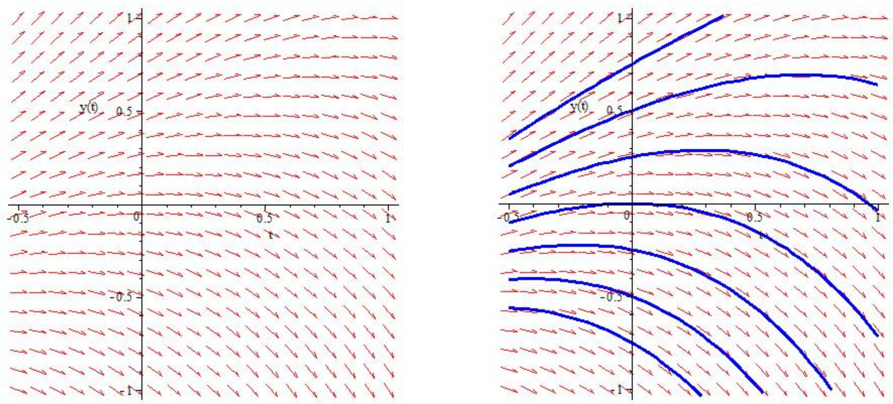
In Section 3.6 slope fields were introduced by the example $\frac{dy}{dt} = \sqrt{1+t^3}$ to picture the antiderivatives of $\sqrt{1+t^3}$. While that discussion was restricted to differential equations of the form $\frac{dy}{dt} = f(t)$, slope fields can be easily drawn when $\frac{dy}{dt}$ is given in terms of t and y. At the point (t, y) one draws a short segment whose slope is the derivative evaluated at (t, y). The next example illustrates the process in the case $\frac{dy}{dt} = y - t$.

EXAMPLE 1 Sketch the direction field for $\frac{dy}{dt} = y - t$.
SOLUTION At the point (t, y) sketch a short segment whose slope is $y - t$. For instance, at $(3, 3)$ the slope is 0. Figure 13.1.1(a) shows a few such segments. After drawing many more such segments one sees the segments seem to form curves, as in Figure 13.1.1(b).

Each such curve is the graph of a solution to the differential equation $\frac{dy}{dt} = y - t$. ◇

Slope fields are helpful in analyzing a differential equation of the form $\frac{dy}{dt} = f(t, y)$ where $f(t, y)$ is a function that involves both t and y. Note that slope fields can be drawn only for differential equations of order one and only when the DE can be written in the form $\frac{dy}{dt} = f(t, y)$.

There are many automatic slope field plotters on almost any type of device with a graphical display: on calculators, on the web, even as an app for a smartphone.



(a) (b)

Figure 13.1.1 (a) Direction field and (b) solution curves for $y' = y - t$.

Separable Equations

In Section 5.7 we solved the differential equation

$$\frac{dP}{dt} = kP \qquad (P > 0). \tag{13.1.6}$$

The first step was to divide (13.1.6) by P:

$$\frac{\frac{dP}{dt}}{P} = k.$$

Both sides of (13.1.6) can be rewritten as derivatives (with respect to t):

$$\frac{d}{dt}\ln(P(t)) = \frac{d}{dt}(kt),$$

and therefore there is a constant C such that

$$\ln(P(t)) = kt + C.$$

From this it follows that

$$P(t) = e^{kt+C}$$

so

$$P(t) = e^{C}e^{kt}$$

Renaming the constant e^C by A, we conclude that

$$P(t) = Ae^{kt}.$$

When a first-order differential equation has the special form $y' = f(x)g(y)$ one may separate the variables by dividing by $g(y)$:

$$\frac{y'}{g(y)} = f(x).$$

All the y's appear on one side of the equation and all the x's on the other. The variables have been separated.

Then, find an antiderivative (with respect to x) of both sides:

$$\int \frac{y'(x)}{g(y)}\ dx = \int f(x)\ dx. \tag{13.1.7}$$

Notice, that $y'(x)\ dx$ in the left-hand side of (13.1.7) is the differential dy. Thus

$$\int \frac{dy}{g(y)} = \int f(x)\ dx. \tag{13.1.8}$$

To go further, one must be able to compute both integrals in (13.1.8).

EXAMPLE 2 (a) Solve $y' = -x^3y^3$. (b) Find the solution with $y(0) = \frac{-1}{4}$.
SOLUTION (a) Here $f(x) = x^3$ and $g(y) = -y^3$. We rewrite $y' = -x^3y^3$ as $\frac{dy}{dx} = -x^3y^3$ and then have

$$\int \frac{-dy}{y^3} = \int x^3\ dx \tag{13.1.9}$$

Since these antiderivatives are easy to evaluate, we find

$$\frac{1}{2}y^{-2} = \frac{1}{4}x^4 + C$$

or

$$\frac{1}{y^2} = \frac{1}{2}x^4 + 2C. \tag{13.1.10}$$

To solve (13.1.10) for y we need to take the reciprocal of each side of this equation. To facilitate this, find a common denominator for the right-hand side:

$$\frac{1}{y^2} = \frac{x^4 + 4C}{2}$$

so

$$y^2 = \frac{2}{x^4 + 4C}.$$

As a result we are led to consider both

$$y = +\sqrt{\frac{2}{x^4 + 4C}} \qquad \text{and} \qquad y = -\sqrt{\frac{2}{x^4 + 4C}}.$$

(b) We are asked to find the solution to the differential equation that satisfies the initial condition $y(0) = \frac{-1}{4}$. Note that because $y(0)$ is negative, we use $y = -\sqrt{2/(x^4 + 4C)}$. And, because $y(0)$ is $-1/4$:

$$\frac{-1}{4} = -\sqrt{\frac{2}{0^4 + 4C}} \qquad \text{or} \qquad \frac{1}{16} = \frac{2}{4C},$$

from which it follows that $C = 8$.

The solution to the initial value problem with $y(0) = \frac{-1}{4}$ is

$$y = -\sqrt{\frac{2}{x^4 + 32}}.$$

◇

In general, to solve $\frac{dy}{dx} = f(x)g(y)$, first multiply by the differential dx and divide by $g(y)$ and then integrate: $\int \frac{dy}{g(y)} = \int f(x)\ dx$.

For example:

1. $y' = x^3y^2$ becomes $\frac{dy}{y^2} = x^3\ dx$ and so $\frac{-1}{y} = \frac{1}{4}x^4 + C$.
2. $y' = (1 + y^2)/x^3$ becomes $\frac{dy}{1+y^2} = \frac{dx}{x^3}$ and so $\arctan(y) = \frac{-1}{2}x^{-2} + C$.
3. $y' = x^3/y^2$ becomes $y^2dy = x^3\ dx$ and so $\frac{y^3}{3} = \frac{1}{4}x^4 + C$.

Summary

Earlier we dealt with several differential equations, namely $P' = kP$, $F'(t) = f(t)$, and $y'' = a$, a a constant. In this section we introduced some terminology of differential equations, in particular, the order and solution of a differential equation.

Usually there is an infinite number of solutions to a differential equation. By imposing initial or boundary conditions we may select one particular solution. We showed how to solve a special type of first-order equation, namely an equation of the form $y' = f(t)g(y)$, where we can "separate the variables" We also generalized slope fields to equations of the form $y' = f(t, y)$.

EXERCISES for Section 13.1

In Exercises 1 to 4 state the order of the differential equation.

1. $(y'')^3 + (y')^2 = y$

2. $t^5(y''')^4 + \cos(y^2)y + 3 = 0$

3. $\sqrt{1+t^3}y' + (\cos(y))^4 y + 3t^2 = 0$

4. $(y'')^3\sqrt{1+(y')^2} = y^6$

In Exercises 5 to 8 sketch the slope fields for the given first-order differential equation.

5. $y' = -y$

6. $y' = 1 - y$

7. $y' = 2t - y$

8. $y' = t - y$

In Exercises 9 to 13 determine if the differential equation is separable.

9. $3t^2y' + 6t = t^3 + \sin t$

10. $y' = \sin(t)/y^3$

11. $y' = \sin\left(\frac{t}{y^3}\right)$

12. $y' = t + y$

13. $t^3 + t^2y' = \ln(t)$

In Exercises 14 to 35 solve the given separable equations

14. $\frac{dy}{dt} = t$

15. $\frac{dy}{dt} = t^2$

16. $\frac{dy}{dt} = y$

17. $\frac{dy}{dt} = y^2$

18. $\frac{dy}{dx} = \frac{\sin(x)}{\cos(y)}$

19. $\frac{dy}{dt} = \sec^2(t)$

20. $\frac{dy}{dt} = \frac{t}{y}$

21. $\frac{dy}{dt} = \frac{y}{t}$

22. $\frac{dy}{dt} = \frac{ty}{t^2+1}$

23. $\frac{dy}{dx} = \frac{\cos(y)}{e^x}$

24. $\frac{dy}{dt} = \sqrt{\frac{1-y^2}{t}}$, $t > 0$

25. $\frac{dy}{dx} = \frac{-e^{y^2}}{2xy}$

26. $\frac{dy}{dx} = (x\ln(x))\left(4+y^2\right)$

27. $\frac{dy}{dt} = \frac{e^x \sin(3x)}{\sqrt{9-4y^2}}$

28. $\frac{dy}{dx} = \sin(3t)\sec(2y)$

29. $\frac{dy}{dt} = y^2 \ln(t)$

30. $\frac{dy}{dt} = \frac{\tan(2t)}{e^{2y}}$

31. $\frac{dy}{dt} = y\sqrt{1+3t}$

32. $\sec(\theta)\frac{d\theta}{dt} - t^2 = 3t$

33. $y^2\cos(\theta)\frac{dy}{d\theta} - \sin(\theta) = \cos(\theta)$

34. $\frac{d\theta}{dt} = \frac{\sin^2(2t)}{\cos^2(3\theta)}$

35. $\frac{dy}{dt} = \frac{e^y \sec^2(2t)}{y}$

36.

(a) What is the general solution of $\frac{d^2y}{dt^2} = -16$?

(b) Find the solution for which $y(0) = 10$ and $y'(0) = 5$.

37. Show that for any constants a, b, c, and d, $y = ae^x + be^{-x} + c + dx + \frac{W}{2}x^2$ satisfies the fourth-order equation for a weight-bearing beam

$$\frac{d^4y}{dx^4} - \frac{d^2y}{dx^2} = -W.$$

38. Assume that $f(x)$ and $g(x)$ are solutions of $2y'' + 3y' + y = 0$. Which of the following are also solutions of the same equation?

(a) $3f(x)$

(b) $g(3x)$

(c) $f(x) + g(x)$

(d) $3f(x) - 2g(x)$

39.

(a) Check that $y = e^t - 1$ is a solution of $y' = 1 + y$.

(b) Find all solutions of the equation in (a).

40.

(a) Check that $y = \tan(x)$ is a solution of $y' = 1 + y^2$.

(b) Find all solutions of the equation in (a).

41. Identify each equation for which $y = \sin(3x)$ is a solution.

(a) $\frac{d^2y}{dx^2} = 9y$

(b) $\frac{d^2y}{dx^2} = -9y$

(c) $\frac{dy}{dx} = \sqrt{1 - y^2}$

(d) $\frac{dy}{dx} = 3\sqrt{1 - y^2}$

In Exercises 42 to 46 do not solve the equation. Instead, answer qualitatively, using information hidden in the differential equation.

42. Assume that $y'(x) = y(x)(1 - y(x))$ for $x \geq 0$ and that $y(0)$ is 1.

(a) What constant function satisfies the equation and initial condition?

(b) Show that there are no other functions that satisfy the equation and initial condition.

The same reasoning applies to $y' = y^m(1 - y^n)$, where m and n are positive integers.

43. Assume that $y'(x) = y(x)(1 - y(x))$ and that $y(0) = 1/2$.

(a) Show that the graph of $y(x)$ has an inflection point at $(0, 1/2)$.

(b) Show that $\lim_{x\to\infty} y(x)$ exists.

(c) Find the limit in (b).

(d) Examine $\lim_{x\to-\infty} y(x)$.

44. Assume the sign of $y'(t)$ is always opposite the sign of $y(t)$.

(a) Must the sign of $y''(t)$ always be the opposite the sign of $y'(t)$?

(b) What is the relationship between $y''(t)$ and $y(t)$?

(c) Give two examples of functions with this property and $y(0) = 1$.

45. Assume that the sign of $y''(t)$ is always opposite the sign of $y(t)$.

(a) What might the graph of $y(t)$ look like?

(b) Give a specific example of such a function.

46. Consider the differential equation $P' = kP(M - P)$ where k and M are positive constants.

(a) Show that $P(t) = 0$ and $P(t) = M$ are constant solutions to this equation.

(b) Show that P is increasing if and only if $0 < P(t) < M$.

(c) To determine concavity of solutions it is necessary to know the sign of P''. Find an expression for P'' that involves P and P' (and k and M). (Differentiate the differential equation with respect to t and remember that $P = P(t)$.)

(d) Use the original differential equation to obtain an expression for P'' that does not involve P'.

(e) Explain why solutions have inflection points when $P = M/2$.

47.

(a) Give an example of a solution to

$$\left(\frac{dy}{dx}\right)^2 = -y^2$$

(b) Find all solutions of the equation in (a).

48. Find a first-order differential equation that has $y = e^{x^2}$ as a solution.

49.

(a) Show directly that if $\frac{dy}{dx} = \sqrt{1 - y^2}$ then $\frac{d^2y}{dx^2} = -y$.

(b) Show that for every constant k, $y = \sin(x + k)$ satisfies both equations in (a).

13.2 First-Order Linear Differential Equations

In this section we will solve differential equations of the form

$$a_1(t)\frac{dy}{dt} + a_0(t)y = f(t). \tag{13.2.1}$$

In Sections 13.3 and 13.4 we will solve some differential equations of the form

$$a_2\frac{d^2y}{dt^2} + a_1\frac{dy}{dt} + a_0y = f(t). \tag{13.2.2}$$

These are special cases of n^{th}-order differential equations of the form

$$a_n(t)\frac{d^ny}{dt^n} + a_{n-1}(t)\frac{d^{n-1}y}{dt^{n-1}} + \cdots + a_1(t)\frac{dy}{dt} + a_0(t)y = f(t). \tag{13.2.3}$$

In (13.2.3) each coefficient $a_i(t)$, $i = 0, 1, \ldots, n$, is a function of the independent variable t. It may be quite complicated or it may be simply a constant function. The coefficient $a_n(t)$ is assumed not to be the zero function. Each summand is of the form "function of t times some derivative of y". In particular (13.2.1) is a **first-order linear differential equation** and (13.2.2) is a **second-order linear differential equation**. These two special types often appear in applications. For instance, see the Calculus is Everywhere section at the end of this chapter.

Equation (13.2.3) is called a **linear differential equation** of order n. The word "linear" reminds us of the equation of a line, $Ax + By = C$, where each variable appears to only the first power.

When the function $f(t)$ on the right-hand side of (13.2.1), (13.2.2), and (13.3.3) is the zero function, the equation is called **homogeneous**. If $f(t)$ is not the zero function the equation is called **nonhomogeneous**.

Solving $a_1(t)\frac{dy}{dt} + a_0(t)y = f(t)$

To solve a first-order linear differential equation we begin by dividing by $a_1(t)$, obtaining the "standard form"

$$\frac{dy}{dt} + p(t)y = q(t) \tag{13.2.4}$$

where $p(t)$ and $q(t)$ are functions that do not depend on the unknown function $y(t)$.

Examples 1 and 2 illustrate the technique for solving (13.2.4).

EXAMPLE 1 Find all solutions to

$$\frac{dy}{dt} + 4t^3y = 0, \qquad \text{where } y > 0. \tag{13.2.5}$$

SOLUTION This (homogeneous) first-order differential equation is separable. Separating the terms with t from those with y yields

$$\frac{dy}{y} = -4t^3\ dt.$$

Finding antiderivatives (with respect to t) produces

$$\ln(y) = -t^4 + K \qquad \text{for any constant } K.$$

Taking the exponential of each side of the equation leads to

$$\begin{array}{lrcl} & e^{\ln(y)} & = & e^{-t^4+K} \\ \text{and so} & y & = & e^{-t^4}\ e^K \end{array}$$

Thus, with e^K replaced by A,

$$y(t) = Ae^{-t^4} \qquad \text{for any positive value of the constant } A.$$

◇

Notice that this approach depends only on finding an antiderivative of the coefficient of y, namely an antiderivative of $4t^3$. This means that any homogeneous first-order linear differential equation is as easy to solve as a separable differential equation.

In Example 1 the number A is called a **parameter**. For each choice of A there is a corresponding solution. This is similar to parameterizing a curve where for each choice of the parameter there is a corresponding point on the curve.

Next, we show how the solution to the homogeneous differential equation (13.2.5) can be used to find a solution to the nonhomogeneous differential equation (13.2.6).

EXAMPLE 2 Find all solutions to

$$\frac{dy}{dt} + 4t^3y = \cos(t)e^{-t^4}. \tag{13.2.6}$$

SOLUTION In Example 1 we found that the solution to the corresponding homogeneous equation is

$$y_h(t) = Ae^{-t^4} \qquad \text{for any value of the parameter } A. \tag{13.2.7}$$

We look for solutions to the nonhomogeneous equation in the form

$$y(t) = A(t)e^{-t^4} \qquad \text{for an as yet unknown function } A(t). \tag{13.2.8}$$

In other words, we replace the constant A by a function $A(t)$, hoping that we will be able to find a solution of (13.2.6). As we show after this example, this gamble will always work if we can find an antiderivative of a function that will appear.

Letting $y(t)$ be $A(t)e^{-t^4}$ in (13.2.6) we obtain

$$\frac{d}{dt}\left(A(t)e^{-t^4}\right) + 4t^3A(t)e^{-t^4} = \cos(t)e^{-t^4}.$$

Using the product rule we find that

$$\frac{dA(t)}{dt}e^{-t^4} - 4t^3A(t)e^{-t^4} + 4t^3A(t)e^{-t^4} = \cos(t)e^{-t^4}.$$

Then, after the cancellation, we have

$$\frac{dA(t)}{dt}e^{-t^4} = \cos(t)e^{-t^4},$$

which, since e^{-t^4} is never zero, reduces to

$$\frac{dA(t)}{dt} = \cos(t).$$

Thus

$$A(t) = \sin(t) + C.$$

Inserting this $A(t)$ back into (13.2.8) produces the general solution to the first-order linear differential equation (13.2.6), namely

$$y(t) = (\sin(t) + C)\, e^{-t^4} = \sin(t)e^{-t^4} + Ce^{-t^4} \qquad \text{for any constant } C.$$

◇

In Example 1 we had to integrate $-4t^3$ and in Example 2 we had to integrate $\cos(t)$. In general, for a first-order linear differential equation, if the antiderivatives encountered are elementary we can give explicit formulas for the general solution.

The approach in Example 2 is called **variation of parameters** because the solutions to the nonhomogeneous case are found by replacing the constant parameter in the solutions of the homogeneous case by a function.

Why Variation of Parameters Works

To solve the nonhomogeneous equation

$$\frac{dy}{dt} + p(t)y = q(t) \qquad (13.2.9)$$

we first solve the associated homogeneous equation

$$\frac{dy}{dt} + p(t)y = 0. \tag{13.2.10}$$

Notice that (13.2.10) is separable, and can be rewritten as

$$\frac{dy}{y} = -p(t)\ dt. \tag{13.2.11}$$

Let $u(t)$ be any solution of (13.2.10). We then seek a solution of the form $A(t)u(t)$ to the nonhomogeneous equation (13.2.9). Substituting $y(t) = A(t)u(t)$ into (13.2.9) gives us

$$\frac{d}{dt}\left(A(t)u(t)\right) + p(t)A(t)u(t) = q(t).$$

Hence

$$\frac{dA(t)}{dt}u(t) + A(t)\frac{du}{dt} + p(t)A(t)u(t) = q(t).$$

Collecting the two terms with $A(t)$, we reach the equation

$$\frac{dA}{dt}u(t) + A(t)\left(\frac{du(t)}{dt} + p(t)u(t)\right) = q(t). \tag{13.2.12}$$

Because $u(t)$ is a solution of (13.2.10), (13.2.12) reduces to

$$\frac{dA(t)}{dt}u(t) = q(t),$$

or

$$\frac{dA(t)}{dt} = \frac{q(t)}{u(t)}. \tag{13.2.13}$$

So, finding $A(t)$ depends on being able to integrate $q(t)/u(t)$.

We illustrate variation of parameters by another example.

EXAMPLE 3 Find all solutions to $(t^2+1)\frac{dy}{dt} - ty = t$, with $y > 0$.

SOLUTION To put this equation in standard form, divide by the coefficient of the derivative: $t^2 + 1$, obtaining

$$\frac{dy}{dt} - \frac{t}{t^2+1}y = \frac{t}{t^2+1}. \tag{13.2.14}$$

The first step is to solve the homogeneous equation:

$$\frac{dy}{dt} - \frac{t}{t^2+1}y = 0. \tag{13.2.15}$$

As expected, (13.2.15) is separable:

$$\frac{dy}{y} = \frac{t}{t^2+1}\,dt. \tag{13.2.16}$$

Taking the antiderivatives of both sides of (13.2.16) leads to

$$\ln y(t) = \frac{1}{2}\ln(t^2+1) + K \qquad \text{for any value of the constant } K. \tag{13.2.17}$$

Solving for $y(t)$, by exponentiating both sides of (13.2.17) and introducing $A = e^K$, produces

$$y(t) = A(t^2+1)^{1/2} \qquad \text{for any value of the constant } A.$$

Using the variation of parameters method, we look for a solution to the nonhomogeneous equation in the form

$$y = A(t)(t^2+1)^{1/2}$$

To simplify the algebra, denote $(t^2+1)^{1/2}$ by $u(t)$, keeping in mind that $u(t)$ is a solution to the homogeneous equation (13.2.15).

Substituting $y(t) = A(t)u(t)$ into (13.2.14) leads to

$$\frac{d}{dt}\left(A(t)u(t)\right) - \frac{t}{t^2+1}A(t)u(t) = \frac{t}{t^2+1},$$

hence

$$A'(t)u(t) + A(t)u'(t) - \frac{t}{t^2+1}A(t)u(t) = \frac{t}{t^2+1}$$

or

$$A'(t)u(t) + A(t)\left(u'(t) - \frac{t}{t^2+1}u(t)\right) = \frac{t}{t^2+1}.$$

Because $u(t)$ solves the homogeneous equation (13.2.15), it follows that

$$A'(t)u(t) = \frac{t}{t^2+1}.$$

Recalling that $u(t) = (t^2+1)^{1/2}$, we have

$$A'(t)\left(t^2+1\right)^{1/2} = \frac{t}{t^2+1},$$

hence

$$A'(t) = \frac{t}{(t^2+1)^{3/2}}. \tag{13.2.18}$$

One antidifferentiation (with respect to t) yields

$$A(t) = \int \frac{t}{\left(t^2+1\right)^{3/2}}\, dt = -\left(t^2+1\right)^{-1/2} + C.$$

The general solution to the nonhomogeneous equation (13.2.14) is therefore

$$\begin{aligned} y(t) &= A(t)u(t) \\ &= \left(-(t^2+1)^{-1/2} + C\right)\sqrt{t^2+1} \\ &= -1 + C\sqrt{t^2+1}. \qquad \text{for any value of the constant } C. \end{aligned}$$

◇

Summary

Any first-order linear differential equation can be put in the form

$$y' + p(t)y = q(t). \tag{13.2.19}$$

The corresponding homogeneous equation, obtained by replacing $q(t)$ by 0, is separable. Its general solution has the form $Au(t)$, where A is a constant.

Replacing the parameter A by a function $A(t)$ provides a candidate $A(t)u(t)$ as a solution for the nonhomogeneous equation (13.2.19). This leads to a differential equation for $A(t)$ that can be solved by evaluating one antiderivative.

A similar two-step approach will be used in Section 13.4 to solve nonhomogeneous second-order linear differential equations.

EXERCISES for Section 13.2

1. Using as few mathematical symbols as possible, describe what is meant by a linear differential equation.

2. What is the difference between a homogeneous and a nonhomogeneous linear differential equation?

In Exercises 3 to 9 decide which differential equations are linear. Rewrite each linear equation in standard form.

3. $\cos(3t)y' + e^{t^2}y + (t^3 - 1) = 0$

4. $3(y')^2 + 4y = t^7$

5. $y' + \sin(y) = t^3$

6. $y'' - 3y' + t^2 y = t^2$

7. $y' + yy' = t$

8. $e^t y' - y\cos(t) = t^4$

9. $y + 5y' + t^3 y'' = e^t$

10. Check that $y(t) = Ae^{-t^4}$ is a solution to equation (13.2.5) in Example 1.

11. Check that $y(t) = (\sin(t) + C)\, e^{-t^4}$ is a solution to equation (13.2.6) in Example 2.

12. Check that $y(t) = -1 + C\left(t^2 + 1\right)^{1/2}$ is a solution to equation (13.2.14) in Example 3.

13. In Example 3 we denoted $\left(t^2 + 1\right)^{1/2}$ by $u(t)$. If we didn't do that, we would substitute $y(t) = A(t)\left(t^2 + 1\right)^{1/2}$ into (13.2.14) instead. Carry out this substitution and note how much more work is involved in arriving at (13.2.18).

In Exercises 14 to 28 find the general solution to the given linear differential equation. If an initial condition is provided, find the corresponding particular solution.

14. $y' + 2y = 1$, $y(0) = 0$

15. $y' - y = 3e^{2t}$

16. $y' = y - t$, $y(0) = 1/2$

17. $ty' + 7y = 5t^2$

18. $ty' + 2y = 5t$, $y(2) = 4$

19. $2ty' + y = 10\sqrt{t}$

20. $ty' - y = t^3$, $y(1) = 7$

21. $2ty' - 5y = 12t^2$

22. $(1+t)y' + y = \sin(t)$

23. $ty' - 2y = t^3 \sec(t)\tan(t)$

24. $y' = 1 + t + y + ty$

25. $y' = 2ty + 9t^2 e^{t^2}$, $y(0) = 0$

26. $ty' + (2t-3)y = 6t^6$

27. $ty' + 2y = \ln(t)$

28. $ty' + 4y = t^{-2}e^t$

13.3 Second-Order Linear Differential Equations: The Homogeneous Case

In the next section we deal with the equation

$$a\frac{d^2y}{dt^2} + b\frac{dy}{dt} + cy = g(t) \tag{13.3.1}$$

where a, b, and c are constants and $g(t)$ is a function. The restriction to *constant* coefficients may seem narrow, but (13.3.1) is broad enough to be of use in many applications. In this section we treat the homogeneous equation obtained by replacing $g(t)$ by 0:

$$a\frac{d^2y}{dt^2} + b\frac{dy}{dt} + cy = 0. \tag{13.3.2}$$

New Solutions from Old

Note that if $y(t)$ is a solution to (13.3.2), then so is $ky(t)$ for any constant k. To see this, check that ky satisfies (13.3.2) whenever y does:

$$\begin{aligned} a(ky)'' + b(ky)' + c(ky) &= aky'' + bky' + cky \\ &= k \cdot (ay'' + by' + cy) \\ &= k \cdot 0 = 0 \end{aligned}$$

Next, if y_1 and y_2 are solutions of (13.3.2), then so is their sum $y_1 + y_2$. We show this by substituting directly into the differential equation:

$$\begin{aligned} a(y_1+y_2)'' + b(y_1+y_2)' + c(y_1+y_2) &= ay_1'' + ay_2'' + by_1' + by_2' + cy_1 + cy_2 \\ &= (ay_1'' + by_1' + cy_1) + (ay_2'' + by_2' + cy_2) \\ &= 0 + 0 = 0. \end{aligned}$$

So $y_1 + y_2$ is a solution, as claimed.

Noting that $y_1 - y_2 = y_1 + (-1)y_2$, we see that the difference $y_1 - y_2$ of solutions is also a solution to the homogeneous equation. (These results depend on the fact that the right-hand side of the homogeneous equation is zero.)

The key idea that we will exploit is a result of the linearity of (13.3.2):

Theorem 13.3.1. *If y_1 and y_2 are solutions to the homogeneous equation* (13.3.2)*, then so is $k_1y_1 + k_2y_2$ for any values of the constants k_1 and k_2.*

Proof

The calculations just made show that k_1y_1 and k_2y_2 are solutions to (13.3.2) and that their sum $k_1y_1 + k_2y_2$ is also a solution to (13.3.2). •

The Main Idea

If y is a solution of (13.3.2), we would expect y' and y'' to resemble y. If they didn't, there would be little hope of the sum of the three terms on the left-hand side of (13.3.2) adding up to 0. For this reason we anticipate that the solutions will involve exponential functions (or possibly sines and cosines).

Pursuing this idea, it is reasonable to look for solutions of the form e^{rt}, where r is some constant to be determined.

If $y = e^{rt}$, we have

$$y = e^{rt}, \qquad y' = re^{rt}, \quad \text{and} \quad y'' = r^2e^{rt}. \tag{13.3.3}$$

Substituting these into (13.3.2) yields an equation for r:

$$0 = a\frac{d^2y}{dt^2} + b\frac{dy}{dt} + cy = ar^2e^{rt} + bre^{rt} + ce^{rt} = (ar^2 + br + c)e^{rt}.$$

Since e^{rt} is never zero we conclude that $y = e^{rt}$ is a solution to (13.3.2) when r is a root of the **characteristic polynomial**, $ar^2 + br + c$; that is, when

$$ar^2 + br + c = 0. \tag{13.3.4}$$

Equation (13.3.4) has the roots:

$$r_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \qquad \text{and} \qquad r_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}. \tag{13.3.5}$$

The sign of $b^2 - 4ac$ determines whether there are two (distinct) real roots ($b^2 - 4ac > 0$), a single repeated real root ($b^2 - 4ac = 0$), or a pair of complex conjugate roots ($b^2 - 4ac < 0$). Each case will be treated separately.

Case 1: Two distinct real roots ($b^2 - 4ac > 0$)

The two functions $y_1(t) = e^{r_1t}$ and $y_2 = e^{r_2t}$ are both solutions to (13.3.2). By Theorem 13.3.1, $y = k_1y_1 + k_2y_2 = k_1e^{r_1t} + k_2e^{r_2t}$ is a solution of (13.3.2) for any constants k_1 and k_2.

It is proved in a more advanced course that there are no other solutions.

EXAMPLE 1 Find the general solution to

$$\frac{d^2y}{dt^2} - \frac{dy}{dt} - 2y = 0. \tag{13.3.6}$$

SOLUTION The characteristic equation is $r^2 - r - 2 = 0$. Since $b^2 - 4ac = (-1)^2 - 4(1)(-2) = 9 > 0$ there are two real roots. Instead of resorting to the quadratic formula, notice that $r^2 - r - 2 = (r-2)(r+1)$, from which we see that $r_1 = 2$ and $r_2 = -1$. The general solution to (13.3.6) is

$$y = c_1 e^{2t} + c_2 e^{-t}.$$

◇

Case 2: One real repeated root ($b^2 - 4ac = 0$)

When $b^2 - 4ac = 0$, then $r_1 = \frac{-b}{2a}$ and $r_2 = \frac{-b}{2a}$; the characteristic equation has only one real root. As a result there is only one exponential solution: $y_1 = e^{\frac{-b}{2a}t}$.

To obtain another solution, try $y_2(t) = A(t)e^{-bt/(2a)}$, a technique similar to the variation of parameters used in Section 13.2. It turns out that $te^{-bt/(2a)}$ is also a solution. Exercise 23 invites you to check that it is by substituting it into the equation (13.3.2) and using the fact that $b^2 - 4ac = 0$ (and $a \neq 0$). Exercise 24 contains the complete derivation of this second solution.

EXAMPLE 2 Find the general solution to $\frac{d^2y}{dt^2} + 4\frac{dy}{dt} + 4y = 0$.
SOLUTION The characteristic equation is $r^2 + 4r + 4 = 0$. Since $b^2 - 4ac = (4)^2 - 4(1)(4) = 0$ there is only a single repeated real root. Since $r^2 + 4r + 4 = (r+2)^2$ we see that $r_1 = r_2 = -2$. Two solutions are $y_1 = e^{-2t}$ and $y_2 = te^{-2t}$ and the general solution is

$$y = c_1 e^{-2t} + c_2 t e^{-2t}.$$

◇

Case 3: Complex conjugate roots ($b^2 - 4ac < 0$)

In this case the roots r_1 and r_2 of the characteristic equation involve $\sqrt{b^2 - 4ac}$, which is the square root of a negative number. But we are interested in finding real-valued solutions of (13.3.2). At the moment we have solutions $e^{r_1 t}$ and $e^{r_2 t}$, which are not real-valued functions. Differentiation of such functions is defined in a course on complex analysis. It turns out that $\frac{d}{dt}\left(e^{r_1 t}\right) = r_1 e^{r_1 t}$ and the properties of derivatives developed in Chapter 3 hold for complex-valued functions.

Happily, the following theorem provides a way to extract two real-valued solutions from a single complex-valued solution.

Theorem 13.3.2. *If $y = u(t) + iv(t)$ is a complex-valued solution to* (13.3.2)*, then the real-valued functions $u(t)$ and $v(t)$ are also solutions of that differential equation.*

The proof is a straight-forward calculation and depends on the fact that if x and y are real and $x + iy = 0$, then both x and y must be 0. This is Exercise 26.

Just as in the case of two distinct real roots (when $b^2 - 4ac > 0$), we have two *complex*-valued solutions when $b^2 - 4ac < 0$. One root is

$$r_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} = \frac{-b}{2a} + \frac{\sqrt{4ac - b^2}}{2a} i = p + qi$$

where $p = \frac{-b}{2a}$ and $q = \frac{\sqrt{4ac-b^2}}{2a}$ are the real and imaginary parts of r_1, respectively.

The solution corresponding to the root $r_1 = p + qi$ is

$$e^{r_1 t} = e^{pt+iqt} = e^{pt} e^{iqt}.$$

The exponential e^{iqt} is simplified by Euler's formula: $e^{i\theta} = \cos(\theta) + i\sin(\theta)$, so we may write

$$e^{r_1 t} = e^{pt}\left(\cos(qt) + i\sin(qt)\right) = e^{pt}\cos(qt) + ie^{pt}\sin(qt).$$

By Theorem 13.3.2, two real-valued solutions to (13.3.2) are

$$y_1 = e^{pt}\cos(qt) \qquad \text{and} \qquad y_2 = e^{pt}\sin(qt).$$

The second root is $r_2 = p - qi$, the complex conjugate of r_1. The corresponding complex-valued solution is (because $\cos(-q) = \cos(q)$)

$$e^{r_2 t} = e^{pt}e^{-iqt} = e^{pt}\cos(qt) - ie^{pt}\sin(qt). \tag{13.3.7}$$

This does not provide new solutions, for the real part of (13.3.7) is the solution y_1 and the imaginary part of (13.3.7) is the negative of y_2. So the second root may be disregarded.

As a result, the general solution to (13.3.2) when the characteristic equation has complex conjugate roots $r_1 = p + qi$ and $r_2 = p - qi$ is

$$y = c_1 e^{pt}\cos(qt) + c_2 e^{pt}\sin(qt)$$

for any constants c_1 and c_2.

EXAMPLE 3 Find the general solution of

$$\frac{d^2y}{dt^2} + 2\frac{dy}{dt} + 10y = 0. \tag{13.3.8}$$

SOLUTION The characteristic equation is $r^2 + 2r + 10 = 0$. Since $b^2 - 4ac =$

$(2)^2 - 4(1)(10) = -36 < 0$ there are complex conjugate roots: $r_1 = \frac{-2+\sqrt{-36}}{2} = -1 + 3i$ and $r_2 = \frac{-2-\sqrt{36}}{2} = -1 - 3i$. According to the analysis just given, the general solution of (13.3.8) is

$$y = c_1 e^{-t} \cos(3t) + c_2 e^{-t} \sin(3t).$$

◇

These three cases cover all possibilities for *homogeneous* second-order linear differential equations with constant coefficients. Solutions to *nonhomogeneous* linear differential equations with constant coefficients will be addressed in the next section.

Summary

The homogeneous second-order linear differential equation with constant coefficients has the form

$$ay'' + by' + cy = 0 \tag{13.3.9}$$

where a, b, and c are constants (and a is not zero). Its general solution is

$$y(t) = c_1 y_1(t) + c_2 y_2(t) \tag{13.3.10}$$

where the functions y_1 and y_2 are shown in Table 13.3.1.

Condition	Classification	Roots	Fundamental Solutions
$b^2 - 4ac > 0$	distinct real roots	r_1, r_2	$e^{r_1 t}$, $e^{r_2 t}$
$b^2 - 4ac = 0$	repeated real root	$r_1 = -\frac{b}{2a}$	$e^{r_1 t}$, $te^{r_1 t}$
$b^2 - 4ac < 0$	complex conjugate roots	$r_1 = p + iq$, $r_2 = p - iq$	$e^{pt} \cos(qt)$, $e^{pt} \sin(qt)$

Table 13.3.1 Summary of the three cases for the second-order linear differential equation with constant coefficients.

There are four special equations whose solutions should be memorized. The first three special cases are $y'' + cy = 0$ where $c = k^2$, $c = -k^2$ and $c = 0$. All four results summarized in Table 13.3.2 are developed in Exercises 6 to 9.

Special Case	Roots	Solutions	
$y'' + k^2y = 0$	$ki, -ki$	$\sin(kt)$	$\cos(kt)$
$y'' - k^2y = 0$	$k, -k$	e^{kt}	e^{-kt}
$y'' = 0$	0 (multiplicity 2)	1	t
$y'' + ky' = 0$	$0, -k$	1	e^{-kt}

Table 13.3.2 Four important and easy-to-remember second-order linear differential equations with constant coefficients.

EXERCISES for Section 13.3

1. Which second-order linear equations ($ay'' + by' + cy = f(t)$) have the zero function ($y = 0$) as a solution?

2. Let y_1 and y_2 be solutions of $ay'' + by' + cy = t^2$.

(a) Is $y = y_1 + y_2$ a solution?

(b) Is $y = 2y_1 - y_2$ a solution?

(c) Is $y = 3y_2$ a solution?

3. Let y_1 and y_2 be solutions of $ay'' + by' + cy = 0$.

(a) Is $y = y_1 + y_2$ a solution?

(b) Is $y = 2y_1 - y_2$ a solution?

(c) Is $y = 3y_2$ a solution?

4. Let y_1 and y_2 be solutions of $a(y'')^2 + by' + cy = 0$.

(a) Is $y = y_1 + y_2$ a solution?

(b) Is $y = 2y_1 - y_2$ a solution?

(c) Is $y = 3y_2$ a solution?

5. Consider the equation $ay'' + by' + cy = 0$ with $b^2 - 4ac = 0$.

(a) Verify that $y = e^{-bt/(2a)}$ is a solution.

(b) Verify that $y = te^{-bt/(2a)}$ is a solution.

6. The differential equation $y'' = 0$ was first encountered in Section 3.7 where it was shown that the general solution has the form $a + bt$ for some constants a and b. Explain how this is consistent with the third line in Table 13.3.2.

7. Let k be a positive constant.

(a) Using Table 13.3.2, find two solutions to $y'' + k^2 y = 0$.

(b) Check that both solutions found in (a) are solutions to $y'' + k^2 y = 0$.

8. Let k be a positive constant.

(a) Using Table 13.3.2, find two solutions to $y'' - k^2 y = 0$.

(b) Check that both solutions found in (a) are solutions to $y'' - k^2 y = 0$.

9. Let k be a positive constant.

(a) Using Table 13.3.2, find two solutions to $y'' + ky' = 0$.

(b) Check that both solutions found in (a) are solutions to $y'' + ky' = 0$.

In each of Exercises 10 to 15 find the general solution of the differential equation.

10. $y'' + 5y' + 6y = 0$

11. $y'' - y' - 6y = 0$

12. $y'' + 9y = 0$

13. $y'' - 4y' + 4y = 0$

14. $y'' - 2y' + 5y = 0$

15. $y'' + 10y' + 25y = 0$

In each of Exercises 16 to 21 solve the initial value problem. That is, find the solution of the differential equation that satisfies the given initial conditions. (Observe that the equations are the same as in Exercises 10 to 15.)

16. $y'' + 5y' + 6y = 0$, $y(0) = 0$, $y'(0) = 2$

17. $y'' - y' - 6y = 0$, $y(0) = 1$, $y'(0) = 2$

18. $y'' + 9y = 0$, $y(0) = 1$, $y'(0) = 3$

19. $y'' - 4y' + 4y = 0$, $y(0) = 0$, $y'(0) = -1$

20. $y'' - 2y' + 5y = 0$, $y(0) = 0$, $y'(0) = 0$

21. $y'' + 10y' + 25y = 0$, $y(0) = 4$, $y'(0) = 0$

22. Verify that $y_1 = e^{-t}\cos(3t)$ and $y_2 = e^{-t}\sin(3t)$ are solutions to the differential equation in Example 3 by substituting them in the equation.

23. Show that $y = te^{-bt/(2a)}$ is a solution to $ay'' + by' + cy = 0$ when $b^2 - 4ac = 0$ (and $a \neq 0$).

24. When $b^2-4ac=0$ (and $a \neq 0$) we found that one solution to $ay''+by'+cy=0$ is $y_1(t)=Ae^{-bt/(2a)}$ and claimed that a second solution is $y_2(t)=te^{-bt/(2a)}$. In this Exercise we show how this second solution was obtained.

(a) Verify that $y_1(t) = Ae^{-bt/(2a)}$ is a solution to $ay'' + by' + cy = 0$ for any constant A.

(b) The variation of parameters idea suggests looking for a second solution in the form $y(t) = A(t)e^{-bt/(2a)}$. Find the second-order differential equation for $A(t)$ that makes $y_2(t) = A(t)e^{-bt/(2a)}$ a solution to $ay'' + by' + cy = 0$.

(c) Solve the differential equation found in (b).

(d) What is the resulting solution y_2?

25.

Sam: In Example 2 the authors say that the general solution is $c_1e^{-2t} + c_2te^{-2t}$.

Jane: What's your point?

Sam: They missed the obvious solution $y = 0$. I'm going to send them an e-mail.

Write the authors' response.

26. Prove Theorem 13.3.2. That is, suppose $y = u(t) + iv(t)$ is a complex-valued solution to (13.3.2) and show that $u(t)$ and $v(t)$ are also solutions of (13.3.2).

27. Show that if $k > 0$ and a_1 and a_2 are constants, then $(a_1 + a_2t)\,e^{-kt/2} \to 0$ as $t \to \infty$.

13.4 Second-Order Linear Differential Equations: The Nonhomogeneous Case

In this section we consider the *nonhomogeneous* second-order linear differential equation with constant coefficients

$$ay'' + by' + cy = f(t), \tag{13.4.1}$$

giving one method to find all solutions. We first show that if we can find one solution of (13.4.1) and all solutions of the **associated homogeneous equation**

$$ay'' + by' + cy = 0, \tag{13.4.2}$$

then we can easily construct *all* solutions of (13.4.1). Then we describe a way to find one solution of (13.4.1). The technique is "intelligent guessing." Its more elegant name is the **method of undetermined coefficients**.

The Key Theorem

The key idea behind this method is the following theorem.

Theorem 13.4.1. *Let u be a particular solution to $ay'' + by' + cy = f(t)$ and let v be a particular solution to $ay'' + by' + cy = g(t)$. Then*

(a) $u + v$ is a solution to $ay'' + by' + cy = f(t) + g(t)$,

(b) ku is a solution to $ay'' + by' + cy = kf(t)$ where k is any constant, and

(c) $u - v$ is a solution to $ay'' + by' + cy = f(t) - g(t)$.

Proof

We prove the first assertion, (a). The assumptions on u and v tell us that

$$\begin{aligned} au'' + bu' + cu &= f(t) \\ av'' + bv' + cv &= g(t). \end{aligned}$$

Then

$$\begin{aligned} a(u+v)'' + b(u+v)' + c(u+v) &= a(u'' + v'') + b(u' + v') + c(u+v) \\ &= (au'' + bu' + cu) + (av'' + bv' + cv) \\ &= f(t) + g(t). \end{aligned}$$

The proofs of (b) and (c) are similar (see Exercises 25 and 26.) •

We will use (a) to prove part 1 and (c) to prove part 2 in the following theorem; (b) will be used later.

Theorem 13.4.2. *Let y_p be a particular solution of* (13.4.1).

1. *If y_h is a solution of the associated homogeneous equation* (13.4.2), *then $y_p + y_h$ is also a solution of* (13.4.1).
2. *Every solution of* (13.4.1) *is of the form $y_p + y_h$ for some solution y_h of the associated homogeneous equation* (13.4.2).

Proof

We first prove (1). Since y_p solves $ay''+by'+cy = f(t)$ and y_h solves $ay''+by'+cy = 0$, part(a) of Theorem 13.4.1 tells us that $y_p + y_h$ solves $ay'' + by' + cy = f(t) + 0 = f(t)$.

Now we prove (2). Let both y_p and y_q be solutions of (13.4.1). That is,

$$ay_p'' + by_p' + cy_p = f(t)$$

and

$$ay_q'' + by_q' + cy_q = f(t).$$

By part (c) of Theorem 13.4.1, $y_q - y_p$ is a solution of

$$ay'' + by' + cy = f(t) - f(t) = 0.$$

Therefore, $y_q - y_p = y_h$, a solution of the homogeneous equation (13.4.2). Thus, $y_q = y_h + y_p$. •

The point of Theorem 13.4.2 is that if we know one solution of (13.4.1) and all solutions of the associated homogeneous equation (13.4.2), we can easily construct *all* solutions of (13.4.1). In this section we describe a way to find one solution of (13.4.1). Section 13.3 described ways to find all solutions of the associated homogeneous equation (13.4.2).

Intelligent Guessing: The Method of Undetermined Coefficients

A few examples illustrate the method of undetermined coefficients.

EXAMPLE 1 Find one solution to

$$y'' - 3y' + 2y = 5e^{3t}. \tag{13.4.3}$$

SOLUTION Influenced by the right-hand side of (13.4.3), we try $y = Ae^{3t}$

for some constant A. Then $y' = 3Ae^{3t}$ and $y'' = 9Ae^{3t}$. Substituting these into (13.4.3) gives

$$y''-3y'+2y = 9Ae^{3t}-3\left(3Ae^3t\right)+2\left(Ae^{3t}\right) = (9-3(3)+2)Ae^{3t} = 2Ae^{3t} = 5e^{3t}.$$

Because e^{3t} is never 0,

$$2A = 5$$

or

$$A = \frac{5}{2}.$$

The steps are reversible, so $y_p = \frac{5}{2}e^{3t}$ is a solution of (13.4.3). ◇

EXAMPLE 2 Find one solution of

$$y'' - 3y' + 2y = 40\cos(2t). \tag{13.4.4}$$

SOLUTION It is reasonable to expect a solution to have the form

$$y = A\sin(2t) + B\cos(2t) \tag{13.4.5}$$

where A and B are constants. So we substitute (13.4.5) in (13.4.4) to see if our guess is right:

$$\begin{aligned}(A\sin(2t) + B\cos(2t))'' - 3(A\sin(2t) + B\cos(2t))' + 2(A\sin(2t) + B\cos(2t))\\ = \;\; 40\cos(2t).\end{aligned}$$

Carrying out the differentiation in each term produces the equation

$$\begin{aligned}(-4A\sin(2t) - 4B\cos(2t)) - 3(2A\cos(2t) - 2B\sin(2t)) + 2(A\sin(2t) + B\cos(2t))\\ = \;\; 40\cos(2t).\end{aligned}$$

Collecting the trigonometric terms yields the equation

$$(-4A + 6B + 2A)\sin(2t) + (-4B - 6A + 2B)\cos(2t) = 40\cos(2t) \tag{13.4.6}$$

To satisfy (13.4.6) for all values of t we must have both

$$-2A + 6B = 0 \qquad \text{and} \qquad -6A - 2B = 40.$$

A little algebra shows that the solution to this system of two linear equations is $A = -6$ and $B = -2$. Thus one solution to (13.4.4) is

$$y_p = -6\sin(2t) - 2\cos(2t).$$

◇

Notice that in Example 2 it would not have sufficed to guess either $A\sin(2t)$ or $B\cos(2t)$ separately; we have to expect contributions from both terms.

EXAMPLE 3 Find one solution to

$$y'' - 3y' + 2y = 7e^{-2t}. \tag{13.4.7}$$

SOLUTION It's reasonable to try

$$y = Ae^{-2t}. \tag{13.4.8}$$

Substituting this into (13.4.7) leads to

$$\left(Ae^{-2t}\right)'' - 3\left(Ae^{-2t}\right)' + 2\left(Ae^{-2t}\right) = 7e^{-2t}$$

or

$$4Ae^{-2t} + 6Ae^{-2t} + 2Ae^{-2t} = 7e^{-2t}.$$

Thus $12A = 7$ or $A = \frac{7}{12}$, and

$$y_p = \frac{7}{12}e^{t}$$

is one solution of (13.4.3), as is easily checked. ◇

EXAMPLE 4 Find one solution to

$$y'' - 3y' + 2y = 5e^{3t} + 40\cos(2t) + 7e^{-2t}. \tag{13.4.9}$$

SOLUTION According to part (a) of Theorem 13.4.1, since the right-hand side of (13.4.9) is the sum of the right-hand sides of (13.4.3), (13.4.4), and (13.4.7), we can just add the solutions found in Examples 1, 2, and 3:

$$y_p = \frac{5}{2}e^{3t} - 6\sin(2t) - 2\cos(2t) + \frac{7}{12}e^{-2t}. \tag{13.4.10}$$

◇

EXAMPLE 5 Find one solution to

$$y'' + 3y' + 2y = 7e^{-2t}. \tag{13.4.11}$$

SOLUTION Inspired by Example 3 we try $y = Ae^{-2t}$, getting

$$\left(Ae^{-2t}\right)'' + 3\left(Ae^{-2t}\right)' + 2\left(Ae^{-2t}\right) = 7e^{-2t}$$

or

$$4Ae^{-2t} - 6Ae^{-2t} + 2Ae^{-2t} = 7e^{-2t}$$

Thus, since e^{-2t} is never zero,

$$4A - 6A + 2A = 7$$

or

$$0 = 7.$$

This gamble failed.

Rather than lose all hope, we try $y = Ate^{-2t}$. Substituting this guess into (13.4.11) gives

$$\left(Ate^{-2t}\right)'' + 3\left(Ate^{-2t}\right)' + 2\left(Ate^{-2t}\right) = 7e^{-2t}.$$

Differentiation, including the product rule, leads to

$$A(-4 + 4t)e^{-2t} + 3A(1 - 2t)e^{-2t} + 2Ate^{-2t} = 7e^{-2t}.$$

Dividing this equation by e^{-2t}, which is never zero, leads to the equation

$$A(-4 + 4t + 3 - 6t + 2t) = 7,$$

or

$$A(-1 + 0t) = 7, \tag{13.4.12}$$

that is

$$-A = 7.$$

Thus $y_p = -7te^{-2t}$ is a solution of (13.4.11). ◇

Note that if the coefficient of t in (13.4.12) had not been 0, this guess would not have worked either.

In Example 5 our first attempt failed because substituting Ae^{-2t} into the left-hand side of (13.4.11) gave the value 0. In other words, e^{-2t} is a solution of the homogeneous equation $y'' + 3y' + 2y = 0$. Had we solved the homogeneous equation first, we would have noticed this coincidence and not tried Ae^{-2t}. The lesson learned from this is that, it is prudent to solve the homogeneous equation first and the non-homogeneous equation second.

EXAMPLE 6 Find one solution to

$$y'' + 6y' + 9y = 11e^{-3t}. \tag{13.4.13}$$

SOLUTION The solutions to the associated homogeneous equation, $y'' + 6y' + 9y = 0$, are $y_1 = e^{-3t}$ and $y_2 = te^{-3t}$. The right-hand side of (13.4.13) suggests the guess Ae^{-3t}. But, since this is already known to be a solution to the associated homogeneous equation, it can't be a solution to the nonhomogeneous equation. For the same reason, neither can Ate^{-3t}. So we try $y = At^2e^{-3t}$. Substituting this guess into (13.4.13) yields

$$\left(At^2e^{-3t}\right)'' + 6\left(At^2e^{-3t}\right)' + 9\left(At^2e^{-3t}\right) = 11e^{-3t} \tag{13.4.14}$$

Computing the derivatives gives the equation

$$A(2 - 12t + 9t^2)e^{-3t} + 6A(2t - 3t^2)e^{-3t} + 9At^2e^{-3t} = 11e^{-3t},$$

which simplifies to

$$A(2 + 0t + 0t^2)e^{-3t} = 11e^{-3t}.$$

Thus $A = \frac{11}{2}$ and we have the particular solution $y_p = \frac{11}{2}t^2e^{-3t}$. ◇

EXAMPLE 7 Find a solution to

$$y'' + 6y' + 9y = 6t^2 - 4. \tag{13.4.15}$$

SOLUTION As in Example 6, the associated homogeneous equation has solutions $y_1 = e^{-3t}$ and $y_2 = te^{-3t}$.

It's judicious to expect a particular solution of (13.4.15) to be a polynomial. Moreover, there is no point in trying a polynomial of degree greater than 2 since the summand $9y$ would have a term of degree greater than 2, but there would be no term in y'' or $6y'$ to cancel it. As there are no polynomial terms in the solution of the associated homogeneous equation, we try $y = At^2 + Bt + C$.

Substituting it in (13.4.15) and collecting like powers of t leads to

$$9At^2 + (12A + 9B)t + (2A + 6B + 9C) = 6t^2 - 4.$$

These two quadratics are equal for all values of t if and only if corresponding coefficients are equal. That is,

$$\begin{aligned} 9A &= 6 \\ 12A + 9B &= 0 \\ 2A + 6B + 9C &= -4 \end{aligned}$$

The solution is

$$A = \frac{2}{3}, \qquad B = \frac{-8}{9}, \quad \text{and} \quad C = 0.$$

Thus, one solution of (13.4.15) is

$$y_p = \frac{2}{3}t^2 - \frac{8}{9}t.$$

◇

Notice that even though the right-hand side in (13.4.15) has only quadratic and constant terms, the particular solution contained only quadratic and linear terms. Because it's nearly impossible to predict which terms will not appear in the particular solution, the intelligent guess is the full polynomial of the same degree as the right-hand side of the equation.

The final example combines the methods of Section 13.3 and this section.

EXAMPLE 8 Find all solutions of

$$y'' + 2y' + 10y = e^{2t} - 3e^{-t}. \tag{13.4.16}$$

SOLUTION All solutions of the associated homogeneous equation $y'' + 2y' + 10y = 0$ were found in Example 3 of Section 13.3. They are

$$y_h = c_1 e^{-t}\cos(3t) + c_2 e^{-t}\sin(3t)$$

for any values of c_1 and c_2. All that remains is to find a single solution of the nonhomogeneous equation (13.4.16).

We find a solution to (13.4.16) in two steps. First we find one solution y_{p_1} to

$$y'' + 2y' + 10y = e^{2t} \tag{13.4.17}$$

and then we find one solution y_{p_2} to

$$y'' + 2y' + 10y = -3e^{-t}. \tag{13.4.18}$$

Then, by part (a) of Theorem 13.4.1, one solution of (13.4.16) is $y_{p_1} + y_{p_2}$ and, by Theorem 13.4.2, the general solution of (13.4.16) is

$$y = c_1 e^{-t}\cos(3t) + c_2 e^{-t}\sin(3t) + (y_{p_1} + y_{p_2}).$$

To find a solution of (13.4.17) we try $y = Ae^{2t}$. This yields

$$\left(Ae^{2t}\right)'' + 2\left(Ae^{2t}\right)' + 10\left(Ae^{2t}\right) = e^{2t},$$

which reduces to

$$4A + 4A + 10A = 1$$

and so $A = \frac{1}{18}$. Thus $y_{p_1} = \frac{1}{18}e^{2t}$.

Next, as e^{-t} is not a solution to the associated homogeneous equation, one solution of (13.4.18) may be found in the form $y = Be^{-t}$. Substituting this guess into (13.4.18) yields

$$\left(Be^{-t}\right)'' + 2\left(Be^{-t}\right)' + 10\left(Be^{-t}\right) = -3e^{-t}.$$

After simplification, this becomes

$$B - 2B + 10B = -3.$$

Thus, $B = \frac{-1}{3}$ and so $y_{p_2} = \frac{-1}{3}e^{-t}$.

Adding, we obtain a particular solution to the two nonhomogeneous equations, namely

$$y_p = y_{p_1} + y_{p_2} = \frac{1}{18}e^{2t} - \frac{1}{3}e^{-t}.$$

So the general solution of (13.4.16) is

$$y = y_h + y_p = c_1e^{-t}\cos(3t) + c_2e^{-t}\sin(3t) + \frac{1}{18}e^{2t} - \frac{1}{3}e^{-t}.$$

◇

Summary

To solve a second-order linear differential equation with constant coefficients,

$$ay'' + by' + cy = f(t), \tag{13.4.19}$$

first solve the associated homogeneous equation by the methods of Section 13.3.

Second, find a particular solution y_p of (13.4.19) by intelligent guessing, guided by the form of $f(t)$.

When solving the nonhomogeneous equation do not use a guess that happens to be a solution of the associated homogeneous equation.

The general solution of (13.4.19) has the form $y = c_1y_1 + c_2y_2 + y_p$, where y_p solves the nonhomogeneous equation and y_1 and y_2 solve the associated homogeneous equation, and c_1 and c_2 are arbitrary constants.

The initial guesses for the form of the particular solution are given in Table 13.4. These forms need to be multiplied by extra factors of t when they happen to contain terms that also appear in the solution of the associated homogeneous equation.

description	**right-hand side:** $F(t)$	**initial guess for** y_p
exponential	Ae^{kt}	Be^{kt}
trigonometric	$A_1\cos(kt) + A_2\sin(kt)$	$B_1\cos(kt) + B_2\sin(kt)$
constant	A	B
nth-degree polynomial	$A_0 + A_1t + A_2t^2 + \cdots + A_nt^n$	$B_0 + B_1t + B_2t^2 + \cdots + B_nt^n$
exponential × trigonometric	$e^{kt}(A_1\cos(kt) + A_2\sin(kt))$	$e^{kt}(B_1\cos(kt) + B_2\sin(kt))$
exponential × polynomial	$e^{kt}(A_0 + A_1t + A_2t^2 + \cdots + A_nt^n)$	$e^{kt}(B_0 + A_1t + B_2t^2 + \cdots + B_nt^n)$

EXERCISES for Section 13.4

In Exercises 1 to 8, what form would be your first guess for y_p if $ay''+by'+cy = f(t)$? Assume that $f(t)$ is not a solution of the homogeneous equation and that there are no repeated roots. *(Do not solve for y_p.)*

1. $f(t) = 3\sin(2t)$

2. $f(t) = 4\cos(5t)$

3. $f(t) = 2e^{-2t}$

4. $f(t) = 3e^{5t}\sin(4t)$

5. $f(t) = 3\sin(2t) + 5\cos(3t)$

6. $f(t) = t^2 + 5$

7. $f(t) = 7t + \cos(t)$

8. $f(t) = e^{7t} + \sin(3t)$

9. Give an appropriate guess for a particular solution of $y'' - 8y' + 15y = f(t)$ if

(a) $f(t) = e^{2t}$

(b) $f(t) = 5\sin(3t)$

(c) $f(t) = e^{3t}$

(d) $f(t) = e^{5t}$

(e) $f(t) = 2\sin(3t) + 4\cos(3t)$

(f) $f(t) = e^{2t} + 6\sin(3t) + 3e^{3t} - 2e^{5t}$?

10. Give an appropriate guess for a particular solution of $y'' + 10y' + 25y = f(t)$ if

(a) $f(t) = e^{2t}$

(b) $f(t) = 5\sin(3t)$

(c) $f(t) = e^{-5t}$

(d) $f(t) = t^2e^{5t}$

(e) $f(t) = e^{-5t}\sin(3t)$

(f) $f(t) = e^{2t} + 6\sin(3t) - 4e^{-5t} + t^2e^{5t} + e^{-5t}\sin(3t)$?

11. Give an the appropriate guess for a particular solution of $y'' + 25y = f(t)$ if

(a) $f(t) = \sin(4t)$

(b) $f(t) = t\cos(4t)$

(c) $f(t) = \sin(5t)$

(d) $f(t) = t\cos(5t)$

(e) $f(t) = e^{5t}$

(f) $f(t) = t\cos(4t) + 2t\cos(5t) + e^{5t}$?

12. Give an appropriate guess for a particular solution of $y'' + y' - 2y = f(t)$ if

(a) $f(t) = 3t + 1$

(b) $f(t) = e^{3t}$

(c) $f(t) = e^{-2t}$

(d) $f(t) = e^{t} + e^{2t}$

(e) $f(t) = e^{t}\sin(t)$?

In Exercises 13 to 20 find a particular solution of the differential equation.

13. $y'' + y' - 6y = e^{t}$

14. $y'' + y' - 6y = e^{2t}$

15. $y'' + 25y = 3\cos(4t)$

16. $y'' + 25y = 20\cos(5t)$

17. $y'' + y' - 12y = e^{3t}$

18. $y'' - 2y' + y = e^{t}$

19. $y'' - 2y' + 3y = 2t^2 + 12$

20. $3y'' + 2y' + y = e^{2t}\left(56\cos(t) - 28\sin(t)\right)$

21. Find all solutions of the $y'' - 2y' - 3 = f(t)$ when

(a) $f(t) = e^{2t}$

(b) $f(t) = e^{t}$

(c) $f(t) = e^{-3t}$

22. Find all solutions of the $y'' + 6y' + 9y = f(t)$ when

(a) $f(t) = e^t$

(b) $f(t) = e^{3t}$

(c) $f(t) = e^{-3t}$

23. Find all solutions of the $y'' + 9y = f(t)$ when

(a) $f(t) = \cos(4t)$

(b) $f(t) = \sin(4t)$

(c) $f(t) = e^t \sin(4t)$

24. Find all solutions of the $y'' - 4y' + 4y = f(t)$ when

(a) $f(t) = 8$

(b) $f(t) = 4t + 4$

(c) $f(t) = -4t^2 + 16t - 6$

25. Prove part (b) of Theorem 13.4.1

26. Prove part (c) of Theorem 13.4.1

27. Which equations of the form $ay'' + by' + cy = f(t)$ have distinct solutions u and v such that $u - v$ is also a solution?

In an earlier section (Exercise 13.1.4 in Section 13.1) we asked the reader verify a solution to the fourth-order beam equation $y^{(iv)} - y'' = -W$ (W a constant). Exercises 28 and 29 present two ways to find this solution.

28. The first approach takes advantage of the fact that the beam equation involves only even derivatives of y to solve the equation in two steps. Each step involves solving a second-order equation.

(a) Show that the substitution $u = y''$ reduces the fourth-order equation to the second-order equation $u'' - u = -W$.

(b) Find the general solution to the second-order equation found in (a).

(c) Solve the second-order equation $y'' = u$, where u is the solution found in (b).

29. The second approach to solving the beam equation extends the ideas of Section 13.3 to higher-order homogeneous equations and then uses the ideas of this section to find a particular solution.

(a) Find all values of r that make $y = e^{rt}$ a solution to $y^{(iv)} - y'' = 0$.

(b) Use the information in (a) (and the ideas in Section 13.3) to find all solutions to $y^{(iv)} - y'' = 0$. (This should involve four independent constants.)

(c) Use the method of intelligent guessing to find a particular solution to the nonhomogeneous equation $y^{(iv)} - y'' = -W$.

30. Consider the differential equation $y'' - 4y' + qy = 4e^{3t}$ where q is some constant.

(a) For which values of q is Ae^{3t} a valid guess for a nonhomogeneous solution?

(b) For the value of q found in (a), what is an a valid guess for a nonhomogeneous solution?

31. Consider the differential equation $y'' + py' + qy = 7e^{3t}$, where p and q are constants.

(a) For which values of p and q do neither Ae^{3t} nor Ate^{3t} provide a solution?

(b) For which values of p and q does the use of At^2e^{3t} provide a solution?

(c) For the values of p and q found in (b), why would it not make sense to guess $(At^2 + Bt + C)e^{3t}$ as the form of a particular solution for this equation?

13.5 Euler's Method

The methods introduced in this chapter apply only to differential equations of special forms: separable (Section 13.1), first-order linear (Section 13.2), and second-order linear with constant coefficients (Sections 13.3 and 13.4). In a differential equations course you would encounter several more special types.

However, many differential equations do not have solutions that can be expressed in terms of elementary functions. This is the case for such seemingly simple equations as

$$\frac{dy}{dt} = e^{-t^2} \qquad \text{or} \qquad \frac{dy}{dt} = \sqrt{1+y^3}.$$

Or it may happen that a differential equation can be solved exactly, but only after much work, for instance,

$$\frac{dy}{dt} = \sin y \qquad \text{and} \qquad \frac{dy}{dt} = 1 + e^y.$$

Fortunately, there is a general and easy-to-apply numerical technique for obtaining approximate solutions. This technique, known as **Euler's method**, involves calculations easily implemented on a calculator, computer, or other programmable device.

Euler's method applies to a general first-order differential equation:

$$\frac{dy}{dt} = h(t,y) \tag{13.5.1}$$

with an initial condition

$$y(t_0) = y_0. \tag{13.5.2}$$

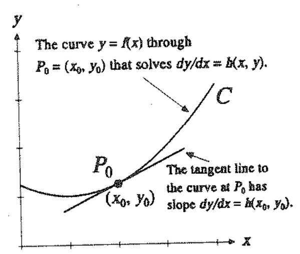


Figure 13.5.1

The solution to (13.5.1) and (13.5.2) is a function $y = f(t)$, whose graph is a curve that we will denote by C. (See Figure 13.5.1.)

The point $P_0 = (t_0, y_0)$ is on C, and at this point the slope of the solution is $m_0 = y'(t_0) = h(t_0, y_0)$. Then we use the tangent line to C at P_0 to construct an estimate at a nearby value of the independent variable, say at $t = t_1$.

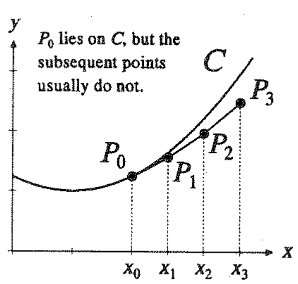


Figure 13.5.2

Because the point $P_1 = (t_1, y_1)$ is on the tangent line to C at (t_0, y_0), $y_1 = y_0 + (t_1 - t_0)m_0$. This is just the linear approximation described in Section 5.5. While P_1 will not generally be on C it will be close to C when t_1 is near t_0.

Then we repeat the process, with P_1 instead of P_0. Using $m_1 = h(t_1, y_1)$ as an estimate of the slope of C at P_1, we choose t_2 near t_1 and compute $y_2 = y_1 + (t_2 - t_1)m_1$. This process can be repeated to construct points P_1, P_2, … that approximate the curve C. (See Figure 13.5.2.)

With each step the estimate generally moves further from the curve C. This tendency can be controlled by choosing t_i closer to t_{i-1}. However, even if the process is automated, you don't want to take too many steps because of the accumulation of arithmetical errors due to rounding.

EXAMPLE 1 Approximate the solution to $y' = y - t$, $y(0) = \frac{1}{2}$ on the interval $0 \leq t \leq 1$ using 4 equal steps.

SOLUTION Here $h(t, y) = y - t$ and $y_0 = \frac{1}{2}$. Since the length of the interval is $1 - 0 = 1$, we have $\Delta t = \frac{1}{4}$.

The following table organizes the work involved in Euler's method. Note that the first row has been filled in using the initial conditions.

i	t_i	$y_i = y_{i-1} + m_{i-1}\Delta t$	$m_i = h(t_i, y_i)$	$P_i = (t_i, y_i)$
0	$t_0 = 0$	$y_0 = \frac{1}{2}$	$m_0 = \frac{1}{2} - 0 = \frac{1}{2}$	$P_0 = (0, 0.5)$
1	$t_1 = \frac{1}{4}$			
2	$t_2 = \frac{1}{2}$			
3	$t_3 = \frac{3}{4}$			
4	$t_4 = 1$			

Table 13.5.1 Initial table for Euler's method for $y' = y - t$, $y(0) = \frac{1}{2}$, on the interval $[0, 1]$ with four steps of size $\Delta t = \frac{1}{4}$.

Table 13.5.2 shows the calculations for the four steps of Euler's method.

i	t_i	$y_i = y_{i-1} + m_{i-1}\Delta x$	$m_i = h(t_i, y_i)$	$P_i = (t_i, y_i)$
0	$t_0 = 0$	$y_0 = \frac{1}{2}$	$m_0 = \frac{1}{2} - 0 = \frac{1}{2}$	$P_0 = (0.00, 0.500)$
1	$t_1 = \frac{1}{4}$	$y_1 = \frac{1}{2} + \frac{1}{2}\frac{1}{4} = \frac{5}{8}$	$m_1 = \frac{5}{8} - \frac{1}{4} = \frac{3}{8}$	$P_1 = (0.25, 0.625)$
2	$t_2 = \frac{1}{2}$	$y_2 = \frac{5}{8} + \frac{3}{8}\frac{1}{4} = \frac{23}{32}$	$m_2 = \frac{23}{32} - \frac{1}{2} = \frac{7}{32}$	$P_2 = (0.50, 0.719)$
3	$t_3 = \frac{3}{4}$	$y_3 = \frac{23}{32} + \frac{7}{32}\frac{1}{4} = \frac{99}{128}$	$m_3 = \frac{99}{128} - \frac{3}{4} = \frac{3}{128}$	$P_3 = (0.75, 0.773)$
4	$t_4 = 1$	$y_4 = \frac{99}{128} + \frac{3}{32}\frac{1}{4} = \frac{399}{512}$	$m_4 = \frac{399}{512} - \frac{4}{4} = \frac{-113}{512}$	$P_4 = (1.00, 0.779)$

Table 13.5.2 Completed table for Euler's method for $y' = y - t$, $y(0) = \frac{1}{2}$, on the interval $[0, 1]$ with four steps of size $\Delta t = \frac{1}{4}$.

Euler's method, with 4 steps of size $\Delta t = \frac{1}{4}$, provides the approximate value 0.779 for $y(1)$. ◇

Increasing the number of steps reduces Δt. That should improve the accuracy of Euler's method. For example, doubling the number of steps in Example 1 reduces Δt to 1/8 and yields the data in Table 13.5.3.

With $n = 8$ steps, the approximate value for $y(1)$ is 0.717. Doubling the number of steps again further refines the estimate to $y(1) = 0.681$.

In this case we can solve for the solution of $y' = y - t$ with $y(0) = 1/2$. The curve C is the graph of the function $y = 1 + x - \frac{1}{2}e^x$ (see Exercise 16 in

i	t_i	$y_i = y_{i-1} + \Delta x m_{i-1}$	$m_i = h(t_i, y_i)$	$P_i = (t_i, y_i)$
0	$t_0 = 0$	$y_0 = \frac{1}{2}$	$m_0 = \frac{1}{2}$	$P_0 = (0.00, 0.500)$
1	$t_1 = \frac{1}{8}$	$y_1 = 0.5625$	$m_1 = 0.4375$	$P_1 = (0.125, 0.563)$
2	$t_2 = \frac{1}{4}$	$y_2 = 0.617$	$m_2 = 0.367$	$P_2 = (0.250, 0.617)$
3	$t_3 = \frac{3}{8}$	$y_3 = 0.663$	$m_3 = 0.288$	$P_3 = (0.375, 0.663)$
4	$t_4 = \frac{1}{2}$	$y_4 = 0.699$	$m_4 = 0.199$	$P_4 = (0.500, 0.699)$
5	$t_5 = \frac{5}{8}$	$y_5 = 0.724$	$m_5 = 0.990$	$P_5 = (0.625, 0.724)$
6	$t_6 = \frac{3}{4}$	$y_6 = 0.736$	$m_6 = -0.014$	$P_6 = (0.750, 0.736)$
7	$t_7 = \frac{7}{8}$	$y_7 = 0.735$	$m_7 = -0.140$	$P_7 = (0.875, 0.735)$
8	$t_8 = 1$	$y_8 = 0.717$	$m_8 = -0.283$	$P_8 = (1.000, 0.717)$

Table 13.5.3 Completed table for Euler's method for $y' = y - t$, $y(0) = \frac{1}{2}$, on the interval $[0, 1]$ with eight steps of size $\Delta t = \frac{1}{8}$.

Section 13.2). Thus $y(1)$ is $2 - \frac{e}{2} \approx 0.641$. Table 13.5.4 shows how the estimates of $y(1)$ approach 0.641 and Figure 13.5.3 shows the convergence visually.

n	Δt	Estimate of $y(1)$	Error
4	$\frac{1}{4}$	0.779	$0.779 - 0.641 = 0.138$
8	$\frac{1}{8}$	0.717	$0.717 - 0.641 = 0.076$
16	$\frac{1}{16}$	0.681	$0.681 - 0.641 = 0.040$

Table 13.5.4 Convergence of approximations of $y(1)$ for $y' = y - t$, $y(0) = \frac{1}{2}$, by Euler's method with 4, 8, and 16 steps. The exact value is $2 - \frac{e}{2} \approx 0.641$.

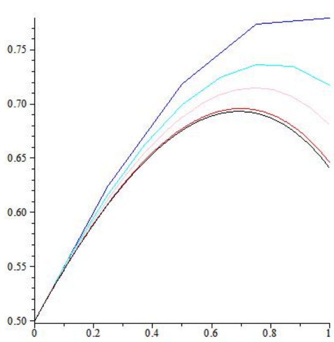


Figure 13.5.3

As Table 13.5.4 suggests, with Euler's method, doubling the number of steps tends to cut the error in half. To put it another way, the error tends to be proportional to Δt.

Note that while we have focused on the approximate values of the solution at the right-hand endpoint of the interval, Euler's method provides estimates of the solution throughout the interval at equally spaced points with spacing Δt. Thus, Euler's method can also be used to graph an approximate solution to an initial value problem on a specific interval. This is how the piecewise linear curves in Figure 13.5.3 were created.

A course on differential equations or numerical analysis will present more sophisticated methods for estimating the solution to $y' = h(t, y)$, $y(t_0) = y_0$. Most of these methods are elaborations of Euler's method.

Summary

Euler's method for approximating the solution to the first-order initial value problem

$$y' = h(t, y), \qquad y(a) = y_0$$

on the interval $a \le t \le b$ can be summarized in the following algorithm:

Given	$h(t, y)$, t_0, y_0,
	a, b, n
Compute	$\Delta t = \frac{b-a}{n}$
Compute	`for` i `= 1, 2, 3, ...,` n `do`
	$m = h(t_{i-1}, y_{i-1})$
	$t_i = t_{i-1} + \Delta t$
	$y_i = y_{i-1} + m\Delta t$
	`end do`

EXERCISES for Section 13.5

In Exercises 1 to 3 estimate $y(1)$ for the problem considered in Example 1 using Euler's method with the indicated number of steps. Also, estimate the error. (This can easily be done with pencil and paper.)

1. $n = 1$ ($\Delta t = 1$).

2. $n = 2$ ($\Delta t = 1/2$).

3. $n = 3$ ($\Delta t = 1/3$).

In Exercises 4 to 9 estimate $y(1)$ for the equation $y' = \frac{y}{4}(8 - y)$ with $y(0) = 1$ using Euler's method with the indicated number of steps. Use the fact that $y(1) = 7.09088$ to five decimal digits to estimate the error in each case. (Exercises 4 to 7 can easily be done with pencil and paper.)

4. $n = 1$ ($\Delta t = 1$).

5. $n = 2$ ($\Delta t = 1/2$).

6. $n = 3$ ($\Delta t = 1/3$).

7. $n = 4$ ($\Delta t = 1/4$).

8. $n = 8$ ($\Delta t = 1/8$).

9. $n = 16$ ($\Delta t = 1/16$).

In Exercises 10 to 17, use Euler's method for the given differential equation, initial condition, and step size to estimate y at the right-hand endpoint of the given interval. Present your estimates both as a table and in a graph.

10. $y' = 2t - 3y$, $y(0) = 1$, $\Delta t = 0.2$, $0 \leq t \leq 1$

11. $y' = t + 4y$, $y(0) = \frac{1}{2}$, $\Delta t = 0.2$, $0 \leq t \leq 1$

12. $y' = 3ty$, $y(1) = 1$, $\Delta t = 0.1$, $1 \leq t \leq 2$

13. $y' = 2t^2 y$, $y(0) = 2$, $\Delta t = 0.1$, $0 \leq t \leq 1$

14. $y' = \cos t - \sin y$, $y(0) = 0$, $\Delta t = 0.1$, $0 \leq t \leq 1$

15. $y' = \tan t \sec y$, $y(0) = 0$, $\Delta t = 0.2$, $0 \leq t \leq 1$

16. $y' = y \ln t$, $y(2) = 1$, $\Delta t = 0.2$, $2 \leq t \leq 4$

17. $y' = e^t - y$, $y(0) = 1$, $\Delta t = 0.1$, $0 \leq t \leq 1$

18. In Example 1 Euler's method with $\Delta t = \frac{1}{4}$ was used to estimate $y(1)$ for the initial value problem $y' = y - t$ with $y(0) = \frac{1}{2}$ with $n = 4$. The estimates with $n = 8$ and $n = 16$ were also given.

(a) Verify the estimate of $y(1)$ reported for $n = 8$.

(b) Verify the estimate of $y(1)$ reported for $n = 16$.

(c) Obtain estimates of $y(1)$ for $n = 32$, $n = 64$, and $n = 128$.

(d) Create a table showing each estimate's error, that is, the difference between it and the exact solution.

(e) What pattern do you see in this table?

19.

Sam: I've a neat trick to save labor when using Euler's method.

Jane: Yes?

Sam: Say that with $n = 4$ I get an estimate E_4 and with $n = 8$ I get an estimate E_8. Then I predict that $2E_8 - E_4$ will be a much better estimate.

Jane: Please give me an example.

Sam: In this section, $E_4 = 0.779$ and $E_8 = 0.717$. So, my estimate $2E_8 - E_4 = 2(0.717) - 0.779$ is 0.655. That's pretty close to the correct value. It's even better than $E_{16} = 0.681$.

Jane: How did you ever get such a smart idea?

Sam: There's a clue in the book.

Explain Sam's reasoning.
Sam has hit on a well-known method for improving Euler's estimates. In this method the error is much smaller, being proportional to $(\Delta t)^2$ (which is much smaller than Δt, at least for small values of Δt).

20. Given $y'' = h(t, y, y')$, $y(t_0) = y_0$, and $y'(t_0) = v_0$, describe how you would modify Euler's method to estimate $y(1)$.

13.S Chapter Summary

A differential equation for a function of one variable, $f(t)$, is called an **ordinary differential equation**, abbreviated ODE. This is in contrast to similar equations for functions of more than one variable, for example $f(t, x)$, which are called **partial differential equations**, abbreviated PDE. There are a few examples of PDEs in the Calculus is Everywhere projects at the end of Chapters 16 and 17.

A solution to a typical ODE is not an elementary function, that is, cannot be expressed in terms of exponentials, logarithms, trigonometric functions, polynomials, and the functions obtained from them by algebraic means. If it is not elementary, we can approximate it by various methods, of which Euler's Method is an example.

The chapter focuses on three special types of linear differential equations:

1. The separable equations of the form
$$\frac{dy}{dt} = f(t)g(y),$$
2. The first-order linear equations
$$a_1(t)\frac{dy}{dt} + a_0(t)y = f(t)$$
which we rewrite as
$$\frac{dy}{dt} + p(t)y = q(t), \tag{13.S.1}$$
and solve if we can evaluate two antiderivatives. If both are elementary we can write out our explicit solution to (13.S.1) and it will involve one parameter.
3. $$a_2y'' + a_1y' + a_0y = f(t) \tag{13.S.2}$$
where the coefficients a_0, a_1, and a_2 are constants. Solving (13.S.2) begins by solving the related homogeneous equation $a_2y''+a_1y'+a_0y = 0$. Solutions are found in the form e^{rt}. The general solution to this second-order linear differential equation will involve two parameters. Then we use the Method of Intelligent Guessing (or Undetermined Coefficients) to find one particular solution of (13.S.2), y_p. By adding y_p to the general solution to the corresponding homogeneous equation we find all solutions to (13.S.2).

For any DE of the form $y' = f(t, y)$ we can compute an approximate numerical solution with Euler's Method, whose error is proportional to the width Δt we choose. There are methods similar to Euler's that converge faster to the exact solution.

Numerous software programs and tables help solve many differential equations.

EXERCISES for 13.S

Find the general solution of the differential equations in Exercises 1 to 8.

1. $y^3 \frac{dy}{dt} = (y^4 + 1)\sin(t)$

2. $\frac{dy}{dt} = 2t\sec(y)$

3. $(t^2 + 1)(\tan(y))y' = x$

4. $t^2 y' = 1 - x^2 + y^2 - x^2 y^2$ (Factor the right-hand side.

5. $y' + 2y = 3te^{-2t}$

6. $y' - 2ty = e^{t^2}$

7. $y'' + 2y' + 5y = e^t \sin(t)$

8. $y'' - 9y = 2t^2 e^{3t} + 5)$

Find the solution of the initial value problems in Exercises 9 to 14.

9. $2y\frac{dy}{dt} = \frac{t}{\sqrt{t^2-16}}$, $y(5) = 2$

10. $\tan(t)\frac{dy}{dt} = y$, $y\left(\frac{\pi}{2}\right) = \frac{\pi}{2}$

11. $y' + 2ty = t$, $y(0) = 2$

12. $(1 + t)y' + y = \cos(t)$, $y(05) = 1$

13. $y'' - 2y' + 2y = t + 1$, $y(0) = 3$, $y'(0) = 0$

14. $y'' + 3y' + 2y = e^t$, $y(0) = 0$, $y'(0) = 3$

15. In this Exercise we explore some of the differences between initial conditions and boundary conditions.

(a) Find the general solution to $y'' + y = 0$.

(b) Show that there is exactly one solution that satisfies the general pair of initial conditions: $y(t_0) = a$, $y'(t_0) = b$, for any values of t_0, a and b.

(c) Show that there is exactly one solution that satisfies the general pair of boundary conditions $y(0) = a$, $y(T) = b$ (for any values of a and b) only when T is not an integer multiple of π.

(d) Show that there is no solution that satisfies the general pair of boundary conditions $y(0) = a$, $y(T) = b$ when T is an integer multiple of π and $b \neq 0$.

(e) Show that there is an infinite number of solutions that satisfies the boundary conditions $y(0) = a$, $y(T) = b$ when T is an integer multiple of π and $b = 0$.

16. Assume that the outdoor temperature increases linearly as a function of t, $h(t) = t + 1$, for simplicity. The temperature of the house is $c < 0$ at time $t = 0$. Then it warms up by Newton's law. That is, if the temperature in the house at time t is $T(t)$, then $T'(t) = k(t - T(t))$.

(a) Find $T(t)$.

(b) Is the graph of $T(t)$ asymptotic to the graph of the outdoor temperature?

17. Let M and k be positive constants. Assume that the population $P(t)$ grows at a rate proportional to $M - P(t)$. That is, we assume

$$\frac{dP}{dt} = k(M - P(t)).$$

(a) Show that $Q(t) = M - P(t)$ grows or shrinks exponentially.

(b) Find $\lim_{t\to\infty} P(t)$.

(c) Interpret the constant M in terms of population.

18. This model has been seen before, in Exercises 35 to 37 in Section 5.7 and in the Hubbert's Peak CIE at the end of Chapter 10. Some species have a theoretical maximum sustainable population, which we call M. Assume the population changes at a rate proportional to itself, $P(t)$, and to the amount left to grow, $M - P(t)$. Then

$$\frac{dP}{dt} = kP(t)(M - P(t)). \tag{13.S.3}$$

Such a species is said to satisfy a **logistic growth** model.
Let M and k be positive constants.

(a) Show that $\frac{1}{P(M-P)}$ can be written as $\frac{1}{M}\left(\frac{1}{P} + \frac{1}{M-P}\right)$.

(b) Use separation of variables to find all solutions to

$$\frac{dP}{dt} = kP(M - P). \tag{13.S.4}$$

(c) Use your result in (b) to find the solution to the logistic growth model that satisfies the initial condition $P(0) = P_0$.

19. Consider the differential equation $y' = 1 - y^{10}$ for $t > 0$ with $y(0) = 0$. This differential equation is not linear.

(a) Is y necessarily bounded above and below? That is, must there be numbers m and M such that $m < y(t) < M$ for all values of t?

(b) Show that $|y| \leq 1$.

(c) Show that $\lim_{t\to\infty} y(t)$ exists and that it is 1.

(d) When is the curve concave up? concave down?

(e) What might the graph of the solution look like?

(f) Give one function that satisfies the equation.

The Method of Undetermined Coefficients is effective when the right-hand side is a sum of one of the special forms we have discussed. This method is useless if the right-hand side involves other functions, for example $\tan(t)$ or $\ln(t+1)$ or $\frac{1}{t^2+1}$. In such cases, the **Method of Variation of Parameters** is much more effective. Exercise 20 outlines the general process for Variation of Parameters and Exercise 21 works through Variation of Parameters for a specific example. Additional problems for which Variation of Parameters is appropriate can be found in Exercises 22 to 29.

20. The basic idea behind the Method of Variation of Parameters for $y'' + by' + cy = f(t)$ is as follows: Let y_1 and y_2 be solutions to the associated homogeneous equation. The general solution to the associated homogeneous equation is $y_h(t) = c_1y_1(t) + c_2y_2(t)$ for any constants c_1 and c_2. We seek a particular solution of the form $y_p(t) = u_1(t)y_1(t) + u_2(t)y_2(t)$ where the unknown coefficients, $u_1(t)$ and $u_2(t)$, are found by the following process:

(a) Compute y_p'. This expression has four terms; to simplify the computation of y_p'', assume $u_1'y_1 + u_2'y_2 = 0$.

(b) Compute y_p'' by differentiating the two-term expression for y_p' found in (a). This expression also has four terms.

(c) Show that when the expressions for y_p, y_p', and y_p'' are substituted into the nonhomogeneous equation the resulting equation is $u_1'(t)y_1'(t) + u_2'(t)y_2'(t) = f(t)$.

This has produced a system of two (linear) equations for $u_1'(t)$ and $u_2'(t)$. Once the solutions to this algebraic system are found, all that remains is to integrate $u_1'(t)$ to find $u_1(t)$ and to integrate $u_2'(t)$ to find $u_2(t)$.

21. Verify each steps involved in applying the Method of Variation of Parameters to $y'' + y = \tan(t)$.

(a) Show that the general solution to the associated homogeneous equation is $y_h = c_1 \sin(t) + c_2 \cos(t)$.

(b) Show that $y_p = u_1(t)\sin(t) + u_2(t)\cos(t)$ is a particular solution to the non-homogeneous equation when

$$u_1'(t)\sin(t) + u_2'(t)\cos(t) = 0 \qquad \text{and} \qquad u_1'(t)\cos(t) - u_2'(t)\sin(t) = \tan(t).$$

(c) Solve the system found in (b) to find

$$u_1'(t) = \sin(t) \qquad \text{and} \qquad u_2'(t) = -\tan(t)\sin(t).$$

(d) Integrate $u_1'(t)$ and $u_2'(t)$ to find

$$u_1(t) = -\cos(t) + K_1 \qquad \text{and} \qquad u_2(t) = \sin(t) - \ln|\sec(t) + \tan(t)|\,.$$

(e) Conclude that $y_p = -\cos(t)\ln|\sec(t) + \tan(t)|$.

In Exercises 22 to 29, use the Method of Variation of Parameters to find a particular solution of the given differential equation.

22. $y'' + 3y' + 2y = 4e^t$

23. $y'' - 2y' + y = t^{-1}e^t$

24. $y'' + 4y' + 4y = e^{-2t}\ln(t)$ $(t > 0)$

25. $y'' + 5y' + 6y = 108t^2$

26. $y'' + 2y' + y = e^{-t}$

27. $y'' + 4y = \sec^2(2t)$

28. $y'' + 9y = \tan^2(3t)$

29. $y'' - 5y' + 24y = 1331t^2e^{8t}$

Calculus is Everywhere # 16

Flow Through a Narrow Pipe: Poiseuille's Law

Poiseuille is pronounced *pwa-záy*

In the 1830's the physician Jean Poiseuille (1797–1869) investigated experimentally the flow of liquids in pipes whose diameters were as small as 0.015 mm to as large as 0.6 mm. His experiments were not easy to carry out. For instance, calibrating a single pipe could take as long as twelve hours. The motivation for this work was his desire to understand the flow of blood in arteries. In 1839 he deposited a sealed packet with the French Academy of Sciences describing his results. (This was the standard way to establish priority at that time.)

Poiseuille concluded that the flow is proportional to the fourth power of the inner radius (R^4), to the difference in pressures at the ends of the pipe (P), and inversely proportional to the length of the pipe (L), so proportional to R^4P/L. At the time it was generally felt that the flow would involve R^3, not R^4. However, Eduard Hagenbach in 1860 confirmed Poiseuille's conjecture, deriving his formula mathematically from basic physical principles. The next Calculus is Everywhere shows how Hagenbach did it.

We will obtain Poiseuille's Law, starting with a form of the Navier-Stokes equation, a fundamental equation in the study of fluid dynamics.

Let R be the inner radius of the pipe. One might expect the flow to be proportional to the cross-sectional area πR^2, hence to the square of the radius. That assumes all the fluid flows at the same speed. But that is not the case. Fluid along the surface does not move at all and the highest velocities are found along the axis. Denoting the distance from the axis to the pipe by r, $0 \leq r \leq R$, we let $v(r)$ be the velocity of the fluid a distance r from the axis. In particular, $v(R) = 0$ and, because the maximum velocity occurs along the axis, $v'(0) = 0$.

Let μ denote the viscosity of the fluid (large for oil, small for water, in between for blood), and A be the pressure gradient, defined as P/L. We will find how $v(r)$ depends on r. Once we know that, it will be be easy to measure the flow.

Our starting point is Newton's Third Law of Motion. In this case, balancing the forces of pressure and viscosity results in the following differential equation:

$$\mu \frac{d^2v}{dr^2} + \frac{\mu}{r}\frac{dv}{dr} = -A. \tag{C.16.1}$$

This equation is linear but not all of its coefficients are constants. The methods of Sections 13.3 and 13.4 do not apply.

Note that the function v does not appear in (C.16.1); only its first and second derivatives play a role. So, letting $w = \frac{dv}{dr}$ (and dividing by μ), we can rewrite (C.16.1) as:

$$\frac{dw}{dr} + \frac{w}{r} = -\frac{A}{\mu}. \tag{C.16.2}$$

The first-order differential equation (C.16.2) is linear. As such it could be solved by the methods of Section 13.2. (See Exercise 5.)

The specific combinations of terms on the left-hand side of (C.16.1) is quite common in problems that involve viscosity. In such settings it is typical to take advantage of the fact that

$$\frac{dw}{dr} + \frac{w}{r} = \frac{1}{r}\frac{d}{dr}\left(rw(r)\right).$$

Thus, (C.16.2) can be rewritten as

$$\frac{1}{r}\frac{d}{dr}\left(rw(r)\right) = \frac{-A}{\mu}. \tag{C.16.3}$$

Multiplying this equation by r results in

$$\frac{d}{dr}(rw(r)) = \frac{-A}{\mu}r,$$

which is easily integrated with respect to r. Thus there is a constant K such that

$$rw(r) = \frac{-A}{2\mu}r^2 + K \qquad \text{for all } r \text{ between 0 and } R. \tag{C.16.4}$$

Note that if $w(r)$ grows fast enough as r approaches 0 (from the right), the term $rw(r)$ might not have the value 0 when $r = 0$. But, in this case we know $w(0) = v'(0) = 0$ and so $\lim_{r\to 0^+}(rw(r)) = 0$. (See also Exercise 8.) This is enough to allow us to conclude that $K = 0$. Thus, replacing $w(r)$ with $\frac{dv}{dr}$ brings us to consider

$$\frac{dv}{dr} = \frac{-A}{2\mu}r.$$

Another integration with respect to r produces

$$v(r) = \frac{-Ar^2}{4\mu} + Q \qquad \text{for some constant } Q. \tag{C.16.5}$$

When $r = R$, v is 0, so (C.16.5) implies that

$$0 = \frac{-AR^2}{4\mu} + Q$$

and we have $Q = \frac{AR^2}{4\mu}$. Thus (C.16.5) becomes

$$v(r) = \frac{-A}{4\mu}r^2 + \frac{A}{4\mu}R^2.$$

In conclusion, the velocity of the fluid in the pipe is given by

$$v(r) = \frac{A}{4\mu}\left(R^2 - r^2\right). \tag{C.16.6}$$

Now that we know how the velocity varies with distance from the axis, we can calculate the total flow of the liquid.

Think of a cross section of the liquid perpendicular to the axis of the pipe as a disk composed of narrow rings of width dr. A typical ring is shown in Figure C.16.1. The area of this typical ring is approximately $2\pi r\ dr$. Combining this with (C.16.6) shows that the rate at which fluid crosses this ring is approximately

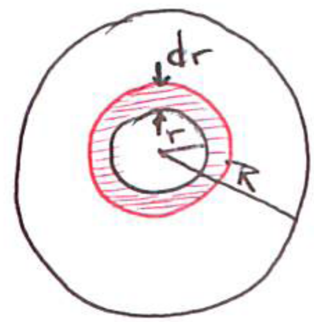


Figure C.16.1

$$\begin{aligned}\text{(velocity)(area of ring)} &= \frac{A}{4\mu}\left(R^2 - r^2\right)(2\pi r\ dr) = \frac{\pi A}{2\mu}\left(R^2 r - r^3\right) \\ &= \frac{\pi A}{2\mu}\left(R^2 r - r^3\right)\ dr.\end{aligned}$$

This fluid moves past the entire disk at the rate

$$\begin{aligned}\int_0^R \frac{\pi A}{2\mu}\left(R^2 r - r^3\right)\ dr &= \frac{\pi A}{2\mu}\int_0^R \left(R^2 r - r^3\right) dr = \frac{\pi A}{2\mu}\left(\frac{R^2 r^2}{2} - \frac{r^4}{4}\right)\Bigg|_0^R \\ &= \frac{\pi A}{2\mu}\left(\frac{R^4}{2} - \frac{R^2}{4}\right) = \frac{\pi A R^4}{8\mu} = \frac{\pi R^4 P}{8\mu L}.\end{aligned}$$

This tells us that the flow is proportional to R^4P/L and Poiseuille's experimentally-based conjecture about the influence of radius, length, and pressure is a consequence of basic physical principles.

Note the strong influence of R^4. Reducing the radius from R to $R/2$, say, reduces the flow by a factor of $2^4 = 16$. Even reducing R by just 20% cuts the flow by almost 60%. That is why "narrowing of the arteries" as a result of atherosclerosis is such a serious medical condition.

Poiseuille's Law also explains how the muscles that circle the arteries can control blood flow by a slight tightening or relaxing.

For a detailed and fascinating description of Poiseuille's experiments, see *The History of Poiseuille's Law*, by Salvatore P. Sutera and Rickard Skalak in The Annual Review of Fluid Mechanics **25** (1993), 1–19.

EXERCISES

An easy way to solve a first-order linear differential equations of the form

$$\frac{dy}{dt} + \frac{k}{t}y = q(t) \qquad \text{with } k \text{ a constant}$$

is to multiply the equation by t^k and then to note that the two terms on the left-hand side are the derivative of $t^k y(t)$. Thus the equation can be rewritten as

$$\frac{d}{dt}\left(t^k y(t)\right) = t^k q(t).$$

Then, the explicit solution is found by integrating both sides and by dividing by t^k. Use this method to find the general solution of the differential equation found in Exercises 1 to 4.

1. $y' + \frac{3}{t}y = 3t$

2. $y' - \frac{1}{2t}y = 1$

3. $y' + \frac{1}{t}y = e^t$

4. $y' - \frac{4}{t}y = t^4 + 2t^3 + 1$

5. In this Exercise we present a different method for solving (C.16.1) with $v(R) = 0$.

(a) Find the general solution of (C.16.2) by the methods described in Section 13.2.

(b) Find the general solution of (C.16.1) by integrating the solution found in (a).

(c) Find all solutions of (C.16.1) that also satisfy $v(R) = 0$.

(d) The solution in (c) should still involve one constant. Find the value of this constant that makes the solution well-defined (that is, is finite) for $r = 0$.

(e) Does the solution found in (d) agree with (C.16.6)? Be sure to explain any differences.

6.

(a) Why is the area of the ring in Figure C.16.1 approximately $2\pi r\ dr$?

(b) What is the exact area of that ring?

7. We assumed that the maximum velocity occurred at the axis in order to assert that $w(0) = 0$. That conclusion can be deduced from our assumption that the velocity depends only on r. To see why, imagine an r-axis perpendicular to the axis of the pipe with its origin at the axis. Imagine graphing v as a function of r for r in $[-R, R]$. What property of the graph implies that $v'(0) = 0$? (Recall that $v' = \frac{dv}{dr}$.)

8. In the context of flow through a narrow pipe, axial symmetry of the velocity was used to conclude that $w(0) = 0$.

(a) Give an example of a function $w(r)$, defined for $r > 0$, for which $rw(r)$ approaches 1 as r approaches 0 (through positive values).

(b) Let k be a non-zero number. Modify your answer to (a) to produce a function $w(r)$ such that $\lim_{r \to 0^+} rw(r) = k$.

(c) Could it happen that $w(r)$ is unbounded near 0, yet $rw(r)$ still approaches 0 as r approaches 0? Justify your answer.

9. We showed that the velocity is proportional to $R^2 - r^2$. If, instead, the velocity is proportional to $R - r$, what power of R would appear in the formula for the flow (instead of R^4)?

10. Let $v(r, R)$ be the velocity of the fluid at a distance r from the axis of a pipe of inner radius R. Assume that there is a differentiable function $f(r)$ such that $v(r, R) = f(R) - f(r)$. Assume that there are constants k and $m \neq 2$ such that for all R:

$$\int_0^R (f(R) - f(r))\, r \; dr = kR^m. \tag{C.16.7}$$

(a) Show that $v(R, R) = 0$.

(b) Show that $v(r, R) = \frac{2km}{m-2}\left(R^{m-2} - r^{m-2}\right)$.

Calculus is Everywhere # 17

Origin of the Equation for Flow in a Narrow Pipe

The preceding section, *Flow Through a Narrow Pipe: Poiseuille's Law*, obtained the fourth-order law for the rate at which fluid flows through a narrow pipe. It depends on finding the solutions to the differential equation

$$\mu \frac{d^2 v}{dr^2} + \frac{\mu}{r}\frac{dv}{dr} + A = 0. \tag{C.17.1}$$

In this section we obtain (C.17.1) from basic physical assumptions.

The Physics

In his monumental *Mathematical Principles of Natural Philosophy and this System of the World*, usually referred to as "The Principia," Newton stated three laws of motion. The second and third laws of motion will be used in the present discussion.

Newton's second law states that if a moving object undergoes no acceleration, then the total force operating on that object is zero.

The third law reads, "To every action there is always opposed an equal reaction; or, the mutual actions of two bodies upon each other are always equal, and directed to contrary parts."

He follows this law by a commentary that begins, "Whatever draws of presses another is as much drawn or pressed by that other. If you press a stone with your finger, the finger is also pressed by the stone. If a horse draws a stone tied to a rope, the horse (if I may so say) will be equally drawn back towards the stone."

We apply the third law immediately in the case of fluid flow.

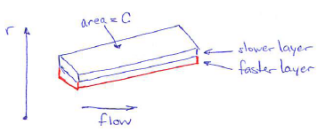


Figure C.17.1

Imagine two thin planar layers of fluid moving from left to right, as in Figure C.17.1. Fluid in the lower one is moving faster than the fluid in the upper one. The area of contact we call C. The velocity v of the fluid depends on r, shown in Figure C.17.1. We are assuming that v is a decreasing function of r.

The faster fluid exerts a force F_1 on the slower fluid, tending to speed it up. The slower fluid exerts a force R_2 on the faster fluid, tending to slow it down. By Newton's third law these two forces are equal but in opposite directions: $F_2 = -F_1$.

The magnitude of F_1 and F_2 is proportional to the area C. It also is proportional to the difference in their velocities, which we will measure by the antiderivative $\frac{dv}{dr}$, viewing v as a function of r, as shown in Figure C.17.1. Thus

$$F_1 \text{ and } F_2 \text{ is proportional to } C \text{ times } \frac{dv}{dr}.$$

The constant of proportionality depends on the particular fluid. Calling this positive constant μ and recalling that v is a decreasing function of r , so $\frac{dv}{dr}$ is negative, we have

$$F_1 = -\mu C \frac{dv}{dr}. \tag{C.17.2}$$

By Newton's third law,

$$F_2 = \mu C \frac{dv}{dr}. \tag{C.17.3}$$

Formulas (C.17.2) and (C.17.3) hold also when the layers are curved, for instance, when they are two concentric thin pipes, the case we will be using.

The constant μ is called the **viscosity** of the fluid. The larger it is the more drag or pull one layer exerts on the other.

Viscosity is a measure of the internal friction of the fluid. The higher the viscosity, the harder it is to make the fluid flow. In one system of units, water has a low viscosity, 0.894, while olive oil has viscosity 81, and blood about 3.5. Temperature affects viscosity. For instance, honey has a very high viscosity at room temperature but flows easily at high temperature.

The Mathematics

Now that we have the necessary physical principles, we are ready to consider a fluid moving from left to right through a narrow cylindrical pipe of inner radius R and length L, as in Figure C.17.2. Fluid at a distance r from the axis has the velocity $v = v(r)$ for $0 \leq r \leq R$, a decreasing function of r.

Imagine breaking up the cylinder of radius R and length L into thin concentric pipes or straws, as we did with the shell technique in Section 7.5. Figure C.17.3 shows one such pipe. Its inner radius is r and its outer radius is $r + \Delta r$.

The flow in this pipe is affected by the flow in the two pipes adjacent to it, shown in Figure C.17.4.

Four forces act on the middle pipe: the force at the left end, the force at the right end, the drag to the left due to slower fluid in the outer pipe, and the pull to the right of the faster fluid in the inner pipe. We calculate each of these forces.

Let P_1 be the pressure at the left end. The force against this end is the product of the pressure P_1 there and the area of the base of the middle pipe,

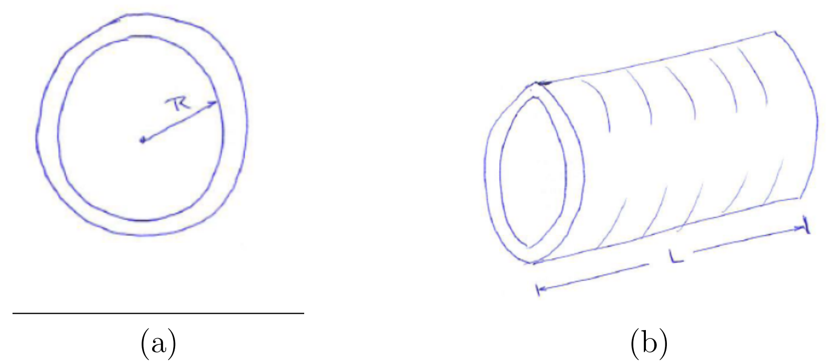


Figure C.17.2 (a) End view and (b) side view of a cylindrical pipe with radius R and length L.

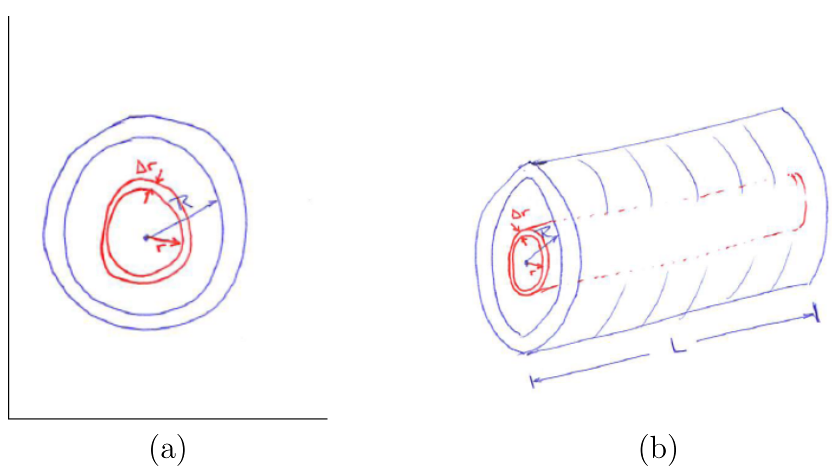
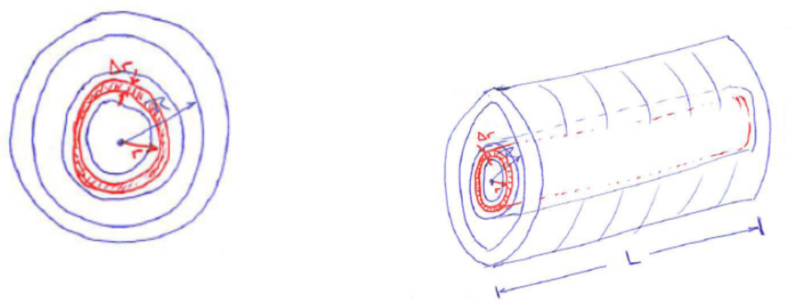


Figure C.17.3 (a) End view and (b) side view of a thin shell with thickness Δr cut from the cylindrical pipe with radius R and length L.



(a) (b)

Figure C.17.4 (a) End view and (b) side view of a cylindrical pipe with radius R and length L.

which is $2\pi\left(r+\frac{\Delta r}{2}\right)\ \Delta r$, as Exercise 1 shows. Thus this force is

$$P_1\left(2\pi\left(r+\frac{\Delta r}{2}\right)\ \Delta r\right).$$

Let the lower pressure at the right end be P_2. The force against the right end of the middle pipe is, similarly,

$$-P_2\left(2\pi\left(r+\frac{\Delta r}{2}\right)\ \Delta r\right).$$

Denoting the difference of the two pressures, P_1-P_2, by P, we see that the net force against the two ends of the middle pipe is

$$P\left(2\pi\left(r+\frac{\Delta r}{2}\right)\ \Delta r\right).$$

The force of the inner pipe on the middle pipe is

$$-\mu\,(\text{Area Shaded}) = -\mu\,(2\pi rL)\,\frac{dv}{dr},$$

where $\frac{dv}{dr}$ is evaluated at r. The minus sign is inserted because $\frac{dv}{dr}$ is negative.

The force of the outer pipe on the middle pipe is

$$2\pi\,(r+\Delta r)\,L\frac{dv}{dr}\ \ (r+\Delta r)\,. \tag{C.17.4}$$

Note that the derivative is evaluated at $r+dr$.

Because the fluid is moving at a steady rate, not being accelerated, the sum of these forces is, by Newton's Law, zero:

$$0 = P(2\pi)\left(r + \frac{\Delta r}{2}\right)\Delta r - \mu(2\pi r L)\frac{dv}{dr} + \mu(2\pi)(r+\Delta r)L\frac{dv}{dr}(r+\Delta r). \quad \text{(C.17.5)}$$

Division by $2\pi L$ yields this simpler equation:

$$\text{(C.17.6)}$$

Using the Taylor polynomial of degree one, the linear approximation, we have the estimate:

$$\frac{dv}{dr}(r + \Delta r) \approx \frac{dv}{dr} + \frac{d^2v}{dr^2}\Delta r, \quad \text{(C.17.7)}$$

where both $\frac{dv}{dr}$ and $\frac{d^2v}{dr^2}$ are evaluated at r. So we replace (C.17.6) by

$$0 = \frac{P}{L}\left(r + \frac{\Delta r}{2}\right)\Delta r - \mu r\frac{dv}{dr} + \mu(r + \Delta r)\left(\frac{dr}{dr} + \frac{d^2v}{dr^2}\right)\Delta r. \quad \text{(C.17.8)}$$

Expanding the product on the right-hand sde of (C.17.8) gives

$$0 = \frac{P}{L}\left(r + \frac{\Delta r}{2}\right)\Delta r - \mu r\frac{dv}{dr} + \mu r\frac{dr}{dr} + \mu r\frac{d^2v}{dr^2} + \mu\frac{dv}{dr}\Delta r + \mu\frac{d^2v}{dr^2}(\Delta r)^2. \quad \text{(C.17.9)}$$

When Δr is small, $(\Delta r)^2$ is much smaller than Δr. Therefore we may omit it, and, after a cancellation, reach:

$$0 = \frac{P}{L}r\ \Delta r + \mu r\frac{d^2v}{dr^2}\Delta r + \mu\frac{dv}{dr}\Delta r. \quad \text{(C.17.10)}$$

Dividing (C.17.10) by r gives us essentially (C.17.1), for $A = P/L$. This was our goal.

EXERCISES

1. Show that the area of a ring of inner radius r and outer radius $r+\Delta r$ is exactly $2\pi\left(r + \frac{\Delta r}{2}\right)\Delta r$.

2. Often $2\pi r\Delta r$ is used as an approximation of the area of the ring in Example 1. Would the argument that derives (C.17.1) still go through with this approximation? Explain.

3. Why is there no minus sign in (C.17.2).

4. Decude (C.17.10) from (C.17.9).

5. Check the claim in the final sentence.

6. Show that (C.17.1) can be written as $\frac{\mu}{r}\frac{d}{dx}\left(r\frac{dv}{dr}\right)+A=0$.

7. In a Taylor expansion the coefficient of the second derivative is $\frac{1}{2}$. However, in (C.17.7) the coefficient of the second derivative is not $\frac{1}{2}$. Why not?

Chapter 14

Vectors

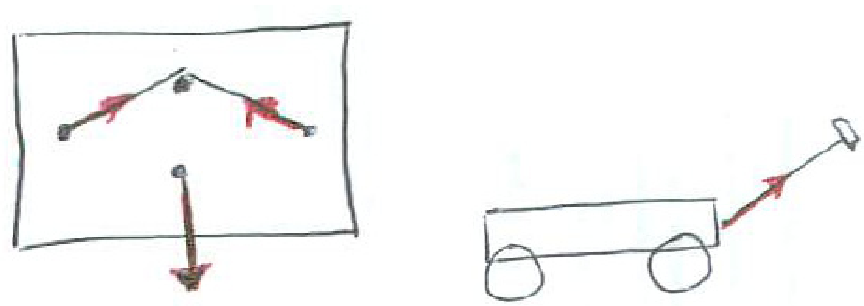
This chapter is part of algebra, not calculus because it involves no limits, derivatives, or integrals.

Section 14.1 introduces vectors, that are usually pictured as arrows. Section 14.2 examines the dot product, a number associated with a pair of vectors. Section 14.3 defines the cross product, a vector perpendicular to two vectors. Applications of vectors and the dot product in Section 14.4 include finding the distance from a point to a line or plane, and giving a parametric description of a line.

14.1 The Algebra of Vectors

SHERMAN: Figure 2 is too wide to put in margin. Text is too brief to have both parts of Figure 1 in margin. Instructions to MAA should mention need to keep this short?

When you hang a picture on wire you deal with three vectors: one describing the downward force of gravity and two describing the force of the wires pulling up, as in Figure 14.1.1(a)



(a) (b)

Figure 14.1.1

When you pull a wagon the force you use is represented by a vector, as in Figure 14.1.1(b). The harder you pull, the larger the vector.

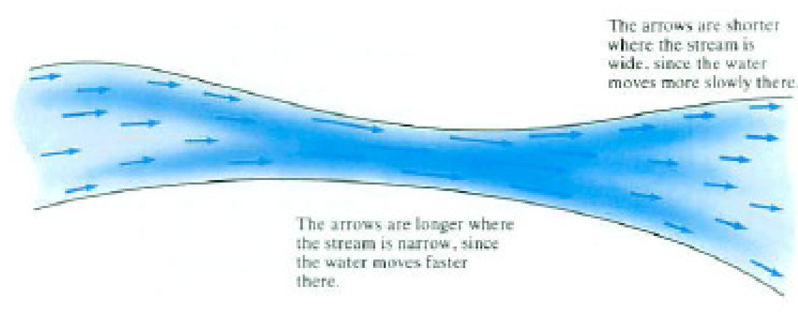


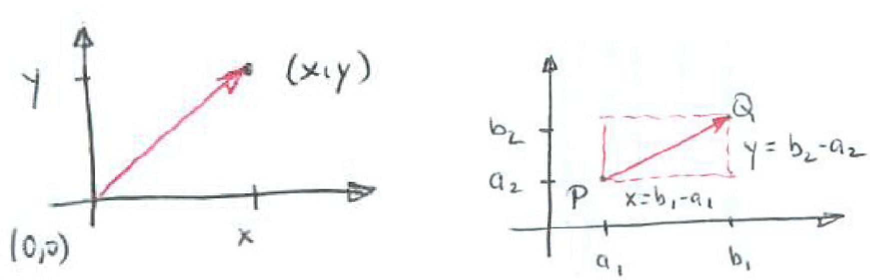
Figure 14.1.2

A vector has a direction and a magnitude. You may think of it as an arrow, whose length and direction carry information. Vectors are of use in describing the flow of a fluid, as in Figure 14.1.2, or the wind, or the strength and direction of a magnetic field.

Vectors in the Plane

A vector in the xy-plane is an ordered pair of numbers x and y, denoted $\langle x, y\rangle$. Its magnitude, or length, is $\sqrt{x^2+y^2}$. Though the notation resembles that for a point, (x, y), we treat vectors differently. We can add them, subtract them and multiply them by a number.

We can represent a vector by an arrow whose tail is at $(0, 0)$ and whose head (or tip) is at (x, y), as in Figure 14.1.3(a).



(a) (b)

Figure 14.1.3 (a) The arrow represents the vector $\langle x, y\rangle$.

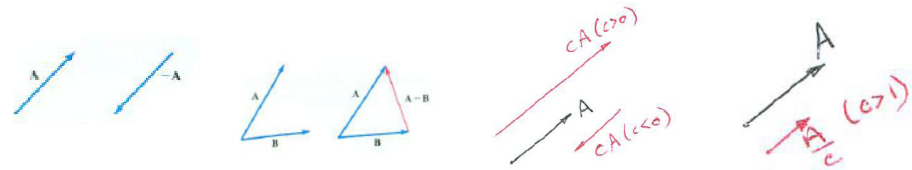
More generally, we can represent $\langle x, y\rangle$ using points $P = (a_1, a_2)$ and $Q = (b_1, b_2)$ if $b_1 - a_1 = x$ and $b_2 - a_2 = y$, as in Figure 14.1.3(b).

We speak then of the vector from P to Q and denote it $\overline{PQ}$. A vector $\langle x, y\rangle$ will be written with bold-face letters, such as **A**, **B**, **r**, **v**, and **a**. In writing they are denoted by using a bar on top, for instance $\overline{AB}$. A vector of length 1 is called a **unit vector** and is topped with a little hat, as in $\widehat{\mathbf{r}}$, which is read as "r hat".

Here is how we operate on vectors. Let $\mathbf{A} = \langle a_1, a_2\rangle$ and $\mathbf{B} = \langle b_1, b_2\rangle$ be vectors and let c be a number.

(a) (b) (c) (d)

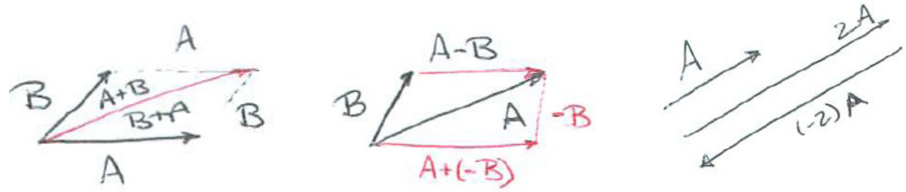
Figure 14.1.4



(a) (b) (c) (d)

Figure 14.1.5

Operation	Definition	Geometry	Comment
$\mathbf{A}+\mathbf{B}$	$\langle a_1+b_1, a_2+b_2\rangle$	Figure 14.1.4	The tail of **B** is placed at the head of **A**
$-\mathbf{A}$	$\langle -a_1, -a_2\rangle$	Figure 14.1.5(a)	$-\mathbf{A}$ points in opposite direction of **A**
$\mathbf{A}-\mathbf{B}$	$\langle a_1-b_1, a_2-b_2\rangle$	Figure 14.1.5(b)	What you add to **B** to get **A**
$c\mathbf{A}$	$\langle ca_1, ca_2\rangle$	Figure 14.1.5(c)	Same direction as **A** and $\lvert c\rvert$ times as long as **A**
$\frac{\mathbf{A}}{c}$	$\left\langle \frac{a_1}{c}, \frac{a_2}{c}\right\rangle$	Figure 14.1.5(d)	Same direction as **A** and $\frac{1}{\lvert c\rvert}$ times as long as **A** $(c \neq 0)$



(a) (b) (c)

Figure 14.1.6

The operation of addition obeys the usual rules of addition of numbers, so $\mathbf{A}+\mathbf{B} = \mathbf{B}+\mathbf{A}$ and $\mathbf{A}+(\mathbf{B}+\mathbf{C}) = (\mathbf{A}+\mathbf{B})+\mathbf{C}$. Figure 14.1.6(a) shows both $\mathbf{A}+\mathbf{B}$ and $\mathbf{B}+\mathbf{A}$, they are equal. In terms of arrows it makes sense. See Figure 14.1.6(a).

Also, $\mathbf{A}-\mathbf{B} = \mathbf{A}+(-\mathbf{B})$ follows from the definitions. This property can also be seen in Figure 14.1.6(b) since $\mathbf{A}-\mathbf{B}$ and $\mathbf{A}+(-\mathbf{B})$ appear as opposite sides of a parallelogram.

When $\mathbf{A} = \langle x, y\rangle$ is a vector and c is a positive constant, $c\mathbf{A}$ is a multiple of **A** that points in the same direction as **A**. If $c > 1$, then $c\mathbf{A}$ is longer than **A** and if $0 < c < 1$, then $c\mathbf{A}$ is shorter than **A**. If $c < 0$, then $c\mathbf{A}$ is the opposite of $|c|\mathbf{A}$. See Figure 14.1.6(c).

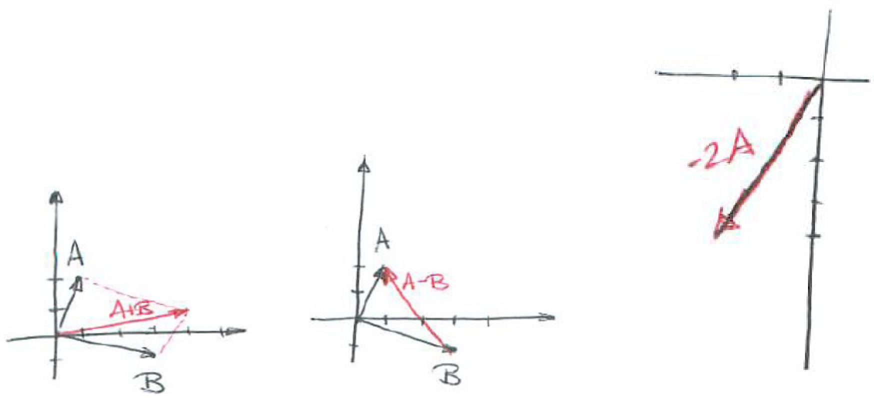
When talking about numbers, such as c, x, and y, in the context of vectors, we call them **scalars**. Thus in $c\mathbf{A}$ the scalar c is multiplying the vector $\mathbf{A}$. Two vectors $\mathbf{A}$ and $\mathbf{B}$ are **parallel** if and only if there is a non-zero scalar c such that $\mathbf{A} = c\mathbf{B}$.

The vector $\langle 0, 0\rangle$ is denoted $\mathbf{0}$ and is called the **zero vector**.

EXAMPLE 1 Let $\mathbf{A} = \langle 1, 2\rangle$, $\mathbf{B} = \langle 3, -1\rangle$ and $c = -2$. Compute $\mathbf{A}+\mathbf{B}$, $\mathbf{A}-\mathbf{B}$, and $c\mathbf{A}$. Draw the corresponding arrows.
SOLUTION

$$\begin{aligned}
A+B &= \langle 1,2\rangle + \langle 3,-1\rangle = \langle 1+3, 2+(-1)\rangle = \langle 4,1\rangle \\
A-B &= \langle 1,2\rangle - \langle 3,-1\rangle = \langle 1-3, 2-(-1)\rangle = \langle -2,3\rangle \\
cA &= -2\langle 1,2\rangle = \langle -2(1), -2(2)\rangle = \langle -2,-4\rangle
\end{aligned}$$

The vectors $\mathbf{A}-\mathbf{B}$ and $\mathbf{A}+\mathbf{B}$ lie on the diagonals of a parallelogram. See



(a) (b) (c)

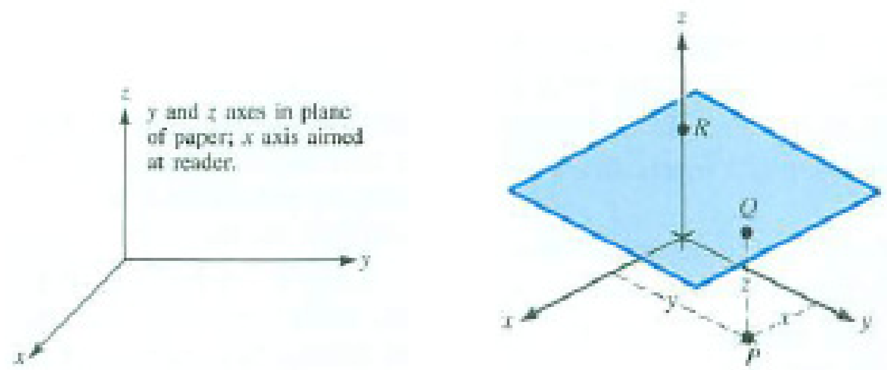
Figure 14.1.7

Figure 14.1.7. ◇

Before we can make definitions for vectors in space, we must introduce an appropriate coordinate system.

Coordinates in Space

Pick a pair of perpendicular intersecting lines to serve as the x- and y-axes. The positive parts of these axes are indicated by arrows. These two lines determine the xy-**plane**. The line perpendicular to the xy-plane and meeting the x- and y-axes at $(0, 0)$ will be called the z-axis. The point where the three axes meet is called the **origin**. The 0 of the z-axis is at the origin. But which half of the z-axis will have positive numbers and which half will have the negative numbers? It is customary to determine this by the **right-hand rule**. Moving in the xy-plane through a right angle from the positive x-axis to the positive y-axis determines a sense of rotation around the z-axis. If the fingers of the right hand curl in that sense, the thumb points in the direction of the positive z-axis, as shown in Figure 14.1.8(a).



(a) (b)

Figure 14.1.8

A point Q in space is now described by three numbers. Two specify the x- and y-coordinates of the point P in the xy-plane directly below (or above) Q and then the height of Q above (or below) the xy-plane is recorded by the z-coordinate of the point R where the plane through Q and parallel to the xy plane meets the z-axis. The point Q is then denoted (x, y, z). See Figure 14.1.8(b).

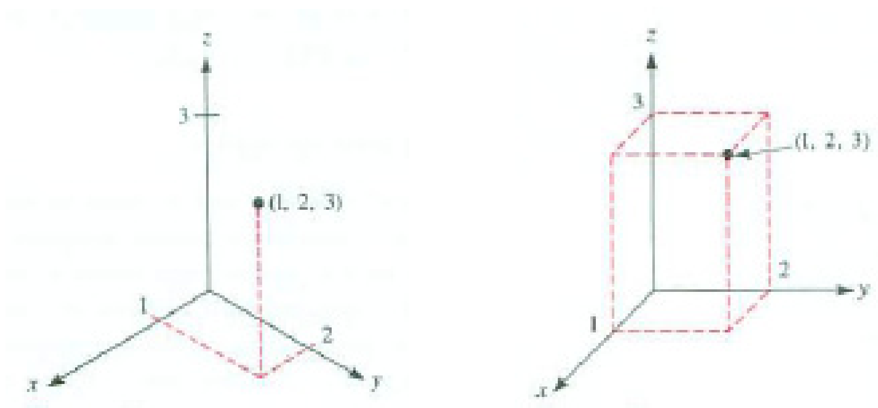
Points (x, y, z) for which $z = 0$ lie in the xy-plane. The points (x, y, z) for which $x = 0$ lie in the plane determined by the y-and z-axes, which is called the yz-**plane**. Similarly, the equation $y = 0$ describes the xz-**plane**. The xy-, xz-, and yz-planes are called the **coordinate planes**.

EXAMPLE 2 Plot the point $(1, 2, 3)$.

SOLUTION One way is to first plot $(1, 2)$ in the xy-plane. Then, on a line perpendicular to the xy-plane at that point, show the point $(1, 2, 3)$ as in Figure 14.1.9(a).

Another way is to draw a box whose edges are parallel to the axes and has the origin $(0, 0, 0)$ and $(1, 2, 3)$ as corners as in Figure 14.1.9(b). $\diamond$

The axes in the xy-plane divide the plane into four quadrants, and the coordinate planes divide space into eight octants.



(a) (b)

Figure 14.1.9

Vectors in Space

The only difference between a vector in space and a vector in the xy-plane is that it has three components, x, y, and z, and is written $\langle x, y, z\rangle$. Its length or magnitude is defined as $\sqrt{x^2+y^2+z^2}$. The definitions of the sum and difference are so similar to the definitions for planar vectors that we will not list them. For instance, $\langle a_1, a_2, a_3\rangle + \langle b_1, b_2, b_3\rangle$ is $\langle a_1+b_1, a_2+b_2, a_3+b_3\rangle$. They are harder to draw, even though they can be suggested by an arrow. It may help to visualize a three-dimensional vector by drawing a box in which it is a main diagonal. For instance, to draw the vector $\langle 2, 3, -1\rangle$ draw the box shown in Figure 14.1.10

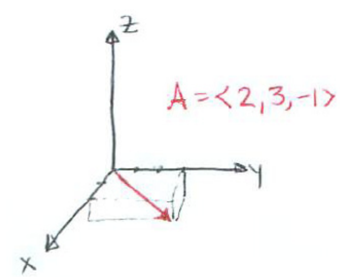


Figure 14.1.10

The Standard Unit Vectors

Three unit vectors indicate the directions of the positive x-, y-, and z-axes. They will be denoted $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$, respectively. For instance, $\mathbf{i} = \langle 1, 0, 0\rangle$. The vector $\langle x, y, z\rangle$ can thus be written as $x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$.

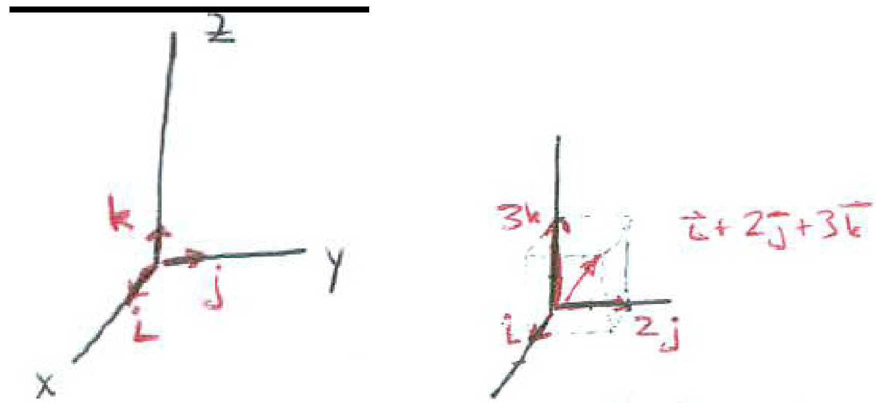
EXAMPLE 3 Draw $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, and $\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$.
SOLUTION Figure 14.1.11(a) shows $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ and Figure 14.1.11(b) shows $\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$. ◇

The magnitude of $\mathbf{A}$ is indicated by $|\mathbf{A}|$, a scalar.

Let $\mathbf{A}$ be a non-zero vector. Then $\frac{\mathbf{A}}{|\mathbf{A}|}$ is a unit vector in the direction of $\mathbf{A}$ because $\mathbf{A}/|\mathbf{A}|$ is the same as the vector $\frac{1}{|\mathbf{A}|}\mathbf{A}$, which has length $\frac{1}{|\mathbf{A}|}|\mathbf{A}| = 1$.

Example 4 shows how vectors can be used to establish geometric properties.

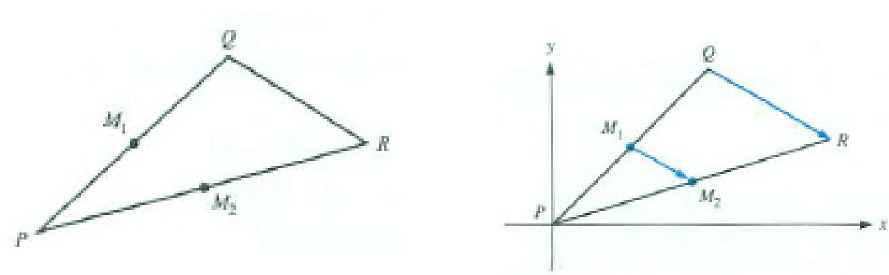
EXAMPLE 4 Prove that the line that joins the midpoints of two sides of a triangle is parallel to the third side and half as long.



(a) (b)

Figure 14.1.11

SOLUTION Let the triangle have vertices P, Q, and R. Let the midpoint of side PQ be M and the midpoint of side PR be N as in Figure 14.1.12(a).



(a) (b)

Figure 14.1.12

Introduce an xy-coordinate system in the plane of the triangle. Its origin could be anywhere in the plane, but we put it at P to simplify the calculations. (See Figure 14.1.12(b).)

We wish to show that the vector $\overline{MN}$ is $\frac{1}{2}\overline{QR}$. We compute $\overline{MN}$ and $\overline{QR}$ in terms of vectors involving P, Q, and R.

Because $\overline{PM} = \frac{1}{2}\overline{PQ}$ and $\overline{PN} = \frac{1}{2}\overline{PR}$ we have

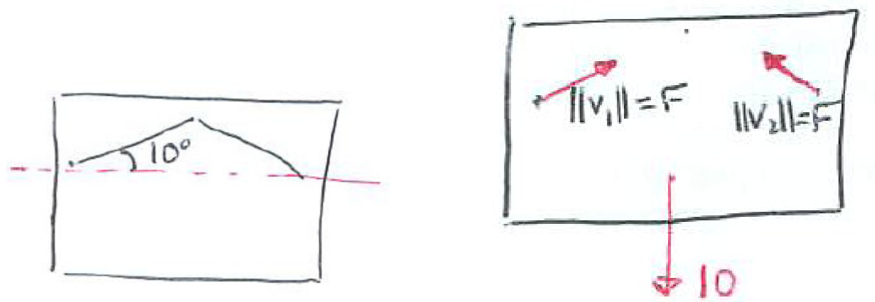
$$\overline{MN} = \frac{1}{2}\overline{PR} - \frac{1}{2}\overline{PQ} = \frac{1}{2}(\overline{PR} - \overline{PQ}) = \frac{1}{2}(\overline{QR}).$$

◇

The next example shows the importance of thinking vectorally. Not thinking that way, one of us had a picture frame fall and break a vase.

EXAMPLE 5 A picture weighing 10 pounds has a wire on the back, which rests on a picture hook, as shown in Figure 14.1.13(a). Find the force (tension) on the wire.

SOLUTION There are three vectors involved. One is straight down, with magnitude 10 pounds and two, v_1 and v_2, are along the wire, with unknown magnitude F: $|\mathbf{v}_1| = F = |\mathbf{v}_2|$.



(a) (b)

Figure 14.1.13

To balance the downward force of gravity, each end of the wire must have a vertical component of 5 pounds. Since the angle with the horizontal is 10° we must have $F\sin(10°) = 5$ or $F = 5/\sin(10°) \approx 29$ pounds. That is greater than the weight of the painting and can pull the screws out of the frame. $\diamond$

Summary

We introduced the notion of vectors $\langle x, y\rangle$ in the xy-plane or $\langle x, y, z\rangle$ in space and defined their addition, subtraction, and we defined the operation of a scalar c on a vector $\langle x, y, z\rangle$, as $\langle cx, cy, cz\rangle$.

We visualized vectors with the aid of arrows, which could be drawn anywhere in the xy-plane or in space.

Each vector in the xy-plane can be written as $x\mathbf{i} + y\mathbf{j}$. Vectors in space can be written as $x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$.

EXERCISES for Section 14.1

1. Draw the vector $2\mathbf{i}+3\mathbf{j}$, placing its tail at (a) $(0,0)$, (b) $(-1,2)$, (c) $(1,1)$.

2. Draw the vector $-\mathbf{i}+2\mathbf{j}$, placing its tail at (a) $(0,0)$, (b) $(3,0)$, (c) $(-2,2)$.

In Exercises 3 to 6 draw the vector $\mathbf{A}$ and enough extra lines to show how it is situated in space.

3. $\mathbf{A}=2\mathbf{i}+\mathbf{j}+3\mathbf{k}$,

(a) tail at $(0,0,0)$

(b) tail at $(1,1,1)$

4. $\mathbf{A}=\mathbf{i}+\mathbf{j}+\mathbf{k}$

(a) tail at $(0,0,0)$

(b) tail at $(2,3,4)$

5. $\mathbf{A}=-\mathbf{i}-2\mathbf{j}+2\mathbf{k}$

(a) tail at $(0,0,0)$

(b) tail at $(1,1,-1)$

6. $\mathbf{A}=\mathbf{j}+\mathbf{k}$

(a) tail at $(0,0,0)$

(b) tail at $(-1,-1,-1)$

In Exercises 7 to 10 plot the points P and Q, draw the vector $\overline{PQ}$, express it in the form $x\mathbf{i}+y\mathbf{j}+z\mathbf{k}$, and find its length.

7. $P=(0,0,0), Q=(1,3,4)$

8. $P=(1,2,3), Q=(2,5,4)$

9. $P=(2,5,4), Q=(1,2,2)$

10. $P = (1, 1, 1), Q = (-1, 3, -2)$

In Exercises 11 and 12 express the vector $\mathbf{A}$ in the form $x\mathbf{i} + y\mathbf{j}$. North is along the positive y-axis and east is along the positive x-axis.

11.

(a) $|\mathbf{A}| = 10$ and $\mathbf{A}$ points northwest

(b) $|\mathbf{A}| = 6$ and $\mathbf{A}$ points south

(c) $|\mathbf{A}| = 9$ and $\mathbf{A}$ points southeast

(d) $|\mathbf{A}| = 5$ and $\mathbf{A}$ points east

12.

(a) $|\mathbf{A}| = 1$ and $\mathbf{A}$ points southwest

(b) $|\mathbf{A}| = 2$ and $\mathbf{A}$ points west

(c) $|\mathbf{A}| = \sqrt{8}$ and $\mathbf{A}$ points northeast

(d) $|\mathbf{A}| = 1/2$ and $\mathbf{A}$ points south

13. The wind is 30 miles per hour to the northeast. An airplane is traveling 100 miles per hour relative to the air, and the vector from the tail of the plane to its front tip points to the south. (See Figure 14.1.14.)

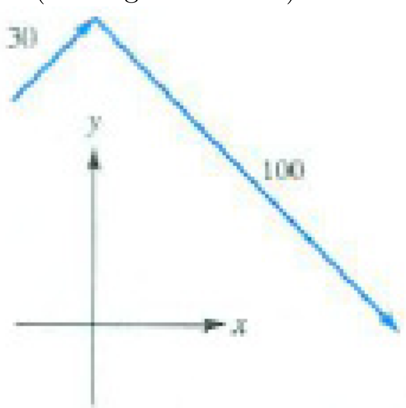


Figure 14.1.14

(a) What is the speed of the plane relative to the ground?

(b) What is the direction of the flight relative to the ground?

14. (See Exercise 13.) The jet stream is moving 200 miles per hour to the southeast. A plane with a speed of 550 miles per hour relative to the air is aimed to the northwest.

(a) Draw the vectors representing the wind and the plane relative to the air. (Choose a scale and make an accurate drawing.)

(b) Using your drawing, estimate the speed of the plane relative to the ground.

(c) Compute the speed exactly.

15. Compute $\mathbf{A} + \mathbf{B}$ and $\mathbf{A} - \mathbf{B}$ if

(a) $\mathbf{A} = \langle -1, 2, 3 \rangle$ and $\mathbf{B} = \langle 7, 0, 2 \rangle$

(b) $\mathbf{A} = 3\mathbf{j} + 4\mathbf{k}$ and $\mathbf{B} = 6\mathbf{i} + 7\mathbf{j}$

16. Compute $\mathbf{A} + \mathbf{B}$ and $\mathbf{A} - \mathbf{B}$ if

(a) $\mathbf{A} = \langle 1/2, 1/3, 1/6 \rangle$ and $\mathbf{B} = \langle 2, 3, -1/3 \rangle$

(b) $\mathbf{A} = 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$ and $\mathbf{B} = -\mathbf{i} + 5\mathbf{j} + 6\mathbf{k}$

17. Compute and sketch $c\mathbf{A}$ if $\mathbf{A} - 2\mathbf{i} + 3\mathbf{j} + \mathbf{k}$ and c is

(a) 2

(b) -2

(c) $\frac{1}{2}$

(d) $-\frac{1}{2}$

18. Express the vectors in the form $c(2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k})$.

(a) $\langle 4, 6, 8 \rangle$

(b) $-2\mathbf{i} - 3\mathbf{j} - 4\mathbf{k}$

(c) $\mathbf{0}$

(d) $\frac{2}{11}\mathbf{i} + \frac{3}{11}\mathbf{j} + \frac{4}{11}\mathbf{k}$

19. If $|\mathbf{A}| = 6$, find the length of

(a) $-2\mathbf{A}$

(b) $\mathbf{A}/3$

(c) $\mathbf{A}/|\mathbf{A}|$

(d) $-\mathbf{A}$

(e) $\mathbf{A} + 2\mathbf{A}$.

20. If $|\mathbf{A}| = 3$, find the length of

(a) $-4\mathbf{A}$

(b) $13\mathbf{A} - 7\mathbf{A}$

(c) $\mathbf{A}/|\mathbf{A}|$

(d) $\mathbf{A}/0.05$

(e) $\mathbf{A} - \mathbf{A}$.

21.

(a) Find a unit vector $\mathbf{u}$ that has the same direction as $\mathbf{A} = \mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$.

(b) Draw $\mathbf{A}$ and $\mathbf{u}$, with their tails at the origin.

22.

(a) Find a unit vector $\mathbf{u}$ that has the same direction as $\mathbf{A} = 2\mathbf{i} - 2\mathbf{j} + 3\mathbf{k}$.

(b) Draw $\mathbf{A}$ and $\mathbf{u}$, with their tails at the origin.

23. Using the definition of addition of vectors $\mathbf{A} = \langle a_1, a_2, a_3\rangle$ and $\mathbf{B} = \langle b_1, b_2, b_3\rangle$, show the $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$ and $\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B})$.

24. Using the definition of addition of vectors show that $\mathbf{A}+(\mathbf{B}+\mathbf{C}) = (\mathbf{A}+\mathbf{B})+\mathbf{C}$.

25. Which unit vector points in the same direction as $2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$?

26. Sketch a unit vector pointing in the same direction as $3\mathbf{i} + 4\mathbf{j}$.

27. *(Midpoint formula)* Let A and B be two points in space. Let M be their midpoint and O the origin. Let $\mathbf{A} = \overline{OA}$, $\mathbf{B} = \overline{OB}$, and $\mathbf{M} = \overline{OM}$.

(a) Show that $\mathbf{M} = \mathbf{A} + \frac{1}{2}(\mathbf{B} - \mathbf{A})$.

(b) Deduce that $\mathbf{M} = (\mathbf{A} + \mathbf{B})/2$. (Draw a picture.)

28. Let A and B be two distinct points in space. Let C be the point on the line segment AB that is twice as far from A as it is from B. Let $\mathbf{A} = \overline{OA}$, $\mathbf{B} = \overline{OB}$, and $\mathbf{C} = \overline{OC}$. Show that $\mathbf{C} = \frac{1}{3}\mathbf{A} + \frac{2}{3}\mathbf{B}$. (Draw a picture.)

29. Show that $2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$ and $6\mathbf{i} + 9\mathbf{j} + 12\mathbf{k}$ are parallel.

30. Show that $\mathbf{i} - 3\mathbf{j} + 6\mathbf{k}$ and $-2\mathbf{i} + 6\mathbf{j} - 12\mathbf{k}$ are parallel.

31. This exercise outlines a proof of the distributive rule, $c(\mathbf{A} + \mathbf{B}) = c\mathbf{A} + c\mathbf{B}$. Write $\mathbf{A}$ and $\mathbf{B}$ in components, and obtain the rule by expressing both $c(\mathbf{A} + \mathbf{B})$ and $c\mathbf{A} + c\mathbf{B}$ in components.

32.

(a) Show that the vectors $\mathbf{u}_1 = \frac{1}{2}\mathbf{i} + \frac{\sqrt{3}}{2}\mathbf{j}$ and $\mathbf{u}_2 = \frac{\sqrt{3}}{2}\mathbf{i} - \frac{1}{2}\mathbf{j}$ are perpendicular unit vectors. (What angles do they make with the x-axis?)

(b) Find scalars x and y such that $\mathbf{i} = x\mathbf{u}_1 + y\mathbf{u}_2$.

33.

(a) Show that the vectors $\mathbf{u}_1 = (\sqrt{2}/2)\mathbf{i} + (\sqrt{2}/2)\mathbf{j}$ and $\mathbf{u}_2 = (-\sqrt{2}/2)\mathbf{i} + (\sqrt{2}/2)\mathbf{j}$ are perpendicular unit vectors. (Draw them.)

(b) Express $\mathbf{i}$ in the form of $x\mathbf{u}_1 + y\mathbf{u}_2$. (Draw $\mathbf{i}$, $\mathbf{u}_1$, and $\mathbf{u}_2$.)

(c) Express $\mathbf{j}$ in the form $x\mathbf{u}_1 + y\mathbf{u}_2$.

(d) Express $-2\mathbf{i} + 3\mathbf{j}$ in the form $x\mathbf{u}_1 + y\mathbf{u}_2$.

34.

(a) Draw a unit vector **u** tangent to the curve $y = \sin x$ at $(\pi/4, \sqrt{2}/2)$.

(b) Express **u** in the form $x\mathbf{i} + y\mathbf{j}$.

35.

(a) Draw a unit vector **u** tangent to the curve $y = x^3$ at $(1, 1)$.

(b) Express **u** in the form $x\mathbf{i} + y\mathbf{j}$.

36.

(a) What is the sum of the five vectors shown in Figure 14.1.15?

(b) Sketch the polygon whose sides, in order, are **A**, **C**, **D**, **E**, **B**.

(c) What is $\mathbf{A} + \mathbf{C} + \mathbf{D} + \mathbf{E} + \mathbf{B}$?

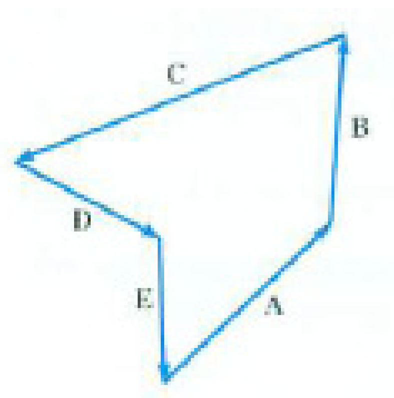


Figure 14.1.15

37. A rectangular box has sides of length x, y, and z. Show that the length of its longest diagonal is $\sqrt{x^2 + y^2 + z^2}$. (Use the Pythogorean Theorem twice.)

38. See Example 5 about hanging a picture. What would be the tension in the wire if it were at an angle of

(a) $60°$ instead of $10°$ to the horizontal?

(b) $5°$ instead of $10°$ to the horizontal?

39.

(a) Draw the vectors $\mathbf{A} = 2\mathbf{i} + \mathbf{j}$, $\mathbf{B} = 4\mathbf{i} - \mathbf{j}$, and $\mathbf{C} = 5\mathbf{i} + 2\mathbf{j}$.

(b) Using it, show that there are scalars x and y such that $\mathbf{C} = x\mathbf{A} + y\mathbf{B}$.

(c) Estimate x and y from the drawing.

(d) Find x and y exactly.

40. (See Exercise 13.) Let $\mathbf{A}$ and $\mathbf{B}$ be two nonzero and nonparallel vectors in the xy-plane. Let $\mathbf{C}$ be a vector in the xy-plane. Show with a sketch that there are scalars x and y such that $\mathbf{C} = x\mathbf{A} + y\mathbf{B}$.

41. Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ be three vectors that do not all lie in one plane. Let $\mathbf{D}$ be a vector in space. Show with a sketch that there are scalars x, y, and z such that $\mathbf{D} = x\mathbf{A} + y\mathbf{B} + z\mathbf{C}$.

42. Let A, B, and C be the vertices of a triangle. Let $\mathbf{A} = \overline{OA}$, $\mathbf{B} = \overline{OB}$, and $\mathbf{C} = \overline{OC}$.

(a) Let P be the point that is on the line segment joining A to the midpoint of the edge BC that is twice as far from A as from the midpoint. Show that $\overline{OP} = (\mathbf{A} + \mathbf{B} + \mathbf{C})/3$.

(b) Use (a) to show that the medians of a triangle are concurrent.

43. The midpoints of a quadrilateral in space are joined to form another quadrilateral. Prove that the second quadrilateral is a parallelogram.

Exercises 44 and 45 discuss a special case of the Cauchy-Schwarz inequality. This was first introduced in the CIE at the end of Chapter 7 (see page 679) and will be proved in Section 16.7 (Exercise 29). It also appears in Exercises 4 and 5 of the Chapter 9 Summary.

44.

(a) Using a diagram, explain why $|\mathbf{A}+\mathbf{B}| \leq |\mathbf{A}|+|\mathbf{B}|$. (This is called the **triangle inequality**.

(b) For what pairs of vectors $\mathbf{A}$ and $\mathbf{B}$ is $|\mathbf{A} + \mathbf{B}| = |\mathbf{A}| + |\mathbf{B}|$?

45. From Exercise 44 deduce that for real numbers x_1, y_1, x_2, and y_2,

$$x_1x_2 + y_1y_2 \leq \sqrt{x_1^2 + y_1^2}\sqrt{x_2^2 + y_2^2}.$$

When does equality hold?

14.2 The Dot Product of Two Vectors

This section introduces the dot product or scalar product, a number defined for pairs of vectors. It begins with the definition, describes a major application, and develops properties of the dot product.

Before we can introduce the dot product we need to define the angle between two vectors.

> **DEFINITION** *(Angle between two nonzero vectors.)* Let $\mathbf{A}$ and $\mathbf{B}$ be two nonparallel and nonzero vectors. They determine a triangle and an angle θ, shown in Figure 14.2.1(a). The **angle between A and B** is θ, $0 \leq \theta \leq \pi$.

(a) (b)

Figure 14.2.1

> If $\mathbf{A}$ and $\mathbf{B}$ are parallel, the angle between them is 0, if they have the same direction, or π, if they have opposite directions.
>
> The angle between the zero vector, $\mathbf{0}$, and any other vector is not defined.

The cosine of the angle between $\mathbf{A}$ and $\mathbf{B}$ is denoted $\cos(\mathbf{A}, \mathbf{B})$.

The Dot Product

That illustrates the dot product of two vectors, which will be introduced after the following definition.

> **DEFINITION** *(Dot product)* Let $\mathbf{A}$ and $\mathbf{B}$ be two nonzero vectors. Their **dot product** is the number
>
> $$|\mathbf{A}||\mathbf{B}|\cos(\theta) = |\mathbf{A}||\mathbf{B}|\cos(\mathbf{A}, \mathbf{B}),$$
>
> where θ is the angle between $\mathbf{A}$ and $\mathbf{B}$. If $\mathbf{A}$ or $\mathbf{B}$ is $\mathbf{0}$, their dot product is 0. The dot product is denoted $\mathbf{A} \cdot \mathbf{B}$. It is a scalar and is also called the **scalar product** of $\mathbf{A}$ and $\mathbf{B}$.

We illustrate how the dot product is useful by one of its applications.

Suppose a rock is pulled along level ground by a constant force $\mathbf{F}$ at an angle θ to the ground, shown in Figure 14.2.2. How much work does it accomplish in moving the rock from the tail to the head of $\mathbf{R}$?

(a) (b)

Figure 14.2.2

The force can be written as $\mathbf{F}_1 + \mathbf{F}_2$, where $\mathbf{F}_1$ is in the direction of motion. The work accomplished by $\mathbf{F}$ is the amount accomplished by $\mathbf{F}_1$. THe perpendicular force $\mathbf{F}_2$ accomplishes no work.

The work accomplished by $\mathbf{F}_1$ is its magnitude $|\mathbf{F}_1|$ times the distance it moves the rock, $|\mathbf{R}|$.

Because $|\mathbf{F}_1| = |\mathbf{F}|\cos(\theta)$, the work accomplished by $\mathbf{F}$ is $|\mathbf{F}_1||]vR|\cos(\theta)$, the dot product of $\mathbf{F}$ and $\mathbf{R}$,

$$\text{Work} = \underbrace{|\mathbf{F}|\cos(\theta)}_{\text{Force in Direction of } \mathbf{R}} \cdot \underbrace{|\mathbf{R}|}_{\text{Distance traveled}} = \mathbf{F}\cdot\mathbf{R}.$$

The dot product satisfies several useful identities that follow from the definition:

$$\begin{array}{rcll} \mathbf{A}\cdot\mathbf{B} & = & \mathbf{B}\cdot\mathbf{A} & \text{(the dot product is commutative)} \\ \mathbf{A}\cdot\mathbf{A} & = & |\mathbf{A}|^2 & \\ (c\mathbf{A})\cdot\mathbf{B} & = & c(\mathbf{A}\cdot\mathbf{B}) = \mathbf{A}\cdot(c\mathbf{B}) & (c \text{ is a scalar}) \\ \mathbf{0}\cdot\mathbf{A} & = & 0. & \end{array}$$

To establish that $\mathbf{A}\cdot\mathbf{A} = |\mathbf{A}|^2$, we calculate

$$\mathbf{A}\cdot\mathbf{A} = |\mathbf{A}||\mathbf{A}|\cos(\theta) = |\mathbf{A}|^2,$$

since the angle θ between $\mathbf{A}$ and $\mathbf{A}$ is 0, and $\cos(0) = 1$.

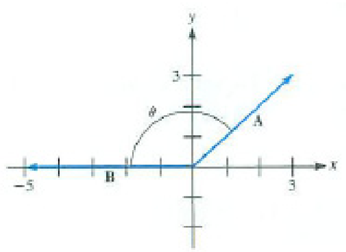


Figure 14.2.3

EXAMPLE 1 Find the dot product $\mathbf{A} \cdot \mathbf{B}$ if $\mathbf{A} = 3\mathbf{i} + 3\mathbf{j}$ and $\mathbf{B} = -5\mathbf{i}$.
SOLUTION Figure 14.2.3 shows that θ, the angle between $\mathbf{A}$ and $\mathbf{B}$, is $3\pi/4$. Also,

$$|\mathbf{A}| = \sqrt{3^2 + 3^2} = \sqrt{18} \text{ and } |\mathbf{B}| = \sqrt{5^2 + 0^2} = 5.$$

Thus

$$\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}||\mathbf{B}| \cos\theta = \sqrt{18} \cdot \left(\frac{-\sqrt{2}}{2} \right) = -15.$$

◇

EXAMPLE 2 Find (a) $\mathbf{i} \cdot \mathbf{j}$, (b) $\mathbf{i} \cdot \mathbf{i}$, and (c) $2\mathbf{k} \cdot (-3\mathbf{k})$.
SOLUTION

(a) The angle between $\mathbf{i}$ and $\mathbf{j}$ is $\pi/2$. Thus

$$\mathbf{i} \cdot \mathbf{j} \quad = \quad |\mathbf{i}||\mathbf{j}| \cos\left(\frac{\pi}{2}\right) = 1 \cdot 1 \cdot 0 = 0.$$

This is a special case of $\mathbf{A} \cdot \mathbf{A} = |\mathbf{A}|^2$.

(b) The angle between $\mathbf{i}$ and $\mathbf{i}$ is 0. Thus

$$\mathbf{i} \cdot \mathbf{i} \quad = \quad |\mathbf{i}||\mathbf{i}| \cos(0) = 1 \cdot 1 \cdot 1 = 1.$$

(c) The angle between $2\mathbf{k}$ and $-3\mathbf{k}$ is π. Thus

$$2\mathbf{k} \cdot (-3\mathbf{k}) \quad = \quad |2\mathbf{k}|\,|-3\mathbf{k}| \cos(\pi) = 2 \cdot 3 \cdot (-1) = -6.$$

◇

Computations like those in Example 2 show that $a\mathbf{i} \cdot b\mathbf{i} = ab$, $a\mathbf{j} \cdot b\mathbf{j} = ab$, and $a\mathbf{k} \cdot b\mathbf{k} = ab$, while $a\mathbf{i} \cdot b\mathbf{j} = 0$, $a\mathbf{i} \cdot b\mathbf{k} = 0$, and $a\mathbf{j} \cdot b\mathbf{k} = 0$.
In particular, $\mathbf{i} \cdot \mathbf{i} = \mathbf{j} \cdot \mathbf{j} = \mathbf{k} \cdot \mathbf{k} = 1$, while $\mathbf{i} \cdot \mathbf{j} = \mathbf{i} \cdot \mathbf{k} = \mathbf{j} \cdot \mathbf{k} = 0$.

By definition, the zero vector is perpendicular to every vector in the xy-plane.

The Geometry of the Dot Product

Let $\mathbf{A}$ and $\mathbf{B}$ be nonzero vectors and θ the angle between them. Their dot product is

$$\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}||\mathbf{B}| \cos(\theta).$$

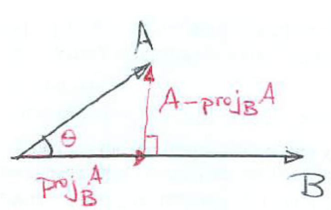


Figure 14.2.4

The quantities $|\mathbf{A}|$ and $|\mathbf{B}|$, being the lengths of vectors, are positive. However, $\cos(\theta)$ can be positive, zero, or negative. When $\cos(\theta) = 0$ $\theta = \pi/2$, so $\mathbf{A}$ and $\mathbf{B}$ are perpendicular. The dot product thus provides a way of telling whether $\mathbf{A}$ and $\mathbf{B}$ are perpendicular:

A Test for Perpendicular Vectors
Let $\mathbf{A}$ and $\mathbf{B}$ be nonzero vectors. If $\mathbf{A}\cdot\mathbf{B} = 0$, then $\mathbf{A}$ and $\mathbf{B}$ are perpendicular. Conversely, if $\mathbf{A}$ and $\mathbf{B}$ are perpendicular, then $\mathbf{A}\cdot\mathbf{B} = 0$.

As Figure 14.2.4 shows, $\mathbf{A}$ can be expressed as the sum of a vector parallel to $\mathbf{B}$ and a vector perpendicular to $\mathbf{B}$. The one parallel to $\mathbf{B}$ is $|\mathbf{A}|\cos(\theta)$ times the unit vector $\mathbf{B}/|\mathbf{B}|$.

DEFINITION *(Projection of* $\mathbf{A}$ *onto* $\mathbf{B}$*)* Given vectors $\mathbf{A}$ and $\mathbf{B}$, as shown in Figure 14.2.4, the component of $\mathbf{A}$ parallel to $\mathbf{B}$ is called the **projection of A on B**

$$\mathbf{proj_B\,A} = \frac{\mathbf{A}\cdot\mathbf{B}}{|\mathbf{B}|}\frac{\mathbf{B}}{|\mathbf{B}|} = \frac{(\mathbf{A}\cdot\mathbf{B})\mathbf{B}}{|\mathbf{B}|^2}.$$

The component of $\mathbf{A}$ perpendicular to $\mathbf{B}$ is then $\mathbf{A} - \mathbf{proj_B\,A}$.

The length of $\mathbf{proj_B\,A}$ is $|\mathbf{A}||\cos(\theta)|$, which equals $\frac{|\mathbf{A}\cdot\mathbf{B}|}{|\mathbf{B}|}$.

If θ is less than $\pi/2$, $\mathbf{proj_B\,A}$ points in the same direction as $\mathbf{B}$. If $\pi/2 < \theta \le \pi$, then $\mathbf{proj_B\,A}$ points in the direction opposite to that of $\mathbf{B}$. In either case, since $\mathbf{B}/|\mathbf{B}|$ is the unit vector in the direction of $\mathbf{B}$, we have

SHERMAN: Should this be a definition, not a boxed formula? Of, should the definition of dot product be boxes? (Yes, I know boxes will be shaded in the actual text.) It seems we have a lot more boxes in the last third than the earlier chapters. Is this a problem?

Let $\mathbf{A}$ and $\mathbf{B}$ be vectors.
If $\mathbf{A}\cdot\mathbf{B}$ is positive, then the angle between the vectors is less than $\pi/2$ and $\mathbf{proj_B\,A}$ points in the same direction as $\mathbf{B}$.
If $\mathbf{A}\cdot\mathbf{B}$ is negative, then the angle between the vectors is greater than $\pi/2$ and $\mathbf{proj_B\,A}$ points in the opposite direction as $\mathbf{B}$.

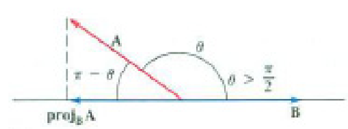


Figure 14.2.5

If $\mathbf{A}\cdot\mathbf{B}$ is negative, then the angle between $\mathbf{A}$ and $\mathbf{B}$ is obtuse (greater than $\pi/2$). Figure 14.2.5 shows this and that $\mathbf{proj_B\,A}$ points in the direction opposite that of $\mathbf{B}$.

Computing $\mathbf{A}\cdot\mathbf{B}$ in Terms of Their Components

We defined $\mathbf{A}\cdot\mathbf{B}$, using the geometric interpretation of $\mathbf{A}$ and $\mathbf{B}$. But what if $\mathbf{A}$ and $\mathbf{B}$ are given in terms of their components along $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$, $\mathbf{A} = \langle a_1, a_2, a_3\rangle$ and $\mathbf{B} = \langle b_1, b_2, b_3\rangle$? How would we find $\mathbf{A}\cdot\mathbf{B}$ in that case?

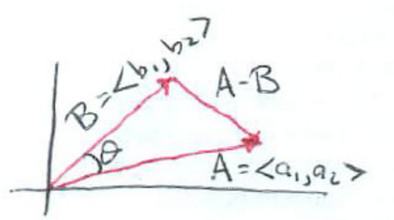
The answer turns out to be simple:

If $\mathbf{A} = \langle a_1, a_2, a_3\rangle$ and $\mathbf{B} = \langle b_1, b_2, b_3\rangle$, then $\mathbf{A}\cdot\mathbf{B} = a_1b_1 + a_2b_2 + a_3b_3$.

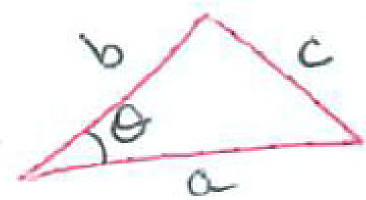
The dot product is the sum of three numbers, the products of corresponding components.

For vectors in the xy-plane,

Component Form of the Dot Product in the xy-plane
If $\mathbf{A} = \langle a_1, a_2 \rangle$ and $\mathbf{B} = \langle b_1, b_2 \rangle$, then $\mathbf{A} \cdot \mathbf{B} = a_1 b_1 + a_2 b_2$.



(a)



(b)

Figure 14.2.6

We establish the second result, which depends on the Law of Cosines. It says that in a triangle whose sides have lengths a, b, and c, and angle θ opposite the side with length c, as in Figure 14.2.6(b), $c^2 = a^2 + b^2 - 2ab\cos(\theta)$.

Thus, in terms of the triangle in Figure 14.2.6(a),

$$|\mathbf{A} - \mathbf{B}|^2 = |\mathbf{A}|^2 + |\mathbf{B}|^2 - 2|\mathbf{A}||\mathbf{B}|\cos(\theta),$$

which tells us that

$$|\mathbf{A} - \mathbf{B}|^2 = |\mathbf{A}|^2 + |\mathbf{B}|^2 - 2\mathbf{A} \cdot \mathbf{B}. \tag{14.2.1}$$

Translating (14.2.1) into components, we have

$$(a_1 - b_1)^2 + (a_2 - b_2)^2 = \left(a_1^2 + a_2^2\right) + \left(b_1^2 + b_2^2\right) - 2\mathbf{A} \cdot \mathbf{B}$$

or

$$a_1^2 - 2a_1b_1 + b_1^2 + a_2^2 - 2a_2b_2 + b_2^2 = a_1^2 + a_2^2 + b_1^2 + b_2^2 - 2\mathbf{A} \cdot \mathbf{B}.$$

Hence

$$-2(a_1b_1 + a_2b_2) = -2\mathbf{A} \cdot \mathbf{B}.$$

And, finally,

$$\mathbf{A} \cdot \mathbf{B} = a_1b_1 + a_2b_2.$$

The argument for space vectors is essentially the same, as doing Exercise 42 will show.

With the aid of the formula for the dot product in terms of components it is a straightforward matter to prove the distributive property,

$$\mathbf{A} \cdot (\mathbf{B} + \mathbf{C}) = \mathbf{A} \cdot \mathbf{B} + \mathbf{A} \cdot \mathbf{C}.$$

See Exercises 40 to 43

EXAMPLE 3 Find $\cos(\mathbf{A}, \mathbf{B})$ when $\mathbf{A} = \langle 6, 3 \rangle$ and $\mathbf{B} = \langle -1, 1 \rangle$.
SOLUTION We know that $\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}||\mathbf{B}| \cos(\mathbf{A}, \mathbf{B})$. Thus

$$6 \cdot (-1) + 3 \cdot (1) = \sqrt{6^2 + 3^2}\sqrt{(-1)^2 + 1^2} \cos(\mathbf{A}, \mathbf{B})$$

or

$$-3 = 3\sqrt{10} \cos(\mathbf{A}, \mathbf{B}),$$

from which we conclude that

$$\cos(\mathbf{A}, \mathbf{B}) = -1/\sqrt{10}.$$

As Figure 14.2.7 shows, the angle between $\mathbf{A}$ and $\mathbf{B}$ is obtuse. A calculator would give a numerical estimate of its value. ◇

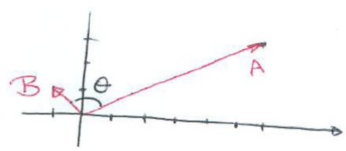


Figure 14.2.7

As Example 3 illustrates, the dot product can be used to find the cosine of the angle between two vectors, and, therefore, the angle itself:

Cosine of the angle between two vectors

$$\cos(\theta) = \cos(\mathbf{A}, \mathbf{B}) = \frac{\mathbf{A} \cdot \mathbf{B}}{|\mathbf{A}||\mathbf{B}|}.$$

SHERMAN: Your notes skip from here to Example 6. I assume you did not intend to omit the solution to an exercise or all of another. I've left that page in, and made the few edits based on what I am seeing in the files.

EXAMPLE 4

(a) Find the projection of $\mathbf{A} = 2\mathbf{i} + \mathbf{j}$ on $\mathbf{B} = -3\mathbf{i} + 2\mathbf{j}$.

(b) Express $\mathbf{A}$ as the sum of a vector parallel to $\mathbf{B}$ and a vector perpendicular to $\mathbf{B}$.

SOLUTION

(a) In this case

$$\begin{aligned}
\mathbf{proj_B\, A} &= \frac{\mathbf{A}\cdot\mathbf{B}}{|\mathbf{B}|}\frac{\mathbf{B}}{|\mathbf{B}|}\\
&= \frac{(2\mathbf{i}+\mathbf{j})\cdot(-3\mathbf{i}+2\mathbf{j})}{|-3\mathbf{i}+2\mathbf{j}|}\frac{-3\mathbf{i}+2\mathbf{j}}{|-3\mathbf{i}+2\mathbf{j}|}\\
&= \frac{(-6+2)}{\sqrt{13}}\frac{(-3\mathbf{i}+2\mathbf{j})}{\sqrt{13}}\\
&= \frac{-4}{13}(-3\mathbf{i}+2\mathbf{j}) = \frac{12}{13}\mathbf{i}-\frac{8}{13}\mathbf{j}.
\end{aligned}$$

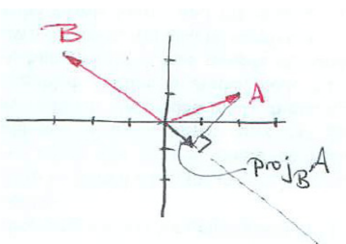


Figure 14.2.8

Figure 14.2.8 shows $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{proj_B\, A}$.

In this case $\mathbf{A}\cdot\mathbf{B}$ is negative, the angle between $\mathbf{A}$ and $\mathbf{B}$ is obtuse, and $\mathbf{proj_B\, A}$ points in the direction opposite the direction of $\mathbf{B}$.

(b) The vector $\mathbf{A}-\mathbf{proj_B\, A}$ is perpendicular to $\mathbf{B}$ and we have

$$\begin{aligned}
vA-\mathbf{proj_B\, A} &= \left(\frac{12}{13}\mathbf{i}-\frac{8}{13}\mathbf{j}\right)+\left(2\mathbf{i}+\mathbf{j}-\left(\frac{12}{13}\mathbf{i}-\frac{8}{13}\mathbf{j}\right)\right)\\
&= \underbrace{\left(\frac{12}{13}\mathbf{i}-\frac{8}{13}\mathbf{j}\right)}_{\text{parallel to }\mathbf{B}}+\underbrace{\left(\frac{14}{13}\mathbf{i}+\frac{21}{13}\mathbf{j}\right)}_{\text{perpendicular to }\mathbf{B}}.
\end{aligned}$$

◇

The scalar $\mathbf{A}\cdot(\mathbf{B}/|\mathbf{B}|)$ is the **scalar component of A in the direction of B**, denoted $\text{comp}_{\mathbf{B}}(\mathbf{A})$. It can be positive, negative, or zero. Its absolute value is the length of $\mathbf{proj_B}(\mathbf{A})$.

EXAMPLE 5 Find $\mathbf{proj_B}(\mathbf{A})$ and $\text{comp}_{\mathbf{B}}(\mathbf{A})$ when $\mathbf{A}=\mathbf{i}+3\mathbf{j}$ and $\mathbf{B}=\mathbf{i}-\mathbf{j}$.

SOLUTION Since $|\mathbf{B}|=\sqrt{1^2+(-1)^2}=\sqrt{2}$ and $\mathbf{A}\cdot\mathbf{B}=1-3=-2$,

$$\mathbf{proj_B}(\mathbf{A})=\frac{\mathbf{A}\cdot\mathbf{B}}{|\mathbf{B}|}\frac{\mathbf{B}}{|\mathbf{B}|}=\frac{-2}{\sqrt{2}}\frac{(\mathbf{i}-\mathbf{j})}{\sqrt{2}}=-\mathbf{i}+\mathbf{j}$$

and $\text{comp}_{\mathbf{B}}(\mathbf{A})=(\mathbf{A}\cdot\mathbf{B})/|\mathbf{B}|=-2/\sqrt{2}=-\sqrt{2}$. This agrees with Figure 14.2.9. ◇

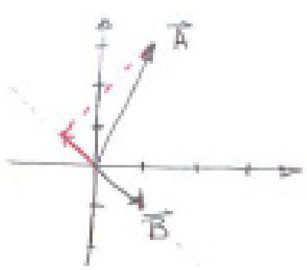


Figure 14.2.9

EXAMPLE 6 Show that the vectors $\langle 1,2,1\rangle$ and $\langle 2,-3,4\rangle$ are perpendicular.

SOLUTION We want to show that the angle θ between the vectors is $\pi/2$. To do this we show $\cos(\theta) = 0$. We have

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{|\mathbf{A}||\mathbf{B}|} = \frac{(1 \cdot 2) + 2(-3) + 1 \cdot 4}{|\mathbf{A}||\mathbf{B}|} = \frac{2 - 6 + 4}{|\mathbf{A}||\mathbf{B}|} = 0,$$

So the vectors are perpendicular. ⋄

Example 6 illustrates the test for two vectors being perpendicular: their dot product is zero.

Summary

We defined the dot (scalar) product of two vectors $\mathbf{A}$ and $\mathbf{B}$ geometrically as $|\mathbf{A}||\mathbf{B}|\cos(\theta)$, where θ is the angle between $\mathbf{A}$ and $\mathbf{B}$. We then obtained a formula for $\mathbf{A} \cdot \mathbf{B}$ in terms of their components: $\langle a_1, a_2 \rangle \cdot \langle b_1, b_2 \rangle = a_1 b_1 + a_2 b_2$ and a similar formula for the dot product of two space vectors.

The dot product enabled us to express a vector $\mathbf{A}$ as the sum of a vector parallel to $\mathbf{B}$, $\mathbf{proj_B}\,\mathbf{A}$, and one perpendicular to $\mathbf{B}$, $\mathbf{A} - \mathbf{proj_B}\,\mathbf{A}$).

When two non-zero vectors have a dot product that is 0, the vectors are perpendicular.

The zero-vector, $\mathbf{0}$, is considered to be perpendicular to every vector.

The dot product can be used to find the angle θ between two vectors:

$$\cos(\theta) = \cos(\mathbf{A}, \mathbf{B}) = \frac{\mathbf{A} \cdot \mathbf{B}}{|\mathbf{A}||\mathbf{B}|}.$$

The Dot Product in Business and Statistics

Imagine that a fast food restaurant sells 30 hamburgers, 20 salads, 15 soft drinks, and 13 orders of french fries. This is recorded by the four-dimensional "vector" $\langle 30, 20, 15, 13\rangle$. A hamburger sells for \$1.99, a salad for \$1.50, a soft drink for \$1.00, and an order of french fries for \$1.10. The "price vector" is $\langle 1.99, 1.50, 1.00, 1.10\rangle$. The dot product of these two vectors,

$$30(1.99) + 20(1.50) + 15(1.00) + 13(1.10),$$

would be the total amount paid for all items.

Descriptions of the economy use "production vectors," "cost vectors," "price vectors," and "profit vectors" with many more than the four componenets of our restaurant example.

In statistics the **coefficient of correlation** is defined in terms of a dot product. For instance, say you have the heights and weights of n persons. Let the height of the ith person be h_i and the weight be w_i. Let h be the average of the n heights and w be the average of the n weights. Let $\mathbf{H} = \langle h_1 - h, h_2 - h, \cdots, h_n - h\rangle$ and $\mathbf{W} = \langle w_1 - w, w_2 - w, \cdots, w_n - w\rangle$. The coefficient of correlation between the heights and weights is defined to be

$$\frac{\mathbf{H}\cdot\mathbf{W}}{|\mathbf{H}||\mathbf{W}|}.$$

In analogy with vectors in the plane or space,

$$\begin{aligned}\mathbf{H}\cdot\mathbf{W} &= \sum_{i=1}^{n}(h_i - h)(w_i - w),\\ |\mathbf{H}| &= \sqrt{\sum_{i=1}^{n}(h_i - h)^2},\\ |\mathbf{W}| &= \sqrt{\sum_{i=1}^{n}(w_i - w)^2}.\end{aligned}$$

It turns out that the coefficient of correlation is simply the cosine of the angle between the vectors $\mathbf{H}$ and $\mathbf{W}$ in n-dimensional space. The closer it is to 1, the closer the angle between the vectors is to $\mathbf{0}$, and the better the weights and heights "correlate."

EXERCISES for Section 14.2

In Exercises 1 to 4 compute $\mathbf{A} \cdot \mathbf{B}$.

1. $\mathbf{A}$ has length 3, $\mathbf{B}$ has length 4, and the angle between them is $\pi/4$.

2. $\mathbf{A}$ has length 2, $\mathbf{B}$ has length 3, and the angle between them is $3\pi/4$.

3. $\mathbf{A}$ has length 5, $\mathbf{B}$ has length $\frac{1}{2}$, and the angle between them is $\pi/2$.

4. $\mathbf{A}$ is the zero vector $\mathbf{0}$, and $\mathbf{B}$ has length 5.

In Exercises 5 to 8 compute $\mathbf{A} \cdot \mathbf{B}$.

5. $\mathbf{A} = -2\mathbf{i} + 3\mathbf{j}$, $\mathbf{B} = 4\mathbf{i} + 4\mathbf{j}$

6. $\mathbf{A} = 0.3\mathbf{i} + 0.5\mathbf{j}$, $\mathbf{B} = 2\mathbf{i} - 1.5\mathbf{j}$

7. $\mathbf{A} = 2\mathbf{i} - 3\mathbf{j} - \mathbf{k}$, $\mathbf{B} = 3\mathbf{i} + 4\mathbf{j} - \mathbf{k}$

8. $\mathbf{A} = \mathbf{i} + \mathbf{j} + \mathbf{k}$,$\mathbf{B} = 2\mathbf{i} + +3\mathbf{j} - 5\mathbf{k}$

9.

(a) Draw the vectors $7\mathbf{i} + 12\mathbf{j}$ and $9\mathbf{i} - 5\mathbf{j}$.

(b) Do they seem to be perpendicular?

(c) Determine whether they are perpendicular by examining their dot product.

10.

(a) Draw the vectors $\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$ and $\mathbf{i} + \mathbf{j} - \mathbf{k}$.

(b) Do they seem to be perpendicular?

(c) Determine whether they are perpendicular by examining their dot product.

11.

(a) Estimate the angle between $\mathbf{A} = 3\mathbf{i} + 4\mathbf{j}$ and $\mathbf{B} = 5\mathbf{i} + 12\mathbf{j}$ by drawing them.

(b) Find the angle between $\mathbf{A}$ and $\mathbf{B}$.

12. Let $P = (6, 1)$, $Q = (3, 2)$, $R = (1, 3)$, and $S = (4, 5)$.

(a) Draw the vectors $\overline{PQ}$ and $\overline{RS}$.

(b) Using the diagram estimate the angle between $\overline{PQ}$ and $\overline{RS}$.

(c) Using the dot product, find $\cos(\overline{PQ}, \overline{RS})$, that is, the cosine of the angle between $\overline{PQ}$ and $\overline{RS}$.

(d) Using a calculator, find the angle.

13. Find the angle between $2\mathbf{i} - 4\mathbf{j} + 6\mathbf{k}$ and $\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$.

14. Find the angle between $\mathbf{i} + \mathbf{j} + 3\mathbf{k}$ and $3\mathbf{i} + 6\mathbf{j} - 3\mathbf{k}$.

15. Find the angle between $\overline{AB}$ and $\overline{CD}$ if $A = (1, 3)$, $B = (7, 4)$, $C = (2, 8)$, and $D = (1, -5)$.

16. Find the angle between $\overline{AB}$ and $\overline{CD}$ if $A = (1, 2, -5)$, $B = (1, 0, 1)$, $C = (0, -1, 3)$, and $D = (2, 1, 4)$.

17. Find the length of the projection of $-4\mathbf{i} + 5\mathbf{j}$ on the line through $(2, -1)$ and $(6, 1)$.

(a) By making a drawing and estimating the length by eye.

(b) By using the dot product.

18.

(a) Find a vector $\mathbf{C}$ parallel to $\mathbf{i} + 2\mathbf{j}$ and a vector $\mathbf{D}$ perpendicular to $\mathbf{i} + 2\mathbf{j}$ such that $-3\mathbf{i} + 4\mathbf{j} = \mathbf{C} + \mathbf{D}$.

(b) Draw the vectors to check that your answer is reasonable.

19.

(a) Find a vector $\mathbf{C}$ parallel to $2\mathbf{i} - \mathbf{j}$ and a vector $\mathbf{D}$ perpendicular to $2\mathbf{i} - \mathbf{j}$ such that $3\mathbf{i} + 4\mathbf{j} = \mathbf{C} + \mathbf{D}$.

(b) Draw the vectors to check that your answer is reasonable.

20. How far is the point $(1, 2, 3)$ from the line through $(1, 4, 2)$ and $(2, 1, -4)$?

21. What is $\mathbf{proj_B}\, \mathbf{A}$ if $\mathbf{A} = 2\mathbf{i} + \mathbf{j} - 3\mathbf{k}$ and $\mathbf{B} = \mathbf{i} + \mathbf{j} + \mathbf{k}$?

22. Express the vector $\mathbf{i} + \mathbf{j} + \mathbf{k}$ as the sum of a vector parallel to $\mathbf{i} - \mathbf{j} + 2\mathbf{k}$ and a vector perpendicular to $\mathbf{i} - \mathbf{j} + 2\mathbf{k}$.

23. Give an example of a non-zero vector in the xy-plane that is perpendicular to $3\mathbf{i} - 2\mathbf{j}$.

24. Give an example of a non-zero vector that is perpendicular to $5\mathbf{i} - 3\mathbf{j} + 4\mathbf{k}$.

Exercises 25 to 29 refer to the cube in Figure 14.2.10. Find

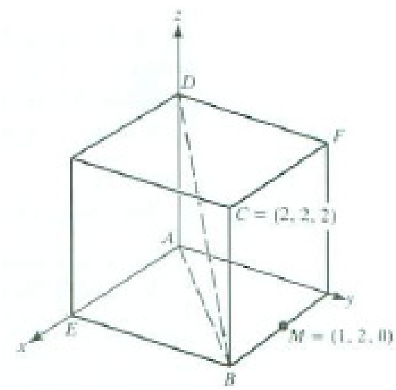


Figure 14.2.10

25. Find $\cos(\overline{AC}, \overline{BD})$.
26. Find $\cos(\overline{AF}, \overline{BD})$.
27. Find $\cos(\overline{AC}, \overline{AM})$.
28. Find $\cos(\overline{MD}, \overline{MF})$.
29. Find $\cos(\overline{EF}, \overline{BD})$.

30. If $\mathbf{A} \cdot \mathbf{B} = \mathbf{A} \cdot \mathbf{C}$ and $\mathbf{A}$ is not $\mathbf{0}$, must $\mathbf{B} = \mathbf{C}$?

31. We found the dot product in terms of components by using the Law of Cosines. We now see why it is true. The proof consists of two applications of the Pythagorean Theorem. Figure 14.2.11 shows a triangle with sides a, b, c, with angle θ opposite side c. (We suppose that θ is less than $\frac{\pi}{2}$.)

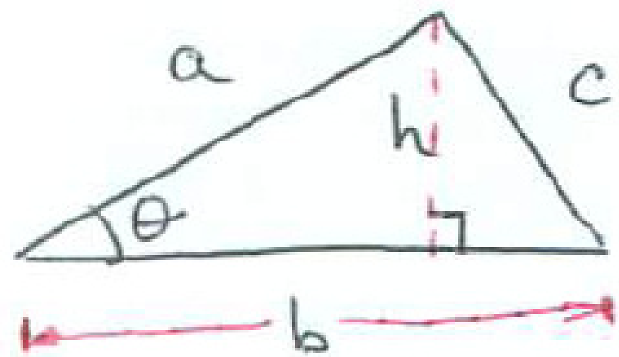


Figure 14.2.11

(a) Show that $h^2 = a^2 - a^2 \cos^2(\theta)$.

(b) Show that $h^2 = c^2 - (b - a\cos(\theta))^2$.

(c) By equating the two expressions obtain the Law of Cosines.

(d) Carry out the proof when θ is greater than $\pi/2$.

32.

(a) Let $\mathbf{A}$ be a vector and $\mathbf{u}_1$ and $\mathbf{u}_2$ are perpendicular unit vectors in that xy-plane. If $\mathbf{A} \cdot \mathbf{u}_1 = 0$ and $\mathbf{A} \cdot \mathbf{u}_2 = 0$, must $\mathbf{A} = \mathbf{0}$?

(b) Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be nonparallel unit vectors in the xy-plane. If $\mathbf{A} \cdot \mathbf{v}_1$ and $\mathbf{A} \cdot \mathbf{v}_2 = 0$, must $\mathbf{A} = \mathbf{0}$?

33.

Jane: I don't like the way the author found how to express $\mathbf{A}$ as the sum of a vector parallel to $\mathbf{B}$ and a vector perpendicular to $\mathbf{B}$.

Sam: It was O.K. for me. But I had to memorize a formula.

Jane: My goal is to memorize nothing. I write $\mathbf{A} = x\mathbf{B} + \mathbf{C}$, when $\mathbf{C}$ is perpendicular to $\mathbf{A}$. Then I dot with $\mathbf{B}$, getting

$$\mathbf{A} \cdot \mathbf{B} = x\mathbf{B} \cdot \mathbf{B} + \mathbf{C} \cdot \mathbf{B}.$$

Since $\mathbf{C}$ is perpendicular to $\mathbf{B}$, $\mathbf{C} \cdot \mathbf{B} = 0$, and lo and behold, I have

$$x = \frac{\mathbf{A} \cdot \mathbf{B}}{\mathbf{B} \cdot \mathbf{B}}.$$

So the vector parallel to $\mathbf{B}$ is $\frac{\mathbf{A} \cdot \mathbf{B}}{\mathbf{B} \cdot \mathbf{B}}\mathbf{B}$.

Sam: Cool. So why did the author go through all that stuff?

Jane: Maybe they wanted to reinforce the definition of the dot product and the role of the angle.

Sam: O.K. But how do I get the vector $\mathbf{C}$ perpendicular to $\mathbf{B}$?

Jane: Simple...

Complete Jane's reply.

34. If $|\mathbf{A}| = 3$ and $|\mathbf{B}| = 5$,

(a) how large can $|\mathbf{A} + \mathbf{B}|$ be?

(b) how small?

35. By taking the dot product of the unit vectors $\mathbf{u}_1 = \cos\theta_1\mathbf{i} + \sin\theta_1\mathbf{j}$ and $\mathbf{u}_2 = \cos\theta_2\mathbf{i} + \sin\theta_2\mathbf{j}$, prove that

$$\cos(\theta_1 - \theta_2) = \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2.$$

36. A tetrahedron, not necessarily regular, has six edges. Show that the line segment joining the midpoints of two opposite edges is perpendicular to the line segment joining another pair of opposite edges if anly only if the remaining two edges have the same length.

37. The output of a firm that manufactures x_1 washing machines, x_2 refrigerators, x_3 dishwashers, x_4 stoves, and x_5 clothes dryers is recorded by the production vector $\mathbf{P} = \langle x_1, x_2, x_3, x_4, x_5\rangle$. Similarly, the cost vector $\mathbf{C} = \langle y_1, y_2, y_3, y_4, y_5\rangle$ records the cost of producing each item.

(a) What is the economic significance of $\mathbf{P}\cdot\mathbf{C} = \langle 20, 0, 7, 9, 15\rangle \cdot \langle 50, 70, 30, 20, 10\rangle$?

(b) If the firm doubles the production of all items, what is its new production vector?

38. Let P_1 be the profit from selling a washing machine and let P_2, P_3, P_4, and P_5 be defined analogously for the firm of Exercise 37. What does it mean to the firm to have $\langle P_1, P_2, P_3, P_4, P_5\rangle$ perpendicular to $\langle x_1, x_2, x_3, x_4, x_5\rangle$?

39. Prove that $\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A}$

(a) using the geometric definition of the dot product,

(b) using the formula for the dot product in terms of components.

Exercises 40 and 43 all deal with the **distributive property for the dot product**:

$$\mathbf{A} \cdot (\mathbf{B} + \mathbf{C}) = \mathbf{A} \cdot \mathbf{B} + \mathbf{A} \cdot \mathbf{C}. \tag{14.2.2}$$

40. Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ be vectors in the xy-plane. Use the formula for the dot product in terms of components to prove that (14.2.2) is true.

41. In Exercise 40 the distributive property is obtained from the formula for the dot product in terms of components. Show that if (14.2.2) is true, then $\langle a_1, a_2 \rangle \cdot \langle b_1, b_2 \rangle = a_1 b_1 + a_2 b_2$. Begin with $(a_1\mathbf{i} + a_2\mathbf{j}) \cdot (b_1\mathbf{i} + b_2\mathbf{j})$.

42. Prove that $\langle a_1, a_2, a_3 \rangle \cdot \langle b_1, b_2, b_3 \rangle = a_1 b_1 + a_2 b_2 + a_3 b_3$ (Look at the proof for planar vectors.)

43. Show that (14.2.2) implies $\mathbf{A} \cdot (\mathbf{B} + \mathbf{C} + \mathbf{D}) = \mathbf{A} \cdot \mathbf{B} + \mathbf{A} \cdot \mathbf{C} + \mathbf{A} \cdot \mathbf{D}$.

44. Let $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ be unit vectors such that any two are perpendicular. Let $\mathbf{A}$ be a vector.

(a) Draw a picture that shows that there are scalars x, y, and z such that $\mathbf{A} = x\mathbf{u}_1 + y\mathbf{u}_2 + z\mathbf{u}_3$.

(b) Find $\mathbf{B}$ such that $\mathbf{A} \cdot \mathbf{B} = x$.

(c) Find $\mathbf{C}$ such that $\mathbf{A} \cdot \mathbf{C} = x - z$.

45. Given the vectors $\mathbf{A}$ and $\mathbf{B}$, we obtained a vector parallel to $\mathbf{B}$, called $\mathbf{proj_B}\, \mathbf{A}$, and then said, on the basis of a picture that $\mathbf{A} - \mathbf{proj_B}\, \mathbf{A}$ is perpendicular to $\mathbf{B}$. Using the dot product, show that $\mathbf{A} - \mathbf{proj_B}\, \mathbf{A}$ and $\mathbf{B}$ are perpendicular.

46. A force $\mathbf{F}$ of 10 newtons is parallel to $2\mathbf{i} + 3\mathbf{j} + \mathbf{k}$ and force pushes an object on a ramp in a straight line from the point $(3, 1, 5)$ to the point $(4, 3, 7)$, where coordinates are measured in meters. How much work does the force accomplish?

47. Show that if the two diagonals of the parallelogram are perpendicular, then the four sides have the same length (forming a rhombus). (Use the dot product.)

48. How far is the point $(2, 3, 5)$ from the line through the origin and $(1, -1, 2)$?

49. Some molecules, such as methane, consist of four atoms arranged as the vertices of a regular tetrahedron, the points labeled A, B, C, and D in Figure 14.2.12. E is the center of the cube.

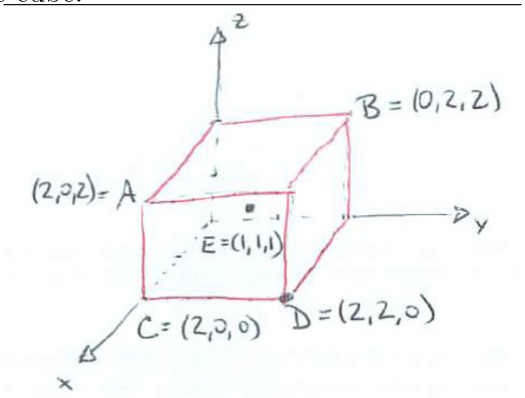


Figure 14.2.12

(a) Show that A, B, C, and D are vertices of a regular tetrahedron. (Show that the four faces are equilateral triangles.)

(b) Chemists are interested in the angle $\theta = AEB$. Show that $\cos(\theta) = -1/3$.

(c) Find θ (approximately).

14.3 The Cross Product of Two Vectors

The dot product of two vectors is a scalar. In this section we define a product of two vectors that is a vector perpendicular to both.

Definition of the Cross Product

Let $\mathbf{A} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ and $\mathbf{B} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$ be two non-zero vectors that are not parallel. We will construct a vector $\mathbf{C}$ that is perpendicular to both $\mathbf{A}$ and $\mathbf{B}$. The vector $\mathbf{C}$ is not unique since any vector parallel to $\mathbf{C}$ is also perpendicular to $\mathbf{A}$ and $\mathbf{B}$.

Let $\mathbf{C} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$. We want $\mathbf{C} \cdot \mathbf{A}$ and $\mathbf{C} \cdot \mathbf{B}$ to be 0. This gives the equations

$$a_1x + a_2y + a_3z = 0 \qquad (14.3.1)$$

$$b_1x + b_2y + b_3z = 0 \qquad (14.3.2)$$

We eliminate x by subtracting b_1 times (14.3.1) from a_1 times (14.3.2), as follows.

$$a_1b_1x + a_1b_2y + a_1b_3z = 0 \qquad (a_1 \text{ times } (14.3.2)) \qquad (14.3.3)$$

$$b_1a_1x + b_1a_2y + b_1a_3z = 0 \qquad (b_1 \text{ times } (14.3.1)) \qquad (14.3.4)$$

Subtracting (14.3.4) from (14.3.3) gives

$$(a_1b_2 - a_2b_1)y + (a_1b_3 - a_3b_1)z = 0 \qquad (14.3.5)$$

This is like solving $2y + 3z = 0$ by letting $y = -3$ and $z = 2$.

A non-zero solution of (14.3.5) is

$$y = -(a_1b_3 - a_3b_1), \quad z = a_1b_2 - a_2b_1.$$

To find x, substitute into (14.3.1). As Exercise 33 shows, algebra yields

$$x = a_2b_3 - a_3b_2.$$

So the vector

$$(a_2b_3 - a_3b_2)\mathbf{i} - (a_1b_3 - a_3b_1)\mathbf{j} + (a_1b_2 - a_2b_1)\mathbf{k} \qquad (14.3.6)$$

is perpendicular to $\mathbf{A}$ and $\mathbf{B}$. It is denoted $\mathbf{A} \times \mathbf{B}$ and is called the **vector product of A and B** or the **cross product of A and B**. This vector is defined even if $\mathbf{A}$ and $\mathbf{B}$ are parallel or if one (or both) of them is the zero vector, $\mathbf{0}$.

Determinants and the Cross Product

The expression (14.3.6) for the cross product is not easy to memorize. Determinants provide a convenient memory aid.

Four numbers arranged in a square form a matrix of order 2, for instance

$$\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}.$$

Its determinant is the number $a_1b_2 - a_2b_1$, denoted

$$\det\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}. \qquad \text{or} \qquad \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$$

Each term in (14.3.6) is the determinant of a matrix of order 2, namely

$$\det\begin{pmatrix} a_2 & a_3 \\ b_2 & b_3 \end{pmatrix}, \det\begin{pmatrix} a_1 & a_3 \\ b_1 & b_3 \end{pmatrix}, \quad \text{and} \quad \det\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}$$

Nine numbers arranged in a square form a matrix of order 3, for instance

$$\begin{pmatrix} c_1 & c_2 & c_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix}.$$

Its determinant is defined with the aid of determinants of order 2:

$$c_1 \det\begin{pmatrix} a_2 & a_3 \\ b_2 & b_3 \end{pmatrix} - c_2 \det\begin{pmatrix} a_1 & a_3 \\ b_1 & b_3 \end{pmatrix} + c_3 \det\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2. \end{pmatrix}$$

The coefficient of each c_i is plus or minus the determinant of the matrix of order 2 that remains when the row and column in which c_i appears are deleted, as shown in Figure 14.3.1 for the coefficient of c_1.

SHERMAN: Why do you want to delete this figure?

Therefore we can write (14.3.6) as a determinant of a matrix, and we have

$$\mathbf{A} \times \mathbf{B} = \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix} \tag{14.3.7}$$

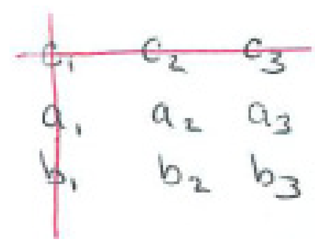


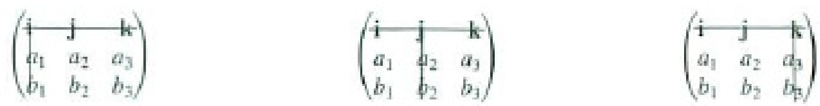
Figure 14.3.1

DEFINITION *(Cross product (vector product).)* The cross product, or vector product, of the vectors

$$\mathbf{A} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k} \quad \text{and} \quad \mathbf{B} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$$

is the vector

$$\begin{aligned} \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix} &= \mathbf{i}\det\begin{pmatrix} a_2 & a_3 \\ b_2 & b_3 \end{pmatrix} - \mathbf{j}\det\begin{pmatrix} a_1 & a_3 \\ b_1 & b_3 \end{pmatrix} + \mathbf{k}\det\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \\ &= (a_2b_3 - a_3b_2)\mathbf{i} - (a_1b_3 - a_3b_1)\mathbf{j} + (a_1b_2 - a_2b_1)\mathbf{k}. \end{aligned}$$



$\mathbf{i}\det$
Delete the two lines through $\mathbf{i}$. The determinant of the remaining square is the coefficient of $\mathbf{i}$ in $\mathbf{A}\times\mathbf{B}$.

$-\mathbf{j}\det$
Delete the two lines through $\mathbf{j}$. The *negative of* the determinant of the remaining square is the coefficient of $\mathbf{j}$ in $\mathbf{A}\times\mathbf{B}$.

$+\mathbf{k}\det$
Delete the two lines through $\mathbf{k}$. The determinant of the remaining square is the coefficient of $\mathbf{k}$ in $\mathbf{A}\times\mathbf{B}$.

Figure 14.3.2 Computation of $\mathbf{A}\times\mathbf{B}$ by expanding a 3×3 determinant along its first row.

SHERMAN: I've made significant changes here. Please check.

Figure 14.3.2 shows how the the determinant for $\mathbf{A}\times\mathbf{B}$ can be found by expanding the 3×3 matrix along its first row.

EXAMPLE 1 Compute $\mathbf{A}\times\mathbf{B}$ if $\mathbf{A} = 2\mathbf{i} - \mathbf{j} + 3\mathbf{k}$ and $\mathbf{B} = 3\mathbf{i} + 4\mathbf{j} + \mathbf{k}$.
SOLUTION By definition,

$$\begin{aligned}\mathbf{A}\times\mathbf{B} &= \begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ 2 & -1 & 3\\ 3 & 4 & 1\end{vmatrix} = \mathbf{i}\begin{vmatrix}-1 & 3\\ 4 & 1\end{vmatrix} - \mathbf{j}\begin{vmatrix}2 & 3\\ 3 & 1\end{vmatrix} + \mathbf{k}\begin{vmatrix}2 & -1\\ 3 & 4\end{vmatrix}\\ &= -13\mathbf{i} + 7\mathbf{j} + 11\mathbf{k}\end{aligned}$$

◇

The cross product has these properties:

The zero vector is, by definition, perpendicular to every vector.

1. $\mathbf{A}\times\mathbf{B}$ is perpendicular to both $\mathbf{A}$ and $\mathbf{B}$.
2. $\mathbf{A}\times\mathbf{B} = -(\mathbf{B}\times\mathbf{A})$ (the cross product is not commutative).
3. $\mathbf{A}\times\mathbf{B} = \mathbf{0}$ if $\mathbf{A}$ and $\mathbf{B}$ are parallel or at least one of them is $\mathbf{0}$.
4. $\mathbf{A}\times(\mathbf{B}+\mathbf{C}) = \mathbf{A}\times\mathbf{B} + \mathbf{A}\times\mathbf{C}$ (the cross product distributes over addition).

See Exercises 27 and 28.

The first property holds because that is how we constructed the cross product. The second and third are established by straightforward computations, using (14.3.7). Exercises 16 and 17 take care of the fourth.

What is the Direction of $\mathbf{A} \times \mathbf{B}$?

We know that $\mathbf{A} \times \mathbf{B}$ is perpendicular to $\mathbf{A}$ and $\mathbf{B}$, but ti could have two directions, as Figure 14.3.3(a) shows.

(a) (b) (c)

Figure 14.3.3 (a) The two possible directions for $\mathbf{A} \times \mathbf{B}$. (b) The right-hand rule for $\mathbf{A} \times \mathbf{B}$. (c) The right-hand rule for $\mathbf{j} \times \mathbf{i}$.

To see which, take a specific case, say $\mathbf{i} \times \mathbf{j}$:

$$\mathbf{i} \times \mathbf{j} = \det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = 0\mathbf{i} - 0\mathbf{j} + \mathbf{k} = \mathbf{k}.$$

This suggests that the direction of $\mathbf{A} \times \mathbf{B}$ is given by the **right hand rule**:

Left-handed people must use their right hand here.

> Curl the fingers of the right hand to go from $\mathbf{A}$ and $\mathbf{B}$. The thumb points in the direction of $\mathbf{A} \times \mathbf{B}$. (See Figure 14.3.3(b).)

EXAMPLE 2 Check that the right hand rule is correct for $\mathbf{j} \times \mathbf{i}$.
SOLUTION

$$\mathbf{j} \times \mathbf{i} = \det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = 0\mathbf{i} - 0\mathbf{j} - \mathbf{k} = -\mathbf{k}.$$

In this case, $\mathbf{j} \times \mathbf{i}$, points downward, the opposite of $\mathbf{i} \times \mathbf{j}$.

The right hand rule for $\mathbf{j} \times \mathbf{i}$ is illustrated in Figure 14.3.3(c). The thumb indeed points downward. ◇

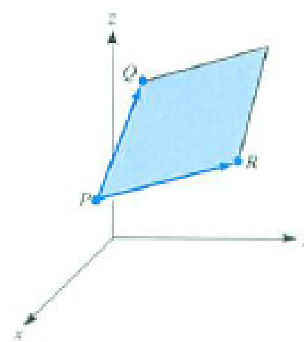


Figure 14.3.4

The next example gives a geometric application of the cross product.

EXAMPLE 3 Find a vector perpendicular to the plane determined by the three points $P = (1, 3, 2)$, $Q = (4, -1, 1)$, and $R = (3, 0, 2)$.
SOLUTION The vectors $\overline{PQ}$ and $\overline{PR}$ lie in a plane (see Figure 14.3.4).

The vector $\mathbf{N} = \overline{PQ} \times \overline{PR}$, being perpendicular to both $\overline{PQ}$ and $\overline{PR}$, is perpendicular to the plane. Because $\overline{PQ} = 3\mathbf{i} - 4\mathbf{j} - \mathbf{k}$ and $\overline{PR} = 2\mathbf{i} - 3\mathbf{j} + 0\mathbf{k}$, thus

$$\mathbf{N} = \det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 3 & -4 & -1 \\ 2 & -3 & 0 \end{pmatrix} = -3\mathbf{i} - 2\mathbf{j} - \mathbf{k}.$$

◇

How Long is $\mathbf{A} \times \mathbf{B}$?

Check these steps by multiplying everything out.

To find a geometric meaning for $|\mathbf{A} \times \mathbf{B}|$ we will find $|\mathbf{A} \times \mathbf{B}|^2$, that is, we will compute $(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{A} \times \mathbf{B})$. Then, by (14.3.6),

$$\begin{aligned} |\mathbf{A} \times \mathbf{B}|^2 &= |(a_2b_3 - a_3b_2)\mathbf{i} + (a_3b_1 - a_1b_3)\mathbf{j} + (a_1b_2 - a_2b_1)\mathbf{k}|^2 \\ &= (a_2b_3 - a_3b_2)^2 + (a_3b_1 - a_1b_3)^2 + (a_1b_2 - a_2b_1)^2 \\ &= a_2^2b_3^2 + a_3^2b_2^2 + a_3^2b_1^2 + a_1^2b_3^2 + a_1^2b_2^2 + a_2^2b_1^2 \\ &\quad -2(a_2a_3b_2b_3 + a_1a_3b_1b_3 + a_1a_2b_1b_2) \\ &= (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) - (a_1b_1 + a_2b_2 + a_3b_3)^2. \end{aligned}$$

The first term is $|\mathbf{A}|^2|\mathbf{B}|^2$ and the second term is the square of $\mathbf{A} \cdot \mathbf{B}$. Let θ denote the angle between $\mathbf{A}$ and $\mathbf{B}$, so $\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}||\mathbf{B}|\cos(\theta)$. Now

$$\begin{aligned} |\mathbf{A} \times \mathbf{B}|^2 &= (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) - (a_1b_1 + a_2b_2 + a_3b_3)^2 \\ &= |\mathbf{A}|^2|\mathbf{B}|^2 - (\mathbf{A} \cdot \mathbf{B})^2 \\ &= |\mathbf{A}|^2|\mathbf{B}|^2 - (|\mathbf{A}||\mathbf{B}|\cos(\theta))^2 \\ &= |\mathbf{A}|^2|\mathbf{B}|^2(1 - \cos^2(\theta)) \\ &= |\mathbf{A}|^2|\mathbf{B}|^2\sin^2(\theta). \end{aligned}$$

Then, because $\sin(\theta)$ is not negative for θ in $[0, \pi]$,

$$|\mathbf{A} \times \mathbf{B}| = |\mathbf{A}||\mathbf{B}|\sin(\theta).$$

We have

We can now give a geometric meaning for the length of $\mathbf{A} \times \mathbf{B}$. The area of the parallelogram is the product of its width and height, see Figure 14.3.5(a). Thus,

$$\text{area of the parallelogram spanned by } \mathbf{A} \text{ and } \mathbf{B} = \underbrace{|\mathbf{A}|}_{\textit{base}}\,\underbrace{|\mathbf{B}|\sin(\theta)}_{\textit{height}}.$$

(a) (b)

Figure 14.3.5 (a) shows the area of a parallelogram is its base times its height.

> **The Length of A × B**
> Let **A** and **B** be nonzero vectors and θ the angle between them. Then
> $$|\mathbf{A} \times \mathbf{B}| = |\mathbf{A}||\mathbf{B}| \sin(\theta).$$
> Geometrically, the length of $\mathbf{A} \times \mathbf{B}$ is the area of the parallelogram spanned by **A** and **B**.

EXAMPLE 4 Find the area of the parallelogram spanned by $\mathbf{A} = a_1\mathbf{i} + a_2\mathbf{j}$ and $\mathbf{B} = b_1\mathbf{i} + b_2\mathbf{j}$.
SOLUTION Write **A** as $a_1\mathbf{i} + a_2\mathbf{j} + 0\mathbf{k}$ and **B** as $b_\mathbf{i} + b_2\mathbf{j} + 0\mathbf{k}$. The area of this parallelogram is the length of $\mathbf{A} \times \mathbf{B}$ and So we compute $\mathbf{A} \times \mathbf{B}$.

$$\mathbf{A} \times \mathbf{B} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & 0 \\ b_1 & b_2 & 0 \end{vmatrix} = (a_1b_2 - a_2b_1)\mathbf{k}.$$

The area is therefore $|a_1b_2 - a_2b_1|$. It is the absolute value of the determinant

$$\det \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}.$$

◇

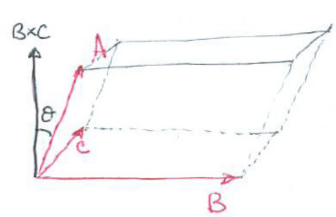


Figure 14.3.6

The Scalar Triple Product

The scalar $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})$ is called the **scalar triple product**. It has an important geometric meaning. (The vector $\mathbf{A} \times (\mathbf{B} \times \mathbf{C})$ is called the **vector triple product**. See Exercises 21 and 22.)

The vectors **A**, **B**, and **C** span a parallelepiped, as shown in Figure 14.3.6. The angle between $\mathbf{B} \times \mathbf{C}$ and **A** is θ (which could be greater than $\pi/2$). The area of the base of the parallelogram is $|\mathbf{B}\times\mathbf{C}|$. The height of the parallelepiped is $|\mathbf{A}||\cos(\theta)|$. Thus its volume is the absolute value of

$$\underbrace{|\mathbf{A}|\cos\theta}_{\text{height}}\ \underbrace{|\mathbf{B}\times\mathbf{C}|}_{\text{area of base}}.$$

This is, by the definition of the dot of product, the dot product of $\mathbf{A}$ and $(\mathbf{B}\times\mathbf{C})$.

$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})$ is plus or minus the volume of the parallelepiped spanned by $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$.

The scalar triple product can be expressed as a determinant. The dot product of $\mathbf{A}$ and $\mathbf{B}\times\mathbf{C}$ is

$$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = a_1 \det\begin{pmatrix} b_2 & b_3 \\ c_2 & c_3 \end{pmatrix} + a_2\left(-\det\begin{pmatrix} b_1 & b_3 \\ c_1 & c_3 \end{pmatrix}\right) + a_3 \det\begin{pmatrix} b_1 & b_2 \\ c_1 & c_2 \end{pmatrix}. \tag{14.3.8}$$

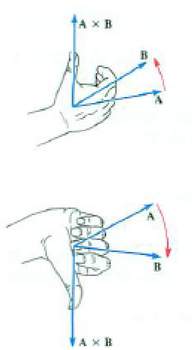


Figure 14.3.7

Or, expressed as the determinant of a matrix of order 3,

$$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

This determinant is plus or minus the volume of the parallelepiped spanned by $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$.

This is like the two-dimensional case where the determinant $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$ is plus or minus the area of the parallelogram spanned by the vectors $\langle a_1, a_2\rangle$ and $\langle b_1, b_2\rangle$.

Summary

We constructed a vector $\mathbf{C}$ perpendicular to vectors $\mathbf{A}$ and $\mathbf{B}$ by requiring that $\mathbf{C}\cdot\mathbf{A} = 0$ and $\mathbf{C}\cdot\mathbf{B} = 0$. A formula for it is

$$\det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3. \end{pmatrix}$$

It is denoted $\mathbf{A}\times\mathbf{B}$ and called the vector product or cross product of $\mathbf{A}$ and $\mathbf{B}$. It also may be described as the vector whose length is the area of the parallelogram spanned by $\mathbf{A}$ and $\mathbf{B}$ and whose direction is given by the right-hand rule (the finger curling from $\mathbf{A}$ and $\mathbf{B}$). It has the properties:

1. $\mathbf{A} \times \mathbf{B} = -(\mathbf{B} \times \mathbf{A})$ (anticommutative)
2. $\mathbf{A} \times (\mathbf{B} \times \mathbf{C})$ is not usually equal to $(\mathbf{A} \times \mathbf{B}) \times \mathbf{C})$ (not associative)
3. $\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{C} \cdot \mathbf{A})\mathbf{B} - (\mathbf{B} \cdot \mathbf{A})\mathbf{C}$ (See Exercises 21 and 22.)
4. $(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{A} \times \mathbf{B}) = (\mathbf{A} \cdot \mathbf{A})(\mathbf{B} \cdot \mathbf{B}) - (\mathbf{A} \cdot \mathbf{B})(\mathbf{A} \cdot \mathbf{B})$
5. $\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \pm$ volume of parallelepiped spanned by $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$

Item 4 appeared in finding the length of $\mathbf{A} \times \mathbf{B}$. It will be used in the next chapter.

EXERCISES for Section 14.3

In Exercises 1 to 4 compute and sketch $\mathbf{A} \times \mathbf{B}$.

1. $\mathbf{A} = \mathbf{k}$, $\mathbf{B} = \mathbf{j}$

2. $\mathbf{A} = \mathbf{i} + \mathbf{j}$, $\mathbf{B} = \mathbf{i} - \mathbf{j}$

3. $\mathbf{A} = \mathbf{i} + \mathbf{j} + \mathbf{k}$, $\mathbf{B} = \mathbf{i} + \mathbf{j}$

4. $\mathbf{A} = \mathbf{k}$, $\mathbf{B} = \mathbf{i} + \mathbf{j}$

In Exercises 5 and 6, find $\mathbf{A} \times \mathbf{B}$ and check that it is perpendicular to $\mathbf{A}$ and $\mathbf{B}$.

5. $\mathbf{A} = 2\mathbf{i} - 3\mathbf{j} + \mathbf{k}$, $\mathbf{B} = \mathbf{i} + \mathbf{j} + 2\mathbf{k}$

6. $\mathbf{A} = \mathbf{i} - \mathbf{j}$, $\mathbf{B} = \mathbf{j} + 4\mathbf{k}$

In Exercises 7 to 10 use the cross product to find the area of.

7. the parallelogram three of whose vertices are $(0,0,0)$, $(1,5,4)$, and $(2,-1,3)$.

8. the parallelogram three of whose vertices are $(1,2,-1)$, $(2,1,4)$, and $(3,5,2)$.

9. the triangle two of whose sides are $\mathbf{i} + \mathbf{j}$ and $3\mathbf{i} - \mathbf{j}$.

10. the triangle two of whose sides are $\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$ and $2\mathbf{i} - \mathbf{j} + 2\mathbf{k}$.

In Exercises 11 to 14 find the volume of the parallelepiped spanned by

11. $\langle 2,1,3\rangle$, $\langle 3,-1,2\rangle$, $\langle 4,0,3\rangle$

12. $3\mathbf{i} + 4\mathbf{j} + 3\mathbf{k}$, $2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$, $\mathbf{i} - \mathbf{j} - \mathbf{k}$.

13. $\overline{PQ}$, $\overline{PR}$, $\overline{PS}$, where $P = (1,1,1)$, $Q = (2,1,-2)$, $R = (3,5,2)$, and $S = (1,-1,2)$.

14. $\overline{PQ}$, $\overline{PR}$, $\overline{PS}$, where $P = (0,0,0)$, $Q = (3,3,2)$, $R = (1,4,-1)$, and $S = (1,2,3)$.

15. Evaluate $\mathbf{A} \cdot (\mathbf{A} \times \mathbf{B})$.

16. Prove that $\mathbf{B} \times \mathbf{A} = -(\mathbf{A} \times \mathbf{B})$ in two ways

(a) using the algebraic definition of the cross product

(b) using the geometric description of the cross product

17. Show that if $\mathbf{B} = c\mathbf{A}$, then $\mathbf{A} \times \mathbf{B} = \mathbf{0}$

(a) using the algebraic definition of the cross product

(b) using the geometric description of the cross product

18. Show that $(0,0,0)$, (x_1,y_1,z_1), (x_2,y_2,z_2) and (x_3,y_3,z_3) lie on a plane if and only if

$$\begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{vmatrix} = 0.$$

19.

(a) If $\mathbf{B}$ is parallel to $\mathbf{C}$, is $\mathbf{A}\times\mathbf{B}$ parallel to $\mathbf{A}\times\mathbf{C}$?

(b) If $\mathbf{B}$ is perpendicular to $\mathbf{C}$, is $\mathbf{A}\times\mathbf{B}$ perpendicular to $\mathbf{A}\times\mathbf{C}$?

20. Let $\mathbf{A}$ be a nonzero vector. If $\mathbf{A}\times\mathbf{B}=\mathbf{0}$ and $\mathbf{A}\cdot\mathbf{B}=0$, must $\mathbf{B}=\mathbf{0}$?

21. Show that $\mathbf{A}\times(\mathbf{A}\times\mathbf{B}) = (\mathbf{A}\cdot\mathbf{B})\mathbf{A} - (\mathbf{A}\cdot\mathbf{A})\mathbf{B}$.

22. Show that $(\mathbf{A}\times\mathbf{B})\times(\mathbf{C}\times\mathbf{D}) = ((\mathbf{A}\times\mathbf{B})\cdot\mathbf{D})\mathbf{C} - ((\mathbf{A}\times\mathbf{B})\cdot\mathbf{C})\mathbf{D}$. (Think of $\mathbf{A}\times\mathbf{B}$ as a single vector, $\mathbf{E}$.)

23.

(a) Give an example of a vector perpendicular to $3\mathbf{i}-\mathbf{j}+\mathbf{k}$.

(b) Give an example of a unit vector perpendicular to $3\mathbf{i}-\mathbf{j}+\mathbf{k}$.

24. Let $\mathbf{u}$ be a unit vector and $\mathbf{B}$ a vector. What happens as you keep crossing by $\mathbf{u}$, and form the sequence $\mathbf{B}$, $\mathbf{u}\times\mathbf{B}$, $\mathbf{u}\times(\mathbf{u}\times\mathbf{B})$ and so on? (See Exercise 21)

25. *(Crystallography)* A crystal is described by three vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. They span a fundamental parallelepiped, whose copies fill out a crystal lattice. (See Figure 14.3.8.) Atoms are at the corners. To study the diffraction of x-rays and light through a crystal, crystallographers work with the reciprocal lattice, whose fundamental parallelepiped is spanned by three vectors, $\mathbf{k}_1$, $\mathbf{k}_2$, and $\mathbf{k}_3$. The vector $\mathbf{k}_1$ is perpendicular to the parallelogram spanned by $\mathbf{v}_2$ and $\mathbf{v}_3$ and has a length equal to the reciprocal of the distance between the parallelogram and the opposite parallelogram of the fundamental parallelepiped. The vectors $\mathbf{k}_2$ and $\mathbf{k}_3$ are defined similarly in terms of the other four faces of the fundamental parallelepiped.

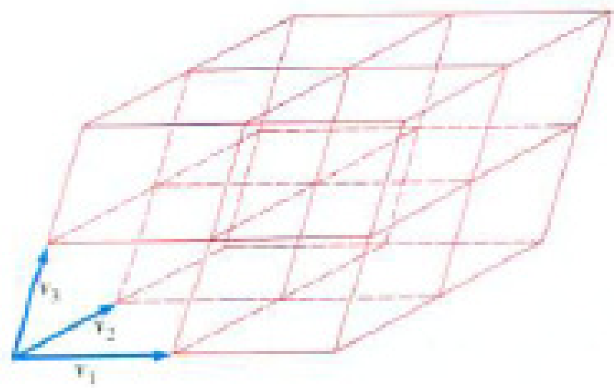


Figure 14.3.8

(a) Show that $\mathbf{k}_1$, $\mathbf{k}_2$, and $\mathbf{k}_3$ may be chosen to be

$$\mathbf{k}_1 = \frac{\mathbf{v}_2 \times \mathbf{v}_3}{\mathbf{v}_1 \cdot (\mathbf{v}_2 \times \mathbf{v}_3)}, \qquad \mathbf{k}_2 = \frac{\mathbf{v}_3 \times \mathbf{v}_1}{\mathbf{v}_1 \cdot (\mathbf{v}_2 \times \mathbf{v}_3)}, \qquad \mathbf{k}_3 = \frac{\mathbf{v}_1 \times \mathbf{v}_2}{\mathbf{v}_1 \cdot (\mathbf{v}_2 \times \mathbf{v}_3)}.$$

(b) Show that the volume of the parallelopiped determined by $\mathbf{k}_1$, $\mathbf{k}_2$, and $\mathbf{k}_3$ is the reciprocal of the volume of the one determined by $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$.

(c) Is the reciprocal of the reciprocal lattice the original lattice? For instance, is

$$\mathbf{v}_1 = \frac{\mathbf{k}_2 \times \mathbf{k}_3}{\mathbf{k}_1 \cdot (\mathbf{k}_2 \times \mathbf{k}_3)}?$$

26. The purpose of this Exercise is to determine the identity known as Jacobi's Identity, $\mathbf{a} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C}$. It will be used in Chapter 18 when dealing with electric currents and magnetic fields.
Let $\mathbf{B}$ and $\mathbf{C}$ be nonzero, nonparallel vectors and $\mathbf{A}$ a vector that is perpendicular neither to $\mathbf{B}$ nor $\mathbf{C}$.

(a) Why are their scalars x and y such that

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = x\mathbf{B} + y\mathbf{C}?$$

(b) Why is $0 = x(\mathbf{A} \cdot \mathbf{B}) + y(\mathbf{A} \cdot \mathbf{C})$?

(c) Using (b), show that there is a scalar z such that

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = z[(\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C}].$$

(d) It would be nice if there were a simple geometric way to show that z is a constant and equals 1. We could show that $z = 1$ by writing $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ in components and making a tedious calculation. Find a simple way to figure out that $z = 1$.

27. Let $\mathbf{A}$ be a nonzero vector and $\mathbf{B}$ be a vector. Let $\mathbf{B}_1$ be the projection of $\mathbf{B}$ on a plane perpendicular to $\mathbf{A}$. Let $\mathbf{B}_2$ be obtained by rotating $\mathbf{B}_1$ through an angle of 90° in the direction given by the right-hand rule with thumb pointing in the same direction as $\mathbf{A}$

(a) Show that $\mathbf{A} \times \mathbf{B} = \mathbf{A} \times \mathbf{B}_1$. (Draw a diagram.)

(b) Show that $\mathbf{A} \times \mathbf{B} = |\mathbf{A}|\mathbf{B}_2$.

28. Using Exercise 27(b), show that for $\mathbf{A}$ not $\mathbf{0}$, $\mathbf{A} \times (\mathbf{B}+\mathbf{C}) = \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C}$. (Draw a large, clear picture.)

29.

(a) From the distributive law $\mathbf{A} \times (\mathbf{B}+\mathbf{C}) = \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C}$, and the property $\mathbf{D} \times \mathbf{E} = -\mathbf{E} \times \mathbf{D}$, deduce the distributive law $(\mathbf{B}+\mathbf{C}) \times \mathbf{A} = \mathbf{B} \times \mathbf{A} + \mathbf{C} \times \mathbf{A}$.

(b) From the distributive law $\mathbf{A} \times (\mathbf{B}+\mathbf{C}) = \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C}$, show that $\mathbf{A} \times (\mathbf{B}+\mathbf{C}+\mathbf{D}) = \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C} + \mathbf{A} \times \mathbf{D}$. (Think of $\mathbf{B}+\mathbf{C}$ as a single vector $\mathbf{E}$.)

30. Check that $-13\mathbf{i}+7\mathbf{j}+11\mathbf{k}$ in Example 1 is perpendicular to $\mathbf{A}$ and to $\mathbf{B}$.

31. Show, using (14.3.7), that $\mathbf{0} \times \mathbf{B} = \mathbf{0}$.

32. Show, using (14.3.7), that $\mathbf{B} \times \mathbf{A} = -\mathbf{A} \times \mathbf{B}$.

33. Use the values for y and z when solving equations (14.3.3) and (14.3.4). Substitute them into (14.3.1) and solve for x.

34. Using components, show that $\mathbf{A} \times (\mathbf{B}+\mathbf{C}) = \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C}$. You may use properties of determinants.

35. Using (14.3.7), show that if $\mathbf{B}$ is parallel to $\mathbf{A}$, then $\mathbf{A} \times \mathbf{B} = \mathbf{0}$. (If $\mathbf{B}$ is parallel to $\mathbf{A}$, there is a scalar t such that $\mathbf{B} = t\mathbf{A}$.)

36. In finding $|\mathbf{A} \times \mathbf{B}|^2$ we stated that

$$a_2^2b_3^2 + a_3^2b_2^2 + a_3^2b_1^2 + a_1^2b_3^2 + a_1^2b_2^2 + a_2^2b_1^2 - 2(a_2a_3b_2b_3 + a_1a_3b_1b_3 + a_1a_2b_1b_2)$$

equals

$$(a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) - (a_1b_1 + a_2b_2 + a_3b_3)^2.$$

Check that.

37.

(a) How could you use cross products to find a vector perpendicular to $2\mathbf{i}+3\mathbf{j}+4\mathbf{k}$? Give an example.

(b) How could you use cross products to find two vectors perpendicular to $2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$ and to each other? Give an example.

38. To understand why you cannot omit the parentheses in $\mathbf{A} \times (\mathbf{B} \times \mathbf{C})$, let $\mathbf{A}$ and $\mathbf{B}$ be non-zero, non-parallel vectors. Show that $\mathbf{A} \times (\mathbf{A} \times \mathbf{B})$ is never equal to $(\mathbf{A} \times \mathbf{A}) \times \mathbf{B}$. This shows that the cross product is not associative.

39. We showed that the direction of $\mathbf{i} \times \mathbf{j}$ is given by the right hand rule. Then we said that the right hand rule holds for any non-zero vectors $\mathbf{A}$ and $\mathbf{B}$. Why is that justified? (Imagine moving a gradually changing pair of vectors through space, starting with $\mathbf{i}$ and $\mathbf{j}$ and ending with $\mathbf{A}$ and $\mathbf{B}$.)

40.

(a) Explain using parallelopipeds why $\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C})$ is plus or minus $\mathbf{B} \cdot (\mathbf{C} \times \mathbf{A})$.

(b) Using properties of determinants, decide if it is plus or minus.

41. In some expositions of the cross product, $\mathbf{A} \times \mathbf{B}$ is defined as the determinant of a matrix of order 3. If we start with this definition, use a property of determents to show that $\mathbf{A} \times \mathbf{B}$ is perpendicular to both $\mathbf{A}$ and $\mathbf{B}$.

14.4 Applications of the Dot Product

This section uses the dot product to deal with lines, planes, and projections onto planes.

Equation of a Plane

First we find an equation of the plane through the point $P_0 = (x_0, y_0, z_0)$ and perpendicular to the vector $A\mathbf{i} + B\mathbf{j} + C\mathbf{k}$, shown in Figure 14.4.1.

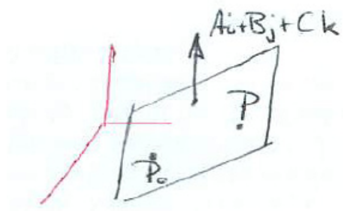


Figure 14.4.1

Let $P = (x, y, z)$ be a point on the plane. The vector $\overline{P_0P}$ is perpendicular to $A\mathbf{i} + B\mathbf{j} + C\mathbf{k}$. Slide it so that P_0 coincides with the tail of $A\mathbf{i} + B\mathbf{j} + C\mathbf{k}$. Thus

$$(A\mathbf{i} + B\mathbf{j} + C\mathbf{k}) \cdot ((x - x_0)\mathbf{i} + (y - y_0)\mathbf{j} + (z - z_0)\mathbf{k}) = 0,$$

so we have

Equation of the Plane

An equation for the plane containing the point (x_0, y_0, z_0) and perpendicular to the vector $A\mathbf{i} + B\mathbf{j} + C\mathbf{k}$ is

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0. \tag{14.4.1}$$

The vector $A\mathbf{i} + B\mathbf{j} + C\mathbf{k}$ is called a **normal** to the plane.

EXAMPLE 1 Find an equation for the plane through $(2, -3, 4)$ and normal to $\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$.

SOLUTION An equation for the plane is

$$1(x - 2) + 2(y - (-3)) + 3(z - 4) = 0$$

which simplifies to

$$x + 2y + 3z - 8 = 0.$$

◇

The graph of $Ax + By + Cz + D = 0$, where not all of A, B, and C are 0, is a plane perpendicular to $A\mathbf{i} + B\mathbf{j} + C\mathbf{k}$. To show this, first pick a point (x_0, y_0, z_0) that satisfies the equation $Ax_0 + By_0 + Cz_0 + D = 0$. Subtracting this from the original equation gives

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0,$$

which is an equation of the plane through (x_0, y_0, z_0) perpendicular to $A\mathbf{i} + B\mathbf{j} + C\mathbf{k}$.

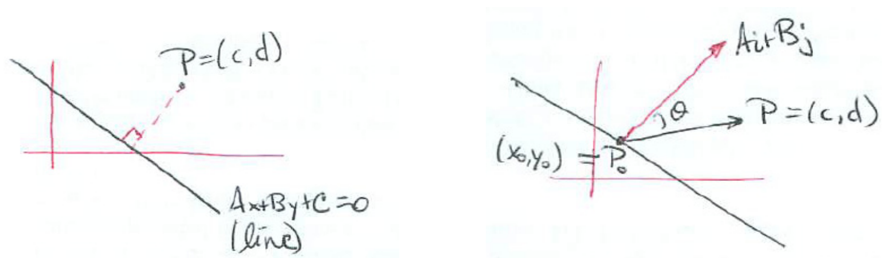
In two dimensions we have

Equation of a Line in the Plane

An equation for the line through (x_0, y_0) and perpendicular to the vector $A\mathbf{i} + B\mathbf{j}$ is

$$A(x - x_0) + B(y - y_0) = 0.$$

Distance From a Point to the Line $Ax + By + C = 0$ or to the Plane $Ax + By + Cz + D = 0$



(a) (b)

Figure 14.4.2

Let us find the distance from $P = (p, q)$ to the line in the xy-plane whose equation is $Ax + By + C = 0$, shown in Figure 14.4.2(a).

Pick a point $P_0 = (x_0, y_0)$ on the line and place $A\mathbf{i} + B\mathbf{j}$ with its tail at P_0, as in Figure 14.4.2(b).

Let θ be the angle between $\overline{P_0P}$ and $A\mathbf{i} + B\mathbf{j}$. Then the distance from P to the line is

Because $0 \le \theta \le 2\pi$, $\cos(\theta)$ could be negative.

$$\begin{aligned}
|P_0P||\cos(\theta)| &= P_0P|\frac{(A\mathbf{i} + B\mathbf{j}) \cdot ((p - x_0)\mathbf{i} + (q - y_0)\mathbf{j})}{P_0P|A\mathbf{i} + B\mathbf{j}|} \\
&= \frac{A(p - x_0) + B(q - y_0)}{\sqrt{A^2 + B^2}} \\
&= \frac{Ap + Bq - (Ax_0 + By_0)}{\sqrt{A^2 + B^2}}.
\end{aligned}$$

Since $Ax_0 + By_0 + C = 0$, we have

Distance from a Point to a Line

The distance from (p, q) to the line $Ax + By + C = 0$ is

$$\frac{|Ap + Bq + C|}{\sqrt{A^2 + B^2}}$$

So, to find the distance substitute the coordinates of the point (p,q) into $Ax+By+C$, divide by $\sqrt{A^2+B^2}$, and take its absolute value.

EXAMPLE 2 How far is the point $(1,3)$ from the line $2x-4y=5$?
SOLUTION Write the equation in the form $2x-4y-5=0$. Then the distance is

$$\frac{|2(1)-4(3)-5|}{\sqrt{2^2+4^2}} = \frac{|-15|}{\sqrt{20}} = \frac{3\sqrt{5}}{2}.$$

◇

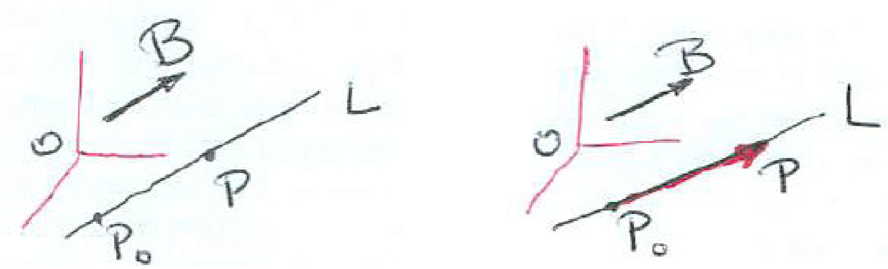
The corresponding formula for the distance from a point $P=(x_0,y_0,z_0)$ to a plane is obtained in Exercise 55.

Distance from a Point to a Plane
The distance from (x_0,y_0,z_0) to the plane $Ax+By+Cz+D=-$ is

$$\frac{|Ax_0+By_0+Cz_0+D|}{\sqrt{A^2+B^2+C^2}}$$

Using Vectors to Parameterize a Line

Let L be the line through the point $P_0=(x_0,y_0,z_0)$ parallel to $\mathbf{B}$, shown in Figure 14.4.3(a).



(a) (b)

Figure 14.4.3

Let P be a point on L. Then the vector $\overline{P_0P}$, which is parallel to $\mathbf{B}$, is $t\mathbf{B}$ for some scalar t. See Figure 14.4.3(b).
So we have $\overline{OP}=\overline{OP_0}+\overline{P_0P}=\overline{OP_0}+t\mathbf{B}$.

Parametric Equation of a Line
The line through P_0 parallel to the vector $\mathbf{B}$ is parameterized by $\overline{OP}=\overline{OP_0}+t\mathbf{B}$. As t varies, the vector from O to P varies, sweeping out the line L.

EXAMPLE 3 The line L passes through the point $(1,1,2)$ and is parallel to the vector $3\mathbf{i}+4\mathbf{j}+5\mathbf{k}$. Parameterize the line.
SOLUTION Because $\overline{OP_0}=\mathbf{i}+\mathbf{j}+2\mathbf{k}$ and $\mathbf{B}=3\mathbf{i}+4\mathbf{j}+5\mathbf{k}$,

$$\begin{aligned}\overline{OP} &= \mathbf{i}+\mathbf{j}+2\mathbf{k}+t(3\mathbf{i}+4\mathbf{j}+5\mathbf{k})\\ &= (3t+1)\mathbf{i}+(4t+1)\mathbf{j}+(5t+2)\mathbf{k}.\end{aligned}$$

If P is (x,y,z), then $\overline{OP}$ is the vector $x\mathbf{i}+y\mathbf{j}+z\mathbf{k}$.
Thus

$$\begin{aligned}x &= 3t+1\\ y &= 4t+1\\ z &= 5t+2.\end{aligned}$$

One vector equation does the work of three scalar equation. ◇

Describing the Direction of Vectors and Lines

The direction of a vector in the plane is described by the angle it makes with the positive x-axis. To specify the direction of a vector in space requires three angles, two of which almost determine the third.

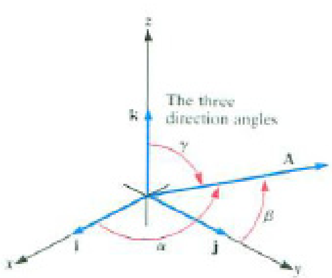


Figure 14.4.4

DEFINITION *(Direction of a vector.)* Let $\mathbf{A}$ be a nonzero vector in space. The angle between

$\mathbf{A}$ and $\mathbf{i}$ is denoted α (alpha),

$\mathbf{A}$ and $\mathbf{j}$ is denoted β (beta),

$\mathbf{A}$ and $\mathbf{k}$ is denoted γ (gamma).

The angles α, β and γ are called the **direction angles of A**. (See Figure 14.4.4.)

DEFINITION *(Direction cosines of a vector)* The **direction cosines** of a vector are the cosines of its direction angles, $\cos(\alpha)$, $\cos(\beta)$, and $\cos(\gamma)$.

EXAMPLE 4 The angle between $\mathbf{A}$ and $\mathbf{k}$ is $\pi/6$. Find γ and $\cos(\gamma)$ for

1. $\mathbf{A}$

2. $-\mathbf{A}$.

SOLUTION

1. By definition, the direction angle γ for $\mathbf{A}$ is $\pi/6$. It follows that $\cos(\gamma) = \cos(\pi/6) = \sqrt{3}/2$.

2. To find γ and $\cos(\gamma)$ for $-\mathbf{A}$, we draw Figure 14.4.5. For $-\mathbf{A}$, $\gamma = 5\pi/6$ and $\cos(\gamma) = \cos(5\pi/6) = -\sqrt{3}/2$.

◇

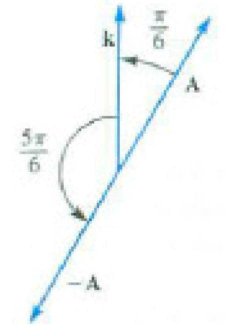


Figure 14.4.5

As Example 4 illustrates, if the direction angles of $\mathbf{A}$ are α, β, γ, then the direction angles of $-\mathbf{A}$ are $\pi - \alpha$, $\pi - \beta$, and $\pi - \gamma$. The direction cosines of $-\mathbf{A}$ are the negatives of the direction cosines of $\mathbf{A}$.

The three direction angles are not independent, as is shown by the next theorem. Two of them determine the third up to sign.

Theorem 14.4.1. *If α, β, γ are the direction angles of $\mathbf{A}$, then* $\cos^2(\alpha) + \cos^2(\beta) + \cos^2(\gamma) = 1$.

Proof

It is no loss of generality to assume that $\mathbf{A}$ is a unit vector. Its component on the y-axis, for instance, is $\cos(\beta)$, as the right triangle $\triangle OPQ$ in Figure 14.4.6 shows. The vector $\mathbf{A}$ lies along the hypotenuse.

Since $\mathbf{A}$ is a unit vector, $|\mathbf{A}|^2 = 1$, and we have $\cos^2(\alpha)+\cos^2(\beta)+\cos^2(\gamma) = 1^2 = 1$. •

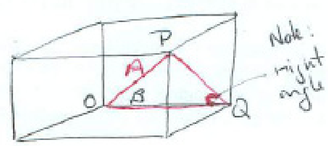


Figure 14.4.6

EXAMPLE 5 If $\mathbf{A}$ makes an angle of 60° with both the x-axis and the y-axis. What angle does it make with the z-axis?

SOLUTION Here $\alpha = 60°$ and $\beta = 60°$; hence

$$\cos(\alpha) = \frac{1}{2} \quad \text{and} \quad \cos(\beta) = \frac{1}{2}.$$

Since

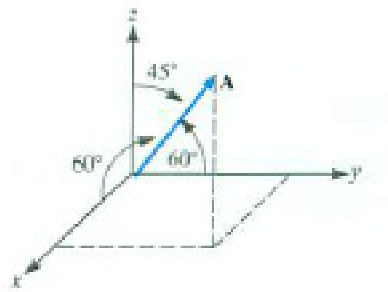
$$\cos^2(\alpha) + \cos^2(\beta) + \cos^2(\gamma) = 1,$$

it follows that

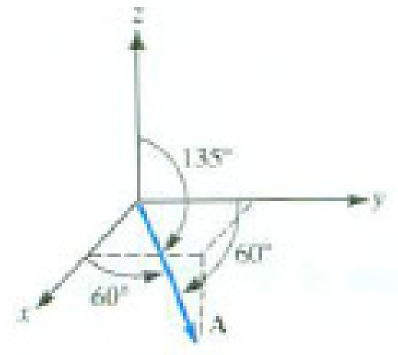
$$\begin{aligned} (\frac{1}{2})^2 + (\frac{1}{2})^2 + \cos^2(\gamma) &= 1, \\ \cos^2(\gamma) &= \frac{1}{2}. \end{aligned}$$

Thus

$$\cos(\gamma) = \frac{\sqrt{2}}{2} \quad \text{or} \quad \cos(\gamma) = -\frac{\sqrt{2}}{2}.$$



(a)



(b)

Figure 14.4.7

Hence

$$\gamma = 45^\circ \quad \text{or} \quad \gamma = 135^\circ.$$

Figures 14.4.7(a) and (b) show the possibilities for **A**. ◇

Dot Products and Flow

Let the vector **v** whose magnitude is v describe the velocity of water flowing down a river, as in Figure 14.4.8(a). Place a stick of length L on the surface

(a)

(b)

(c)

Figure 14.4.8

of the water. The amount of water crossing the stick depends on its position. If the stick is parallel to **v**, no water crosses it. If the stick is perpendicular to **v**, water crosses it. How does the angle at which we place the stick affect the amount of water that crosses it?"

To answer this, we introduce a unit vector **n** perpendicular to the stick and record its position, as in Figure 14.4.8(b). Let the angle between **n** and **v** be θ.

The amount of water that crosses the stick during time Δt is proportional to the area of the parallelogram in Figure 14.4.8(c). The base of the parallelogram has length $v\Delta t$ (speed times time). The height is $L\cos(\theta)$. The area of the parallelogram is therefore

$$vL\cos(\theta)\Delta t.$$

Then $vL\cos(\theta)$ measures the amount of water that crosses the stick in one unit of time.

But $vL\cos(\theta)$ is equal to $\mathbf{v}\cdot\mathbf{n}$. So $\mathbf{v}\cdot\mathbf{n}$ measures the tendency of water to cross the stick.

As a check, when the stick is parallel to $\mathbf{v}$, $\theta = \pi/2$ and $\cos(\pi/2) = 0$. Then $\mathbf{v}\cdot\mathbf{n} = 0$ and no water crosses the stick. When the stick is perpendicular to $\mathbf{v}$, $\theta = 0$, and $\mathbf{v}\cdot\mathbf{n} = v$.

EXAMPLE 6 When a stick is perpendicular to $\mathbf{v}$, water crosses it at the rate of 100 cubic feet per second. When the stick is placed at an angle of $\pi/6$ to $\mathbf{v}$ at what rate does water cross it?

SOLUTION Figure 14.4.9 shows the position of the stick $\overline{PQ}$.

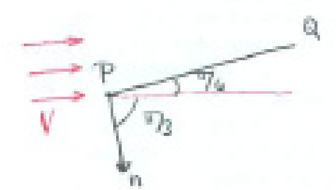


Figure 14.4.9

The angle between the normal to the stick, $\mathbf{n}$, and $\mathbf{v}$ is $\pi/2 - \pi/6 = \pi/3$. Let x be the rate at which the water crosses the stick. Since the rate of flow across the stick is proportional to $v\cos(\theta)$, where θ is the angle between the normal $\mathbf{n}$ and $\mathbf{v}$, we have

$$\frac{100}{v\cos(0)} = \frac{x}{v\cos(\pi/3)}.$$

This tells us that

$$\frac{100}{v} = \frac{x}{(v)(1/2)},$$

hence $x = 50$. The flow is half the maximum possible. ◇

Summary

We used the dot product to obtain an equation of a plane in space (or of a line in the xy-plane) and to find the distance from a point to a line or plane. We also showed how to parameterize a line using a vector parallel to the line.

Direction angles and cosines of a vector were defined. We showed how the dot product describes the rate of flow across a line segment. This concept will be needed in Chapters 17 and 18, where we deal with flows across curves and surfaces.

EXERCISES for Section 14.4

In Exercises 1 to 4 find an equation of the line in the xy-plane through the point and perpendicular to the vector.

1. $(2, 3)$, $4\mathbf{i} + 5\mathbf{j}$

2. $(1, 0)$, $2\mathbf{i} - \mathbf{j}$

3. $(4, 5)$, $1\mathbf{i} + 3\mathbf{j}$

4. $(2, -1)$, $\mathbf{i} + 3\mathbf{j}$

In Exercises 5 to 8 find a vector in the xy-plane that is perpendicular to the line.

5. $2x - 3y + 8 = 0$

6. $\pi x - \sqrt{2}y = 7$

7. $y = 3x + 7$

8. $2(x - 1) + 5(y + 2) = 0$

In Exercises 9 to 12 find an equation of the line in xyz-space with the indicated properties.

9. Contains $(2, 3, 4)$ and perpendicular to $4\mathbf{i} + 5\mathbf{j}$ and $3\mathbf{i} + 5\mathbf{j} + 6\mathbf{j}$

10. Belongs to both $x + 2y - z = 4$ and $-2x + 2y + 5z = -2$

11. Contains $(4, 0, 5)$ and parallel to $1\mathbf{i} + 3\mathbf{j} - 2\mathbf{j}$

12. Contains the three points $(2, -1, -1)$, $(0, 1, 3)$ and $(2, -1, 2)$

13. Find a vector perpendicular to the plane through $(2, 1, 3)$, $(4, 5, 1)$ and $(-2, 2, 3)$.

14. How far is the point $(1, 2, 2)$ from the plane through $(0, 0, 0)$, $(3, 5, -2)$, and $(2, -1, 3)$?

15. How far is the point $(1, 2, 3)$ from the line through $(-2, -1, 3)$, and $(4, 1, 2)$?

16.

(a) Describe how you would find an equation for the plane through $P_1 = (x_1, y_1, z_1)$, $P_2 = (x_2, y_2, z_2)$, and $P_3 = (x_3, y_3, z_3)$.

(b) Find an equation for the plane through $(2, 2, 1)$, $(0, 1, 5)$ and $(2, -1, 0)$.

17.

(a) Describe how you would decide whether the line through $P_1 = (x_1, y_1, z_1)$ and $P_2 = (x_2, y_2, z_2)$, is parallel to the line through $P_3 = (x_3, y_3, z_3)$ and $P_4 = (x_4, y_4, z_4)$.

(b) Is the line through $(1, 2, -3)$ and $(5, 9, 4)$ parallel to the line through $(-1, -1, 2)$ and $(1, 3, 5)$?

18.

(a) Describe how you would decide whether the line through $P_1 = (x_1, y_1, z_1)$ and $P_2 = (x_2, y_2, z_2)$ is parallel to the plane $Ax + By + Cz + D = 0$.

(b) Is the line through $(1, -2, 3)$ and $(5, 3, 0)$ parallel to the plane $2x-y+z+3 = 0$?

19.

(a) Describe how you would decide whether the line through P_1 and P_2 is parallel to the plane through Q_1, Q_2, and Q_3.

(b) Is the line through $(0, 0, 0)$ and $(1, 1, -1)$ parallel to the plane through $(1, 0, 1)$, $(2, 1, 0)$, and $(1, 3, 4)$?

20.

(a) How would you decide whether the plane through P_1, P_2 and P_3 is parallel to the plane through Q_1, Q_2, and Q_3.

(b) Is the plane through $(1, 2, 3)$, $(4, 1, -1)$, and $(2, 0, 1)$ parallel to the plane through $(2, 3, 4)$, $(5, 2, 0)$, and $(3, 1, 2)$?

21. Find the parametric equations of the line through $(1, 1, 2)$ and perpendicular to the plane $3x - y + z = 6$.

22. Find an equation of the plane through $(1, 2, 3)$ that contains the line given parametrically as $\overline{OP} = 2\mathbf{i} - \mathbf{j} + 3\mathbf{k} + t(3\mathbf{i} + 2\mathbf{j} + \mathbf{k})$.

23. Is the point $(21, -3, 28)$ on the line given parametrically as $\overline{OP} = \mathbf{i} + 2\mathbf{j} + 3\mathbf{k} + t(4\mathbf{i} - \mathbf{j} + 5\mathbf{k})$?

24. Find the angle between the line through $(3, 2, 2)$ and $(4, 3, 1)$ and the line through $(3, 2, 2)$ and $(5, 2, 7)$.

25. The angle between two planes is the angle between their normals. Find the angle between the planes $2x + 3y + 4z = 11$ and $3x - y + 2z = 13$.

26.

(a) How many unit vectors are perpendicular to the plane $Ax+By+Cz+D=0$?

(b) How would you find one of them?

(c) Find a unit vector perpendicular to the plane $3x-2y+4z+6=0$.

27.

(a) How would you find a point on the plane $Ax+By+Cz+D=0$?

(b) Give the coordinates of a point that lies on the plane $3x-y+z+10=0$.

28.

(a) How would you find a point that lies on both planes $A_1x+B_1y+C_1z+D_1=0$ and $A_2x+B_2y+C_2z+D_2=0$?

(b) Find a point that lies on both planes $3x+z+2=0$ and $x-y-z+5=0$.

29.

(a) Let $\mathbf{A}$ and $\mathbf{B}$ be vectors in space. How would you find the area of the parallelogram they span?

(b) Find the area of the parallelogram spanned by $(2,3,1)$ and $(4,-1,5)$.

In Exercises 30 to 33 find the distance from the point to the plane.

30. The point $(0,0,0)$ to the plane $2x-4y+3z+2=0$

31. The point $(1,2,3)$ to the plane $x+2y-3z+5=0$.

32. The point $(2,2,-1)$ to the plane that passes through $(1,4,3)$ and has a normal $2\mathbf{i}-7\mathbf{j}+2\mathbf{k}$.

33. The point $(0,0,0)$ to the plane that passes through $(4,1,0)$ and is perpendicular to the vector $\mathbf{i}+\mathbf{j}+\mathbf{k}$.

34. Find the direction cosines of $2\mathbf{i}+3\mathbf{j}+4\mathbf{k}$.

35. Find the direction cosines of the vector from $(1,3,2)$ to $(4,-1,5)$.

36. Let $P_0 = (2,1,5)$ and $P_1 = (3,0,4)$. Find the direction cosines and direction angles of

(a) $\overline{P_0P_1}$

(b) $\overline{P_1P_0}$.

37. Give parametric equations for the line through $(1/2, 1/3, 1/2)$ with direction numbers 2, -5, and 8 in

(a) scalar form

(b) vector form

38. Give parametric equations for the line through $(1,2,3)$ and $(4,5,7)$ in

(a) scalar form

(b) vector form

39. Give parametric equations for the line through $(7,-1,5)$ and $(4,3,2)$.

40. A vector $\mathbf{A}$ has direction angles $\alpha = 70°$ and $\beta = 80°$. Find the third direction angle γ and show the possibile angles for γ on a diagram.

41. Find where the line through $(1,2)$ and $(3,5)$ meets the line through $(1,-1)$ and $(2,3)$.

42. Where does the line through $(1,2,4)$ and $(2,1,-1)$ meet the plane $x+2y+5z = 0$?

43. Give parametric equations for the line through $(1,3,-5)$ that is perpendicular to the plane $2x - 3y + 4z = 11$.

44. How far is the point $(1,5)$ from the line through $(4,2)$ and $(3,7)$? (Draw a picture and think in terms of vectors.)

45. How far is the point $(1,2,-3)$ from the line through $(2,1,4)$ and $(1,5,-2)$?

46. Give parametric equations for the line through $(1, 3, 4)$ that is parallel to the line through $(2, 4, 6)$ and $(5, 3, -2)$.

47.

(a) If you know the coordinates of point P and parametric equations of line L, how would you find an equation of the plane that contains P and L? Assume P is not on L.

(b) Find an equation for the plane through $(1, 1, 1)$ that contains the line parameterized by

$$\begin{aligned} x &= 2+t \\ y &= 3-t \\ z &= 4+2t. \end{aligned}$$

48.

(a) Sketch four points P, Q, R, and S, not all in one plane, such that $\overline{PQ}$ and $\overline{RS}$ are not parallel. Explain way there is a unique pair of parallel planes one of which contains P and Q and one of which contains R and S.

(b) Express a normal vector to the planes in terms of P, Q, R, and S.

49. Find an equation for the plane through P_1 that is parallel to the non-parallel segments P_2P_3 and P_4P_5.

50. Find where the line through $P_0 = (2, 1, 3)$ and $P_1 = (4, -2, 5)$ meets the plane whose equation is $2x + y - 4z + 5 = 0$.

51. Find where the line through $(1, 2, 1)$ and $(2, 1, 3)$ meets the plane that is perpendicular to $2\mathbf{i} + 5\mathbf{j} + 7\mathbf{k}$ and passes through the point $(1, -2, -3)$.

52. Are the points $(1, 2, -3)$, $(1, 6, 2)$, and $(7, 14, 11)$ on a single line?

53. If α, β, and γ are direction angles of a vector, what is $\sin^2(\alpha)+\sin^2(\beta)+\sin^2(\gamma)$?

54. Find the angle between the line through $(1, 3, 2)$ and $(4, 1, 5)$ and the plane $x - y - 2z + 15 = 0$.

55. We showed that the distance from (p,q) to the line $Ax+By+C=0$ is $\frac{|Ap+Bq+C|}{\sqrt{A^2+B^2}}$. Show, using a similar argument, that the distance from (p,q,r) to the plane $Ax+By+Cx+D=0$ is $\frac{|Ap+Bq+Cr+D|}{\sqrt{A^2+B^2+C^2}}$.

56. How far apart are the planes $Ax+By+Cz+D=0$ and $Ax+By+Cz+E=0$? Explain.

57.

(a) Sketch a parabola and a line in the xy-plane that does not meet the parabola.

(b) Identify, graphically, the point on the parabola closest to the line.

(c) Find, analytically, the point on the parabola $y=x^2$ closest to the line $y=x-3$.

(d) The tangent to the parabola at the point found in (a) looks as if it might be parallel to the line in (a). Is it?

58. The planes $A_1x+B_1y+C_1z+D_1=0$ and $A_2x+B_2y+C_2z+D_2=0$ intersect in a line L. Find the direction cosines of a vector parallel to L.

59. How far apart are the lines given parametrically by $2\mathbf{i}+\mathbf{j}-3\mathbf{k}+t(3\mathbf{i}-5\mathbf{j}+2\mathbf{k})$ and $3\mathbf{i}+\mathbf{j}+5\mathbf{k}+s(2\mathbf{i}+6\mathbf{j}+7\mathbf{k})$? We use different letters, s and t, for the parameters because they are independent of each other.

60.

(a) Using properties of determinants, show that

$$\det\begin{pmatrix} x & y & 1 \\ a_1 & a_2 & 1 \\ b_1 & b_2 & 1 \end{pmatrix} = 0$$

is the equation of a line through (a_1,a_2) and (b_1,b_2).

(b) What determinant of order 4 would give an equation for the plane through three points?

61. A disk of radius a lies in the plane $x+3y+4z=5$. What is the area of its projection on $2x+y-z=6$?

62. Does the line through $(5, 7, 10)$ and $(3, 4, 5)$ meet the line through $(1, 4, 0)$ and $(3, 6, 4)$? If so, where? (Use parametric equations but give the parameters of the lines different names, such as t and s.)

63. Develop a formula for determining the distance from $P_1 = (x_1, y_1, z_1)$ to the line through $P_0 = (x_0, y_0, z_0)$ that is parallel to $\mathbf{A} = a_1\mathbf{i} + z_2\mathbf{j} + a_3\mathbf{k}$. The formula should be in terms of $\overline{P_0P_1}$ and $\mathbf{A}$.

64. How far is $(1, 2, -1)$ from the line through $(1, 3, 5)$ and $(2, 1, -3)$?

(a) Solve by calculus, minimizing a function.

(b) Solve by vectors.

65. How small can the largest of three direction angles ever be?

66. The plane π in Figure 14.4.10 is tilted at an angle θ to a horizontal plane. A convex region R in π has area A. Show that the area of its projection on the horizontal plane is $A\cos(\theta)$. Assume that the rays of light are perpendicular to the horizontal plane.

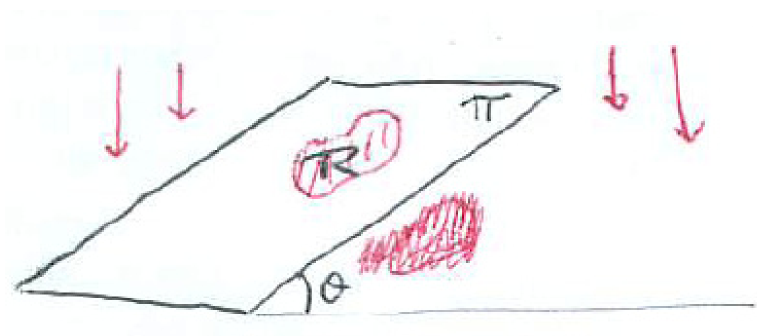


Figure 14.4.10

67. A square with side a lies in the plane $2x + 3y + 2z = 8$. What is the area of its projection

(a) on the xy-plane?

(b) on the yz-plane?

(c) on the xz-plane?

68. How would you decide whether the origin and $P = (x_0, y_0, z_0)$ are on the same side or opposite sides of $Ax + Bx + Cz + D = 0$?

69. How would you decide whether the points P and Q are on the same side or opposite sides of $Ax + By + Cz + D = 0$?

70. Devise a procedure for determining whether $P = (x, y)$ is inside the triangle whose three vertices are $P_1 = (x_1, y_1)$, $P_2 = (x_2, y_2)$, and $P_3 = (x_3, y_3)$.

71.

(a) Let L_1 be the line through P_1 and Q_1 and let L_2 be the line through P_2 and Q_2. Assume that L_1 and L_2 are **skew lines** (that is, not parallel and not intersecting). How would you find the point R_1 on L_1 and point R_2 on L_2 such that $\overline{R_1R_2}$ is perpendicular to both L_1 and L_2?

(b) Find R_1 and R_2 when $P_1 = (3, 2, 1)$, $Q_1 = (1, 1, 1)$, $P_2 = (0, 2, 0)$, $R_2 = (2, 1, -1)$.

72. (Contributed by Melvyn Kopald Stein.) An industrial hopper is shaped as shown in Figure 14.4.11. Its top and bottom are squares of different sizes. The angle between the plane ABD and the plane BDC is $70°$. The angle between the plane ABD and the plane ABC is $80°$. What is the angle between plane ABC and plane BCD? The angle is needed during the fabrication of the hopper, since the planes ABC and BCD are made from a single piece of sheet metal bent along the edge BC.

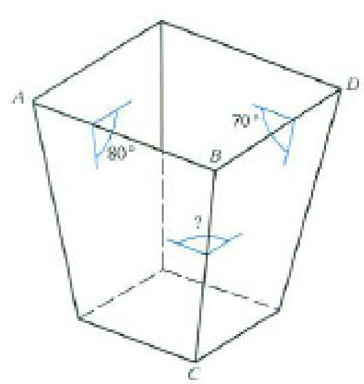


Figure 14.4.11

14.S Chapter Summary

The following tables and list summarize the chapter.
Assume $\mathbf{A} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$, $\mathbf{B} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$, and $\mathbf{C} = c_1\mathbf{i} + c_2\mathbf{j} + c_3\mathbf{k}$.

For plane vectors, disregard the third component.

Symbol	Name	Comment	Algebraic Formula
$\mathbf{A}$	Vector	has both direction and magnitude	$a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ or $\langle a_1, a_2, a_3\rangle$
$\lvert\mathbf{A}\rvert$	Length	also called magnitude or norm	$\sqrt{a_1^2 + a_2^2 + a_3^2}$
$-\mathbf{A}$	Negative, or opposite, of $\mathbf{A}$	points in opposite direction as $\mathbf{A}$	$-a_1\mathbf{i} - a_2\mathbf{j} - a_3\mathbf{k}$ or $\langle -a_1, -a_2, -a_3\rangle$
$\mathbf{A}+\mathbf{B}$	Sum of $\mathbf{A}$ and $\mathbf{B}$	place the tail of $\mathbf{B}$ at the head of $\mathbf{A}$	$(a_1+b_1)\mathbf{i} + (a_2+b_2)\mathbf{j} + (a_3+b_3)\mathbf{k}$
$\mathbf{A}-\mathbf{B}$	Difference of $\mathbf{A}$ and $\mathbf{B}$	add $-\mathbf{B}$ to $\mathbf{A}$	$(a_1-b_1)\mathbf{i} + (a_2-b_2)\mathbf{j} + (a_3-b_3)\mathbf{k}$
$c\mathbf{A}$	Scalar multiple of $\mathbf{A}$	same direction as $\mathbf{A}$, $\lvert c\rvert$ times as long as $\mathbf{A}$	$ca_1\mathbf{i} + ca_2\mathbf{j} + ca_3\mathbf{k}$
$\mathbf{A}\cdot\mathbf{B}$	Dot, or scalar, product	$\lvert\mathbf{A}\rvert\lvert\mathbf{B}\rvert\cos(\theta)$	$a_1b_1 + a_2b_2 + a_3b_3$
$\mathbf{A}\times\mathbf{B}$	Cross, or vector, product	magnitude: area of parallelogram spanned by $\mathbf{A}$ and $\mathbf{B}$, $\lvert\mathbf{A}\rvert\lvert\mathbf{B}\rvert\sin(\theta)$ direction: perpendicular to $\mathbf{A}$ and $\mathbf{B}$, direction by right-hand rule	$\det\begin{pmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\end{pmatrix}$
$\mathbf{proj_B}\,\mathbf{A}$	(Vector) Projection of $\mathbf{A}$ on $\mathbf{B}$	$\mathbf{A} - \mathbf{proj_B}\,\mathbf{A}$ is perpendicular to $\mathbf{B}$	$(\mathbf{A}\cdot\mathbf{u})\mathbf{u}$, where $\mathbf{u} = \mathbf{B}/\lvert\mathbf{B}\rvert$
$\mathrm{comp}_{\mathbf{B}}\,\mathbf{A}$	(Scalar) Component of $\mathbf{A}$ on $\mathbf{B}$	$\lvert\mathbf{proj_B}\,\mathbf{A}\rvert$; can be positive, negative, or zero	$\mathbf{A}\cdot\mathbf{u}$, where $\mathbf{u} = \mathbf{B}/\lvert\mathbf{B}\rvert$
$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})$	Scalar triple product	$\pm$ volume of parallelepiped spanned by $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$	$\det\left(\begin{vmatrix} a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{vmatrix}\right)$
$\mathbf{A}\times(\mathbf{B}\times\mathbf{C})$	Vector triple product	memory device: 1. first write ___$\mathbf{B}$ - ___$\mathbf{C}$. 2. then fill in the dot products.	$(\mathbf{A}\cdot\mathbf{C})\mathbf{B} - (\mathbf{A}\cdot\mathbf{B})\mathbf{C}$

Table 14.S.1 Basic combinations of vectors.

Dot Product	Cross Product
$\mathbf{A} \cdot \mathbf{B}$	$\mathbf{A} \times \mathbf{B}$
$\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A}$	$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A}$
$\lvert\mathbf{A} \cdot \mathbf{B}\rvert = \lvert\mathbf{A}\rvert\lvert\mathbf{B}\rvert\lvert\cos(\theta)\rvert$	$\lvert\mathbf{A} \times \mathbf{B}\rvert = \lvert\mathbf{A}\rvert\lvert\mathbf{B}\rvert\sin(\theta)$
$\mathbf{A} \cdot \mathbf{B} = 0$ is a test for perpendicular vectors	$\mathbf{A} \times \mathbf{B} = \mathbf{0}$ is a test for parallel vectors
formula in components involves $a_i b_i$ (same indices)	formula in components involves $a_i b_j$ (unequal indices)

Table 14.S.2 Comparison of the dot product and vector product.

Several applications were discussed. The following list gives the main ideas in most applications.

1. The angle θ between $\mathbf{A}$ and $\mathbf{B}$ satisfies $\cos(\theta) = \frac{\mathbf{A}\cdot\mathbf{B}}{|\mathbf{A}||\mathbf{B}|}$ $(0 \leq \theta \leq \pi)$.

2. Non-zero vectors $\mathbf{A}$ and $\mathbf{B}$ are perpendicular when $\mathbf{A} \cdot \mathbf{B} = 0$.

3. The plane through (x_0, y_0, z_0) perpendicular to $\mathbf{A}$ is $\mathbf{A} \cdot \langle x - x_0, y - y_0, z - z_0\rangle =$

4. The distance from the point (p, q) to the line $Ax + By + C = 0$ is $\dfrac{|Ap + Bq + C|}{\sqrt{A^2 + B^2}}$.

5. The distance from the plane with equation $Ax + By + Cz + D = 0$ to the point (p, q, r) is
$$\frac{|Ap + Bq + Cr + D|}{\sqrt{A^2 + B^2 + C^2}}.$$

6. The line through $P_0 = (x_0, y_0, z_0)$ parallel to $\mathbf{A} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ is given parametrically as $\overline{OP} = \overline{OP_0} + t\mathbf{A}$ or, component-wise
$$\begin{aligned} x &= x_0 + a_1 t \\ y &= y_0 + a_2 t \\ z &= z_0 + a_3 t \end{aligned}$$

7. The direction cosines of $\mathbf{A}$ are the numbers $\cos(\alpha)$, $\cos(\beta)$, and $\cos(\gamma)$ where α, β, and γ are the direction angles between $\mathbf{A}$ and $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$, respectively, and $\cos^2(\alpha) + \cos^2(\beta) + \cos^2(\gamma) = 1$.

EXERCISES for 14.S

1. Find a vector perpendicular to the plane determined by $(1, 2, 1)$, $(2, 1, -3)$, and $(0, 1, 5)$.

2. Find a vector perpendicular to the plane determined by $(1,3,-1)$, $(2,1,1)$, and $(1,3,4)$.

3. Find a vector perpendicular to the line through $(3,6,1)$ and $(2,7,2)$ and to the line through $(2,1,4)$ and $(1,-2,3)$.

4. Find a vector perpendicular to the line through $(1,2,1)$ and $(4,1,0)$ and also to the line through $(3,5,2)$ and $(2,6,-3)$.

5. How far apart are the lines whose vector equations are $2\mathbf{i}+4\mathbf{j}+\mathbf{k}+t(\mathbf{i}+\mathbf{j}+\mathbf{k})$ and $\mathbf{i}+3\mathbf{j}+2\mathbf{k}+s(2\mathbf{i}-\mathbf{j}-\mathbf{k})$?

6. Find the direction cosines of the vector **A** shown in Figure 14.S.1. (First draw a large diagram.)

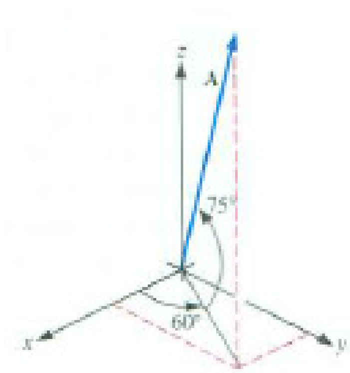
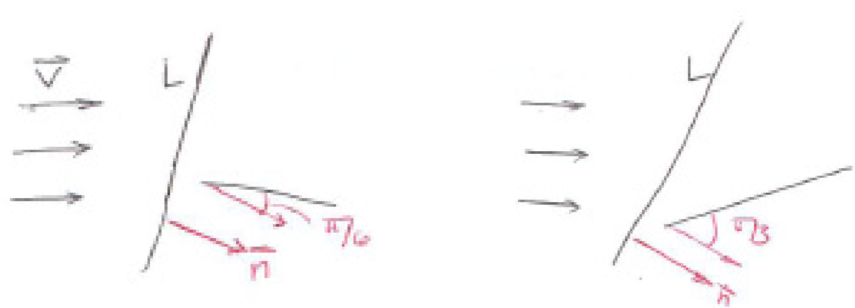


Figure 14.S.1

7. Why is the angle θ shown in Figure 14.S.2(a) and (b) the same as the angle between $\overline{v}$ and $\widehat{\mathbf{n}}$?

8. What is the ratio of the flows across the two sticks in Figure 14.S.2(a) and (b)?



(a) (b)

Figure 14.S.2

9. Find the point on the line through $(1, 2, 1)$ and $(2, -1, 3)$ that is closest to the line that goes through $(3, 0, 3)$ and is parallel to the vector $\mathbf{i} + 2\mathbf{j} + 5\mathbf{k}$.

In Exercises 10 and 11, find the distance from the point to the line.

10. The point $(0, 0)$ to $3x + 4y - 10 = 0$.

11. The point $(3/2, 2/3)$ to $2x - y + 5 = 0$.

In Exercises 12 and 13 find a normal and a unit normal to the given planes. (Recall that normal means perpendicular.)

12. $2x - 3y + 4z + 11 = 0$

13. $z = 2x - 3y + 4$

14. Show that the line through $(1, 1, 1)$ and $(2, 3, 4)$ is perpendicular to the plane $x + 2y + 3z + 4 = 0$.

15.

(a) Find the point on the curve $y = \sin(x)$, $0 \leq x \leq \pi$, that is nearest the line $y = x/2 + 2$.

(b) Check your answer by sketching the curve and the line.

(Use vectors.)

16.

(a) Find the point on the curve $y = \sin(x)$, $0 \leq x \leq \pi$, that is nearest the line $y = 2x + 4$.

(b) Check your answer by sketching the curve and the line.

(Use vectors.)

17.

(a) How would you find the angle between the planes $A_1x + B_1y + C_1z + D_1 = 0$ and $A_2x + B_2y + C_2z + D_2 = 0$?

(b) Find the angle between $x - y - z - 1 = 0$ and $x + y + z + 2 = 0$.

18. A line segment has projections of lengths a, b, and c on the coordinates axes. What, if anything, can be said about its length?

19. Suppose that the direction angles of a vector are equal. What can they be? Draw the cases.

20. What point on the line through $(1, 2, 5)$ and $(3, 1, 1)$ is closest to $(2, -1, 5)$?

21. Three points $P_1 = (x_1, y_1, z_1)$, $P_2 = (x_2, y_2, z_2)$, and $P_3 = (x_3, y_3, z_3)$ are the vertices of a triangle.

(a) What is the area of the triangle?

(b) What is the area of its projection of that triangle on the xy-plane?

22. How can you decide whether the line through P and Q is parallel to the plane $Ax + By + Cz + D = 0$?

23. Find where the line through $(1, 1)$ and $(2, 3)$ meets the line $x + 2y + 3 = 0$.

24.

(a) Give an example of a vector perpendicular to the plane $2x + 3y - z + 4 = 0$.

(b) Give an example of a vector parallel to it.

25. A plane contains the points P_0, P_1, and P_2, which do not lie on a line. Find a vector perpendicular to it.

26.

Sam: Just because $\mathbf{i} \times \mathbf{j}$ obeys the right-hand rule that doesn't mean $\mathbf{A} \times \mathbf{B}$ does in general. I'm not convinced.

Jane: Oh, but it does settle the general case.

Sam: How so?

Jane: Slowly move and alter $\mathbf{i}$ and $\mathbf{j}$ so they become $\mathbf{A}$ and $\mathbf{B}$, never letting either one become $\mathbf{0}$ or letting them be parallel. If it's the right hand rule at the start, it can't shift to the left hand rule.

Sam: Why not?

Jane: Think about it.

Explain what Jane is thinking.

27. Assume that the planes $A_1x+B_1y+C_1z+D_1 = 0$ and $A_2x+B_2y+C_2z+D_2 = 0$ meet in a line L.

(a) How would you find a vector parallel to L?

(b) How would you find a point on L?

(c) Find parametric equations for the line that is the intersection of the planes $2x - y + 3z + 4 = 0$ and $3x + 2y + 5z + 2 = 0$.

28.

(a) How far is the point P from the line through Q and R?

(b) How far is $(2, 1, 3)$ from the line through $(1, 5, 2)$ and $(2, 3, 4)$?

29.

(a) How would you decide whether the points $P_1 = (x_1, y_1, z_1)$, $P_2 = (x_2, y_2, z_2)$, $P_3 = (x_3, y_3, z_3)$ and $P_4 = (x_4, y_4, z_4)$ lie in a plane?

(b) Do the points $(1, 2, 3)$, $(4, 1, -5)$, $(2, 1, 6)$, and $(3, 5, 3)$ lie in a plane?

30. What is the angle between the line through $(1, 2, 1)$ and $(-1, 3, 0)$ and the plane $x + y - 2z = 0$?

31. Explain why the projection of a circle is an ellipse. (Set up coordinate systems in the plane of the circle and in the plane of its shadow, which we can take to be the xy-plane. Choose the axes for the coordinate systems to be as convenient as possible. Then express the equation of the shadow in terms of x and y by utilizing the equation of the circle.)

32. Figure 14.S.3 shows a tetrahedron $OABC$ with edges of lengths 4, 5, and 6.

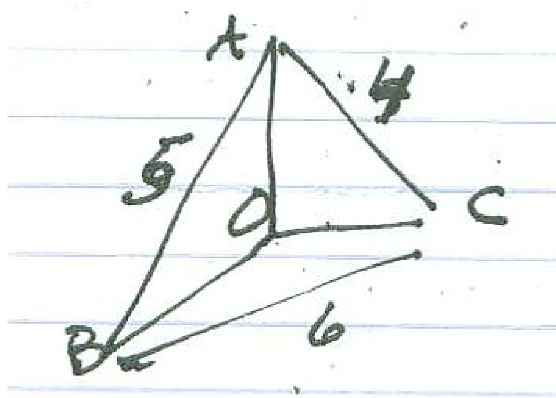


Figure 14.S.3

(a) Find the coordinates of A, B, and C.

(b) Find the volume of the tetrahedron.

(c) Find the area of triangle ABC.

(d) Find the distance from O to the plane in which triangle ABC lies.

(e) Find the cosine of angle ABC.

33. Let f be a differentiable function and L a line that does not meet the graph of f. Let P_0 be the point on the graph that is nearest the line.

(a) Using calculus, show that the tangent there is parallel to L.

(b) Why is the result to be expected?

34. Review the Folium of Descartes in Exercise 38 in Section 9.3. Show that its part in the fourth quadrant is asymptotic to $x + y + 1 = 0$.

Calculus is Everywhere # 18

Space Flight: The Gravitational Slingshot

In a slingshot or gravitational assist a spacecraft picks up speed as it passes near a planet. For instance, New Horizons, launched on January 19, 2006, enjoyed a gravitational assist as it passed by Jupiter, February 27, 2007 on its long journey to Pluto. Its speed increased from 47,000 to 50,000 miles per hour (mph). It will arrive near Pluto in 2015, instead of 2018.

Before we see how this works, look at a situation involving a truck on Earth that illustrates the idea. Later we will replace the truck with a planet's gravitational field.

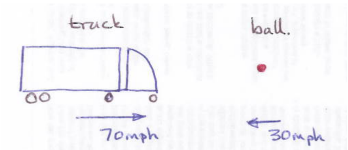


Figure C.18.1

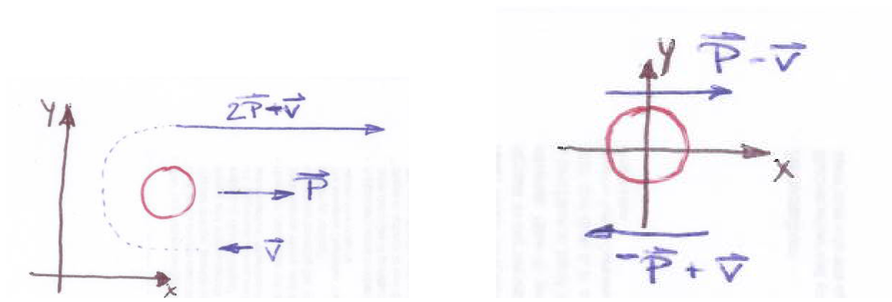
A playful lad throws a perfectly elastic tiny ball at 30 mph directly at a truck approaching him at 70 miles per hour, as shown in Figure C.18.1. The truck driver sees the ball coming toward her at $70 + 30 = 100$ mph. The ball hits the windshield and, because the ball is perfectly elastic, the driver sees it bounce off at 100 mph in the opposite direction. Because the truck is moving in the same direction as the ball, the ball is moving through the air at $100 + 70 = 170$ mph as it returns to the boy. The ball has gained 140 mph, twice the speed of the truck.

Instead of a truck, think of a planet whose velocity relative to the solar system is represented by the vector $\mathbf{P}$. A spacecraft, moving in the opposite direction with the velocity $\mathbf{v}$ relative to the solar system, comes close to the planet.

An observer on the planet sees the spacecraft approaching with velocity $-\mathbf{P}+\mathbf{v}$. The spacecraft swings around the planet as gravity controls its orbit and sends it off in the opposite direction. Whatever speed it gained as it arrived, it loses as it exits. Its velocity vector when it exits is $-(-\mathbf{P}+\mathbf{v}) = \mathbf{P} - \mathbf{v}$ as viewed by the observer on the planet. Since the planet is moving through the solar system with velocity vector $\mathbf{P}$, the spacecraft is now moving through the solar system with velocity $\mathbf{P}+(\mathbf{P}-\mathbf{v}) = 2\mathbf{P}-\mathbf{v}$. See Figure C.18.2.

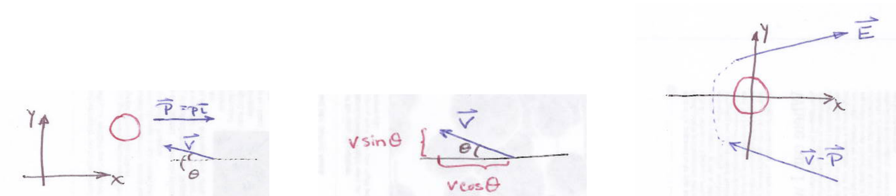
If $\mathbf{P} = 70\mathbf{i}$ and $\mathbf{v} = -30\mathbf{i}$, we have the vector $2(70\mathbf{i}) - (-30\mathbf{i}) = 170\mathbf{i}$, the case of the ball and truck.

The direction of the spacecraft as it arrives may not be exactly opposite the direction of the planet. To treat the general case, assume that $\mathbf{P} = p\mathbf{i}$, where p is positive and $\mathbf{v}$ makes an angle θ, $0 \le \theta \le \pi/2$, with $-\mathbf{i}$, as shown in Figure C.18.3(a). Let $v = |\mathbf{v}|$ be the speed of the spacecraft relative to the solar system. We will assume that the spacecraft's speed (relative to the planet) as it exits is the same as its speed relative to the planet on its arrival. Figure C.18.3(b) shows the arrival vector, $\mathbf{v} - \mathbf{P}$, and the exit vector, $\mathbf{E}$. The y-components of $\mathbf{E}$ and $\mathbf{v} - \mathbf{P}$ aree the same, but the x-component of $\mathbf{E}$ is the negative of the x-component of $\mathbf{v} - \mathbf{P}$.



(a) (b)

Figure C.18.2 (a) The velocity vector relative to the solar system. (b) The velocity vector relative to the planet.



(a) (b) (c)

Figure C.18.3

Figure C.18.3(c) shows the arrival vector relative to the solar system. So, $\mathbf{v} = -v\cos(\theta)\mathbf{i} + v\sin(\theta)\mathbf{j}$.

Relative to the planet we have

$$\begin{array}{lrcl} \text{Arrival Vector:} & \mathbf{v} - \mathbf{P} & = & -p\mathbf{i} + (-v\cos(\theta)\mathbf{i} + v\sin(\theta)\mathbf{j}) \\ \text{Exit Vector:} & \mathbf{E} & = & p\mathbf{i} + v\cos(\theta)\mathbf{i} + v\sin(\theta)\mathbf{j} \end{array}$$

The exit vector relative to the solar system is therefore

$$\mathbf{E} = (2p + v\cos(\theta))\mathbf{i} + v\sin(\theta)\mathbf{j}.$$

The magnitude of $\mathbf{E}$ is

$$\sqrt{(2p + v\cos(\theta))^2 + (v\sin(\theta))^2} = \sqrt{v^2 + 4pv\cos(\theta) + 4p^2}.$$

When $\theta = 0$, we have the truck and ball or the planet and spacecraft in Figure C.18.2. Then $\cos(\theta) = 1$ and $|\mathbf{E}| = \sqrt{v^2 + 4pv + 4p^2} = v + 2p$, in agreement with our observations.

The scientists controlling a slingshot carry out more extensive calculations, that take into consideration the masses of the spacecraft and the planet, and involve an integration while the spacecraft is near the planet. “Near” for the slingshot around Jupiter means 1.4 million miles. If the spacecraft gets too close, the atmosphere slows down or destroys the craft. The diameter of Jupiter is 86,000 miles.

The gravity assist was proposed by Michael Minovitch in 1963 when he was a graduate student at UCLA. Before then it was felt that to send a spacecraft to the outer solar system and beyond would require launch vehicles with nuclear reactors to achieve the necessary thrust.

Calculus is Everywhere # 19
How to Find Planets around Stars

Astronomers have discovered that stars other than the sun have planets circling them. How do they do this,since the planets are too small to be seen? They combine some vector calculus with observations of the star.

Suppose there is a star S and a planet P in orbit around it. To describe the situation, we are tempted to choose a coordinate system attached to the star. In that case the star would appear motionless, hence having no acceleration. However, the planet exerts a gravitational force F on the star and the equation force = mass $\times$ acceleration would be violated.

Let $\mathbf{X}$ be the position vector of the planet P and $\mathbf{Y}$ be the position vector of the star S, relative to our standard coordinates, called an **inertial system**. We will introduce a second inertial system later.

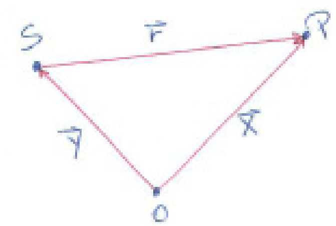


Figure C.19.1

Let M be the mass of the sun and m the mass of planet P. Let $\mathbf{r} = \mathbf{X} - \mathbf{Y}$ be the vector from the star to the planet, as shown in Figure C.19.1.

The gravitational pull of the star on the planet is proportional to the product of their masses and the reciprocal of the square of the distance between them:

$$\mathbf{F} = \frac{-GmM\widehat{\mathbf{r}}}{r^2} = \frac{-GmM\mathbf{r}}{r^3}.$$

where G is a constant. Equating the force with mass times acceleration, we have

$$M\mathbf{X}'' = \frac{-GmM\mathbf{r}}{r^3}.$$

Thus

$$\mathbf{X}'' = \frac{-Gm\mathbf{r}}{r^3}.$$

By calculating the force that the planet exerts on the star, we have

$$\mathbf{Y}'' = \frac{Gm\mathbf{r}}{r^3}.$$

The center of gravity of the system consisting of the planet and the star, which we will denote C (see Figure C.19.2), is given by

$$\mathbf{C} = \frac{M\mathbf{Y} + m\mathbf{X}}{M + m}.$$

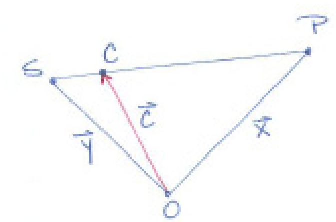


Figure C.19.2

The center of gravity is closer to the star than to the planet. For our sun and Earth, the center of gravity is 300 miles from the center of the sun.

The acceleration of the center of gravity is

$$\mathbf{C}'' = \frac{M\mathbf{Y}'' + m\mathbf{X}''}{M+m} = \frac{1}{M+m}\left(M\left(\frac{Gm\mathbf{r}}{r^3}\right) + m\left(\frac{-GM\mathbf{r}}{r^3}\right)\right) = \mathbf{0}.$$

Because the center of gravity has no acceleration, $va = \mathbf{0}$, it is moving at a constant velocity relative to the coordinate system. Therefore a coordinate system rigidly attached to the center of gravity may also serve as a system in which the laws of physics still hold.

We now describe the position of the star and planet in the new coordinate system. Star S has the vector $\mathbf{y}$ from $\mathbf{C}$ to it and planet P has the vector $\mathbf{x}$ from $\mathbf{C}$ to it, so $vr = \mathbf{x} + \mathbf{y}$, as shown in Figure C.19.3.

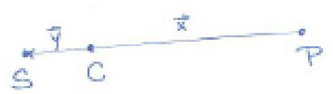


Figure C.19.3

To obtain a relation between $\mathbf{x}$ and $\mathbf{y}$, we express them in terms of $\mathbf{r}$. We have

$$\mathbf{y} = \mathbf{Y} - \overline{OC} = \mathbf{Y} - \frac{M\mathbf{Y} + m\mathbf{X}}{M+m} = \frac{m}{M+m}\mathbf{Y} - \frac{m}{M+m}\mathbf{X}.$$

Letting $k = m/M$, a small quantity, we have

$$\mathbf{y} = \frac{k}{1+k}(\mathbf{Y} - \mathbf{X}) = \frac{-k}{1+k}\mathbf{r}. \qquad \text{(C.19.1)}$$

Since $\mathbf{r} = \mathbf{x} - \mathbf{y}$, it follows that $\mathbf{x} = \mathbf{r} + \mathbf{y}$, hence

$$\mathbf{x} = \mathbf{r} + \left(\frac{-k}{1+k}\right)\mathbf{r} = \frac{1}{1+k}\mathbf{r}. \qquad \text{(C.19.2)}$$

SHERMAN: This box needs a title. Can you help?

Combining (C.19.1) and (C.19.2) shows that

$$\boxed{\mathbf{y} = -k\mathbf{x}. \qquad \text{(C.19.3)}}$$

This tells us

1. The star and planet remain on opposite sides of C on a straight line through C.
2. The star is always closer to C than the planet is.
3. The orbit of the star is similar in shape to the orbit of the planet, but smaller and reflected through C.
4. If the orbit of the star is periodic so is the orbit of the planet, with the same period.

Equation (C.19.3) is the key to the discovery of planets around stars. The astronomers look for a star that wobbles which is the sign that it is in orbit around the center of gravity. The time it takes for the planet to orbit the star is the time it takes for the star to oscillate back and forth once.

The reference cited below shows that the star and the planet sweep out elliptical orbits in the second coordinate system (the one relative to C).

Astronomers have found hundreds of stars with planets, some with several planets. A registry of these **exoplanets** is maintained at `http://exoplanets.org/`.

Reference: Robert Osserman, Kepler's Laws, Newton's Laws, and the search for new planets, A*American Mathematics Monthly* **108** (2001), pp. 813–820.

EXERCISES

1. The mass of the sun is about 330,000 times that of Earth. The closest Earth gets to the sun is about 91,341,000 miles, and the farthest is about 94,448,000 miles. What is the closest the center of the sun gets to the center of gravity of the sun-Earth system? What is the farthest it gets from it? (It always lies within the sun.)

2. The mass of Earth is $1/330,000$ the mass of the sun. What would it's mass have to be so that the center of gravity of the sun-Earth system would lie outside the sun? The diameter of the sun is about $87,000$ miles.

3. Find the condition that must be satisfied by a planet if the center of gravity of a sun-planet system lies outside the sun. (The diameter of the sun is about 870,000 miles.)

Chapter 15

Derivatives and Integrals of Vector Functions

In Section 9.3 we studied parametric curves in the plane. Using calculus we saw how to compute arc length, speed, and curvature. We defined curvature as the rate at which an angle changes as a function of arc length.

In this chapter we examine curves in the plane or in space. Of particular interest will be velocity and acceleration. For a particle moving along a straight line, say the x-axis, these are dx/dt and d^2x/dt^2. For a particle moving in space, velocity and acceleration involve both magnitude and direction. How should we calculate them?

How can we define curvature for a curve that does not lie in a plane? While arc length still makes sense, there is no angle to differentiate with respect to arc length.

While we could answer these questions using the component notation for a parameterization $\langle x(t), y(t)\rangle$ or $\langle x(t), y(t), z(t)\rangle$, we will use vector notation, where a curve is denoted by one letter. We will sometimes resort to the component notation to carry out computations or a proof.

15.1 The Derivative of a Vector Function: Velocity and Acceleration

For motion on a horizontal line the derivative of position with respect to time is sufficient to describe the motion of the particle. If it is positive, the particle is moving to the right. If it is negative, the particle is moving to the left. The speed is the absolute value of the derivative. For motion in the plane or in space we need the derivative of a vector function. This section introduces the calculus of a vector function and applies it to motion along a curve in a plane or in space.

Defining the Derivative

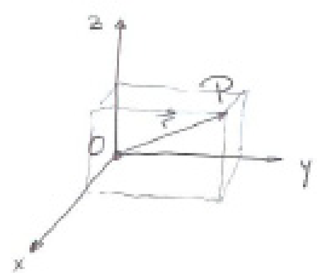


Figure 15.1.1

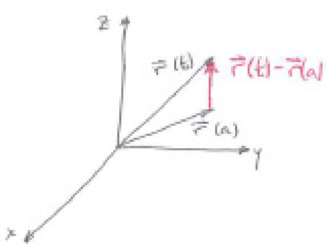


Figure 15.1.2

Assume that a curve in the plane is parameterized as $\langle x(t), y(t)\rangle$ or, in space, by $\langle x(t), y(t), z(t)\rangle$. Let $P = P(t)$ be the point corresponding to t, which we may think of as time, though it can be any parameter, such as arc length.

The **position vector**, $\mathbf{r} = \mathbf{r}(t)$, has its tail at the origin O and its tip at P. Then $\mathbf{r} = \overrightarrow{OP}$, as shown in Figure 15.1.1.

We will assume that $\mathbf{r}(t)$ is continuous, in that each of its components is continuous. The vector function $\mathbf{r}(t)$ is said to approach the vector $\mathbf{L}$ as t approaches a if $\lim_{t\to a} |\mathbf{r}(t) - \mathbf{L}| = 0$. We write $\lim_{t\to a} \mathbf{r}(t) = \mathbf{L}$. This is equivalent to the assertion that each of the scalar components of $\mathbf{r}(t)$ has a limit and $\mathbf{L} = \langle \lim_{t\to a} x(t), \lim_{t\to a} y(t), \lim_{t\to a} z(t)\rangle$.

Figure 15.1.2 shows this geometrically. As t approaches a, $\mathbf{r}(t) - \mathbf{r}(a)$ gets shorter as it approaches the zero vector $\mathbf{0}$.

We will say that $\mathbf{r}(t)$ is differentiable at $t = a$ if its components are differentiable at $t = a$. Then the derivative of $\mathbf{r}(t)$ is defined as the vector.

$$\langle x'(a), y'(a), z'(a)\rangle.$$

In vector notation,

$$\mathbf{r}'(a) = \lim_{t\to a} \frac{\mathbf{r}(t) - \mathbf{r}(a)}{t - a} \quad \text{or} \quad \mathbf{r}'(a) = \lim_{\Delta t\to 0} \frac{\mathbf{r}(a + \Delta t) - \mathbf{r}(a)}{\Delta t}.$$

and, if $\Delta\mathbf{r} = \mathbf{r}(r + \Delta t) - \mathbf{r}(r)$, $\mathbf{r}'(a) = \lim_{\Delta t\to 0} \frac{\Delta \mathbf{r}}{\Delta t}$. When t is near a (or Δt is near 0) the vector in the numerator will be short. It is divided by $t - a$ (or Δt), which is small, so the quotient could be a vector of any size.

Some Derivative Formulas

We state some useful identities:

If $\mathbf{r}$ and $\mathbf{s}$ are differentiable vector functions, and f is a differentiable scalar function, then

$$\begin{array}{rcll} (\mathbf{r}+\mathbf{s})' &=& \mathbf{r}'+\mathbf{s}' & \text{differentiate a sum} \\ (f\mathbf{r})' &=& f'\mathbf{r}+f\mathbf{r}' & \text{product rule (} f \text{ is a scalar function)} \\ (\mathbf{r}\times\mathbf{s})' &=& \mathbf{r}'\times\mathbf{s}+\mathbf{r}\times\mathbf{s}' & \text{differentiate a cross product} \\ (\mathbf{r}\cdot\mathbf{s})' &=& (\mathbf{r}'\cdot\mathbf{s})+(\mathbf{r}\cdot\mathbf{s}') & \text{differentiate a dot product} \\ (\mathbf{r}(f(t)))' &=& \mathbf{r}'(f(t))f'(t) & \text{Chain Rule.} \end{array}$$

The proofs are straightforward. We prove the formula for $(\mathbf{r}\cdot\mathbf{s})'$ in both component and vector notation. For convenience, assume $\mathbf{r}(t)$ and $\mathbf{s}(t)$ are vectors in the xy-plane so $\mathbf{r}(t)=\langle x(t),y(t)\rangle$ and $\mathbf{s}(t)=\langle u(t),v(t)\rangle$.

Proof

Using components:

$$\begin{array}{rcl} (\mathbf{r}\cdot\mathbf{s})' &=& (x(t)u(t)+y(t)v(t))' = x'u+xu'+y'v+yv' \\ &=& (x'u+y'v)+(xu'+yv') = \mathbf{r}'\cdot\mathbf{s}+\mathbf{r}\cdot\mathbf{s}'. \end{array}$$

Now, the same proof, but in vectors:

$$\begin{array}{rcl} (\mathbf{r}\cdot\mathbf{s})' &=& \lim\limits_{\Delta t\to 0}\dfrac{\mathbf{r}(a+\Delta t)\cdot\mathbf{s}(a+\Delta t)-\mathbf{r}(a)\cdot\mathbf{s}(a)}{\Delta t} \\ &=& \lim\limits_{\Delta t\to 0}\dfrac{(\mathbf{r}(a)+\Delta\mathbf{r})\cdot(\mathbf{s}(a)+\Delta\mathbf{s})-\mathbf{r}(a)\cdot\mathbf{s}(a)}{\Delta t} \\ &=& \lim\limits_{\Delta t\to 0}\dfrac{\mathbf{r}(a)\cdot\mathbf{s}(a)+\Delta\mathbf{r}\cdot\mathbf{s}(a)+\mathbf{r}(a)\cdot\Delta\mathbf{s}+\Delta\mathbf{r}\cdot\Delta\mathbf{s}-\mathbf{r}(a)\cdot\mathbf{s}(a)}{\Delta t} \\ &=& \lim\limits_{\Delta t\to 0}\dfrac{\Delta\mathbf{r}}{\Delta t}\cdot\mathbf{s}(a)+\mathbf{r}(a)\cdot\dfrac{\Delta\mathbf{s}}{\Delta t}+\Delta\mathbf{r}\cdot\dfrac{\Delta\mathbf{s}}{\Delta t} \\ &=& \mathbf{r}'(a)\cdot\mathbf{s}(a)+\mathbf{r}(a)\cdot\mathbf{s}'(a)+\mathbf{0}\cdot\mathbf{s}'(a) \\ &=& \mathbf{r}'(a)\cdot\mathbf{s}(a)+\mathbf{r}(a)\cdot\mathbf{s}'(a). \end{array}$$

This resembles the proof for the derivative of the product in Section 4.3. •

EXAMPLE 1 At time t, a particle has the position vector $\mathbf{r}(t)=3\cos(2\pi t)\mathbf{i}+3\sin(2\pi t)\mathbf{j}+5t\mathbf{k}$. Describe its path.

SOLUTION At time t the particle is at the point (x,y,z) with

$$x=3\cos(2\pi t),\quad y=3\sin(2\pi t),\quad \text{and}\quad z=5t.$$

Because $x^2+y^2=(3\cos(2\pi t))^2+(3\sin(2\pi t))^2=9$, the particle is always above or below the circle

$$x^2+y^2=9.$$

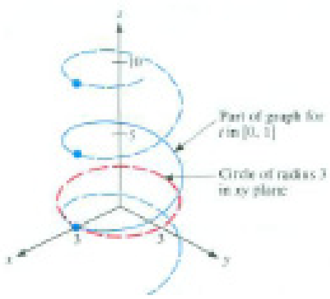


Figure 15.1.3

As t increases, $z = 5t$ increases.

The path is thus a spiral sketched in Figure 15.1.3. When t increases by 1, the angle $2\pi t$ increases by 2π, and the particle goes around the spiral once. The path is called a **helix**. $\diamond$

The Meaning of $\mathbf{r}'$ and $\mathbf{r}''$

The vector $\mathbf{r}'(a)$ is the limit of

$$\frac{\mathbf{r}(a+\Delta t) - \mathbf{r}(a)}{\Delta t}$$

as $\Delta t \to 0$. The numerator $\mathbf{r}(a+\Delta t) - \mathbf{r}(a) = \Delta\mathbf{r}$ is shown in Figure 15.1.4.

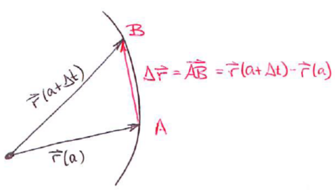


Figure 15.1.4

Since $\Delta\mathbf{r}$ coincides with a chord, it points almost along the tangent line at the head of $\mathbf{r}(a)$ when Δt is small. Dividing a vector by a scalar produces a parallel vector. The position vector is $\mathbf{r}(t)$, so

$$\frac{\mathbf{r}(a+\Delta t) - \mathbf{r}(a)}{\Delta t}$$

approximates a vector tangent to the curve at a. We conclude that

$$\mathbf{r}'(a) = \lim_{\Delta t \to 0} \frac{\mathbf{r}(t+\Delta t) - \mathbf{r}(a)}{\Delta t}$$

is a vector tangent to the curve at $\mathbf{r}(a)$. That is the geometric meaning of the derivative $\mathbf{r}'$.

$\mathbf{r}'$ is tangent to the curve.

To see what $\mathbf{r}'$ means when t is time, we compute its length.

Since $\mathbf{r}'(t) = \langle x'(t), y'(t), z'(t)\rangle$, its length is

$$\sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2}.$$

This is the natural extension to three dimensions of the speed on a planar curve in Section 9.4.

The length of $\mathbf{r}'(t)$, $\lvert\mathbf{r}'(t)\rvert$, is the speed.

We can also see that the magnitude of $\mathbf{r}'(t)$ is the speed by using vector language. For small Δt, the vector $\Delta \mathbf{r}$ lies on a short chord of the curve and its length is close to the length of arc swept out during that short interval of time. (See Figure 15.1.4.) Thus the magnitude of $\Delta \mathbf{r}/\Delta t$ approximates the speed.

Since $\mathbf{r}'(t)$ points in the direction of motion and its length is the speed, we call $\mathbf{r}'(t)$ the **velocity vector**. Note that velocity is a vector, while speed is a scalar. Velocity carries much more information than speed: it also tells the direction of the motion.

The velocity $\mathbf{r}'(t)$ is also denoted $\mathbf{v}$ or $\mathbf{v}(t)$.
The speed is $|\mathbf{v}|$, denoted v or $v(t)$.

The **acceleration vector**, $\mathbf{a}(t)$, is the derivative of the velocity vector.

The acceleration is $\mathbf{a}(t) = \mathbf{v}'(t) = \frac{d\mathbf{v}}{dt} = \mathbf{r}''(t) = \frac{d^2\mathbf{r}}{dt^2}$.

EXAMPLE 2 Let $\mathbf{r}(t) = \langle t, t^3 \rangle$.

(a) Draw and label $\mathbf{r}$, $\mathbf{v}$, and $\mathbf{a}$ at $t = 1$.

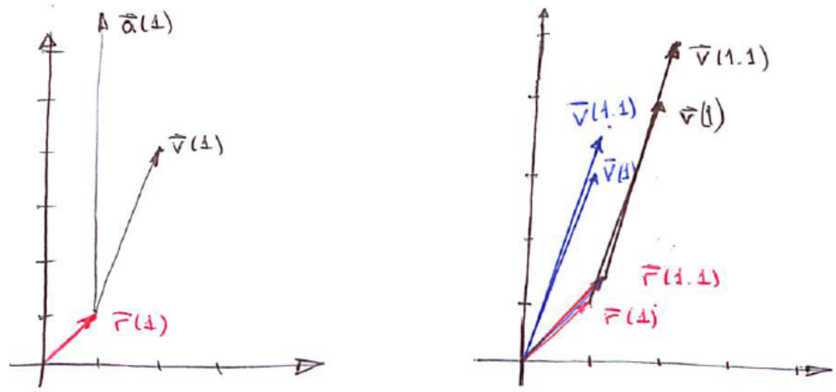
(b) Draw $\mathbf{v}(1.1)$.

SOLUTION

(a) $\mathbf{r}(t) = \langle t, t^3 \rangle$, $\mathbf{v}(t) = \langle 1, 3t^2 \rangle$, and $\mathbf{a} = \langle 0, 6t \rangle$. So $\mathbf{r}(1) = \langle 1, 1 \rangle$, $\mathbf{v}(1) = \langle 1, 3 \rangle$ and $a(1) = \langle 0, 6 \rangle$. We show these in Figure 15.1.5(a).

(b) Before we compute $\mathbf{v}(1.1)$, let us predict how it may change from $\mathbf{v}(1)$. The acceleration vector represents a force. Since $\mathbf{a}(1)$ is almost in the direction of $\mathbf{v}(1)$, the particle is speeding up. That is, $\mathbf{v}(1.1)$ should be longer than $\mathbf{v}(1)$.

Also, it would tend to rotate the velocity vector counterclockwise. So the direction of $\mathbf{v}(1.1)$ should be a bit counterclockwise from that of $\mathbf{v}(1)$. To check, we compute $\mathbf{v}(1.1) = \langle 1.1, 3(1.1)^2 \rangle = \langle 1.1, 3.63 \rangle$.



(a) (b)

Figure 15.1.5

It is longer than $\mathbf{v}(1) = \langle 1, 3\rangle$ since $\sqrt{(1.1)^2 + (3.63)^2}$ is larger than $\sqrt{1+3^2}$. Figure 15.1.5(b) shows that it is turned a bit counterclockwise, as expected. Its tail is placed at the head of

$$\mathbf{r}(1.1) \quad = \quad \langle 1.1, 1.331\rangle = 1.1\mathbf{i} + 1.331\mathbf{j}.$$

◇

EXAMPLE 3 Find the speed at time t of the particle described in Example 1.

SOLUTION From Example 1, the position is $\mathbf{r}(t) = \langle 3\cos(2\pi t), 3\sin(2\pi t), 5t\rangle$. So the velocity is $\mathbf{v}(t) = \mathbf{r}'(t) = \langle -6\pi\sin(2\pi t), 6\pi\cos(2\pi t), 5\rangle$ and

$$\begin{aligned} \text{speed} = |\mathbf{r}'(t)| &= \sqrt{(-6\pi\sin 2\pi t)^2 + (6\pi\cos 2\pi t)^2 + 5^2} \\ &= \sqrt{36\pi^2(\sin^2 2\pi t + \cos^2 2\pi t) + 25} = \sqrt{36\pi^2 + 25}. \end{aligned}$$

The particle travels at a constant speed along its helical path. In t units of time it travels the distance $\sqrt{36\pi^2 + 25}\; t$.

The velocity vector is not constant because its direction always changes. However, its length remains constant, and so the speed is constant. ◇

EXAMPLE 4 Sketch the path of a particle whose position vector at time $t \geq 0$ is $\mathbf{r}(t) = \cos(t^2)\mathbf{i} + \sin(t^2)\mathbf{j}$. Find its speed at time t.

SOLUTION Because $|\mathbf{r}(t)| = \sqrt{\cos^2(t^2) + \sin^2(t^2)} = 1$ the path of the particle is on the circle of radius 1 and center $(0, 0)$. Its speed is

$$\begin{aligned} |\mathbf{v}(t)| = |\mathbf{r}'(t)| &= |-2t\sin(t^2)\mathbf{i} + 2t\cos(t^2)\mathbf{j}| \\ &= \sqrt{(-2t\sin(t^2))^2 + (2t\cos(t^2))^2} \\ &= 2t\sqrt{\sin^2(t^2) + \cos^2(t^2)} = 2t. \end{aligned}$$

The particle travels faster and faster around the circle. ◇

EXAMPLE 5 If the acceleration vector is always perpendicular to the velocity vector, show that the speed is constant.

SOLUTION The speed is $|\mathbf{v}|$. We will show that the square of the speed, $|\mathbf{v}|^2$, is constant by showing that its derivative with respect to time is zero. Since $|\mathbf{v}|^2 = \mathbf{v} \cdot \mathbf{v}$, we have

$$\frac{d}{dt}\left(|\mathbf{v}|^2\right) = \frac{d}{dt}(\mathbf{v} \cdot \mathbf{v}) \quad = \quad \mathbf{v}' \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{v}' = 2\,\mathbf{v} \cdot \mathbf{v}' = 2\,\mathbf{v} \cdot \mathbf{a}.$$

Because $\mathbf{a}$ is perpendicular to $\mathbf{v}$ we know that $\mathbf{v} \cdot \mathbf{a} = 0$.

Thus $\mathbf{v} \cdot \mathbf{v}$ is constant, so the speed is constant. ◇

The calculation in Example 5 implies that if $\mathbf{r}(t)$ is always perpendicular to $\mathbf{r}'(t)$, then the length of $\mathbf{r}(t)$ is constant. The converse is also true:

> If the length of $\mathbf{r}(t)$ is constant,
> then $\mathbf{r}'(t)$ is perpendicular to $\mathbf{r}(t)$.

This is not surprising. If $\mathbf{r}(t)$ is constant, the $\mathbf{r}(t)$ lies on a sphere of radius $|\mathbf{r}(t)| = c$. A tangent to the curve at P is tangent to the sphere. The tangent to a sphere is perpendicular to its radius, as indicated in Figure 15.1.6(a), and the result follows.

EXAMPLE 6 Is the particle shown in Figure 15.1.6(b) speeding up or slowing down? Is its direction turning clockwise or counterclockwise?

SOLUTION Represent $\mathbf{a}$ as the sum of two vectors, one parallel to $\mathbf{v}$, and the other perpendicular to $\mathbf{v}$, as shown in Figure 15.1.6(c). Since $\mathbf{b}$ is in the same direction as $\mathbf{v}$, the particle is speeding up. The direction of $\mathbf{c}$ indicates that the direction of $\mathbf{v}$ is shifting counterclockwise. ◇

(a) (b) (c)

Figure 15.1.6

Summary

Instead of parameterizing a curve by displaying its components $\langle x(t), y(t)\rangle$ or $\langle x(t), y(t), z(t)\rangle$, we introduced the position vector $\overrightarrow{OP} = \mathbf{r}(t)$. If $\mathbf{r}(t)$ describes the position of a moving particle at time t, then $\mathbf{r}'(t)$ is the velocity of the particle and $|\mathbf{r}'(t)|$ is its speed. The acceleration $\mathbf{a}(t)$ is the second derivative of $\mathbf{r}(t)$: $\mathbf{a} = \mathbf{r}''$. It is proportional to the force operating on the particle.

We showed that if $\mathbf{r}(t)$ and $\mathbf{r}'(t)$ are perpendicular, then the length of $\mathbf{r}(t)$, $|\mathbf{r}(t)|$, is constant. The converse holds: If $\mathbf{r}(t)$ has constant length, then $\mathbf{r}'(t)$ is perpendicular to $\mathbf{r}(t)$, and $\mathbf{r}(t) \cdot \mathbf{r}'(t) = 0$.

EXERCISES for Section 15.1

1. At time t a particle has the position vector $\mathbf{r}(t) = t\mathbf{i} + r^2\mathbf{j}$.

(a) Compute and draw $\mathbf{r}(1)$, $\mathbf{r}(2)$, and $\mathbf{r}(3)$.

(b) Show that the path is a parabola.

2. At time t a particle has the position vector $\mathbf{r}(t) = (2t+1)\mathbf{i} + 4t\mathbf{j}$.

(a) Compute and draw $\mathbf{r}(0)$, $\mathbf{r}(1)$, and $\mathbf{r}(2)$.

(b) Show that the path is a straight line.

3. Let $\mathbf{r}(t) = 2t\mathbf{i} + t^2\mathbf{j}$.

(a) Compute and draw $\mathbf{r}(1.1)$, $\mathbf{r}(1)$, and their difference $\Delta\mathbf{r} = \mathbf{r}(1.1) - \mathbf{r}(1)$.

(b) Compute and draw $\Delta\mathbf{r}/0.1$.

(c) Compute and draw $\mathbf{r}'(1)$. Use one set of axes for all the graphs.

4. Let $\mathbf{r}(t) = 3t\mathbf{i} + t^2\mathbf{j}$.

(a) Compute and draw $\Delta\mathbf{r} = \mathbf{r}(2.01) - \mathbf{r}(2)$.

(b) Compute and draw $\Delta\mathbf{r}/0.01$.

(c) Compute and draw $\mathbf{r}'(2)$. Use one set of axes for all the graphs.

5. At time t the position vector of a ball is $\mathbf{r}(t) = 32t\mathbf{i} - 16t^2\mathbf{j}$.

(a) Draw $\mathbf{r}(1)$ and $\mathbf{r}(2)$.

(b) Sketch the path.

(c) Compute and draw $\mathbf{v}(0)$, $\mathbf{v}(1)$, and $\mathbf{v}(2)$. Place the tail of the vector at the head of the corresponding position vector.

6. At the time $t \geq 0$ a particle is at the point $x = 2t$, $y = 4t^2$.

(a) What is the position vector $\mathbf{r}(t)$ at time t?

(b) Sketch the path.

(c) How fast is the particle moving when $t = 1$?

(d) Draw $\mathbf{v}(1)$ with its tail at the head of $\mathbf{r}(1)$.

7. Let $\mathbf{r}(t)$ describe the path of a particle moving in the xy plane. If $\mathbf{r}(1) = 2.3\mathbf{i} + 4.1\mathbf{j}$ and $\mathbf{r}(1.2) = 2.31\mathbf{i} + 4.05\mathbf{j}$, estimate

(a) how much the position of the particle changes during the time interval $[1, 1.2]$.

(b) the slope of the tangent vector at $\mathbf{r}(1)$.

(c) the velocity vector $\mathbf{r}'(1)$.

(d) the speed of the particle at time $t = 1$.

8. Let $\mathbf{r}(t)$ describe the path of a particle moving in space. If $\mathbf{r}(2) = 1.7\mathbf{i} + 3.6\mathbf{j} + 8\mathbf{k}$ and $\mathbf{r}(2.01) = 1.73\mathbf{i} + 3.59\mathbf{j} + 8.02\mathbf{k}$, estimate

(a) how far the particle travels during the time interval $[2, 2.01]$.

(b) the velocity vector $\mathbf{r}'(2)$.

(c) the speed of the particle at time $t = 1$.

In Exercises 9 and 12 compute the velocity vector and speed.

9. $\mathbf{r}(t) = \cos 3t\mathbf{i} + \sin 3t\mathbf{j} + 6t\mathbf{k}$

10. $\mathbf{r}(t) = 3\cos 5t\mathbf{i} + 2\sin 5t\mathbf{j} + t^2\mathbf{k}$

11. $\mathbf{r}(t) = \ln(1 + t^2)\mathbf{i} + e^{3t}\mathbf{j} + \frac{\tan t}{1+2t}\mathbf{k}$

12. $\mathbf{r}(t) = \sec^2 3t\mathbf{i} + \sqrt{1 + t^2}\mathbf{j}$

13. At time t the position vector of a particle is

$$\mathbf{r}(t) = 2\cos(4\pi t)\mathbf{i} + 2\sin(4\pi t)\mathbf{j} + t\mathbf{k}.$$

(a) Sketch its path.

(b) Find its speed.

(c) Find a unit tangent vector to its path at time t.

In Exercises 14 to 21 the figure shows a velocity vector and an acceleration vector. Decide whether (a) the particle is speeding up, slowing down, or neither, (b) the velocity vector is turning clockwise, counter-clockwise, or neither.

14. Figure 15.1.7(a)

15. Figure 15.1.7(b)

16. Figure 15.1.7(c)

17. Figure 15.1.7(d)

18. Figure 15.1.8(a)

19. Figure 15.1.8(b)

20. Figure 15.1.8(c)

21. Figure 15.1.8(d)

(a) (b) (c) (d)

Figure 15.1.7

(a) (b) (c) (d)

Figure 15.1.8

22. At time t a particle is at $(4t, 16t^2)$.

(a) Show that the particle moves on the curve $y = x^2$.

(b) Draw $\mathbf{r}(t)$ and $\mathbf{v}(t)$ for $t = 0, 1/4, 1/2$.

(c) What happens to $|\mathbf{v}(t)|$ and the direction of $\mathbf{v}(t)$ for large t?

23. At time $t \geq 1$ a particle is at the point $(x, y) = (t, t^{-1})$.

(a) Draw the path of the particle.

(b) Draw $\mathbf{r}(1)$, $\mathbf{r}(2)$, and $\mathbf{r}(3)$.

(c) Draw $\mathbf{v}(1)$, $\mathbf{v}(2)$ and $\mathbf{v}(3)$.

(d) As t increases, what happens to dx/dt, dy/dt, $|\mathbf{v}|$, and $\mathbf{v}$?

24. At time t a particle is at $(2\cos(t^2), \sin(t^2))$.

(a) Show that it moves on an ellipse.

(b) Compute $\mathbf{v}(t)$.

(c) How does $|\mathbf{v}(t)|$ behave for large t? What does this say about the particle?

25. An electron travels at constant speed clockwise in a circle of radius 100 feet 200 times a second. At time $t = 0$ it is at $(100, 0)$.

(a) Compute $\mathbf{r}(t)$ and $\mathbf{v}(t)$.

(b) Draw $\mathbf{r}(0)$, $\mathbf{r}(1/800)$, $\mathbf{v}(0)$, $\mathbf{v}(1/800)$.

(c) How do $|\mathbf{r}(t)|$ and $|\mathbf{v}(t)|$ behave as t increases?

26. A ball is thrown up at an initial speed of 200 feet per second and at an angle of 60° from the horizontal. At time t it is at $(100t, 100\sqrt{3}t - 16t^2)$. Compute and draw $\mathbf{r}(t)$ and $\mathbf{v}(t)$ (a) when $t = 0$, (b) when the ball reaches its maximum height, and (c) when the ball strikes the ground.

27. A particle moves in a circular orbit of radius a. At time t its position vector is

$$\mathbf{r}(t) = a\cos(2\pi t)\mathbf{i} + a\sin(2\pi t)\mathbf{j}.$$

(a) Draw its position vector when $t = 0$ and when $t = \frac{1}{4}$.

(b) Draw its velocity when $t = 0$ and when $t = \frac{1}{4}$.

(c) Show that its velocity vector is always perpendicular to its position vector.

28. Use a computer or graphing calculator to graph $\mathbf{r} = \mathbf{r}(t) = (2\cos(t) + \cos(3t))\mathbf{i} + (3\sin(t) + \sin(3t))\mathbf{j}$, $0 \leq t \leq 2\pi$.

29. If $\mathbf{r}(t)$ is the position vector, $\mathbf{v}$ the velocity vector, and $\mathbf{a}$ the acceleration vector, show that $\frac{d}{dt}(\mathbf{r} \times \mathbf{v}) = \mathbf{r} \times \mathbf{a}$.

30. Let $\mathbf{r}(t) = t^2\mathbf{i} + t^3\mathbf{j}$.

(a) Sketch the vector $\Delta\mathbf{r} = \mathbf{r}(1.1) - \mathbf{r}(1)$.

(b) Sketch the vector $\Delta\mathbf{r}/\Delta t$ for $\Delta t = 0.1$.

(c) Sketch $\mathbf{r}'(1)$.

(d) Find $|\Delta\mathbf{r}/\Delta t - \mathbf{r}'(1)|$ for $\Delta t = 0.1$.

31. Instead of t, use the arc length s along the path as a parameter, so $\mathbf{r} = \mathbf{r}(s)$.

(a) Sketch $\Delta\mathbf{r}$ and the arc of length Δs. Why is it reasonable that $|\Delta\mathbf{r}/\Delta s|$ is near 1 when Δs is small?

(b) Show that $d\mathbf{r}/ds$ is a unit vector.

32. A particle at time $t = 0$ is at the point (x_0, y_0, z_0). It moves on the line through that point in the direction of the unit vector $\mathbf{u} = \cos(\alpha)\mathbf{i} + \cos(\beta)\mathbf{j} + \cos(\gamma)\mathbf{k}$. It travels at the constant speed of 3 feet per second.

(a) Give a formula for its position vector $\mathbf{r} = \mathbf{r}(t)$.

(b) Find its velocity vector $\mathbf{v} = \mathbf{r}'(t)$.

33. A rock is thrown up at an angle θ from the horizontal and at a speed v_0.

(a) Show that
$$\mathbf{r}(t) = (v_0\cos(\theta))t\mathbf{i} + ((v_0\sin(\theta))t - 16t^2)\mathbf{j}.$$
At time $t = 0$, the rock is at $(0, 0)$ and the x-axis is horizontal. Time is in seconds and distance is in feet.

(b) Show that the horizontal distance that the rock travels by the time it returns to its initial height is the same whether the angle is θ or $\pi/2 - \theta$.

(c) What value of θ maximizes the horizontal distance traveled?

(This is similar to Exercise 24 in Section 9.3, but uses vectors.)

34.

(a) Solve Example 5 by writing the speed as $\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}$ and differentiating.

(b) Which way do you prefer? The vector method in Example 5 or the component method in (a)?

35. The force of a magnetic field on a moving electron is always perpendicular to the velocity vector of the electron, $\mathbf{v}$. What does this imply about $\mathbf{v}$?

36. Show that if $\mathbf{r}(t) = \langle x(t), y(t), z(t)\rangle$ and $\lim_{t\to a} |\mathbf{r}(t) - \langle p, q, r\rangle| = 0$, then $\lim_{t\to a} x(t) = p$, $\lim_{t\to a} y(t) = q$, and $\lim_{t\to a} z(t) = r$.

37. Show the converse of the preceding exercise, namely that if the scalar components have limits, so does $\mathbf{r}(t)$.

38. At time t the position vector of a particle is

$$\mathbf{r}(t) = t\cos(2\pi t)\mathbf{i} + t\sin(2\pi t)\mathbf{j} + t\mathbf{k}.$$

Sketch its path.

39. A spaceship is on the path $\mathbf{r}(t) = t^2\mathbf{i} + 3t\mathbf{j} + 4t^3\mathbf{k}$. At time $t = 1$ it shuts off its rockets and coasts along the tangent line to the curve at that point.

(a) Where is it at time $t > 1$?

(b) Does it pass through the point $(9, 15, 50)$?

(c) If not, how close does it get to that point? At what time?

40. A particle traveling on the curve $\mathbf{r}(t) = \ln(t)\mathbf{i} + \cos(3t)\mathbf{j}$, $t \geq 1$, leaves the curve when $t = 2$ and travels on the xy-plane along the tangent to the curve at $\mathbf{r}(2)$. Where is it when $t = 3$?

41. Drawing a picture of $\mathbf{r}(t)$, $\mathbf{r}(t+\Delta t)$, and $\Delta\mathbf{r}$, explain why $\left|\frac{\Delta\mathbf{r}}{\Delta t}\right|$ is an estimate of the speed of a particle moving on the curve $\mathbf{r}(t)$.

42. The moment a ball is dropped straight down from a tall tree, you shoot an arrow directly at it. Assume that there is no air resistance. Show that the arrow will hit the ball, assuming that the ball does not hit the ground first.

(a) Solve using the formulas in Exercise 33.

(b) Solve with a maximum of intuition and a minimum of computation.

43.

(a) At time t a particle has the position vector $\mathbf{r}(t)$. Show that for small Δt the area swept out by the position vector is approximately $\frac{1}{2}|\mathbf{r}(t)\times\mathbf{v}(t)|\Delta t$. (See Figure 15.1.9.) ($\mathbf{v}(t)$ is approximated by $\Delta\mathbf{r}/\Delta t$.)

(b) Assume that the curve in (a) is parameterized over the time interval $[a, b]$. Show that the area swept out is $\frac{1}{2}\int_a^b |\mathbf{r}\times\mathbf{v}|\ dt$.

(c) Must the curve lie in a plane for the formula in (b) to hold?

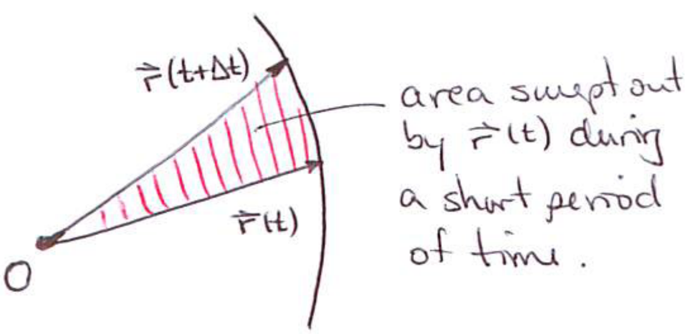


Figure 15.1.9

In Exercises 44 to 50 $\mathbf{v}(t)$ is the velocity vector at time t for a moving particle and $\mathbf{r}(0)$ is the particle's position at time $t = 0$. Find $\mathbf{r}(t)$, the position vector of the particle at time t.

44. $\mathbf{v}(t) = \sin^2(3t)\mathbf{i} + \dfrac{t}{3t^2+1}\mathbf{j}$, $\mathbf{r}(0) = \mathbf{j}$

45. $\mathbf{v}(t) = \dfrac{t}{t^2+t+1}\mathbf{i} + \tan^{-1}(3t)\mathbf{j}$, $\mathbf{r}(0 = \mathbf{i}+\mathbf{j}$

46. $\mathbf{v}(t) = \dfrac{t^3}{t^4+1}\mathbf{i} + \ln(t+1)\mathbf{j}$, $\mathbf{r}(0) = \mathbf{0}$

47. $\mathbf{v}(t) = e^{2t}\sin(3t)\mathbf{i} + \dfrac{t^3}{3t+2}\mathbf{j}$, $\mathbf{r}(0) = \mathbf{i} + 3\mathbf{j}$

48. $\mathbf{v}(t) = \dfrac{t}{(t+1)(t+2)(t+3)}\mathbf{i} + \dfrac{t^2}{(t+2)^3}\mathbf{j}$, $\mathbf{r}(0) = \mathbf{i} - \mathbf{j}$

49. $\mathbf{v}(t) = \dfrac{(\ln(t+1))^3}{t+1}\mathbf{i} + \dfrac{1}{\sqrt{1-4t^2}}\mathbf{j} + \sec^2(3t)\mathbf{k}$, $\mathbf{r}(0) = \mathbf{i} + \mathbf{j} + \mathbf{k}$

50. $\mathbf{v}(t) = t^3e^{-t}\mathbf{i} + (1+t)(2+t)\mathbf{j}$, $\mathbf{r}(0) = 2\mathbf{i} - \mathbf{j}$

15.2 Curvature and Components of Acceleration

In Section 9.6 we defined the curvature of a plane curve as the absolute value of the derivative $d\phi/ds$, where ϕ is the angle the tangent makes with the x-axis and s is the arc length. This definition does not work for a curve that does not lie in a plane. (Why not?) In this section we use vectors to define curvature for curves in space and then use curvature to analyze the acceleration vector.

Definition of Curvature

A particle whose position vector at the time t is $\mathbf{r}(t)$ has velocity $\mathbf{v}(t)$. When $\mathbf{v}(t)$ is not the zero vector, the unit vector in the direction of $\mathbf{v}(t)$ is

$$\mathbf{T}(t) = \frac{\mathbf{v}(t)}{|\mathbf{v}|} = \frac{\mathbf{r}'(t)}{|\mathbf{r}'(t)|}, \text{ or}$$

$$\boxed{\mathbf{T} = \frac{\mathbf{v}}{|\mathbf{v}|} = \frac{\mathbf{r}'}{|\mathbf{r}'|} \qquad \text{(assuming } \mathbf{v} \neq \mathbf{0}\text{)}}$$

The **unit tangent vector**, $\mathbf{T}$, records the direction of motion.

As the particle moves along the curve the direction of $\mathbf{T}$ changes most rapidly where the curve is curviest. So we define curvature of a curve in the plane or in space to be the length of the rate of change with respect to arclength of the unit tangent vector

Curvature of a Curve

$$\kappa = \left|\frac{d\mathbf{T}}{ds}\right|$$

where s denotes the length of arc of a curve, measured from a fixed starting point.

We check in Example 1 that the definition of curvature agrees with the definition for curvature for plane curves in Section 9.6.

EXAMPLE 1 Show that the definition of curvature as $|d\mathbf{T}/ds|$ agrees with the definition $|d\phi/ds|$ given earlier for plane curves.

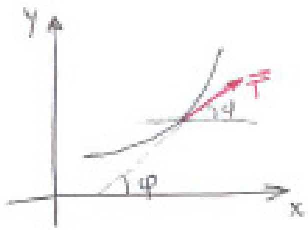


Figure 15.2.1

SOLUTION As Figure 15.2.1 shows, ϕ is the angle that $\mathbf{T}$ makes with the x-axis. Since $\mathbf{T}$ is a unit vector, $\mathbf{T} = \cos(\phi)\mathbf{i} + \sin(\phi)\mathbf{j}$. Thus

$$\begin{aligned}\kappa = \left|\frac{d\mathbf{T}}{ds}\right| &= \left|\frac{d(\cos\phi\mathbf{i} + \sin\phi\mathbf{j}}{ds}\right| = \left|\frac{d(\cos\phi\mathbf{i} + \sin\phi\mathbf{j})}{d\phi}\frac{d\phi}{ds}\right| \\ &= \left|(-\sin\phi\mathbf{i} + \cos\phi\mathbf{j})\frac{d\phi}{ds}\right| = |(-\sin\phi\mathbf{i} + \cos\phi\mathbf{j})|\left|\frac{d\phi}{ds}\right| = \left|\frac{d\phi}{ds}\right|\end{aligned}$$

so that

$$\left|\frac{d\mathbf{T}}{ds}\right| = \left|\frac{d\phi}{ds}\right|.$$

◇

Define the **radius of curvature** as the reciprocal of κ.

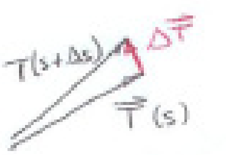
EXAMPLE 2 Compute the curvature of the helix

$$\mathbf{r}(t) = \cos(t)\mathbf{i} + \sin(t)\mathbf{j} + 3t\mathbf{k}.$$

SOLUTION To find $\mathbf{T}$ we compute $\mathbf{v} = -\sin(t)\mathbf{i} + \cos(t)\mathbf{j} + 3\mathbf{k}$ and $|\mathbf{v}| = \sqrt{(-\sin(t))^2 + (\cos(t))^2 + 3^2} = \sqrt{10}$. Thus

$$\mathbf{T} = \frac{1}{\sqrt{10}}(-\sin(t)\mathbf{i} + \cos(t)\mathbf{j} + 3\mathbf{k}).$$

Because speed is both $v = |\mathbf{v}|$ and the rate of change of arc length, the curvature equals



$$\left|\frac{d\mathbf{T}}{ds}\right| = \frac{\left|\frac{d\mathbf{T}}{dt}\right|}{\left|\frac{ds}{dt}\right|} = \frac{\left|\frac{d\mathbf{T}}{dt}\right|}{v} = \frac{\left|\frac{1}{\sqrt{10}}(-\cos(t)\mathbf{i} - \sin(t)\mathbf{j})\right|}{\sqrt{10}}$$

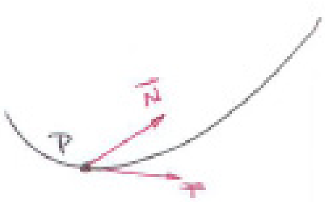
The curvature is $1/10$ and the radius of curvature is 10. For any helix the curvature and radius of curvature are both constant. ◇

Figure 15.2.2

The Unit Normal N

Since $\mathbf{T}(t)$ has constant length, $d\mathbf{T}/ds$ is perpendicular to $\mathbf{T}$. By considering small Δs and $\Delta\mathbf{T}$, as in Figure 15.2.2, we see that $d\mathbf{T}/ds$ points in the direction in which $\mathbf{T}$ is turning. Since the length of $d\mathbf{T}/ds$ is the curvature κ, we may write

$$\frac{d\mathbf{T}}{ds} = \kappa\mathbf{N}$$



where $\mathbf{N}$ is a unit normal called the **principal normal**. Because κ is positive, $d\mathbf{T}/ds$ and $\mathbf{N}$ point in the same direction. The vectors $\mathbf{T}$ and $\mathbf{N}$, if placed with their tails at a point P on the curve, determine a plane. The part of the curve near P stays close to the plane. (See Figure 15.2.3.)

Figure 15.2.3

The Acceleration Vector and T and N

The acceleration vector, $\mathbf{a}$, is defined as the second derivative of the position vector, $\mathbf{r}$. We will show that $\mathbf{a}$ is parallel to the plane determined by $\mathbf{T}$ and $\mathbf{N}$, so $\mathbf{a}$ can be written in the form $c_1\mathbf{T} + c_2\mathbf{N}$, where c_1 and c_2 are scalars.

Since $\mathbf{a} = \frac{d\mathbf{v}}{dt}$, we express $\mathbf{v}$ in terms of $\mathbf{T}$ and $\mathbf{N}$. By the definition of $\mathbf{T}$, $\mathbf{v} = v\mathbf{T}$, where $v = |\mathbf{v}|$, is the speed. $\mathbf{N}$ is not needed to express the velocity vector $\mathbf{v}$.

Thus

$$\begin{aligned}
\mathbf{a} = \tfrac{d\mathbf{v}}{dt} &= \tfrac{d(v\mathbf{T})}{dt} \\
&= \tfrac{dv}{dt}\mathbf{T} + v\tfrac{d\mathbf{T}}{dt} && \text{(product rule)} \\
&= \tfrac{d^2s}{dt^2}\mathbf{T} + v\tfrac{d\mathbf{T}}{ds}\tfrac{ds}{dt} && (v = \tfrac{ds}{dt} \text{ and chain rule}) \\
&= \tfrac{d^2s}{dt^2}\mathbf{T} + v^2\tfrac{d\mathbf{T}}{ds}.
\end{aligned}$$

Replacing $\frac{d\mathbf{T}}{ds}$ with $\kappa\mathbf{N}$, we find

Acceleration in terms of T, N, and κ

$$\mathbf{a} = \frac{d^2s}{dt^2}\mathbf{T} + v^2\kappa\mathbf{N}$$

If κ is not 0, then we have

$r = 1/\kappa$ ($\kappa \neq 0$)

Acceleration in terms of T, N, and radius of curvature

$$\mathbf{a} = \frac{d^2s}{dt^2}\mathbf{T} + \frac{v^2}{r}\mathbf{N}.$$

The tangential component of acceleration, $\frac{d^2s}{dt^2}$, is positive if the particle is speeding up and is negative if it is slowing down. The normal component of acceleration, v^2/r, is always positive.

Tangential component of acceleration: $\mathbf{a}\cdot\mathbf{T} = \frac{d^2s}{dt^2}$

Figure 15.2.4 shows how $\mathbf{a}$ may look relative to $\mathbf{T}$ and $\mathbf{N}$. In both cases $\mathbf{T}$ turns in the direction of $\mathbf{N}$. In Figure 15.2.4 that means that $\mathbf{T}$ is turning counterclockwise.

Normal component of acceleration: $\mathbf{a}\cdot\mathbf{N} = \frac{v^2}{r}$

Computing Curvature

We can compute the curvature directly from its definition. There is also a shorter formula for κ. To develop it we compute

$$\mathbf{T}\times\mathbf{a} = \mathbf{T}\times\left(\frac{d^2s}{dt^2}\mathbf{T} + v^2\kappa\mathbf{N}\right). \qquad (15.2.1)$$

(a) (b)

Figure 15.2.4 The tangential and normal components of acceleration: (a) $d^2s/dt^2 > 0$ and (b) $d^2s/dt^2 < 0$.

We do this for two reasons. First, $\mathbf{T} \times \mathbf{T} = 0$. Second $|\mathbf{T} \times \mathbf{N}| = 1$, since $\mathbf{T}$ and $\mathbf{N}$ span a unit square. By (15.2.1), we then have

$$\mathbf{T} \times \mathbf{a} = \kappa v^2 (\mathbf{T} \times \mathbf{N}).$$

Thus

$$|\mathbf{T} \times \mathbf{a}| = \kappa v^2.$$

Because $\mathbf{T} = \mathbf{v}/v$, we have

$$\frac{|\mathbf{v} \times \mathbf{a}|}{v} = \kappa v^2$$

and thus

Curvature in terms of speed, velocity and acceleration

$$\kappa = \frac{|\mathbf{v} \times \mathbf{a}|}{v^3}. \tag{15.2.2}$$

We illustrate (15.2.2) by applying it to the helix of Example 2.

EXAMPLE 3 Use (15.2.2) to compute the curvature of the helix $\mathbf{r}(t) = \cos(t)\mathbf{i} + \sin(t)\mathbf{j} + 3t\mathbf{k}$.

SOLUTION We compute $\mathbf{v}$, v and $\mathbf{a}$. Because $\mathbf{v} = -\sin t\mathbf{i} + \cos t\mathbf{j} + 3\mathbf{k}$, $v = \sqrt{(-\sin(t))^2 + (\cos(t))^2 + 3^2} = \sqrt{10}$. Then

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = -\cos(t)\mathbf{i} - \sin(t)\mathbf{j}.$$

Next we compute $\mathbf{v} \times \mathbf{a}$:

$$\begin{aligned}
\det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -\sin(t) & \cos(t) & 3 \\ -\cos(t) & -\sin(t) & 0 \end{pmatrix} &= 3\sin(t)\mathbf{i} - 3\cos(t)\mathbf{j} + (\sin^2(t) + \cos^2(t))\mathbf{k} \\
&= 3\sin(t)\mathbf{i} - 3\cos(t)\mathbf{j} + \mathbf{k}.
\end{aligned}$$

Thus

$$\begin{aligned}
k = \frac{|\mathbf{v} \times \mathbf{a}|}{v^3} &= \frac{|3\sin(t)\mathbf{i} - 3\cos(t)\mathbf{j} + \mathbf{k}|}{(\sqrt{10})^3} \\
&= \frac{\sqrt{(3\sin(t))^2 + (-3\cos(t))^2 + 1^2}}{\sqrt{10}^3} \\
&= \frac{\sqrt{10}}{(\sqrt{10})^3} = \frac{1}{10}.
\end{aligned}$$

◇

Though curvature is defined as a derivative with respect to arc length s, there are two reasons s is rarely used in computations. First, we seldom can obtain a formula for the arc length. Second, if the curve is described in terms of a parameter t, such as time or angle, then we can use the chain rule to express $d\mathbf{T}/ds$ as the directly calculatable

$$\frac{\frac{d\mathbf{T}}{dt}}{\frac{ds}{dt}}.$$

The Meaning of the Components of a

If no force acts on a moving particle it would move in a line at a constant speed. But if there is a force $\mathbf{F}$, then, according to Newton's Laws, it is related to the acceleration vector $\mathbf{a}$ by $\mathbf{F} = m\mathbf{a}$.

If $\mathbf{F}$ is parallel to $\mathbf{T}$, the particle of mass m moves in a line with an acceleration $dv/dt = d^2s/dt^2$. So we would expect $\mathbf{a}$ to equal $d^2s/dt^2\mathbf{T}$.

If $\mathbf{F}$ is perpendicular to $\mathbf{T}$, it would not change the particle's speed, but it would push it away from a straight path, as shown in Figure 15.2.5

Figure 15.2.5

If you spin a pail of water at the end of a rope you can feel the force. It is proportional to the square of the speed and inversely proportional to the length of the rope. Driving a car around a sharp curve too fast can cause it to skid because the friction of the tires against the road cannot supply the necessary force, whose magnitude is the speed squared divided by the radius of the turn, to prevent skidding.

The Third Unit Vector, B

The vector $\mathbf{T}\times\mathbf{N}$ has length 1 and is perpendicular to both $\mathbf{T}$ and $\mathbf{N}$. We may think of it as a normal to the plane parallel to $\mathbf{T}$ and $\mathbf{N}$ through a point P on the curve. The unit vector $\mathbf{T}\times\mathbf{N}$ is denoted $\mathbf{B}$ and is called the **binormal**. It is shown in Figure 15.2.6

Figure 15.2.6

The three vectors, $\mathbf{T}$, $\mathbf{N}$, and $\mathbf{B}$, may change direction as P moves along the curve. However, they remain a rigid frame, where $\mathbf{T}$ indicates the direction of motion, $\mathbf{N}$ the direction of turning, and $\mathbf{B}$ the tilt of the **osculating plane**, the plane that contains P and the vectors $\mathbf{T}$ and $\mathbf{N}$.

More formulas for κ are found in Exercises 21, 22, and 23. When given explicit formulas for a curve, it can be easiest to use a computer algebra system such as Mathematica or Maple to find the curvature, $\mathbf{T}$, $\mathbf{N}$, and $\mathbf{B}$.

Summary

We defined the curvature of a curve in space (or in the xy-plane) using vectors. The definition agrees with the definition of curvature for curves in the xy-plane given in Section 9.6. The curvature, or its reciprocal, the radius of the curvature, appears in the normal component of the acceleration vector.

The section concluded with the definition of the binormal, $\mathbf{B}=\mathbf{T}\times\mathbf{N}$, which indicates the tilt of the plane determined by $\mathbf{T}$ and $\mathbf{N}$.

EXERCISES for Section 15.2

In Exercises 1 to 4, $\mathbf{v}$ denotes the velocity and $\mathbf{a}$ denotes the acceleration. Evaluate the dot product.

1. $\mathbf{v} \cdot \mathbf{T}$

2. $\mathbf{a} \cdot \mathbf{T}$

3. $\mathbf{v} \cdot \mathbf{N}$

4. $\mathbf{a} \cdot \mathbf{N}$

5.

(a) Why is $\mathbf{T} \times \mathbf{N}$ a unit vector?

(b) Why is $\mathbf{N}$ perpendicular to $\mathbf{T}$?

In Exercises 6 and 7, $\mathbf{v}$ and $\mathbf{a}$ are given at an instant. Find the (i) curvature, (ii) radius of curvature, and (iii) d^2s/dt^2.

6. $\mathbf{v} = 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$, $\mathbf{a} = \mathbf{i} - \mathbf{j} + \mathbf{k}$

7. $\mathbf{v} = \mathbf{i} + \mathbf{j} + \mathbf{k}$, $\mathbf{a} = -\mathbf{i} + \mathbf{j} + \mathbf{k}$

In Exercises 8 and 9 compute the curvature using $\kappa = |d\mathbf{T}/dt|/v$.

8. $\mathbf{r}(t) = t\mathbf{i} + t^2\mathbf{j} + t^3\mathbf{k}$

9. $\mathbf{r}(t) = 3\cos(2t)\mathbf{i} + 3\sin(2t)\mathbf{j} + 4t\mathbf{k}$

In Exercises 10 and 11, compute the curvature using the speed, velocity, and acceleration, that is, using $\kappa = |\mathbf{v} \times \mathbf{a}|/v^3$.

10. $\mathbf{r}(t) = t\mathbf{i} + t^2\mathbf{j} + t^3\mathbf{k}$

11. $\mathbf{r}(t) = 3\cos(2t)\mathbf{i} + 3\sin(2t)\mathbf{j} + 4t\mathbf{k}$

12. To emphasize the value of the vector approach, compute $d|\mathbf{v}|/dt$ in two ways.

(a) Differentiate $|\mathbf{v}|^2 = \mathbf{v} \cdot \mathbf{v}$ to conclude that $d|\mathbf{v}|/dt = \mathbf{v} \cdot \mathbf{a}/|\mathbf{v}|$.

(b) Derive the result starting with

$$|\mathbf{v}| = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}.$$

13. Let a and b be constants. A particle moves on the helix path described by

$$\mathbf{r}(t) = 3\cos(at)\mathbf{i} + 3\sin(at)\mathbf{j} + bt\mathbf{k}$$

(a) Compute its curvature.

(b) As $b \to \infty$ what happens to the curvature?

(c) Why is the answer to (b) reasonable?

(d) As $a \to \infty$, what happens to the curvature?

(e) Why is the answer to (d) reasonable?

14. Show that for $\mathbf{r}(x) = x\mathbf{i} + f(x)\mathbf{j} + 0\mathbf{k}$ the formula $\frac{|(\mathbf{v}\times\mathbf{a})|}{v^3}$ gives the formula in Section 9.6 for curvature of $y = f(x)$.

15. Show that $\frac{d\mathbf{r}}{ds}$ is a unit vector,

(a) by drawing $\mathbf{r}(s + \Delta s)$ and $\mathbf{r}(s)$ and considering $\frac{\mathbf{r}(s+\Delta s)-\mathbf{r}(s)}{\Delta s}$.

(b) by writing it as $(d\mathbf{r}/dt)/(ds/dt)$.

16. Express the area of the parallelogram spanned by $\mathbf{v}$ and $\mathbf{a}$ in terms of the curvature and speed.

17. If a particle reaches a maximum speed at time t_0, must d^2s/dt^2 be 0 at t_0? Must $d^2\mathbf{r}/dt^2$ be $\mathbf{0}$ at t_0? Assume the time interval is $(-\infty, \infty)$.

18. If $\mathbf{r}(t)$ is the position vector and s denotes arc length, is $d^2\mathbf{r}/dt^2$ parallel to $d^2\mathbf{r}/ds^2$?

In Exercises 19 and 20 the Figure 15.2.7 shows the velocity and acceleration vectors at a point P on a curve. Find (i) v, (ii) d^2s/dt^2, and (iii) κv^2. Then (iv) find r, the radius of the curvature, (v) draw the osculating circle, and (vi) using the osculating circle, draw an approximation of a short piece of the path near P.

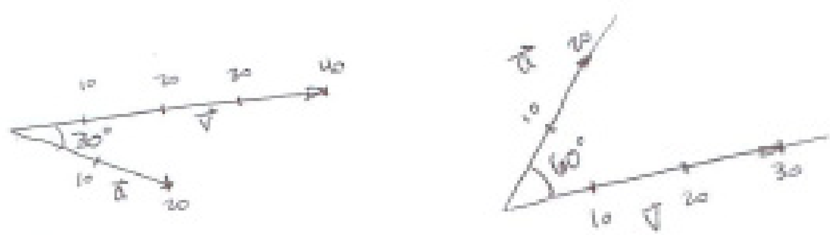
19. Figure 15.2.7(a)

20. Figure 15.2.7(b)

21.

Jane: After doing Exercises 19 and 20, I have a simpler way to get a formula for curvature. Just look at the right triangle whose hypotenuse has length $|\mathbf{a}|$ and one leg is the component of $\mathbf{a}$ along $\mathbf{v}$. By trigonometry,

$$\kappa v^2 = |\mathbf{a}||\sin(\mathbf{a}, \mathbf{v})|. \tag{15.2.3}$$



(a) (b)

Figure 15.2.7

All that's left is getting $\sin(\mathbf{a}, \mathbf{v})$ out and $\cos(\mathbf{a}, \mathbf{v})$ in because we know how to express $\cos(\mathbf{a}, \mathbf{v})$ in terms of a dot product. Squaring (15.2.3) gives

$$\kappa^2 v^4 = |\mathbf{a}|^2 \left(1 - \cos^2(\mathbf{a}, \mathbf{v})\right).$$

If you use the fact that

$$\cos(\mathbf{a}, \mathbf{v}) = \frac{\mathbf{a} \cdot \mathbf{v}}{|\mathbf{a}| v}.$$

and a little algebra, you get

$$\kappa^2 = \frac{(\mathbf{v} \cdot \mathbf{v})(\mathbf{a} \cdot \mathbf{a}) - (\mathbf{a} \cdot \mathbf{v})^2}{v^6} \qquad (15.2.4)$$

My way is simpler than using the cross product. I guess the authors don't understand trigonometry.

(a) Fill in the missing steps.

(b) Check that Jane's formula agrees with (15.2.2).

22.

Sam: You used trigonometry. I can do it with just the Pythagorean Theorem. Look at the triangle with hypotenuse $|\mathbf{a}|$. Its legs have lengths $|d^2s/dt^2|$ and κv^2. So

$$|\mathbf{a}|^2 = \left(\frac{d^2 s}{dt^2}\right)^2 + (\kappa v^2)^2.$$

Solve this for κ.

Jane: But you have to express everything in vectors. We're in the chapter on vectors.

Sam: O.K. First $|\mathbf{a}|^2 = \mathbf{a} \cdot \mathbf{a}$ and $v^2 = |\mathbf{v}|^2 = \mathbf{v} \cdot \mathbf{v}$.

Jane: But d^2s/dt^2?

Sam: That's dv/dt. So I differentiate both sides of $v^2 = \mathbf{v} \cdot \mathbf{v}$, getting $2v\frac{dv}{dt} = 2\mathbf{v} \cdot \mathbf{a}$. So $dv/dt = (\mathbf{v} \cdot \mathbf{a})/v$. So $(dv/dt)^2 = (\mathbf{v} \cdot \mathbf{a})^2/v^2$. So

$$\mathbf{a} \cdot \mathbf{a} = \frac{(\mathbf{v} \cdot \mathbf{a})^2}{v^2} + \kappa^2 (v^2)^2.$$

Then solve for κ^2.

I get the same result that you got in Exercise 21. It seems quite straightforward. The authors should have used my formula.

Jane: There were so many "so's" that I got lost.

Show that the formula for curvature that Sam obtained agrees with Jane's formula in Exercise 21.

23. Here is another way to find a formula for curvature. For the right triangle whose hypotenuse is $|\mathbf{a}|$ and whose legs are parallel to $\mathbf{T}$ and $\mathbf{N}$, show that

$$\kappa^2 v^4 = (\ddot{x})^2 + (\ddot{y})^2 + (\ddot{z})^2 - (\ddot{s})^2.$$

Two dots over a variable denotes its second derivative with respect to t.

24. Assume that you are prone to car sickness on curvy roads. Which matters more to you, $|d\mathbf{T}/ds|$ where s is arc length or $|d\mathbf{T}/dt|$ where t is time? Explain the difference in the two quantities.

25. Let $\mathbf{r} = \mathbf{r}(s)$, where s is arc length. Show that the curvature is $\kappa = \left|\frac{d^2\mathbf{r}}{ds^2}\right|$.

26. Consider curves situated on the surface of a sphere $\mathcal{S}$ of radius a. (A sphere is the surface of a ball.)

(a) Show that there are curves on $\mathcal{S}$ that have large curvature.

(b) Exhibit a curve whose curvature is as small as $1/a$.

(c) Show that there are no curves with curvature smaller than $1/a$. (See Exercise 25 and start with $\mathbf{r} \cdot \mathbf{r} = a^2$.)

The Frenet formulas are:

$$\frac{d\mathbf{T}}{ds} = \kappa\mathbf{N}, \quad \frac{d\mathbf{B}}{ds} = \tau\mathbf{N}, \quad \frac{d\mathbf{N}}{ds} = -\kappa\mathbf{T} + \tau\mathbf{B}.$$

Here κ is curvature and τ is torsion, the measure of the tendency of the plane through $\mathbf{T}$ and $\mathbf{N}$ to turn. We already have the first one, while Exercises 27 and 28 develop the formulas for $d\mathbf{B}/ds$ and $d\mathbf{N}/ds$.

27.

(a) Why is $d\mathbf{B}/ds$ perpendicular to $\mathbf{B}$?

(b) Why are there scalars p and q such that $\frac{d\mathbf{B}}{ds} = p\mathbf{T} + q\mathbf{N}$?

(c) Using the fact that $\mathbf{B}$ and $\mathbf{T}$ are always perpendicular show that

$$(p\mathbf{T} + q\mathbf{N}) \cdot \mathbf{T} = \mathbf{0}.$$

(d) From (c) show that $p = 0$. Thus $d\mathbf{B}/ds = q\mathbf{N}$. The scalar function q is usually denoted τ (tau). Thus $d\mathbf{B}/ds = \tau\mathbf{N}$.

28.

(a) Why are there scalars c and d such that $\frac{d\mathbf{N}}{ds} = c\mathbf{T} + d\mathbf{B}$?

(b) Using the fact that $\mathbf{B}$ and $\mathbf{N}$ are always perpendicular, show that $\tau\mathbf{N} \cdot \mathbf{N} + \mathbf{B} \cdot (c\mathbf{T} + d\mathbf{B}) = 0$.

(c) From (b) show that $d = \tau$.

(d) Similarly, starting with $\mathbf{T} \cdot \mathbf{N} = 0$, show that $c = -\kappa$. Thus $\frac{d\mathbf{N}}{ds} = -\kappa\mathbf{T} + \tau\mathbf{N}$.

29. A pail of water is being swung at the end of a rope. The amount of rope is slowly increased until the radius of the circle the pail sweeps out doubles. Does the force of your pull remain the same? Increase? Decrease? Explain.

30. In Example 1 we used calculus to show that for a plane curve $|d\mathbf{T}/ds| = |d\phi/ds|$, when ϕ is the angle that $\mathbf{T}$ makes with the x-axis. This suggests that for small values of Δs, $|\Delta\phi| = |\phi(s + \Delta s) - \phi(s)|$ is a good approximate of $|\Delta\mathbf{T}| = |\mathbf{T}(s + \Delta s) - \mathbf{T}(s)|$.

(a) Draw $\mathbf{T}(s + \Delta s)$ and $\mathbf{T}(s)$ with their tails at the origin.

(b) Using the diagram, show why you would expect $|\Delta\mathbf{T}|$ and $|\Delta\phi|$ to be close to each other in the sense that $|\Delta\mathbf{T}/\Delta\phi|$ would be near 1.

31. Show that a curve that has a constant curvature $\kappa = 0$ is a line. Don't say, "Oh, it's a curve with infinite radius of curvature, so it must be a line". (Start with the definition, $\kappa = |d\mathbf{T}/ds|$.)

32. Express $d\mathbf{T}/ds$ in terms of the curvature and $\mathbf{N}$.

33.

Jane: I don't like the way the authors got the formula for curvature. I'm sure they didn't need to drag in the components of the acceleration vector. It's not elegant.

Sam: They're trying to save space. Calculus books are too long.

Jane: My way is neat and short: just calculate

$$\left|\frac{d\mathbf{T}}{ds}\right| = \frac{\left|\frac{d\mathbf{T}}{dt}\right|}{|\mathbf{v}|}.$$

To begin I write $\mathbf{T}$ as $\mathbf{v}/|\mathbf{v}|$. Then I differentiate the quotient $\mathbf{v}/|\mathbf{v}|$. Along the way I'll need $d|\mathbf{v}|/dt$, but I get that by differentiating $|\mathbf{v}|^2 = \mathbf{v} \cdot \mathbf{v}$. That will give me

$$\frac{d\mathbf{T}}{ds} = \frac{v^2\mathbf{a} - (\mathbf{a} \cdot \mathbf{v})\mathbf{v}}{v^4} \tag{15.2.5}$$

Sam: That's a nice formula but there's no cross product.

Jane: If you like cross products, then use (15.2.5) to find $(d\mathbf{T})/ds \cdot d\mathbf{T}/ds$ and call on that identity that appeared when getting the length of the cross product $|\mathbf{A} \times \mathbf{B}|$ (see (14.3) in Section 14.3). I'll let you fill in the steps. I don't want to deprive you of a little fun.

Fill in the missing steps.

34. Using (15.2.5), obtain the formula in Exercise 15.2.4 for κ^2.

15.3 Line Integrals and Conservative Vector Fields

In Section 6.2, we defined the integral of $f(x)$ over an interval $[a, b]$ as the limit of sums of the form $\sum_{i=1}^{n} f(c_i)\Delta x_i$. Now we use similar definitions for integrals over curves. In the next section we apply them to work, fluid flow, and geometry.

The Integral with Respect to Arc Length s

Let $\mathbf{r}(t)$ be the position vector corresponding to a parameter value t in $[a, b]$. Assume that $\mathbf{r}(t)$ sweeps out a curve C with a continuous unit tangent vector $\mathbf{T}(t)$. Let f be a scalar-valued function defined on C. We will define the integral of f over C with respect to arc length.

Partition $[a, b]$ by $t_0 = a$, t_1, …, $t_n = b$ and let $\mathbf{r}(t_0) = \overrightarrow{OP_0}$, $\mathbf{r}(t_1) = \overrightarrow{OP_1}$, …, $\mathbf{r}(t_n) = \overrightarrow{OP_n}$ be the position vectors as shown in Figure 15.3.1. The points P_0, P_1, …, P_n break the curve into n shorter curves of lengths Δs_1, Δs_2, …, Δs_n. Form the Riemann sum

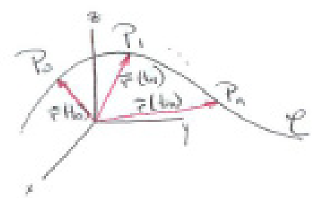


Figure 15.3.1

$$\sum_{i=1}^{n} f(P_i)\Delta s_i.$$

The limit of sums of this form as all the lengths Δs_i are chosen smaller and smaller is denoted $\int_C f(P)ds$. That is,

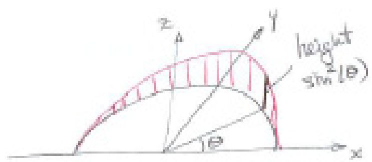
$$\int_C f(P)\ ds = \lim_{\Delta s \to 0} \sum_{i=1}^{n} f(P_i)\Delta s_i.$$

The limit does not depend on the parameterization and so it does not depend on the direction in which the curve is swept out. To compute the integral when the curve is parameterized by t we can use

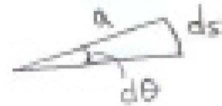
$$\int_C f(P)\ ds = \int_a^b f(P)\left|\frac{ds}{dt}\right| dt.$$

EXAMPLE 1 A fence is built as a semicircle of radius a with center at the origin. The height of the fence is $\sin^2(\theta)$ where θ is the angle made with the positive x-axis, as in Figure 15.3.2(b)). What is the area of one side of the fence?

SOLUTION Let $f(P)$ be the height of the fence at $P = (r, \theta)$ in polar coordinates. Then $f(r, \theta) = \sin^2(\theta)$. The parameter θ ranges from 0 to π. Let



(a)



(b)

Figure 15.3.2

$s = a\theta$ be the arc length subtended by the angle θ, as in Figure 15.3.2(b). Then $ds = ad\theta$ and we have

$$\text{area} = \int\limits_C \sin^2(\theta)\ ds \ \ = \ \ \int\limits_0^\pi \sin^2(\theta)a\ d\theta = 2\int\limits_0^{\pi/2} \sin^2(\theta)a\ d\theta = 2a\frac{\pi}{4} = \frac{\pi a}{2}.$$

$\diamond$

The Integrals with Respect to x, y, or z

The integral with respect to arc length is similar to an integral over an interval. The integrals with respect to x, y, or z are different.

Let C be a parameterized curve and let f be a scalar function defined on C. Divide the interval $[a, b]$ into n sections by $t_0 = a, t_1, \cdots, t_n = b$. For convenience, take the sections to be of equal lengths.

Let $\mathbf{r}(t_i) = \langle x(t_i), y(t_i), z(t_i)\rangle$. Instead of considering the arc length Δs_i of each short interval we consider instead the change in the x-coordinate, $x_{i+1} - x_i = \Delta x_i$. It can be positive or negative. We then make the definition.

The integral of f over the curve C with respect to x is the limit of sums

$$\sum_{i=1}^{n} f(x(t_i), y(t_i), z(t_i))\Delta x_i$$

as n approaches infinity. It is denoted

$$\int_C f\ dx, \qquad \int_C f(x,y,z)\ dx, \quad \text{or} \quad \int_C f(P)\ dx.$$

The integrals $\int_C f(P)\ dy$ and $\int_C f(P)\ dz$ are defined similarly.

Each of $\int_C f(P)\ dx$, $\int_C f(P)\ dy$, and $\int_C f(P)\ dz$ is called a **line integral** of f over the curve C. Another line integral is the line integral with respect to arc length: $\int_C f(P)\ ds$. While it would be more natural to call them curve integrals, tradition dictates that they be known as line integrals.

To compute a line integral such as $\int_C f(P)\ dx$ over a parameterized curve C, express it as $\int_a^b f(x(t), y(t), z(t))\frac{dx}{dt}\ dt$. In the same way

$$\int_C f(P)\ ds = \int_a^b f(x(t), y(t), z(t))\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\ dt.$$

A more general line integral is a sum of three types,

$$\int_C \left(P(x,y,z)\ dx + Q(x,y,z)\ dy + R(x,y,z)\ dz\right). \tag{15.3.1}$$

The integrand for this line integral, $Pdx + Qdy + Rdz$, is sometimes referred to as a **differential form**. This language will be encountered again in Chapter 18. The differential form can also be written as a dot product of the vector field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ and $d\mathbf{r} = dx\mathbf{i} + dy\mathbf{j} + dz\mathbf{k}$, $\mathbf{F} \cdot d\mathbf{r}$.

In contrast to an integral with respect to arc length, the value of $\int_C f(P)\ dx$ depends on the orientation in which the curve is swept out. If we reverse the orientation, the expression $x_{i+1} - x_i$ changes sign. For instance, if x is an increasing function of the parameter in one parameterization, then $\Delta x_i = x_{i+1} - x_i$ is positive; but in the reverse orientation x is a decreasing function of the parameter, so $\Delta x_i = x_{i+1} - x_i$ is negative.

If $-C$ denotes the curve C swept out in the opposite orientation, then

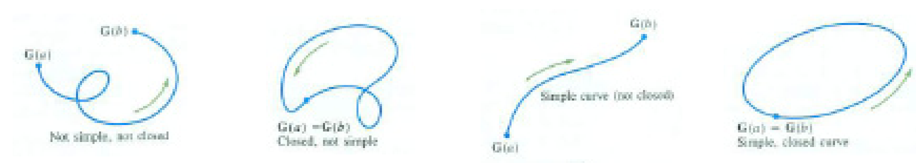
$$\int_{-C} f(P)\ dx = -\int_C f(P)\ dx.$$

When evaluating line integrals $\int_C f(P)\ dx$, $\int_C f(P)\ dy$, $\int_C f(P)\ dz$, hence $\int_C \mathbf{F}(P) \cdot d\mathbf{r}$, it is necessary to pay attention to the orientation of C.

closed curve

simple curve

A **closed curve** is a curve that starts and ends at the same point. If the curve does not intersect itself except perhaps at its endpoints, we call the curve **simple**. These are independent: a curve can be neither closed nor simple, closed but not simple, simple but not closed, or both simple and closed. (See Figure 15.3.3.)



(a) (b) (c) (d)

Figure 15.3.3

When C is a closed curve we will usually use the notation $\oint_C \mathbf{F} \cdot\ d\mathbf{r}$ for a line integral over C.

EXAMPLE 2 Let C be a smooth closed convex curve without line segments that is situated in the first quadrant, as shown in Figure 15.3.4. Find $\oint_C y\ dx$ if C is oriented counterclockwise.

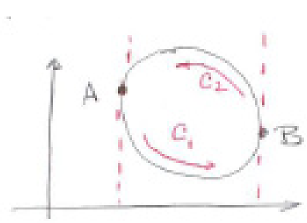


Figure 15.3.4

SOLUTION Let A and B be the contact points of the vertical tangents to C. Break C into a lower curve C_1 from A to B and an upper curve C_2 from B to A, both swept out counterclockwise.

On C_1, $\Delta x_i = x_{i+1} - x_i$ is positive. Let y_i be the y-coordinate of a point on C_1 above $[x_i, x_{i+1}]$. Then $(x_{i+1} - x_i)y_i$ approximates the area under C_1 and above $[x_i, x_{i+1}]$, as shown in Figure 15.3.5(a)

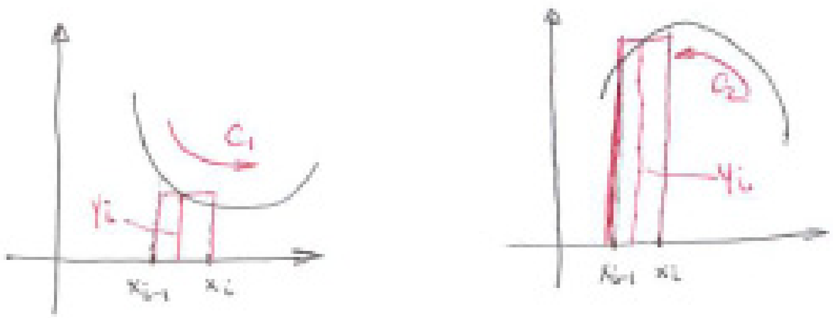
We may think of $y\ dx$ as the local approximation to the area under C_1. Thus

$$\int\limits_{C_1} y\ dx = \text{area below } C_1 \text{ and above the } x\text{-axis.}$$

On C_2, x is a decreasing function of the parameter and $\Delta x_i = x_{i+1} - x_i$ is negative, as Figure 15.3.5(b) shows.

Let y_i be the y-coordinate of a point on C_2 above the interval whose ends are x_i and x_{i+1}. Because $(x_{i+1} - x_i)y_i$ is the negative of an approximation of the area below C_2 and above the x-axis, we conclude that

$$\int\limits_{C_2} y\ dx = \text{negative of the area below } C_2 \text{ and above the } x\text{-axis.}$$



(a) (b)

Figure 15.3.5

Since $\int_C y\ dx = \int_{C_1} y\ dx + \int_{C_2} y\ dx$, it follows that when C is oriented counterclockwise

$$\int\limits_C y\ dx = \text{negative of the area inside } C.$$

◇

For curves in the xy-plane, then the most general line integral would be

$$\int\limits_C \left(P(x,y,z)\ dx + Q(x,y,z)\ dy\right). \tag{15.3.2}$$

Both (15.3.1) and (15.3.2) can be expressed in the compact language of vectors, as we will now show.

Vector Fields

A **vector field** assigns a vector to each point in some region of space (or the plane). The function that assigns to each point the vector that describes the direction and speed of the wind is an example of a vector field. The use of field instead of function is in deference to physicists and engineers, who speak of magnetic field and electric field, which are also examples of vector fields.

A function that assigns a scalar (real number) to each point in a region in space (or the plane) is called a **scalar field**. The function that assigns the temperature at a point in space is a scalar function as is the function that gives the density of an object at a point.

A typical vector field in space is $\mathbf{F}(x,y,z) = P(x,y,z)\mathbf{i} + Q(x,y,z)\mathbf{j} + R(x,y,z)\mathbf{k}$ where $P(x,y,z)$, $Q(x,y,z)$, and $R(x,y,z)$ are scalar fields. A vector field $\mathbf{F}$ in the plane can be described by two scalar fields with $\mathbf{F}(x, y) = P(x, y)\mathbf{i} + Q(x, y)\mathbf{j}$.

To take advantage of vector notation, we write $d\mathbf{r} = dx\mathbf{i} + dy\mathbf{j} + dz\mathbf{k}$. Then (15.3.1) can be written in vector notation as

$$\int_C \mathbf{F}(x, y, z) \cdot d\mathbf{r}, \qquad \int_C \mathbf{F}(\mathbf{r}) \cdot d\mathbf{r}, \qquad \text{or} \qquad \int_C \mathbf{F} \cdot d\mathbf{r}.$$

For computation we may write it as

$$\int_a^b \mathbf{F}(x(t), y(t), z(t)) \cdot \frac{d\mathbf{r}}{dt}\, dt \qquad \text{or} \qquad \int_a^b \mathbf{F} \cdot \frac{d\mathbf{r}}{dt} dt.$$

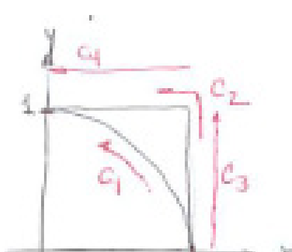
Line integrals in the plane are expressed in the same way.

Another standard notation uses the unit vector $\mathbf{T} = \frac{d\mathbf{r}}{ds}$. Writing $d\mathbf{r}$ as $\mathbf{T}\, ds$ we can rewrite $\int_C \mathbf{F} \cdot d\mathbf{r}$ as $\int_C \mathbf{F} \cdot \mathbf{T}\, ds$.

The integrand depends on the orientation of the curve because switching it changes $\mathbf{T}$ to $-\mathbf{T}$.

Conservative Vector Fields

The next example shows that different paths with the same initial point and terminal point may yield different integrals.



We used $u = \sin(\theta)$ to find an antiderivative of $\sin^2(\theta)\cos(\theta)$.
Figure 15.3.6

EXAMPLE 3 Let C_1 be the path from $(1, 0)$ to $(0, 1)$ along the unit circle with center at the origin. Let C_2 be the path that starts at $(1, 0)$, goes to $(1, 1)$ on the line $x = 1$, and then to $(0, 1)$ on the line $y = 1$. Compute $\int_{C_1} xy\ dx$ and $\int_{C_2} xy\ dx$.

SOLUTION Figure 15.3.6 shows the two paths C_1 and C_2, together with two more curves, C_3 and C_4, that together make up C_2.

To compute $\int_{C_1} xy\ dx$, we parameterize the circle with the angle θ in $[0, \pi/2]$. Thus $x = \cos(\theta)$, $y = \sin(\theta)$, and $dx = \frac{dx}{d\theta} d\theta = -\sin(\theta)\ d\theta$, so

$$\begin{aligned} \int_{C_1} xy\ dx &= \int_0^{\pi/2} (\cos(\theta))(\sin(\theta))(-\sin(\theta))\ d\theta \\ &= -\int_0^{\pi/2} \sin^2(\theta)\cos(\theta)\ d\theta = -\frac{\sin^3(\theta)}{3}\bigg|_0^{\pi/2} = -\frac{1}{3}. \end{aligned}$$

Why could we have predicted that the integral over C_1would be negative?

To calculate $\int_{C_2} xy\ dx$ we break C_2 into two curves: C_3 from $(1,0)$ to $(1,1)$ and C_4 from $(1,1)$ to $(0,1)$. (See Figure 15.3.6.)

On C_3, $x = 1$ and $dx = 0$. Thus $\int_{C_3} xy\ dx = 0$.

On C_4, $y = 1$ and x begins at 1 and ends at 0. A parameterization of C_4 is $x = 1 - t$, $y = 1$ for $0 \leq t \leq 1$. Then

$$\int\limits_{C_4} xy\ dx = \int\limits_0^1 (1-t)(1)(-dt) = \int\limits_0^1 (t-1) = \left.\frac{t^2}{2} - t\right|_0^1 = -\frac{1}{2}.$$

On C_4 we could have used the parameter x itself, which starts at 1 and decreases to 0. Then we would have $\int_{C_4} xy\ dx = \int_1^0 x\ dx = \left.\frac{x^2}{2}\right|_1^0 = -\frac{1}{2}$.

On C_2, made up of C_3 followed by C_4, we have $\int_{C_2} xy\ dx = 0 + (-1/2) = -1/2$.

The line integrals $\int_{C_1} xy\ dx$ and $\int_{C_2} xy\ dx$ are not equal even though they start at the same point $(1,0)$, end at the same point $(0,1)$, and have the same integrand. ◇

As Example 3 shows, $\int_C xy\ dx$ is not determined by the end points of the curve C. However, some line integrals are. The next example presents a case where the line integral depends only on the endpoints.

EXAMPLE 4 Compute $\int_C \frac{x\ dx + y\ dy}{x^2+y^2}$ on the paths, C_1 and C_2, in Example 3.

SOLUTION On the circular path C_1 we use θ as a parameter and have

$$\int\limits_{C_1} \frac{x\ dx + y\ dy}{x^2+y^2} = \int\limits_0^{\pi/2} \frac{(\cos(\theta))(-\sin(\theta)\ d\theta) + \sin(\theta)(\cos(\theta))\ d\theta}{\cos^2(\theta)+\sin^2(\theta)} = \int\limits_0^{\pi/2} \frac{0}{1} d\theta = 0.$$

Next we compute the integral on C_2. The path from $(1,0)$ to $(1,1)$ is C_3. There $x = 1$, so $dx = 0$. Therefore, using y as the parameter, we find that

$$\begin{aligned}\int\limits_{C_3} \frac{x\ dx + y\ dy}{x^2+y^2} &= \int\limits_{C_3} \frac{1\cdot 0 + y\ dy}{1+y^2} = \int\limits_{C_3} \frac{y}{1+y^2}\ dy \\ &= \int\limits_0^1 \frac{y\ dy}{1+y^2} = \left.\frac{\ln(1+y^2)}{2}\right|_0^1 = \frac{\ln 2}{2}.\end{aligned}$$

On the path C_4, from $(1,1)$ to $(0,1)$, we use x as the parameter starting at $x = 1$ and $y = 1$, so $dy = 0$ and we have

$$\int\limits_{C_4} \frac{x\ dx + y\ dy}{x^2+y^2} = \int\limits_1^0 \frac{x\ dx}{x^2+1} = \left.\frac{\ln(x^2+1)}{2}\right|_1^0 = -\frac{\ln 2}{2}.$$

Thus $\int_{C_2} \frac{x\ dx + y\ dy}{x^2+y^2} = -\frac{\ln 2}{2} + \frac{\ln 2}{2} = 0$. This is the same value as the integral over C_1. ◇

In Section 18.1 we will show that $\int_C \frac{x\ dx+y\ dy}{x^2+y^2}$ depends only on the end points of C. That is, if C_1 and C_2 are any two curves from point A to point B then

$$\int\limits_{C_1} \frac{x\ dx + y\ dy}{x^2 + y^2} = \int\limits_{C_2} \frac{x\ dx + y\ dy}{x^2 + y^2}.$$

A differential form $P(x, y, z)\ dx + Q(x, y, z)\ dy + R(x, y, z)\ dz$ is called a **conservative form** if its line integrals depend only on the endpoints of the curves over which the integration takes place. Likewise, the vector field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ is called a **conservative field** when $\int_C \mathbf{F} \cdot d\mathbf{r}$ depends only on the end points of the curve C. In Section 18.6 we will develop a criterion for identifying conservative fields and see that conservative fields are easier to work with in applications.

Summary

We defined four integrals for curves in space (three for curves in the xy-plane). The first, $\int_C f(P) ds$, is the limit of sums of the form $\sum_{i=1}^n f(P_i)\Delta s_i$, which is an integral defined in Chapter 6. The other three, $\int_C f(P)\ dx$, $\int_C f(P)\ dy$, $\int_C f(P)\ dz$, are different. For instance, the first is the limit of sums of the form $\sum_{i=1}^n f(P_i)\Delta x_i$, where x is the x-coordinate of a point on the curve. Putting them together we have the general line integral

$$\int\limits_C (P(x, y, z)\ dx + Q(x, y, z)\ dy + R(x, y, z)\ dz) = \int\limits_C (P\ dx + Q\ dy + R\ dz).$$

Introducing the vector field $\mathbf{F}(x, y, z) = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$, we developed three compact notations for a line integral, $\int_C \mathbf{F} \cdot d\mathbf{r}$, $\int_C \mathbf{F} \cdot \mathbf{r}' dt$, and $\int_C \mathbf{F} \cdot \mathbf{T}\ ds$.

EXERCISES for Section 15.3

1. Following the approach in Example 2, show that if C were oriented clockwise, then $\oint_C y\ dx$ would equal the area inside C.

2. Let C in Example 2 be oriented counterclockwise. Show why $\oint_C x\ dy$ equals the area inside C.

3. Show that the area within a convex curve C is $\frac{1}{2}\oint_C (x\ dy - y\ dx)$ if C is oriented counterclockwise.

4. (See Example 3.) Compute $\int_C xy\ dx$ on the straight-line path that goes from $(1, 0)$ to $(0, 0)$, and from there to $(0, 1)$.

5. If $\mathbf{F}(P)$ is perpendicular to the curve C at every point P on C, what is $\int_C \mathbf{F}\cdot d\mathbf{r}$?

6. If $\mathbf{F}(P)$ equals $\mathbf{T}(P)$ for every point P on the curve C, what is $\int_C \mathbf{F} \cdot d\mathbf{r}$?

7. Let a and b be positive numbers. Let C be the curve bounding the rectangle with vertices $(0, 0)$, $(a, 0)$, (a, b), and $(0, b)$. By calculating $\oint_C x\ dy$ with C oriented counterclockwise, confirm the result of Example 2. That is, check that the line integral over the closed curve C equals the area of the rectangle.

8. Let a and b be positive numbers. Let C be the curve bounding the triangle with vertices $(0, 0)$, $(a, 0)$, and $(0, b)$. By calculating $\oint_C y\ dx$ with C oriented clockwise, show that the integral equals the area of the triangle.

9. Let C be the curve bounding the circle of radius a with center at the origin. By calculating $\int_C x\ dy$ counterclockwise, check that the integral equals the area of the circle.

10. Let $\mathbf{F}(x, y, z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} = \mathbf{r}$. Let C be a curve starting at (x_0, y_0, z_0) and ending at (x_1, y_1, z_1). Calculate $\int_C \mathbf{F} \cdot d\mathbf{r}$ by rewriting it as $\int_a^b (\mathbf{F} \cdot \mathbf{r}')\ dt$. Note that $\int_C \mathbf{F} \cdot d\mathbf{r}$ depends only on the endpoints of C.

In Exercises 11 to 14, sketch the curve and label its start and finish.

11. $\mathbf{r}(t) = t\mathbf{i} + t^2\mathbf{j}$, t in $[0, 1]$

12. $\mathbf{r}(t) = (1 - t)\mathbf{i} + (1 - t)^2\mathbf{j}$, t in $[0, 1]$

13. $\mathbf{r}(t) = (2t + 1)\mathbf{i} + 3t\mathbf{j}$, t in $[0, 2]$

14. $\mathbf{r}(t) = 4\cos t\mathbf{i} + 5\sin t\mathbf{j}$, t in $[0, 1]$

In Exercises 15 to 18, parameterize the curve with the indicated orientation.

15. Figure 15.3.7(a)

16. Figure 15.3.7(b)

17. Figure 15.3.7(c)

18. Figure 15.3.7(d)

(a) (b) (c) (d)

Figure 15.3.7

In Exercises 19 to 22, evaluate

19. $\int_C xy\ dx$, where C is the straight line from $(1, 1)$ to $(3, 3)$.

20. $\int_C x^2\ dy$, where C is the straight line from $(2, 0)$ to $(2, 5)$.

21. $\int_C x^2\ dy$, where C is the straight line from $(3, 2)$ to $(7, 2)$.

22. $\int_C (xy\ dx + x^2\ dy)$, where C is the straight line from $(1, 0)$ to $(0, 1)$.

In Exercises 23 to 26 evaluate the integral with minimum effort. C is a counterclockwise curve bounding a region of area 5.

23. $\oint_C 3y\ dx$

24. $\oint_C (2y\ dx + 6x\ dy)$

25. $\oint_C [2x\ dx + (x + y)\ dy]$

26. $\oint_C [(x + 2y + 3)\ dx + (2x - 3y + 4)\ dy]$

In Exercise 10, the value of the line integral depends only on the endpoints, not on the path that joins them. Exercises 27 and 28 are examples where the path matters.

27. Evaluate $\int_C (xy\ dx + 7\ dy)$ on

(a) the straight path from $(1, 1)$ to $(2, 4)$

(b) the path from $(1, 1)$ to $(2, 4)$ that lies on the parabola $y = x^2$.

28. Evaluate $\int_C x\ dy$ on

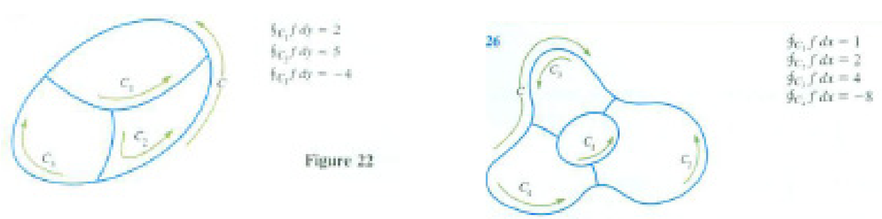
(a) the straight path from $(0, 0)$ to $(\pi/2, 1)$

(b) the path from $(0, 0)$ to $(\pi/2, 1)$ that lies on the curve $y = \sin(x)$.

In Exercises 29 and 30, the values of some line integrals are given for curves oriented as shown. Use this information to find $\int_C f\ dy$. (Pay attention to the orientations.)

29. Figure 15.3.8(a)

30. Figure 15.3.8(b)



(a) (b)

Figure 15.3.8

31. Let the closed curve C bound the region $\mathcal{R}$, which is broken into regions $\mathcal{R}_i$, $1 \le i \le n$, with each $\mathcal{R}_i$ bounded by a closed curve C_i. Let $\mathbf{F}$ be a vector field on $\mathcal{R}$. If all the curves are swept out counterclockwise, show that $\oint_C \mathbf{F}\cdot d\mathbf{r} = \sum_{i=1}^n \oint_{C_i} \mathbf{F}\cdot d\mathbf{r}$.

32. Show that $\int_C \frac{x\ dx}{x^2+y^2}$ is not conservative by calculating the line integral on two paths joining $(1,0)$ to $(1,1)$ for which the integrals are not equal.

33. Let k be a constant. Show that $\oint_C k\ dy = 0$.

34. Let $\mathbf{r} = \mathbf{r}(t)$ describe a curve C in the plane or in space. What is the geometric interpretation of

$$\frac{1}{2}\int\limits_C |\mathbf{r} \times \mathbf{T}|\ ds?$$

This is used in the first Calculus is Everywhere section (CIE 20) at the end of this chapter.

35. If t represents time and $\mathbf{r}(t)$ describes a curve C, what is the meaning of $\int_C \mathbf{T}\cdot d\mathbf{r}$? (Draw a picture of a small section of the curve.)

15.4 Four Applications of Line Integrals

In the last section we defined line integrals and showed that $\oint_C y\,dx$ and $\oint_C x\,dy$ in the plane are related to the area of the region bounded by the closed curve C. In this section we show how line integrals occur in the study of work, fluid flow, and of the angle subtended by a planar curve.

In each application we divide the domain into smaller pieces, approximate a quantity on each piece, add the contribution from the pieces, and take a limit as the pieces get smaller and smaller.

Work Along a Curve

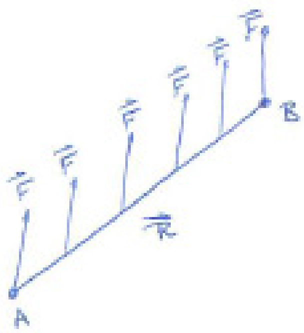


Figure 15.4.1

A force $\mathbf{F}$ remains constant (in direction and magnitude) and pushes a particle in a straight line from A to B. The work accomplished by $\mathbf{F}$ is defined as $\mathbf{F}\cdot\mathbf{R}$, where $\mathbf{R} = \overrightarrow{AB}$:

$$\text{work} = \mathbf{F}\cdot\mathbf{R}.$$

This is the product of the scalar component of $\mathbf{F}$ in the direction of $\mathbf{R}$ and the distance the particle moves. (See Figure 15.4.1)

What if $\mathbf{F}$ varies and pushes the particle along a curve that is not straight? (See Figure 15.4.2(a).)

(a) (b)

Figure 15.4.2

Assume the curve, C, is parameterized by $\mathbf{r}(t)$ for t in $[a,b]$. Partition $[a,b]$ by $t_0 = a,\ t_1,\ \ldots,\ t_n = b$ and let $\mathbf{r}(t_0) = \overrightarrow{OP_0}$, $\mathbf{r}(t_1) = \overrightarrow{OP_1}$, $\ldots$, $\mathbf{r}(t_n) = \overrightarrow{OP_n}$, be the corresponding position vectors. (See Figure 15.4.2(b).) The points P_0, P_1, $\ldots$, P_n break the curve into n shorter curves. The work done by $\mathbf{F}$ along C between P_i and P_{i+1} is approximately $\mathbf{F}(P_i)\cdot\Delta\mathbf{r}_i$ where $\Delta\mathbf{r}_i = \overrightarrow{P_iP_{i+1}}$. The total work done by $\mathbf{F}$ along C is approximated by

$$\sum_{i=1}^{n} \mathbf{F}(P_i)\cdot\Delta\mathbf{r}_i.$$

Taking the limit as the largest $|\Delta\mathbf{r}_i|$ approaches 0, we conclude

$$\text{work done by } \mathbf{F} \text{ along } C \text{ is } \int\limits_C \mathbf{F} \cdot\ d\mathbf{r}. \qquad (15.4.1)$$

If $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$, where P and Q are functions of x and y and $d\mathbf{r} = dx\mathbf{i} + dy\mathbf{j}$, then

$$\text{Work done by } P\mathbf{i} + Q\mathbf{j} \text{ along } C \text{ is } \int\limits_C (P\ dx + Q\ dy).$$

Physicists and engineers commonly use (15.4.1) as a starting point when expressing work.

The vector notation $\mathbf{F} \cdot d\mathbf{r}$ is far more suggestive than the scalar notation $P\ dx + Q\ dy$. It says that work is the dot product of force and displacement. That implies that only the component of the force in the direction of motion accomplishes work.

EXAMPLE 1 How much work is accomplished by the force $\mathbf{F}(x, y) = xy\mathbf{i}+ y\mathbf{j}$ in pushing a particle from $(0, 0)$ to $(3, 9)$ along the parabola $y = x^2$?
SOLUTION

Figure 15.4.3 shows the path of the particle. Call it C. Then

$$\text{Work} = \int\limits_C \mathbf{F} \cdot d\mathbf{r} = \int\limits_C (xy\mathbf{i} + y\mathbf{j}) \cdot (dx\mathbf{i} + dy\mathbf{j}) = \int\limits_C (xy\ dx + y\ dy).$$

To evaluate the line integral, we use x as the parameter with x in $[0, 3]$. Then $y = x^2$ and $dy = 2x\ dx$, so

$$\int\limits_C (xy\ dx + y\ dy) = \int\limits_0^3 \left(x \cdot x^2\ dx + x^2(2x\ dx)\right) = \int\limits_0^3 3x^3\ dx = \frac{243}{4}.$$

◇

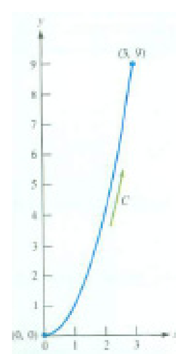


Figure 15.4.3

If we write dr as $\mathbf{T}\ ds$ then the work integral becomes $\int_C \mathbf{F} \cdot \mathbf{T}\ ds$. This says "Work is the integral of the tangential component of the force."

Circulation of a Fluid

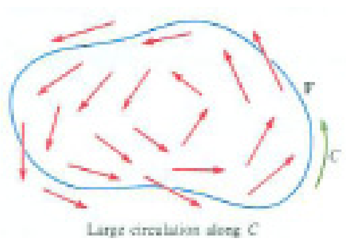


Figure 15.4.4 circulation

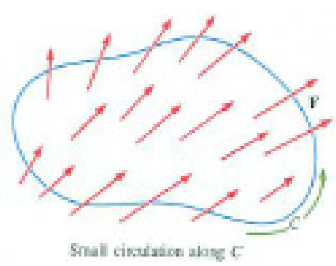


Figure 15.4.5

Draw a closed closed curve C as in Figure 15.4.4 or Figure 15.4.5. In Figure 15.4.4, C surrounds a whirlpool and there is a tendency for fluid to flow along C rather than across it. In Figure 15.4.5 most of the fluid flow is across C rather than parallel to it. The component of $\mathbf{F}$ parallel to the tangent vector determines the tendency of the fluid to flow along C. Because $\mathbf{F}\cdot d\mathbf{r}$ represents flow in the direction of $d\mathbf{r}$, $\oint_C \mathbf{F}\cdot d\mathbf{r}$ represents the tendency of the fluid to flow along C. If C is oriented counterclockwise and $\oint_C \mathbf{F}\cdot d\mathbf{r}$ is positive, the flow of $\mathbf{F}$ along C would be counterclockwise as well. If $\oint_C \mathbf{F}\cdot d\mathbf{r}$ is negative, the flow would tend to be clockwise. The line integral $\oint_C \mathbf{F}\cdot d\mathbf{r}$ is called the **circulation** of $\mathbf{F}$ along C.

The same integral, $\oint_C \mathbf{F}\cdot d\mathbf{r}$, occurs in the study of work and in the study of fluids.

EXAMPLE 2 Find the circulation of the planar flow $\mathbf{F}(x,y) = xy\mathbf{i} + y\mathbf{j}$ around the closed curve that follows $y = x^2$ from $(0,0)$ to $(3,9)$, then horizontally to $(0,9)$, and straight down to $(0,0)$.

SOLUTION The closed curve C comes in three parts: $C = C_1 + C_2 + C_3$ where C_1 is $y = x^2$ for $0 \le x \le 3$, $-C_2$ is $y = 9$, $0 \le x \le 3$, and $-C_3$ is $x = 0$, $0 \le y \le 9$.

The circulation is

$$\begin{aligned}\oint\limits_C \mathbf{F}\cdot d\mathbf{r} &= \int\limits_{C_1} \mathbf{F}\cdot d\mathbf{r} + \int\limits_{C_2} \mathbf{F}\cdot d\mathbf{r} + \int\limits_{C_3} \mathbf{F}\cdot d\mathbf{r} \\ &= \int\limits_{C_1} \mathbf{F}\cdot d\mathbf{r} - \int\limits_{-C_2} \mathbf{F}\cdot d\mathbf{r} - \int\limits_{-C_3} \mathbf{F}\cdot d\mathbf{r}.\end{aligned}$$

We use $-C_2$ and $-C_3$ because they are easier to parameterize than C_2 and C_3.

By Example 1, $\int_{C_1} \mathbf{F}\cdot d\mathbf{r} = \frac{243}{4}$. By direct calculation:

$$\int\limits_{-C_2} \mathbf{F}\cdot d\mathbf{r} = \int\limits_0^3 \langle 9x, 9\rangle \cdot \langle dx, 0\rangle = \int\limits_0^3 9x\ dx = \frac{81}{2}$$

and

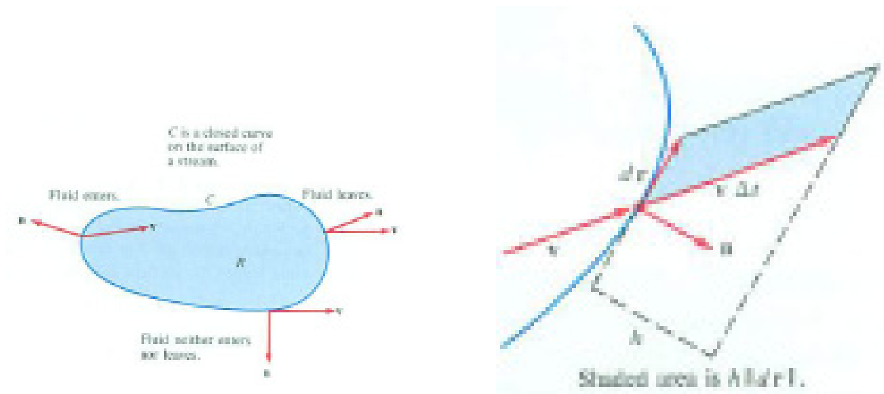
$$\int\limits_{-C_3} \mathbf{F}\cdot d\mathbf{r} = \int\limits_0^9 \langle 0, y\rangle \cdot \langle 0, dy\rangle = \int\limits_0^9 y\ dy = \frac{81}{2}.$$

The circulation of $\mathbf{F}$ around C is $\oint_C \mathbf{F}\cdot d\mathbf{r} = \frac{243}{4} - \frac{81}{2} - \frac{81}{2} = \frac{-81}{4}$. ◇

That a negative circulation means this vector field tends to circulate counterclockwise around this curve is hinted at in Exercise 15 and investigated in more detail in Chapter 18.

Loss or Gain of a Fluid (Flux)

Draw a closed curve C, fixed in space, on the surface of a moving fluid. In Figure 15.4.6(a), at what rate is fluid escaping or entering the region R whose boundary is C?



(a) (b)

Figure 15.4.6

If the fluid tends to escape, then it is thinning out in R, becoming less dense at some points. If the fluid tends to accumulate, it is becoming denser at some points. (Think of this ideal fluid as a gas rather than a liquid; gases can vary in density while liquids tend to have constant density.)

Since the fluid is escaping or entering R only along its boundary, it suffices to consider the total loss or gain across C. Where $\mathbf{v}$, the fluid velocity, is tangent to C, fluid neither enters nor leaves. Where $\mathbf{v}$ is not tangent to C, fluid is either entering or leaving across C, as indicated in Figure 15.4.6(a).

The rate at which fluid crosses C depends not only on its velocity but also on its density, which we denote σ. So the vector field of interest is $\mathbf{F} = \sigma\mathbf{v}$.

The vector $\mathbf{n}$ is a unit vector perpendicular to C and pointing away from the region it bounds. It is called the **exterior normal** or **outward normal**.

To find the total loss or gain of fluid past C, let us look at a short section of C, which we will view as a vector $d\mathbf{r}$. How much fluid crosses $d\mathbf{r}$ in a short interval of time Δt?

During time Δt the fluid moves a distance $|\mathbf{v}|\Delta t$ across $d\mathbf{r}$. The fluid that crosses $d\mathbf{r}$ during the time Δt forms approximately the parallelogram shown in Figure 15.4.6(b).

Its area is the product of its height and its base $|d\mathbf{r}|$. That is,

$$\text{area of parallelogram} \quad = \quad |\,\mathbf{proj_n}(\mathbf{v}\Delta t)||d\mathbf{r}| = (\mathbf{v}\Delta t)\cdot\mathbf{n}|d\mathbf{r}|.$$

Since the density of the fluid is σ,

$$\text{mass in parallelogram} = \sigma(\mathbf{v}\Delta t)\cdot\mathbf{n}|d\mathbf{r}| = (\sigma\mathbf{v})\cdot\mathbf{n}|d\mathbf{r}|\Delta t = \mathbf{F}\cdot\mathbf{n}|d\mathbf{r}|\Delta t.$$

Thus the rate at which fluid crosses $d\mathbf{r}$ per unit time is approximately

$$\frac{\mathbf{F}\cdot\mathbf{n}|d\mathbf{r}|\Delta t}{\Delta t} = \mathbf{F}\cdot\mathbf{n}|d\mathbf{r}|.$$

Since $d\mathbf{r}$ approximates a short piece of the curve, its length $|d\mathbf{r}|$ approximates the arc length ds. Therefore, the rate at which the fluid crosses a short part of C of length ds is approximately

$$\mathbf{F}\cdot\mathbf{n}\ ds.$$

Hence the line integral

$$\oint\limits_C \mathbf{F}\cdot\mathbf{n}\ ds$$

represents the rate of net loss or gain of fluid inside R. If it is positive, fluid tends to leave R, and the mass of fluid in R decreases. If it is negative, fluid tends to enter R, and the mass of fluid in R increases.

rate of net loss or gain of fluid inside the region bounded by C is $\oint\limits_R \mathbf{F}\cdot\mathbf{n}\ ds$.

Flux comes from the Latin *fluxus* (flow), from which we also get influx, reflux and fluctuate, but, oddly, not flow, which comes from the Latin *pluere* (to rain).

The quantity $\oint_C \mathbf{F}\cdot\mathbf{n}\ ds$ is called the **flux** of $\mathbf{F}$ across C. Contrast this with $\int \mathbf{F}\cdot\mathbf{T}\ ds$, the integral of the tangential component, which describes circulation and work.

$$\text{circulation} = \int\limits_C \mathbf{F}\cdot\mathbf{T}\ ds$$
$$\text{flux} = \int\limits_C \mathbf{F}\cdot\mathbf{n}\ ds.$$

Flux is the integral of the normal component of **F**. Circulation is the integral of the tangential component of **F**.

EXAMPLE 3 Let $\mathbf{F} = (2+x)\mathbf{i}$ describe the flow of a fluid in the xy plane. Does the amount of fluid within the circle C of radius 2 and center $(0,0)$ tend to increase or decrease?

Before doing any calculations, what is your answer?

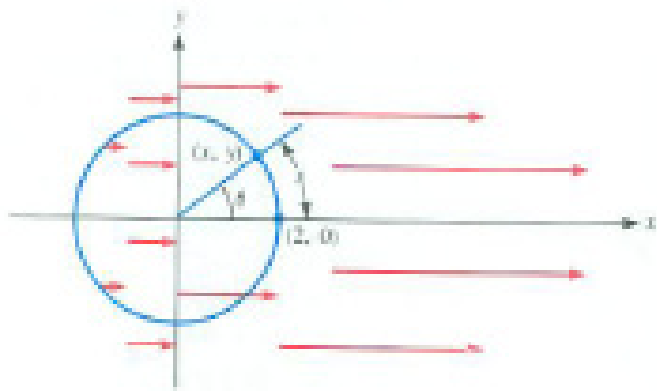


Figure 15.4.7

SOLUTION Figure 15.4.7 shows the circle and a few of the vectors of **F**. Since the flow increases as we move to the right, there appears to be more fluid leaving the disk than entering it. We expect the flux $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$ to be positive. To compute $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$, introduce θ as the parameter. Then

$$x = 2\cos(\theta), \qquad y = 2\sin(\theta) \qquad \text{for } 0 \le \theta \le 2\pi.$$

Since the circle has radius 2, $s = 2\theta$ and therefore

$$ds = 2d\theta.$$

The unit normal is parallel to the radius vector $x\mathbf{i} + y\mathbf{j}$. Therefore,

$$\mathbf{n} = \frac{x\mathbf{i}+y\mathbf{j}}{|x\mathbf{i}+y\mathbf{j}|} = \frac{2\cos(\theta)\mathbf{i} + 2\sin(\theta)\mathbf{j}}{2} = \cos(\theta)\mathbf{i} + \sin(\theta)\mathbf{j},$$

which leads to

Evaluate these definite integrals in your head. Why are $\int_0^{2\pi}\cos\theta\,d\theta = 0$ and $\int_0^{2\pi}\cos^2\theta\,d\theta = \pi$?

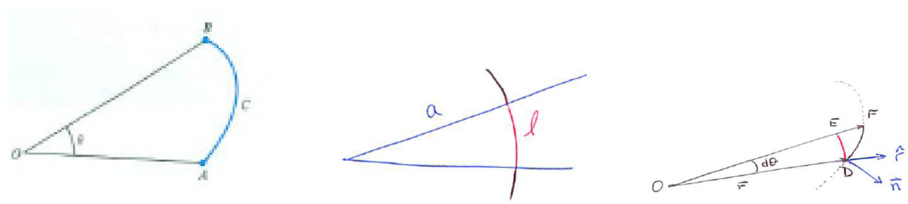
$$\begin{aligned}
\text{flux} &= \oint_C \mathbf{F}\cdot\mathbf{n}\,ds = \int_0^{2\pi} \underbrace{[(2+x)\mathbf{i}\cdot\mathbf{n}]}_{\mathbf{F}\cdot\mathbf{n}}\underbrace{2d\theta}_{ds} \\
&= \int_0^{2\pi} (2+2\cos(\theta))\mathbf{i}\cdot(\cos(\theta)\mathbf{i}+\sin(\theta)\mathbf{j})2\,d\theta = \int_0^{2\pi}(4\cos(\theta)+4\cos^2(\theta))d\theta \\
&= \int_0^{2\pi}(4\cos(\theta)+2+2\cos(2\theta))d\theta = (4\sin(\theta)+2\theta+\sin(2\theta))\Big|_0^{2\pi} = 4\pi.
\end{aligned}$$

As expected, the flux is positive since there is a net flow out of the disk. ◇

The Angle Subtended by a Curve

Our fourth illustration of a line integral concerns the angle subtended at a point O by a curve C in the plane. We assume that each ray from O meets C in at most one point. We include this example as background for the solid angle subtended by a surface, which appears in Chapter 18.

The curve C in Figure 15.4.8(a) subtends an angle θ at the point O. We will show that θ can be expressed as a line integral. Of course, we do not need such an integral to find θ, knowing the points A, O, and B is enough. That θ can be expressed as a line integral. It is this idea that generalizes from a curve to a surface, where the concept is useful in the theory of gravity and electromagnetism.



(a) (b) (c)

Figure 15.4.8

In Figure 15.4.8(b). the circle with radius a and center at O has an arc of length ℓ intercepted by the angle. The ratio ℓ/a is the radian measure of the angle.

To express θ in Figure 15.4.8(a) as an integral over C we develop a local estimate, $d\theta$, of the radians subtended by a part of the curve of length ds, as shown in Figure 15.4.8(c). $\overline{OD}$ is a vector, $\mathbf{r}$, of length r, and $\widehat{\mathbf{r}}$ is the unit vector in the direction of $\mathbf{r}$. $\stackrel{\frown}{DF}$ is part of the curve, and $\stackrel{\frown}{DE}$ is part of the circle. Because they are almost straight, we have

$$\overline{DE} \approx \overline{DF}\ \cos(\widehat{\mathbf{r}}, \mathbf{n}) = \overline{DF}\ \frac{\widehat{\mathbf{r}} \cdot \mathbf{n}}{|\widehat{\mathbf{r}}||\mathbf{n}|} = \overline{DF}\ \widehat{\mathbf{r}} \cdot \mathbf{n} \approx \widehat{\mathbf{r}} \cdot \mathbf{n}\ ds.$$

Thus

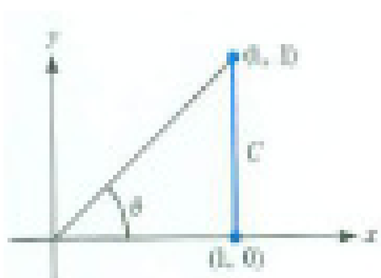
$$d\theta = \frac{\overline{DE}}{r} \approx \frac{\widehat{\mathbf{r}} \cdot \mathbf{n}}{r}\ ds.$$

From the local estimate we conclude that

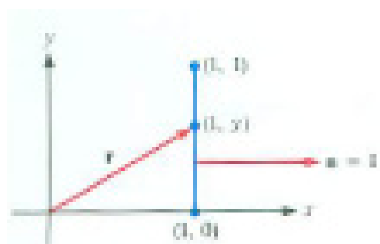
$$\text{angle } \theta \text{ subtended by arc } C \text{ is } \int_C \frac{\widehat{\mathbf{r}} \cdot \mathbf{n}}{r}\, ds. \tag{15.4.2}$$

Therefore, the angle subtended by C is the integral with respect to arc length of the normal component of the vector function $\widehat{\mathbf{r}}/|\mathbf{r}|$. That is, it is the flux of the vector field $\widehat{\mathbf{r}}/r$ (in the plane).

EXAMPLE 4 Verify (15.4.2) for the angle subtended at the origin by the line segment that joins $(1, 0)$ and $(1, 1)$.
SOLUTION The subtended angle θ is shown in Figure 15.4.9(a). Obviously $\theta = \pi/4$.



(a)



(b)

Figure 15.4.9

Let us evaluate the integral in (15.4.2) . Figure 15.4.9(b) shows that $\mathbf{n} = \mathbf{i}$ and $\mathbf{r} = \mathbf{i} + y\mathbf{j}$. Using $s = y$,

$$\begin{aligned} \theta &= \int_C \frac{\mathbf{n} \cdot \widehat{\mathbf{r}}}{|\mathbf{r}|}\, ds = \int_C \frac{\mathbf{i} \cdot \left(\frac{\mathbf{i} + y\mathbf{j}}{\sqrt{1+y^2}} \right)}{\sqrt{1+y^2}}\, ds = \int_C \frac{1}{1+y^2}\, ds \\ &= \int_0^1 \frac{1}{1+y^2}\, dy = \tan^{-1}(y)\Big|_0^1 = \frac{\pi}{4}. \end{aligned}$$

This agrees with our observation. ◇

Summary

The basic concepts introduced in this section are summarized in the following table.

Application	Work	Circulation	Flux
Integral	$\int_C \mathbf{F}\cdot\mathbf{T}\,ds = \int_C \mathbf{F}\cdot d\mathbf{r}$	$\oint_C \mathbf{F}\cdot\mathbf{T}\,ds$	$\int_C \mathbf{F}\cdot\mathbf{n}\,ds$
Description	integral of tangential component of force $\mathbf{F}$ along C	integral of tangential component of flow $\mathbf{F}$ around closed curve C	integral of normal component of flow $\mathbf{F}$ along curve C

When the vector field $\mathbf{F}$ is in the xy-plane, $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$, the integral of the tangential component of $\mathbf{F}$ along C is $\int_C \mathbf{F}\cdot\mathbf{T}\,ds = \int_C (P\,dx + Q\,dy)$ and the integral of the normal component of $\mathbf{F}$ along C is $\int_C \mathbf{F}\cdot\mathbf{n}\,ds$.

EXERCISES for Section 15.4

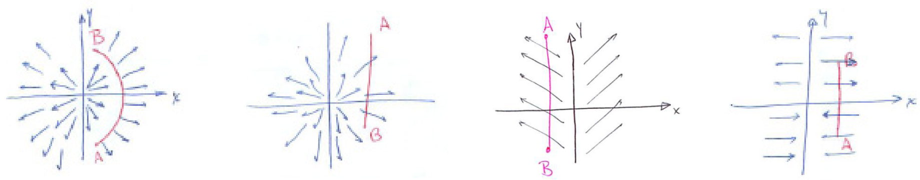
In Exercises 1 to 4 decide whether the work accomplished by the vector field in moving a particle along the curve from A to B is positive, negative, or zero.

1. Figure 15.4.10(a)

2. Figure 15.4.10(b)

3. Figure 15.4.10(c)

4. Figure 15.4.10(d)



(a) (b) (c) (d)

Figure 15.4.10

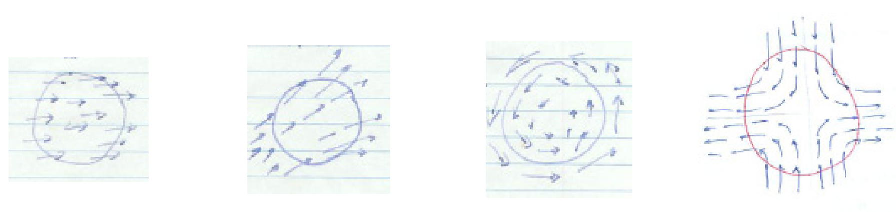
In Exercises 5 to 8 decide whether fluid is tending to leave, or enter or neither.

5. Figure 15.4.11(a)

6. Figure 15.4.11(b)

7. Figure 15.4.11(c)

8. Figure 15.4.11(d)



(a) (b) (c) (d)

Figure 15.4.11

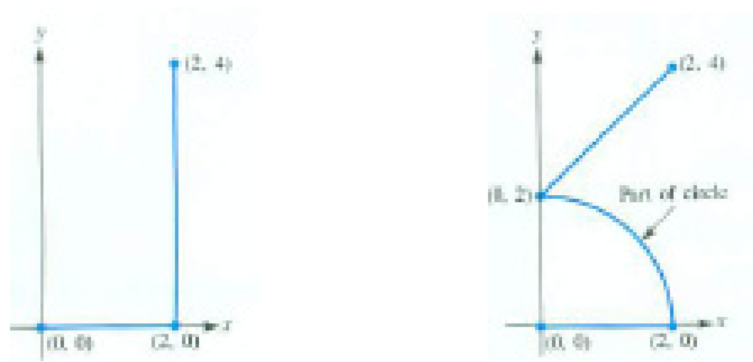
In Exercises 9 to 12 compute the work accomplished by the force $\mathbf{F} = x^2y\mathbf{i} + y\mathbf{j}$ along the curve.

9. From $(0,0)$ to $(2,4)$ along the parabola $y = x^2$.

10. From $(0,0)$ to $(2,4)$ along the line $y = 2x$.

11. From $(0,0)$ to $(2,4)$ along the path in Figure 15.4.12(a).

12. From $(0,0)$ to $(2,4)$ along the path in Figure 15.4.12(b).



(a) (b)

Figure 15.4.12

13. Verify (15.4.2) for the angle subtended at the origin by the line segment that joins $(2, 0)$ to $(2, 3)$.

14. Verify (15.4.2) for the angle subtended at the origin by the line segment that joins $(1, 0)$ to $(0, 1)$.

15. Let $\mathbf{F} = xy\mathbf{i} + y\mathbf{j}$ and C be the closed curve along $y = x^2$ from $(0, 0)$ to $(3, 9)$, then horizontally to $(0, 9)$, and straight down to $(0, 0)$.

(a) Draw $\mathbf{F}$ at a few points on each part of C.

(b) Use (a) to determine if the flow of $\mathbf{F}$ along C is clockwise or counterclockwise.

(c) Does this agree with the result in Example 2?

16. Find the work done by the force $-3\mathbf{j}$ in moving a particle from $(0, 3)$ to $(3, 0)$ along

(a) The circle of radius 3 with center at the origin.

(b) The straight path from $(0, 3)$ to $(3, 0)$.

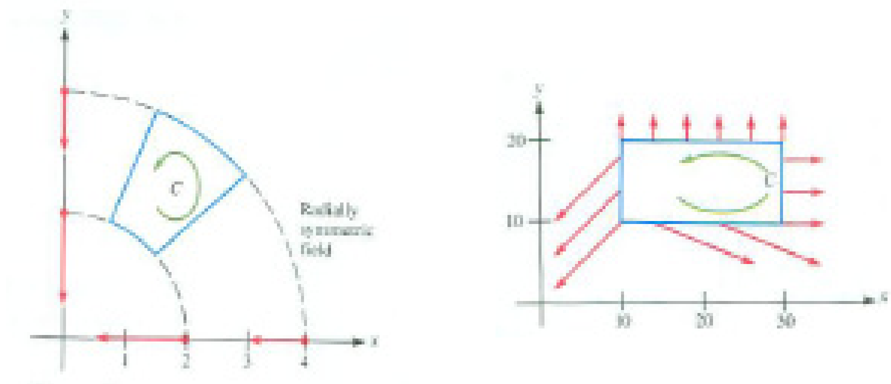
(c) The answers to (a) and (b) are the same. Will they by the same for all curves from $(0, 3)$ to $(3, 0)$?

17. Figure 15.4.13(a) shows some vectors for the vector field $\mathbf{F}$ and curve C. Use them to estimate

(a) the circulation of **F** along the boundary curve C.

(b) the flux of **F** across C.

(Since you have no formula for **F**, there is a range of correct answers.)



(a) (b)

Figure 15.4.13

18. Repeat Exercise 17 for the vector field represented in Figure 15.4.13(b).

19. The gravitational force **F** of Earth, which is located at the origin $(0, 0)$ of a rectangular coordinate system, on a particle at (x, y) is

$$\frac{-x\mathbf{i}}{(\sqrt{x^2+y^2})^3} + \frac{-y\mathbf{j}}{(\sqrt{x^2+y^2})^3} = \frac{-\mathbf{r}}{|\mathbf{r}|^3} = \frac{-\widehat{\mathbf{r}}}{r^2}.$$

where $\widehat{\mathbf{r}} = \mathbf{r}/|\mathbf{r}|$. Compute the total work done by **F** if the particles goes from $(2, 0)$ to $(0, 1)$ along

(a) the portion of the ellipse $x = 2\cos(t)$, $y = \sin(t)$ in the first quadrant;

(b) the line parameterized as $x = 2 - 2t$, $y = t$.

20.

(a) Let $W(b)$ be the work done by the force in Exercise 19 in moving a particle along the straight line from $(1, 0)$ to $(b, 0)$.

(b) What is $\lim_{b\to\infty} W(b)$?

21. Let the vector field describing a fluid flow have the value $(x+1)^2\mathbf{i} + y\mathbf{j}$ at the point (x, y). Let C be the unit circle with parametric equations $x = \cos(t)$, $y = \sin(t)$, for t in $[0, 2\pi]$.

(a) Draw $\mathbf{F}$ at eight equally spaced points on the circle.

(b) Is fluid tending to leave or enter the region bounded by C. That is, is the net outward flow positive or negative? Answer on the basis of your diagram).

(c) Compute the net outward flow using a line integral.

22. Repeat Exercise 21 where $\mathbf{F}(x, y) = (2 - x)\mathbf{i} + y\mathbf{j}$ and C is the square with vertices $(0, 0)$, $(1, 0)$, $(1, 1)$, and $(0, 1)$.

23. Let $\mathbf{F}(x, y) = \sigma\mathbf{v}$ be fluid flow, and let C be a closed curve in the xy-plane. If $\oint_C \mathbf{F} \cdot d\mathbf{r}$ is positive and C is counterclockwise, does the motion along C tend to be clockwise or counterclockwise?

24. Let $\mathbf{F}(x, y) = \sigma\mathbf{v}$ be fluid flow, and let C be a closed curve in the xy-plane. If $\oint_C \mathbf{F} \cdot \mathbf{n}\ ds$ is positive, is fluid tending to leave the region bounded by C or to enter it?

25. Let C be a closed convex curve that encloses the point O. Let $\mathbf{r}$ be the position vector $\overrightarrow{OP}$ for points P on the curve. Determine the value of $\oint_C (\widehat{\mathbf{r}} \cdot \mathbf{n})/r\ ds$ where $\mathbf{n}$ is the outward unit normal to C.

26. Write in your own words and diagrams why $\int_C \mathbf{F} \cdot d\mathbf{r}$ represents the work done by force $\mathbf{F}$ along the curve C.

27. Write in your own words and diagrams why $\int_C \mathbf{F} \cdot \mathbf{n}\ ds$ represents the net loss of fluid across C if $\mathbf{F}$ is the fluid flow and $\mathbf{n}$ is a unit external normal to C. Include the definition of $\mathbf{F}$.

28. Explain why $\int_C \mathbf{F} \cdot d\mathbf{r}$ represents the tendency of a fluid to move along C, if $\mathbf{F}$ is the fluid flow.

29. Explain why $\int_C (\widehat{\mathbf{r}} \cdot \mathbf{n})/r\ ds$ represents the angle subtended by a curve C at the origin. Assume that each ray from the origin meets C at most once.

30. Let C be a curve in space and C^* its projection on the xy plane. Assume that distinct points of C project onto distinct points of C^*. The line integral $\int_C 1\ ds$ equals the arc length of C. What integral over C equals the arc length of C^*?

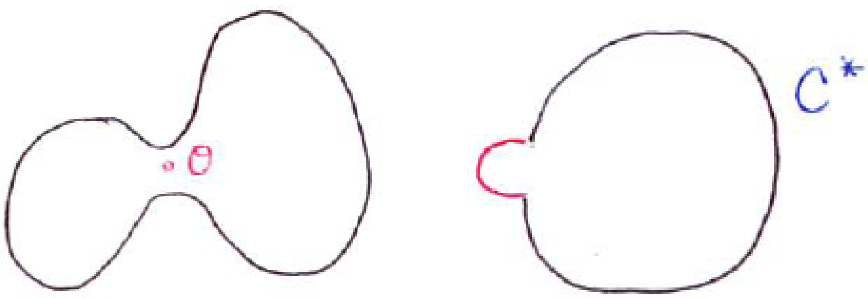
31. Sam, Jane, and Sarah are debating a delicate issue.

Sam: Let C be the circle in the xy-plane whose polar equation is $r = 2\cos(\theta)$. It is a unit circle that passes through the origin O. Let $\mathbf{F}$ be the field $\widehat{\mathbf{r}}/r$. What is the flux of $\mathbf{F}$ across C?

Jane: The field blows up at O, so the flux is an improper integral.

Sam: Yes, but if I move C rigidly just a tiny bit so O is inside it, the flux is 2π. So I say the flux across C is 2π.

Sarah: I say it's π. Just draw a figure 8 made of two copies of C joined smoothly to form one curve, as in Figure 15.4.14(a).



(a) (b)

Figure 15.4.14

The flux across the curve is 2π. Each half must have flux π. Since each half looks like C, the flux across C must be π.

Settle the issue by

(a) Evaluating the integral $\oint_C \mathbf{F}\cdot\mathbf{n}\, ds$ by the Fundamental Theorem of Calculus.

(b) Considering the flux across the curve C^* obtained from C by replacing the small part of C near O by a semicircle C, as in Figure 15.4.14(b).

(c) By considering the angle the curve C subtends at O.

32. Let $\mathbf{F}(P) = \sigma(P)\mathbf{v}(P)$ represent the flow of a fluid. Let C be a closed curve that bounds the region R. Let $Q(t)$ be the total mass of the fluid in R at time t. Express dQ/dt in terms of a line integral.

33. Let $\mathbb{C}$ be a convex curve in the xy-plane and O a point in the xy-plane outside of C. Let $\mathbf{r} = \overrightarrow{OP}$. For P on C, show that $\int_C \frac{\widehat{\mathbf{r}}\cdot\mathbf{n}}{r}\, ds = 0$. (Think about when $\mathbf{r}\cdot\mathbf{n}$ is negative and when it is positive.)

34. Let a, b and c be positive constants. Verify each antiderivative formula by showing that the derivative of the right-hand side of the equation is the integrand on the left-hand side.

(a) $\displaystyle\int \frac{x}{ax^2+c}\,dx = \frac{1}{2a}\ln\left(ax^2+c\right)$

(b) $\displaystyle\int x\sqrt{ax+b}\,dx = \frac{2(3ax-2b)}{15a^2}\sqrt{(ax+b)^3}$

(c) $\displaystyle\int \cos^3(ax)\,dx = \frac{1}{a}\sin(ax) - \frac{1}{3a}\sin^3(ax)$

(d) $\displaystyle\int \tan^2(ax)\,dx = \frac{1}{a}\tan(ax) - x$

(e) $\displaystyle\int x\cos(ax)\,dx = \frac{1}{a^2}\cos(ax) + \frac{x}{a}\sin(ax)$

(f) $\displaystyle\int \arctan(ax)\,dx = x\arctan(ax) - \frac{1}{2a}\ln(1+a^2x^2)$

15.S Chapter Summary

This chapter concerned the derivatives of vector functions and integrals over curves.

Let $\mathbf{r}(t) = \langle x(t), y(t), z(t)\rangle$ be the position vector from the origin to a point on a curve. We defined its derivative, $\mathbf{r}'(t)$, in terms of the derivatives of its components. We could just as well define it as

$$\mathbf{r}'(t) = \lim_{\Delta t \to 0} \frac{\mathbf{r}(t+\Delta t) - \mathbf{r}(t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{\Delta \mathbf{r}}{\Delta t}. \qquad (15.S.1)$$

This definition reveals the underlying geometry, as Figure 15.S.1 shows. For small Δt, the direction of $\Delta\mathbf{r}$ is almost along the tangent. The length of $\Delta\mathbf{r}$ is almost the same as the scalar length Δs along the curve. Thus, $\Delta\mathbf{r}/\Delta t$ is a vector pointing almost in the direction of motion and with a magnitude approximating the instantaneous speed.

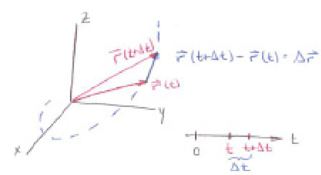
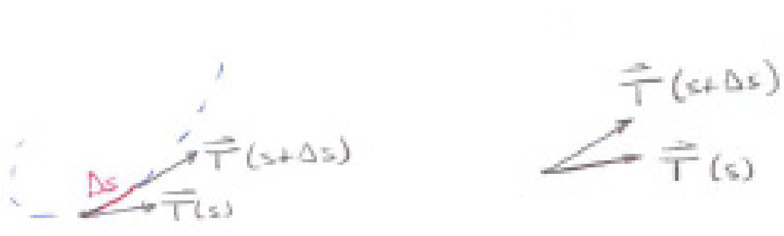


Figure 15.S.1

The limit in (15.S.1) is called the derivative of the function $\mathbf{r}(t)$. If we think of t as time, then $\mathbf{r}'$ is called the velocity vector, denoted $\mathbf{v}$. The derivative of $\mathbf{v}$ is the acceleration vector: $\mathbf{v}' = \mathbf{a}$.

The vector $\mathbf{T} = \mathbf{r}'/|\mathbf{r}'|$ is a unit tangent vector. The magnitude of its derivative with respect to arc length, s, is the curvature of the path, κ, as suggested by Figure 15.S.2. Though the curve may not lie in a plane, the figure resembles Figure 15.2.6 in Section 15.2.



(a) (b)

Figure 15.S.2

It was shown that curvature equals $|\mathbf{v}\times\mathbf{a}|/|\mathbf{v}|^3$.

The vector $d\mathbf{T}/ds$ is perpendicular to $\mathbf{T}$. (Why?) The unit vector $\mathbf{N} = \frac{d\mathbf{T}/ds}{|d\mathbf{T}/ds|}$ is called the principal normal to the curve at the given point. The vector $\mathbf{T}\times\mathbf{N} = \mathbf{B}$ is the third unit vector forming a frame that moves along the curve, with $\mathbf{T}$ and $\mathbf{N}$ indicating the plane in which the curve locally almost lies.

The acceleration vector $\mathbf{a}$, even for space curves, can be expressed relative to $\mathbf{T}$ and $\mathbf{N}$ ($\mathbf{B}$ is not involved):

$$\mathbf{a} = \frac{d^2s}{dt^2}\mathbf{T} + \frac{v^2}{r}\mathbf{N}$$

where $r = 1/\kappa$ is the radius of curvature. The second coefficient shows that the force needed to keep the particle in the path is proportional to the square of the velocity and inversely proportional to the radius of curvature.

This chapter then introduced four integrals involving a curve C:

$$\int\limits_C f(P)\ ds, \qquad \int\limits_C f(P)\ dx, \qquad \int\limits_C f(P)\ dy, \qquad \text{and} \qquad \int\limits_C f(P)\ dz,$$

whose definitions resemble those in Chapter 6 for definite integrals. In the last three the orientation of the curve matters: switching the direction in which the curve is swept out changes the sign of dx, dy, and dz, and thus that of the integral.

For a closed curve taken counterclockwise $\oint_C y\ dx$ is the negative of the area enclosed by the curve. (Why?) On the other hand, $\int_C x\ dy$ taken counterclockwise is the area enclosed.

The most general integral considered was

$$\int\limits_C \left(P(x, y, z)\ dx + Q(x, y, z)\ dy + R(x, y, z)\ dz\right).$$

whose integrand is called a differential form. For $\mathbf{F} = \langle P, Q, R\rangle$, it can be written as $\int_C \mathbf{F}\cdot\ d\mathbf{r}$. However, in proofs or computations we often need the differential form.

If $\int_C \mathbf{F}\cdot d\mathbf{r}$ depends only on the ends of C, $\mathbf{F}$ is called a conservative vector field, which will be important in Chapter 18.

Line integrals were applied to work, circulation, flux, and the angle subtended by a curve (the last in preparation for the solid angle subtended by a surface).

EXERCISES for 15.S

In Exercises 1 to 6, evaluate $\int_C \mathbf{F}\cdot d\mathbf{r}$ for the vector field $\mathbf{F}$ and curve C.

1. $\mathbf{F}(x, y) = 2x\ \mathbf{i}$ and C is a semicircle, $\mathbf{r}(\theta) = 3\cos\theta\ \mathbf{i} + 3\sin\theta\ \mathbf{j}$, $0 \le \theta \le \pi$.

2. $\mathbf{F}(x, y) = x^2\ \mathbf{i} + 2xy\ \mathbf{j}$ and C is a line segment, $\mathbf{r}(t) = 2t^2\ \mathbf{i} + 3t^2\ \mathbf{j}$, $1 \le t \le 2$.

3. $\mathbf{F}(x, y, z) = x\ \mathbf{i} + y\ \mathbf{j} + z\ \mathbf{k}$ and C is a helix, $\mathbf{r}(t) = \cos(t)\ \mathbf{i} + \sin(t)\ \mathbf{j} + 3t\ \mathbf{k}$, $0 \le t \le 4\pi$.

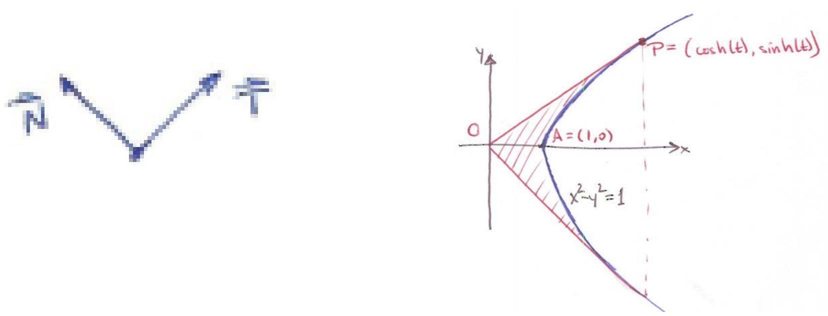
4. $\mathbf{F}(x, y, z) = x^2\ \mathbf{i} + xy\ \mathbf{j} + 3\ \mathbf{k}$ and C is a line segment, $\mathbf{r}(t) = 2t\ \mathbf{i} + (3t+1)\ \mathbf{j} + t\ \mathbf{k}$,

$1 \le t \le 2$.

5. $\mathbf{F}(\mathbf{r}) = \widehat{\mathbf{r}}/|\mathbf{r}|^2$ and C is a line, $\mathbf{r}(t) = 2t\ \mathbf{i} + 3t\ \mathbf{j} + 4t\ \mathbf{k}$, $1 \le t \le 2$.

6. $\mathbf{F}(\mathbf{r}) = \mathbf{r}$ and C is the circle $\mathbf{r}(t) = \cos(\theta)\ \mathbf{i} + \sin(\theta)\ \mathbf{j} + 2\ \mathbf{k}$, $0 \le \theta \le 2\pi$.

7. Figure 15.S.3(a) shows $\mathbf{T}$ and $\mathbf{N}$ for a point P on a curve C. The curve is not shown. Sketch what a short part of C may look like.



(a) (b)

Figure 15.S.3

8.

(a) Express the area under the hyperbola $x^2 - y^2 = 1$ and above the interval $[1, \cosh(t)]$ as a line integral.

(b) Evaluate it).

(c) What is the area of the shaded region in Figure 15.S.3(b)?

See also Exercises 64 in Section 6.5 and 77 in Section 8.6.

The CIE at the end of Chapter 3 developed the reflection properties of parabolas and ellipses. Exercises 9 and 10 show how vectors provide a shorter way to obtain them.

9. A parabola consists of the points P equidistant from a point F and a line L, as in Figure 15.S.4.

Let O be a point on L and let $\mathbf{u}$ be a unit vector perpendicular to L aimed toward P. Let $\mathbf{r} = \overrightarrow{OP}$ and $\mathbf{F} = \overrightarrow{OF}$. We assume the curve is parameterized so that there is a well-defined tangent vector, $\mathbf{r}'$.

(a) Show that $|\mathbf{r} - \mathbf{F}| = \mathbf{r} \cdot \mathbf{u}$.

(b) From (a) deduce that

$$\frac{\mathbf{r} - \mathbf{F}}{|\mathbf{r} - \mathbf{F}|} \cdot \mathbf{r}' = \mathbf{r}' \cdot \mathbf{u}.$$

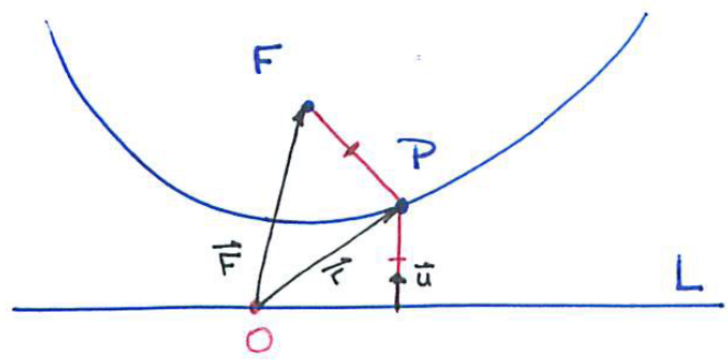


Figure 15.S.4

(c) From (b) deduce that

$$|\mathbf{r}'|\cos(\mathbf{r}', \mathbf{r} - \mathbf{F}) = |\mathbf{r}|\cos(\mathbf{r}', \mathbf{u}).$$

(Let $\mathbf{s}$ is a vector function. To differentiate $|\mathbf{s}|$, start with $\mathbf{s} \cdot \mathbf{s} = \|\mathbf{s}\|^2$.)

(d) From (c) deduce the reflection principle of a parabola.

This proof, which starts with the geometric definition of a parabola rather than the equation $y = x^2$, appears in Harley Flanders', "The Optical Properties of the Conics," *American Mathematical Monthly*, 1968, p. 399.

10. This exercise develops the reflection property of an ellipse. Start with its geometric definition as the locus of points such that the sum of whose distances from two fixed points is constant. Let $\mathbf{p}$ and $\mathbf{q}$ be the position vectors of the fixed points and $\mathbf{r}$ the position vector for a point P on the ellipse, which is parameterized so we may speak of $\mathbf{r}'$, a tangent vector.

(a) Differentiate both sides of $|\mathbf{r} - \mathbf{p}| + |\mathbf{r} - \mathbf{q}| = c$, a constant.

(b) Let $\mathbf{u}_1$ be the unit vector in the direction of $\mathbf{r} - \mathbf{p}$ and $\mathbf{u}_2$ be the unit vector in the direction of $\mathbf{r} - \mathbf{q}$. Show that $\mathbf{u}_1 \cdot \mathbf{r}' + \mathbf{u}_2 \cdot \mathbf{r}' = \mathbf{0}$.

(c) Show that $\mathbf{u}_1 + \mathbf{u}_2$ is normal to the curve at P.

(d) Show that $\mathbf{u}_1$ and $\mathbf{u}_2$ make equal angles with $\mathbf{u}_1 + \mathbf{u}_2$.

(e) From (d) deduce the reflection property of an ellipse.

In Exercises 11 to 13, $\mathbf{a}(t)$ is the acceleration vector at time t for a particle and $\mathbf{r}(t_0)$ and $\mathbf{a}(t_0)$ are the particle's position and acceleration at time $t = t_0$. Find the velocity and position vectors, $\mathbf{v}(t)$ and $\mathbf{r}(t)$, of the particle at time t.

11. $\mathbf{a}(t) = 108t\,(\ln(t))^2\,\mathbf{i} + \ln(1+t^2)\mathbf{j} + t\arctan(t)\mathbf{k}$, $\mathbf{r}(1) = 19\mathbf{i} - \mathbf{j} + ((\pi - ln(2))/6 - 2)\mathbf{k}$, and $\mathbf{v}(1) = 27\mathbf{i} - 2\mathbf{j} + \frac{1}{4}(\pi - 2)\mathbf{k}$

12. $\mathbf{a}(t) = \dfrac{\tan(t) + \sin(t)}{\sec(t)}\mathbf{i} + \dfrac{t^4}{t^2+4}\mathbf{j} + \dfrac{2t-4}{t^2+2t+1}\mathbf{k}$, $\mathbf{r}(0) = \frac{1}{4}\mathbf{i} + \mathbf{j} - 8\mathbf{k}$, and $\mathbf{v}(0) = \frac{-3}{2}\mathbf{i} + 4\mathbf{j} + 6\mathbf{k}$

13. $\mathbf{a}(t) = (t^2 + 4t + 5)^{-1}\mathbf{i} + t^2\cos(t)\mathbf{j} + \dfrac{1}{t^2+4}\mathbf{k}$, $\mathbf{r}(0) = 6\mathbf{j}$, and $\mathbf{v}(0) = \arctan(2)\mathbf{i}$

Calculus is Everywhere # 20

Newton's Law Implies Kepler's Three Laws

After hundreds of pages of computation based on observations by the astronomer Tycho Brahe (1546—1601) in the last thirty years of the sixteenth century, plus lengthy detours and lucky guesses, Kepler (1571–1630) arrived at his three laws of planetary motion:

Kepler's Three Laws

1. Every planet travels around the sun in an elliptical orbit such that the sun is situated at one focus (discovered in 1605, published in 1609).
2. The velocity of a planet varies so that the line joining the planet to the sun sweeps out equal areas in equal times (discovered 1602, published 1609).
3. The square of the time required by a planet for one revolution around the sun is proportional to the cube of its mean distance from the sun (discovered 1618, published 1619).

The work of Kepler shattered the crystal spheres that for 2,000 years had been thought to carry the planets. Before Kepler astronomers used only circular motion and motion composed of circular motions. Copernicus (1473–1543) used five circles to describe the motion of Mars.

Ellipses were not welcomed. In 1605 Kepler complained to a skeptical astronomer:

> You have disparaged my oval orbit .... If you are enraged because I cannot take away oval flight how much more you should be enraged by the motions assigned by the ancients, which I did take away .... You disdain my oval, a single cart of dung, while you endure the whole stable. (If indeed my oval is a cart of dung.)

The astronomical tables that Kepler based on his Laws, published in 1627, proved to be more accurate than any others, and ellipses gradually gained acceptance.

The three laws stood as mysteries alongside a related question: If there are no crystal spheres, what propels the planets? Bullialdus (1605–1694), a French astronomer and mathematician, suggested in 1645:

> The force with which the sun seizes or pulls the planets, a physical force which serves as hands for it, is sent out in straight lines into all the world's space . . . ; since it is physical it is decreased in greater space; . . . the ratio of this distance is the same as that for light, namely as the reciprocal of the square of the distance.

The inverse square law was conjectured.

In 1666, Hooke (1635–1703), more of an experimental scientist than a mathematician, wondered:

> why the planets should move about the sun . . . being not included in any solid orbs . . . nor tied to it . . . by any visible strings . . . . I cannot imagine any other likely cause besides these two: The first may be from an unequal density of the medium . . . ; if we suppose that part of the medium, which is farthest from the centre, or sun, to be more dense outward, than that which is more near, it will follow, that the direct motion will be always deflected inwards, by the easier yielding of the inwards . . . .
>
> But the second cause of inflecting a direct motion into a curve may be from an attractive property of the body placed in the center; whereby it continually endeavours to attract or draw it to itself. For if such a principle be supposed all the phenomena of the planets seem possible to be explained by the common principle of mechanic motions. . . . By this hypothesis, the phenomena of the comets as well as of the planets may be solved.

In 1675, Hooke, in an announcement to the Royal Society, went further:

> All celestial bodies have an attraction towards their own centres, whereby they attract not only their own parts but also other celestial bodies that are within the sphere of their activity . . . . All bodies that are put into direct simple motion will so continue to move forward in a single line till they are, by some other effectual powers, deflected and bent into a motion describing a circle, ellipse, or some other more compound curve . . . . These attractive powers are much more powerful in operating by how much the nearer the body wrought upon is to their own centers . . . . It is a notion which if fully prosecuted as it ought to be, will mightily assist the astronomer to reduce all the celestial motions to a certain rule . . . .

Trying to interest Newton in the question, Hooke wrote on November 24, 1679: "I shall take it as a great favor if . . . you will let me know your thoughts of that of compounding the celestial motion of planets of a direct motion by the tangent and an attractive motion toward the central body." But four days later Newton replied:

Hooke pressed Newton to work on the problem.

> My affection to philosophy [science] being worn out, so that I am almost as little concerned about it as one tradesman used to be about another man's trade or a countryman about learning. I must acknowledge myself averse from spending that time in writing about it which I think I can spend otherwise more to my own content and the good of others ....

In a letter to Newton on January 17, 1680, Hooke returned to the problem of planetary motion:

> It now remains to know the properties of a curved line (not circular ...) made by a central attractive power which makes the velocities of descent from the tangent line or equal straight motion at all distances in a duplicate proportion to the distances reciprocally taken. I doubt not that by your excellent method you will easily find out what that curve must be, and its properties, and suggest a physical reason for this proportion.

Hooke succeeded in drawing Newton back to science, as Newton admitted in his *Philosophiae Naturalis Principia*, usually referred to as the *Principia*, published in 1687: "I am beholden to him only for the diversion he gave me from the other studies to think on these things and for his dogmaticalness in writing as if he had found the motion in the ellipse, which inclined me to try it."

It seems that Newton then obtained a proof, perhaps containing a mistake (the history is not clear), that if the motion is elliptical, the force varies as the inverse square. In 1684, at the request of the astronomer Halley, Newton provided a correct proof. With Halley's encouragement, Newton spent the next year and a half writing the *Principia*.

Halley, of Halley's comet, paid for publication of the *Principia*.

In the *Principia*, which develops the science of mechanics and applies it to celestial motions, Newton began with two laws:

1. Every body continues in its state of rest, or of uniform motion in a straight line, unless it is compelled to change this state by forces impressed upon it.

2. The change of momentum is proportional to the motive force impressed, and is made in the direction of the straight line in which that force is impressed.

To state these in the language of vectors, let $\mathbf{v}$ be the velocity of the body, $\mathbf{F}$ the force, and m the mass of the body. The first law asserts that $\mathbf{v}$ is constant if $\mathbf{F}$ is $\mathbf{0}$. **Momentum** is $m\mathbf{v}$ so the second law asserts that

$$\mathbf{F} = \frac{d}{dt}(m\mathbf{v}).$$

If m is constant, this reduces to

$$\mathbf{F} = m\mathbf{a},$$

where $\mathbf{a}$ is the acceleration vector.

Newton assumed a universal **law of gravitation**, that a particle P exerts an attractive force on any other particle Q, and the direction of the force is from Q toward P. Then assuming that the orbit of a planet moving about the sun is an ellipse, he deduced that the force is inversely proportional to the square of the distance between the particles P and Q.

Nowhere in the *Principia* does he deduce from the inverse-square law of gravity that the planets' orbits are ellipses. (Though there are theorems in *Principia* on the basis of which this deduction could have been made.) In the *Principia* he showed that Kepler's second law (concerning areas) was equivalent to the assumption that the force acting on a planet is directed toward the sun. He also deduced Kepler's third law.

Newton's universal law of gravitation asserts that a particle, of mass M, exerts a force on any other particle, of mass m, and that the magnitude of the force is proportional to the product of the masses, mM, inversely proportional to the square of the distance between them, and is directed toward the particle with the larger mass. (Here, we assume $M > m$.)

Assume that the sun has mass M and is located at O and that the planet has mass m and is located at P. (See Figure C.20.1.) Let $\mathbf{r} = \overrightarrow{OP}$ and $r = |\mathbf{r}|$. Then the sun exerts a force $\mathbf{F}$ on the planet given by

$$\mathbf{F} = -\frac{GmM}{r^2}\mathbf{r}, \tag{C.20.1}$$

where G is a universal constant and $\mathbf{u} = \mathbf{r}/r$ is the unit vector that points in the direction of $\mathbf{r}$.

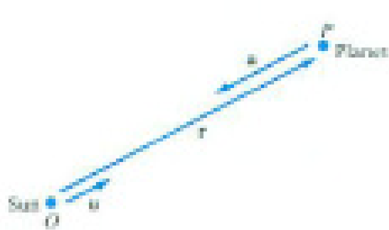


Figure C.20.1

Now, $\mathbf{F} = m\mathbf{a}$, where $\mathbf{a}$ is the acceleration vector of the planet. Thus

$$m\mathbf{a} = -\frac{GmM}{r^2}\mathbf{u},$$

from which it follows that

$$\mathbf{a} = -\frac{q\mathbf{u}}{r^2}, \tag{C.20.2}$$

where $q = GM$ is independent of the planet.

The vectors $\mathbf{u}$, $\mathbf{r}$, and $\mathbf{a}$ are in Figure C.20.1.

The following exercises show how to obtain Kepler's laws from the single law of Newton, $\mathbf{a} = -q\mathbf{u}/r^2$.

EXERCISES

Exercises 1 to 3 obtain Kepler's area law.

1. Let $\mathbf{r}(t)$ be the position vector of a planet at time t. Let $\Delta\mathbf{r} = \mathbf{r}(t + \Delta t) - \mathbf{r}(t)$. Show that for small Δt,

$$\frac{1}{2}\left|\mathbf{r} \times \Delta\mathbf{r}\right|$$

approximates the area swept out by the position vector during the small interval of time Δt. (Draw a picture.)

Figure C.20.2

2. From Exercise 1 deduce that $\frac{1}{2}\left|\mathbf{r} \times \frac{d\mathbf{r}}{dt}\right|$ is the rate at which the position vector $\mathbf{r}$ sweeps out area. (See Figure C.20.2.)

Let $\mathbf{v} = d\mathbf{r}/dt$.

3. With the aid of (C.20.2), show that $\mathbf{r} \times \mathbf{v}$ is constant.

Since $\mathbf{r} \times \mathbf{v}$ is constant, $\frac{1}{2}\left|\mathbf{r} \times \mathbf{v}\right|$ is constant. In view of Exercise 2, it follows that the radius vector of a given planet sweeps out area at a constant rate.

To put it another way, the radius vector sweeps out equal areas in equal times. This is Kepler's second law.

Introduce an xyz-coordinate system such that the unit vector $\mathbf{k}$, which points in the direction of the positive z-axis, has the same direction as the constant vector $\mathbf{r} \times \mathbf{v}$. Thus there is a positive constant h such that

$$\mathbf{r} \times \mathbf{v} = h\mathbf{k}. \tag{C.20.3}$$

Exercises 4 to 13 obtain Kepler's ellipse law.

4. Show that h in (C.20.3) is twice the rate at which the position vector of the planet sweeps out area.

5. Show that the planet remains in the plane perpendicular to $\mathbf{k}$ that passes through the sun.

By Exercise 5, the orbit of the planet is planar. We may assume that it lies in the xy-plane; for convenience, locate the origin of the xy-coordinates at the sun. Introduce polar coordinates in the plane, with the pole at the sun and the polar axis along the positive x axis, as in Figure C.20.3.

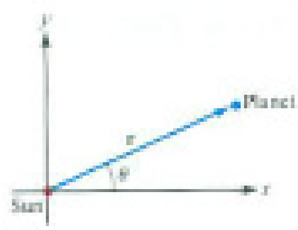


Figure C.20.3

6.

(a) Show that during the time interval $[t_0, t]$ the position vector of the planet sweeps out the area
$$\frac{1}{2}\int_{t_0}^{t} r^2 \frac{d\theta}{dt} dt.$$

(b) Deduce that the radius vector sweeps out area at the rate $\frac{1}{2}r^2\frac{d\theta}{dt}$.

We use the dot notation for differentiation with respect to time, so $\dot{\mathbf{r}} = \mathbf{v}$, $\dot{\mathbf{v}} = \mathbf{a}$, and $\dot{\theta} = \frac{d\theta}{dt}$.

7. Show that $\mathbf{r} \times \mathbf{v} = r^2\dot{\theta}\mathbf{k}$.

8. Show that $\dot{\mathbf{u}} = \frac{d\mathbf{u}}{d\theta}\dot{\theta}$ and is perpendicular to $\mathbf{u}$. ($\mathbf{u}$ is defined as $\mathbf{r}/|\mathbf{r}|$.)

9. From $\mathbf{r} = r\mathbf{u}$, show that $h\mathbf{k} = r^2(\mathbf{u} \times \dot{\mathbf{u}})$.

10. Using (C.20.2) and Exercise 9, show that $\mathbf{a} \times h\mathbf{k} = q\dot{\mathbf{u}}$. (What is the vector identity for $\mathbf{A} \times (\mathbf{B} \times \mathbf{C})$?)

11. Deduce from Exercise 10 that $\mathbf{v} \times h\mathbf{k}$ and $q\mathbf{u}$ differ by a constant vector.

By Exercise 11, there is a constant vector $\mathbf{C}$ such that

$$\mathbf{v} \times h\mathbf{k} = q\mathbf{u} + \mathbf{C}.$$

Then the angle between $\mathbf{r}$ and $\mathbf{C}$ is the angle θ of polar coordinates.
The next exercise requires the vector identity $(\mathbf{A} \times \mathbf{B}) \cdot \mathbf{C} = \mathbf{A} \cdot (\mathbf{B} \times \mathbf{C})$, which is valid for any three vectors.
12.

(a) Show that $(\mathbf{r} \times \mathbf{v}) \cdot h\mathbf{k} = h^2$.

(b) Show that $\mathbf{r} \cdot (\mathbf{v} \times h\mathbf{k}) = rq + \mathbf{r} \cdot \mathbf{C}$.

(c) Combining (a) and (b), deduce that $h^2 = rq + rc\cos(\theta)$, where $c = |\mathbf{C}|$

It follows from Exercise 12 that the polar equation for the orbit of the planet is given by

$$r(\theta) = \frac{h^2}{q + c\cos(\theta)}. \tag{C.20.4}$$

13. By expressing (C.20.4) in rectangular coordinates, show that it describes a conic section.

Since the orbit of a planet is bounded and is also a conic section, it must be an ellipse. This establishes Kepler's first law.
Kepler's third law asserts that the square of the time required for a planet to complete one orbit is proportional to the cube of its mean distance from the sun.
For Kepler **mean distance** meant the average of the shortest distance and the longest distance from the planet to the sun. Let us compute the average for the ellipse with semimajor axis a and semiminor axes b, shown in Figure C.20.4. The sun is at the focus F, which is also the pole of the polar coordinate system we are using. The line through the foci contains the polar axis.

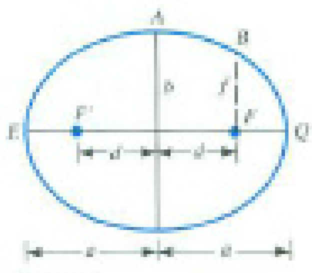


Figure C.20.4

An ellipse is the set of points P such that the sum of the distances from P to the foci F and F' is constant, $2a$. The shortest from the planet to the sun is $\overline{FQ} = a - d$ and the longest distance is $\overline{EF} = a + d$. Thus Kepler's mean distance is

$$\frac{(a-d)+(a+d)}{2} = a.$$

Let T be the time required by the given planet to complete one orbit. Kepler's third law asserts that T^2 is proportional to a^3. Exercises 14 to 18 establish this by showing that T^2/a^3 is the same for all planets.

14. Using the fact that the area of the ellipse in Figure C.20.4 is πab, show that $Th/2 = \pi ab$, hence that

$$T = \frac{2\pi ab}{h}. \tag{C.20.5}$$

The rest of the argument depends only on (C.20.4) and (C.20.5) and the fixed sum of two distances property of an ellipse.

15. Using (C.20.4), show that f in Figure C.20.4 equals h^2/q.

16. Show that $b^2 = af$:

(a) From $\overline{F'A} + \overline{FA} = 2a$, deduce that $a^2 = b^2 + d^2$.

(b) From $\overline{F'B} + \overline{FB} = 2a$, deduce that $d^2 = a^2 - af$.

(c) From (a) and (b), deduce that $b^2 = af$.

17. From Exercises 15 and 16, deduce that $b^2 = ah^2/q$.

18. Combining (C.20.5) and Exercise 17, show that

$$\frac{T^2}{a^3} = \frac{4\pi^2}{q}.$$

Since $4\pi^2/q$ is a constant, the same for all points, Kepler's third law is established.

References

1. M. Caspar, *Kepler*, Hellman, New York, 1959.

2. J. Kepler, Copernican astronomy, *Great Books of the Western World*, Encyclopedia Britannica, Chicago, 1952.

3. A. Koestler, *The watershed*, Doubleday, New York, 1960.

4. I. Newton, *Sir Isaac Newton's Principles of Natural Philosophy and His System of the World*, translated by Andrew Motte (1729), revised by Florian Cajori. Berkeley: University of California Press, 1947.

5. V. Frederick Rickey, Isaac Newton: man, myth, and mathematician, *College Mathematics Journal*, 18 (1987), 362—388.

6. J. L. Russell, Kepler's laws of planetary motion 1609–1666, *British Journal for the History of Science*, 2 (1964), 1—24.

7. S. Stein, "Exactly how did Newton deal with his planets?", *Mathematical Intelligencer*, 18 (1996), 6—11.

Calculus is Everywhere # 21

The Suspension Bridge and the Hanging Cable

In a suspension bridge the roadway hangs from a cable, as shown in Figure C.21.1. We will use calculus to find the shape of the cable. We assume that the weight of a section of the roadway is proportional to its length. That is, there is a constant k such that x feet of the roadway weighs kx pounds. We will assume that the cable is weightless. That is justified for it weighs little in comparison to the roadway.

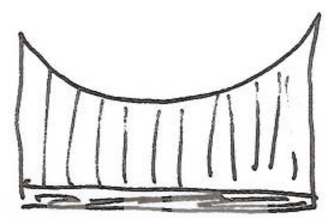
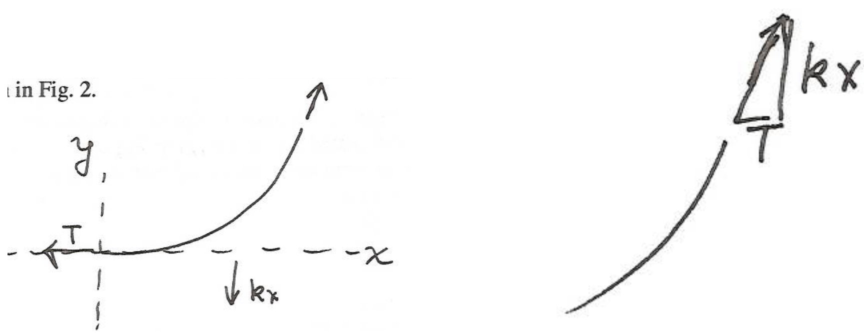


Figure C.21.1

We introduce an xy-coordinate system with origin at the lowest point of the cable, and consider the section of the cable, that goes from $(0,0)$ to (x,y), as shown in Figure C.21.2(a). Three forces act on it. The force at $(0,0)$ is



(a) (b)

Figure C.21.2

horizontal and pulls the cable to the left. Call its magnitude T. Gravity pulls the cable down with a force whose magnitude is kx, the weight of the roadway beneath it. At the top of the section, at (x,y), the cable above it pulls to the right and upward along the tangent line to the cable.

The section does not move. The horizontal part of the force at (x,y) must have magnitude T and the vertical part of the force has magnitude kx, as shown in Figure C.21.2(b).

Since the force at the point (x,y) is directed along the tangent line there, we have

$$\frac{dy}{dx} = \frac{kx}{T}.$$

Therefore

$$y = \frac{kx^2}{T} + C$$

for some constant C. Since $(0,0)$ is on the curve, $C=0$, and the cable has the equation

$$y=\frac{kx^2}{T}.$$

The cable forms a parabola.

What if we have the cable but no roadway? That is also the case for a laundry line, or a telephone wire, or a hanging chain. In these cases the downward force is due to the weight of the cable. If s feet of cable weighs ks pounds, reasoning similar to that for the suspension bridge leads to

$$\frac{dy}{dx}=\frac{ks}{T}.$$

Since

$$s=\int_0^x \sqrt{1+\left(\frac{dy}{dx}\right)^2}\,dx$$

we have

$$\frac{dy}{dx}=\frac{k}{T}\int_0^x \sqrt{1+\left(\frac{dy}{dx}\right)^2}\,dx. \tag{C.21.1}$$

If we differentiate both sides of (C.21.1), and use the second part of the Fundamental Theorem of Calculus, we get

$$\frac{d^2y}{dx^2}=\frac{k}{T}\sqrt{1+\left(\frac{dy}{dx}\right)^2}. \tag{C.21.2}$$

It can be shown that

$$y=\frac{k}{T}\left(e^{\frac{kx}{T}}+e^{\frac{-kx}{T}}\right)-2\frac{k}{T}.$$

The curve is called a **catenary**, after the Latin catena, meaning chain. It may look like a parabola, but it is not. The 630-foot tall Gateway Arch in St. Louis, completed October 28, 1965, is a famous catenary.

EXERCISES

1. Check that the solution to

$$\frac{d^2y}{dx^2}=\frac{k}{T}\sqrt{1+\left(\frac{dy}{dx}\right)^2}$$

that passes through $(0,0)$ is

$$y=\frac{k}{T}\left(e^{\frac{kx}{T}}+e^{\frac{-kx}{T}}\right)-2\frac{k}{T}. \tag{C.21.3}$$

Calculus is Everywhere # 22

The Path of the Rear Wheel of a Scooter

When the front wheel of a scooter follows a certain path, what is the path of its rear wheel? This question could be asked for a bicycle or car, but the scooter is more convenient for carrying out experiments.

The **tractrix problem** is the special case when the front wheel moves in a straight line. (See Exercise 62 in the Section 8.5.) Now, using vectors, we will look at the case when the front wheel sweeps out a circular path.

The Basic Equation

Figure C.22.1 shows the geometry at any instant. Let s denote the arc length of the path swept out by the rear wheel as measured from its starting point. Let a be the length of the wheel base, that is, the distance between the front and rear axles. The vector $\mathbf{r}(s)$ gives the position of the rear wheel and $\mathbf{f}(s)$ gives the position of the front wheel. Because the rear wheel is parallel to $\mathbf{f}(s) - \mathbf{r}(s)$, the unit vector $\mathbf{r}'(s)$ points directly toward the front wheel or directly away from it.

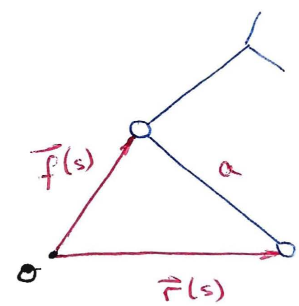


Figure C.22.1

Thus

$$\mathbf{f}(s) = \mathbf{r}(s) + a\mathbf{r}'(s)$$

or

$$\mathbf{f}(s) = \mathbf{r}(s) - a\mathbf{r}'(s).$$

In short, we will write $\mathbf{f}(s) = \mathbf{r}(s) \pm a\mathbf{r}'(s)$.

Assume that the front wheel is moving, say, counterclockwise and tracing out a circular path with center O and radius c. Because

$$\mathbf{f}(s) \cdot \mathbf{f}(s) = c^2$$

we have

$$(\mathbf{r}(s) \pm a\mathbf{r}'(s)) \cdot (\mathbf{r}(s) \pm a\mathbf{r}'(s)) = c^2.$$

By distributivity of the dot product,

$$\mathbf{r}(s) \cdot \mathbf{r}(s) + a^2\mathbf{r}'(s) \cdot \mathbf{r}'(s) \pm 2a\mathbf{r}(s) \cdot \mathbf{r}'(s) = c^2. \qquad \text{(C.22.1)}$$

Letting $r(s) = |\mathbf{r}(s)|$, we may rewrite (C.22.1) as

$$(r(s))^2 + a^2 \pm 2a\mathbf{r}(s) \cdot \mathbf{r}'(s) = c^2. \qquad \text{(C.22.2)}$$

Differentiate $\mathbf{r}(s) \cdot \mathbf{r}(s) = r(s)^2$ to obtain $\mathbf{r}(s) \cdot \mathbf{r}'(s) = r(s)r'(s)$, which changes (C.22.1) to an equation involving the scalar function $r(s)$. If we write $r(s)$ as r and $r'(s)$ as r' we get

$$r^2 + a^2 \pm 2arr' = c^2. \tag{C.22.3}$$

This is the equation we will use to analyze the path of the rear wheel of a scooter.

The Direction of r′

We first find when $\mathbf{r}'$ points towards the front wheel and when it points away from the front wheel.

The movement of the back wheel is determined by the projection of $\mathbf{f}'$ on the line of the scooter, which is the same as $\mathbf{r}'$.

Thus, when the angle θ between the front wheel and the line of the scooter is obtuse, as in Figure C.22.2(a), $\mathbf{r}'$ points towards the front wheel. When θ is acute, the scooter backs up and $\mathbf{r}'$ points away from the front wheel, as shown in Figure C.22.2(b).

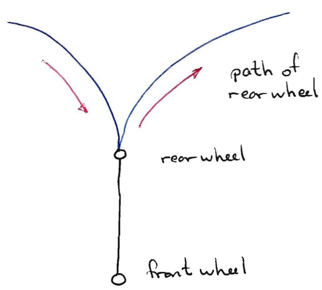
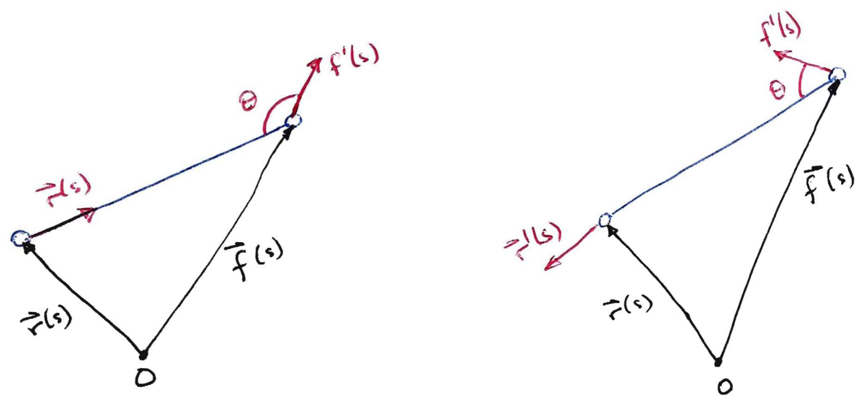


Figure C.22.3



(a) (b)

Figure C.22.2 The direction of $\mathbf{r}'$ depends on the angle θ between the front wheel and the line of the scooter. (a) θ is obtuse, (b) θ is acute.

When the direction of $\mathbf{r}'$ shifts from pointing towards the front wheel to pointing away from it, the path of the rear wheel also changes, as shown in Figure C.22.3.

The path of the rear wheel is continuous but the unit tangent vector $\mathbf{r}'$ is not defined where its direction suddenly shifts. The path is said to contain a cusp and the point at which $\mathbf{r}'(s)$ shifts direction by the angle π is the vertex of the cusp.

The Path of the Rear Wheel for a Short Scooter

Assume the wheelbase a is less than the radius of the circle c, θ is obtuse, and r^2 is less than $c^2 - a^2$. Thus, $c^2 - a^2 - r^2$ is positive. (Exercise 4 shows the significance of $c^2 - a^2$.)

We write $c^2 = a^2 + r^2 + 2rr'a$ in the form

$$\frac{-2rr'}{c^2 - a^2 - r^2} = \frac{-1}{a}. \tag{C.22.4}$$

Integration of both sides of (C.22.4) with respect to arc length s shows that there is a constant k such that

$$\ln(c^2 - a^2 - r^2) = \frac{-s}{a} + k,$$

so

$$c^2 - a^2 - r^2 = e^k e^{-s/a}. \tag{C.22.5}$$

Equation (C.22.5) tells us that r^2 increases but remains less than $c^2 - a^2$, and approaches $c^2 - a^2$ as s increases. Thus the rear wheel traces a spiral path that gets arbitrarily close to the circle of radius $\sqrt{c^2 - a^2}$ and center O, as in Figure C.22.4.

The Path of the Rear Wheel for a Long Scooter

Assume that the wheelbase is longer than the radius of the circle on which the front wheel moves, that is, $a > c$. Assume that initially the scooter is moving forward, so we again have

$$c^2 = a^2 + r^2 + 2rr'a. \tag{C.22.6}$$

The initial position is indicated in Figure C.22.5(a).

Now $c^2 - a^2 - r^2$ is negative, and we have

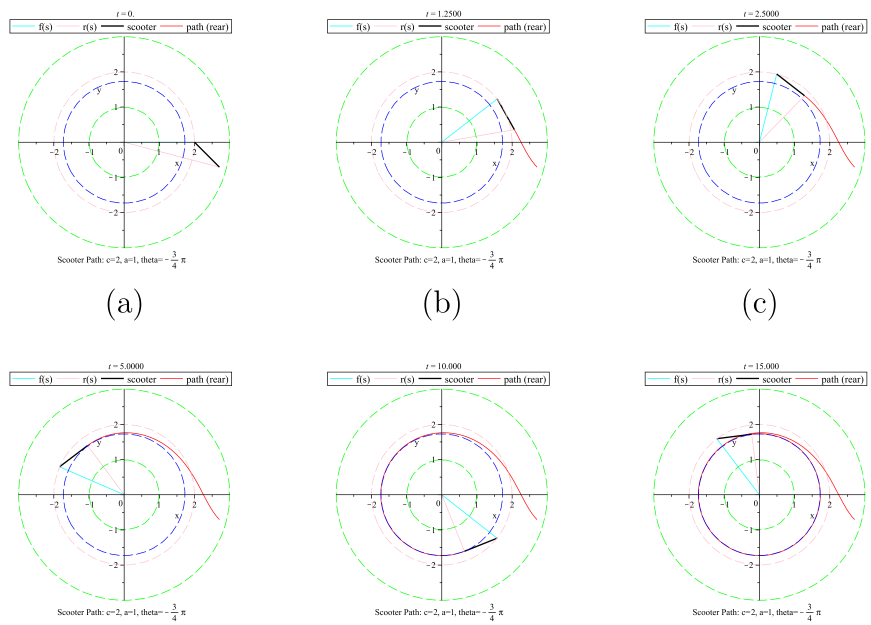
$$\frac{2rr'}{a^2 + r^2 - c^2} = \frac{-1}{a},$$

where the denominator on the left-hand side is positive. Thus there is a constant k such that

$$a^2 + r^2 - c^2 = e^k e^{-s/a}. \tag{C.22.7}$$

As s gets arbitrarily large, (C.22.7) implies that r^2 approaches $c^2 - a^2$. But, $c^2 - a^2$ is negative, so this cannot happen. Our assumption that (C.22.6) holds for all s must be wrong. There must be a cusp and the equation switches to

$$c^2 = a^2 + r^2 - 2arr'.$$



(a) (b) (c)

(d) (e) (f)

Figure C.22.4 The path of the rear wheel of a scooter with length $a = 1$, whose front wheel moves counterclockwise around the circle with radius $c = 2$ from the point $(2, 0)$ with the line of the scooter at an angle $\theta = -3\pi/4$ with the front wheel. The snapshots are taken when (a) $s = 0$, (b) $s = 1.25$, (c) $s = 2.50$, (d) $s = 5.0$, (e) $s = 10.0$, and (f) $s = 15.0$. Because this is a short scooter ($a < c$), the rear wheel approaches the circle with radius $r = \sqrt{c^2 - a^2} = \sqrt{3}$. (Recall that s is the arclength of the rear wheel's path.)

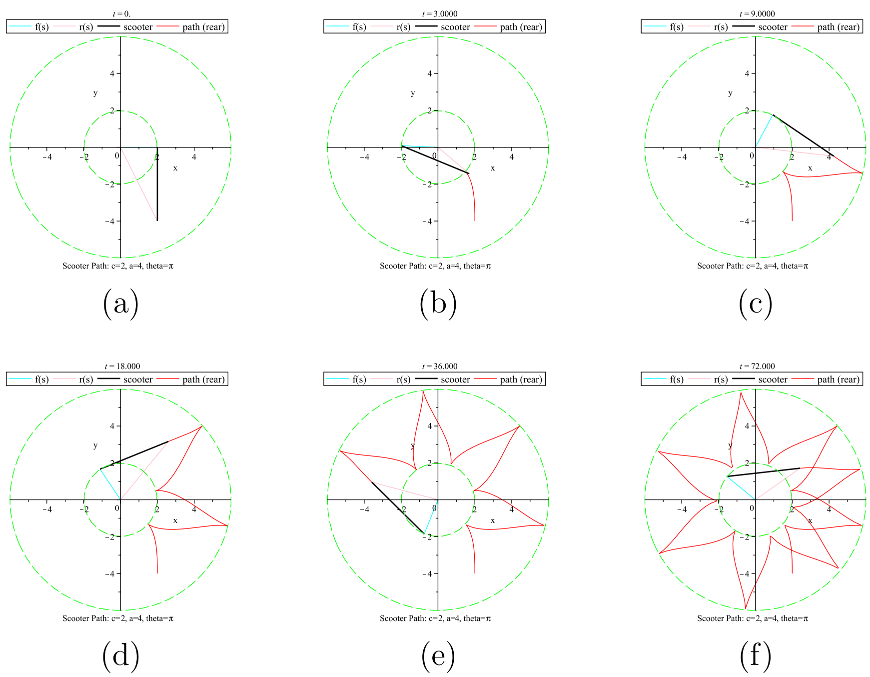


Figure C.22.5 The path of the rear wheel of a scooter with length $a = 4$, whose front wheel moves counterclockwise around the circle with radius $c = 2$ from the point $(2, 0)$ with the line of the scooter at an angle $\theta = \pi$ with the front wheel. The snapshots are taken when (a) $s = 0$, (b) $s = 3$, (c) $s = 9$, (d) $s = 18$, (e) $s = 36$, and (f) $s = 72$. Because the scooter is long ($a > c$), the rear wheel travels along a path that has cusps when $r = c + a$ and $r = |c - a|$. (s is the arclength of the rear wheel's path.)

This leads to

$$a^2 + r^2 - c^2 = e^k e^{s/a},$$

which implies that as s increases r becomes arbitrarily large. However, r can never exceed $c + a$. So, another cusp must form.

It can be shown that the cusps occur when $r = a - c$ (assuming $a > c$) and $r = a + c$. At the vertex of a cusp, $\mathbf{r}'$ is not defined; it changes direction by π.

Figure C.22.5(b) shows the shape of the path of the rear wheel for a long scooter, $a > c$. (For $a > 2c$, that path remains outside the circle.)

EXERCISES

1. When bus or car (or scooter) turns a corner why does a rear tires sometimes go over the curb even though a front tire does not?

2. It is a belief among many bicyclists that the rear tire of a bicycle wears out more slowly than the front tire. Decide whether the belief is justified. (Assume both tires support the same weight.)

3.

(a) Is $\frac{d\mathbf{r}}{ds}$ a unit vector?

(b) Is $\frac{d\mathbf{f}}{ds}$ a unit vector?

4.

(a) Assume a and c are positive with $c > a$ and that the front wheel moves on a circle of radius c. Show that when the front wheel moves along a circle of radius c the rear wheel could remain on a concentric circle of radius $b = \sqrt{c^2 - a^2}$.

(b) Draw the triangle whose sides are a, b, and c and explain why the result in (a) is plausible.

5. We assumed for a short scooter that initially $r^2 < c^2 - a^2$. Examine the case in which initially $r^2 > c^2 - a^2$. Assume that initially the scooter is not backing up.

6. We assumed in the case of the short scooter that initially $r^2 < c^2 - a^2$ and that the scooter is not backing up. Investigate what happens when we assume that initially $r^2 < c^2 - a^2$ and the scooter is backing up.

(a) Draw such an initial position.

(b) Predict what will happen.

(c) Carry out the mathematics.

7. Show that if the path of the front wheel of a scooter is a circle and a cusp forms in the path of the rear wheel, the scooter at that moment lies on a line through the center of the circle.

8. For a long scooter, $a > c$, do cusps always form, whatever the initial value of r and θ?

9. Extend the analysis of the scooter to the case when $a = c$.

10. Assume that the path of the front wheel is a straight line. For convenience, choose that line as the x-axis. The path of the rear wheel is called a **tractrix**. This case appeared in Section 8.6, Exercise 62.
Write $\mathbf{r}(s)$ as $x(s)\mathbf{i} + y(s)\mathbf{j}$.

(a) Show that $y(s) + y'(s)a = 0$.

(b) Deduce that there is a constant k such that $y(s) = ke^{-s/a}$. Thus the distance from the rear wheel to the x-axis decays exponentially as a function of arclength.

11. Show that

$$y(s) = ke^{-s/a},$$

where s is arc length, satisfies

$$\left(\frac{dy}{dx}\right)^2 = \frac{y^2}{a^2 - y^2}.$$

(Differentiate both sides of the equation with respect to x.)

Chapter 16

Partial Derivatives

So far we have been concerned mainly with functions whose domains are part or all of a line or curve. This chapter generalizes the derivative to functions whose domains are part or all of a plane or space, called **functions of two or three variables**. Chapter 17 does the same for definite integrals.

The first seven sections generalize Chapters 1 to 4: picturing functions of two or three variables (Section 16.1), their derivatives (Section 16.2), the chain rule (Section 16.3), more on their derivatives (Section 16.4), the tangent plane to a surface (Section 16.5), and finding extrema (Sections 16.6 and 16.7). In preparation for extending the method of substitution for evaluating definite integrals, the magnification of a function is introduced in Section 16.8. Section 16.9 obtains some fundamental equations in introductory thermodynamics.

16.1 Picturing a Function of Several Variables

The graph of $y = f(x)$, a function of one variable, x, is a curve in the xy-plane. The graph of a function of two variables, $z = f(x, y)$ is a surface in space. It consists of the points (x, y, z) for which $z = f(x, y)$. For instance, if $z = 2x + 3y$, the graph is the plane $z = 2x + 3y$.

This section describes some ways of picturing a scalar-valued functions of two or three variables.

Contour Lines

This is similar to what we did for vector fields.

For $z = f(x, y)$, the simplest method is to attach at (x, y) the value of the function, $z = f(x, y)$. Figure 16.1.1 illustrates this for $z = xy$. It gives a

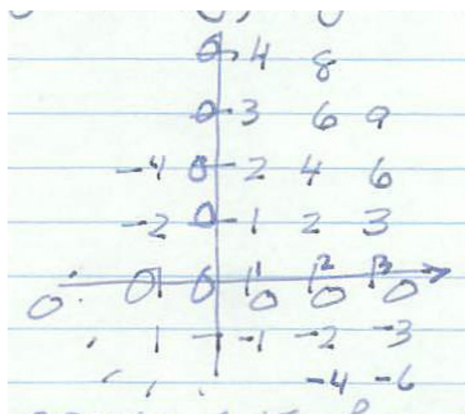


Figure 16.1.1

sense of the function. Its values are positive in the first and third quadrants, negative in the second and fourth. For (x, y) far from the origin near the lines $y = x$ or $y = -x$ the values are large.

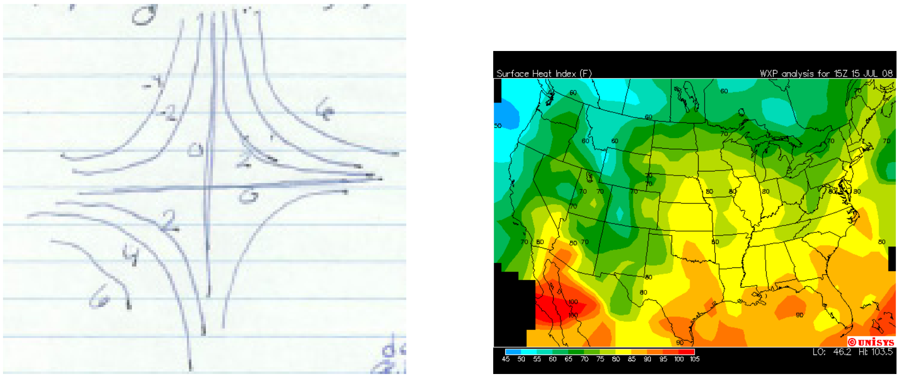
Rather than attach the values at points, we could indicate points where the function has a specific fixed value. We could graph, for a constant k, all the points (x, y) where $f(x, y) = k$, called a **contour** or **level curve**.

contours and level curves

For $z = xy$, contours are hyperbolas $xy = k$. In Figure 16.1.2(a) the contours corresponding to $k = 2, 4, 6, 0, -2, -4, -6$ are shown.

Many newspapers publish a daily map showing temperatures using contour lines. In Figure 16.1.2(b) is an example; in this case the contour lines are the boundary curves between the differently colored regions.

At a glance you can see where it is hot or cold and in what direction to travel to warm up or cool off.



(a) (b)

Figure 16.1.2

Traces

Another way to see the surface $z = f(x, y)$ is to sketch its intersections of various planes with the surface. They are cross sections that are called **traces**.

Figure 16.1.3 shows a trace created by the plane $z = k$, which is parallel to the xy-coordinate plane, namely, the plane $z = k$ The curve is a copy of copy of the contour $f(x, y) = k$.

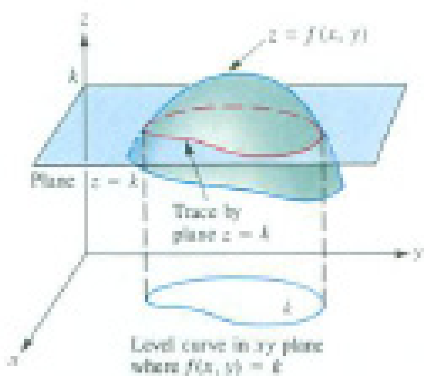


Figure 16.1.3

EXAMPLE 1 Sketch the traces of the surface $z = xy$ with the planes (a) $z = 1$, (b) $x = 1$, (c) $y = x$, (d) $y = -x$, and (e) $x = 0$.
SOLUTION

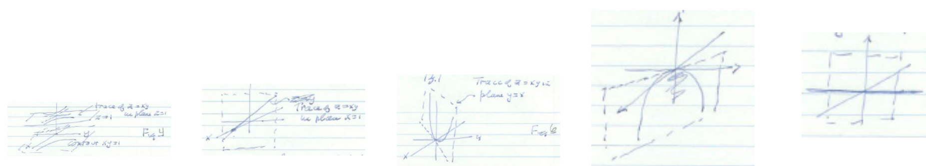
(a) The trace with the plane $z = 1$ is shown in Figure 16.1.4. For points (x, y, z) on it, $xy = 1$, a hyperbola. It is the contour line $xy = 1$ in the xy-plane raised by one unit, as in Figure 16.1.4(a)

(b) The trace in the plane $x = 1$ satisfies the equation $z = 1 \cdot y = y$. It is a straight line, shown in Figure 16.1.4(b)

(c) The trace in the plane $y = x$ satisfies the equation $z = x^2$. It is the parabola shown in Figure 16.1.4(c).

(d) The trace in the plane $y = -x$ satisfies the equation $z = x(-x) = -x^2$. It is an upside-down parabola, shown in Figure 16.1.4(d).

(e) The intersection with the coordinate plane $x = 0$ satisfies the equation $z = 0 \cdot y = 0$. It is the y-axis, shown in Figure 16.1.4(e).



(a) (b) (c) (d) (e)

Figure 16.1.4

The surface can be viewed as made up of lines, of parabolas, or of hyperbolas. It is shown in Figure 16.1.5 with some of the traces drawn on it. ◇

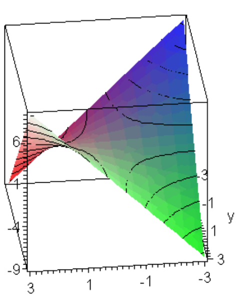
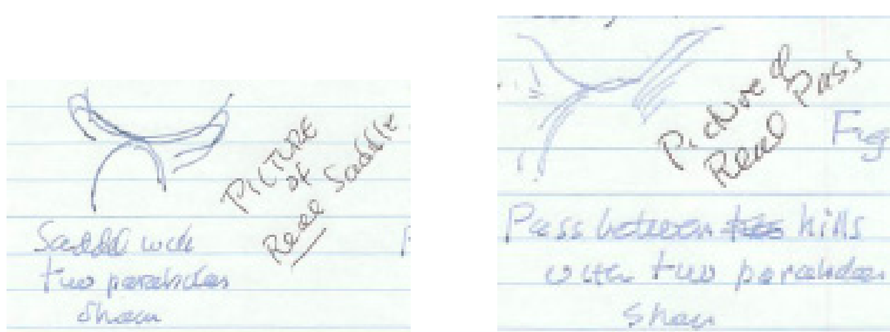


Figure 16.1.5

The surface $z = xy$ looks like a saddle or the pass between two hills, as shown in Figure 16.1.6.



(a) (b)

Figure 16.1.6

Functions of Three Variables

The graph of $y = f(x)$ consists of points in xy-plane. The graph of $z = f(x, y)$ consists of points in the xyz space. What if we have a function of three variables, $u = f(x, y, z)$? (The volume V of a box of sides x, y, z is given by the equation $V = xyz$.) We cannot graph the set of points (x, y, z, u) where $u = f(x, y, z)$ since we cannto draw graphs in four dimensions. What we could do is pick a constant k and draw a level surface, the set of points where $f(x, y, z) = k$. Varying k may give an idea of the function's behavior, just as varying k in $f(x, y) = k$ yields information about the behavior of a function of two variables.

For example, let $T = f(x, y, z)$ be the temperature (Fahrenheit) at the point (x, y, z). Then the level surface

$$68 = f(x, y, z)$$

consists of all points in space where the temperature is 68°.

EXAMPLE 2 Describe the level surfaces of $u = x^2 + y^2 + z^2$.
SOLUTION For each k we examine $k = x^2 + y^2 + z^2$. If k is negative, there are no points on the level surface. If $k = 0$, there is only one, the origin $(0, 0, 0)$. If $k = 1$, the equation is $1 = x^2 + y^2 + z^2$, which describes a sphere of radius 1, with center $(0, 0, 0)$. If k is positive, the level surface $f(x, y, z) = k$ is a sphere of radius $\sqrt{k}$, with center $(0, 0, 0)$. See Figure 16.1.7 ◇

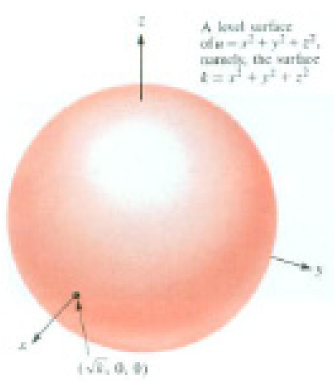


Figure 16.1.7

History of Contours

The use of contour lines goes back to the 1774. Surveyors had collected a large number of the elevations of points on Mount Schiehalli in Scotland so as to estimate its mass and, by its gravitational attraction, the mass of the earth. They asked the mathematician Charles Hutton for help in using the data. Hutton saw that if he connected points on the map that showed the same elevation, the resulting curves, contour lines, suggested the shape of the mountain.

Reference: Bill Bryson, *A Short History of Nearly Everything*, Broadway Books, New York, 2003, p. 57.

Summary

We introduced the idea of a function of two variables $z = f(P)$ defined for points P in a region in the xy-plane. The graph of $z = f(P)$ is usually a surface. It is often more useful to sketch a few of its level curves than to sketch the surface. A level curve in the xy-plane is a copy of a trace by a plane parallel to thet plane. At all points (x, y) on a level curve the function has the same value, so it is constant on a level curve.

We used level curves to analyze the function $z = xy$ whose graph is a saddle.

For functions of three variables $u = f(x, y, z)$, which could also be written as $u = f(P)$. We defined level surfaces $k = f(x, y, z)$ on which f is constant.

EXERCISES for Section 16.1

In Exercises 1 to 10, graph the given function. That is, graph $z = f(x, y)$.

1. $f(x, y) = y$

2. $f(x, y) = x + 1$

3. $f(x, y) = 3$

4. $f(x, y) = -2$

5. $f(x, y) = x^2$

6. $f(x, y) = y^2$

7. $f(x, y) = x + y + 1$

8. $f(x, y) = 2x - y + 1$

9. $f(x, y) = x^2 + 2y^2$

10. $f(x, y) = \sqrt{x^2 + y^2}$

In Exercises 11 to 14 draw the level curves corresponding to the values -1, 0, 1, and 2 if they are not empty.

11. $f(x, y) = x + y$

12. $f(x, y) = x + 2y$

13. $f(x, y) = x^2 + 2y^2$

14. $f(x, y) = x^2 - 2y^2$

In Exercises 15 to 18 draw the level curves that pass through the given points.

15. $f(x, y) = x^2 + y^2$ through $(1, 1)$ (Compute $f(1, 1)$.)

16. $f(x, y) = x^2 + 3y^2$ through $(1, 2)$

17. $f(x, y) = x^2 - y^2$ through $(3, 2)$

18. $f(x, y) = x^2 - y^2$ through $(2, 3)$

19.

(a) Draw the level curves for $f(x, y) = x^2 + y^2$ corresponding to $k = 0, 1, \ldots, 9$.

(b) By inspection of the curves in (a), decide where the function changes most rapidly. Explain why you think so.

20. Sketch three level curves in the xy-plane for the polar functions.

(a) $f(r) = r$

(b) $f(r) = r^2$

(c) $f(r) = e^{-r}$

(d) $f(r) = \ln(r)$

(e) $f(r) = \frac{1}{r}$

(f) $f(r) = \sin(r)$

See also Exercise 24.

21. Let $f(P)$ be the average daily solar radiation at point P, measured in langleys. The level curves corresponding to 350, 400, 450, and 500 langleys are shown in Figure 16.1.8.

Figure 16.1.8

(a) What can be said about the ratio between the maximum and minimum solar radiation at points in the United States?

(b) Why are there sharp bends in the level curves in two areas?

22. Let $u = g(x, y, z)$ be a function of three variables. Describe the level surface $g(x, y, z) = 1$ if $g(x, y, z)$ is

(a) $x + y + z$

(b) $x^2 + y^2 + z^2$

(c) $x^2 + y^2 - z^2$

(d) $x^2 - y^2 - z^2$ ((c) and (d) are examples of quadric surfaces.)

23.

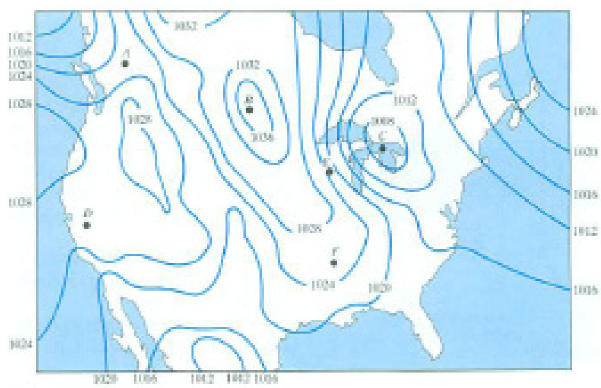


Figure 16.1.9

A weather map, Figure 16.1.9, shows level curves of constant baroemetric pressure (called *isobars*).

(a) Where is the lowest pressure?

(b) Where is the highest pressure?

(c) Where do you think the wind at ground level is the fastest? Why?

24. Compare and contrast the level curves drawn in Exercise 20. How are the similar? How are they different?

25.

(a) Sketch the surface $z = x^2 + y^2$.

(b) Show that traces by planes parallel to the xz plane are parabolas.

(c) Show that the parabolas in (b) are congruent, So the surface is made up of identical parabolas.

(d) What kind of curve is a trace in a plane parallel to the xy plane?

26. For the surface $z = x^2 + 4y^2$, what type of curve is produced by a trace by a plane parallel to

(a) the xy plane?

(b) the xz palne?

(c) the yz plane?

16.2 Limits, Continuity, and Partial Derivatives

Limit, continuity, and derivative carry over with similar definitions from functions of one variable to functions of several variables. But there are differences. A differentiable function $f(x)$ has one first derivative, $f'(x)$ but $f(x,y)$ has many first derivatives.

Limits and Continuity of $f(x,y)$

The **domain** of a function $f(x,y)$ is the set of points where it is defined. The domain of $f(x,y) = x + y$ is the entire xy-plane. The domain of $f(x,y) = \sqrt{1-x^2-y^2}$ is smaller because, for the square root of $1-x^2-y^2$ to be defined, $1-x^2-y^2$ must not be negative. So, $x^2+y^2 \leq 1$. The domain is the disk bounded by the circle $x^2+y^2=1$, shown in Figure 16.2.1.

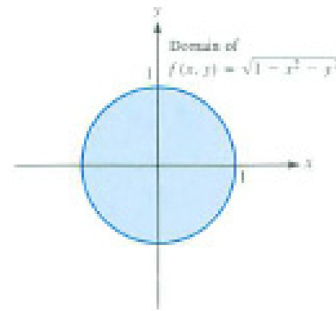


Figure 16.2.1

The domain of $g(x,y) = 1/\sqrt{1-x^2-y^2}$ is even smaller. Now $1-x^2-y^2$ cannot be 0 or negative. The domain of $1/\sqrt{1-x^2-y^2}$ consists of points (x,y) such that $x^2+y^2<1$. It is the disk in Figure 16.2.1 without its boundary.

The function $h(x,y) = 1/(y-x)$ is defined everywhere except on the line $y-x=0$. Its domain is the xy plane from which the line $y=x$ is removed. (See Figure 16.2.2.)

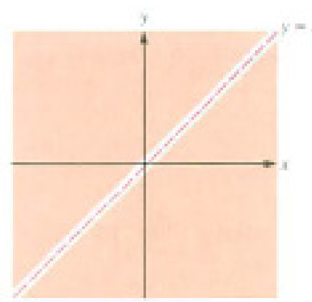
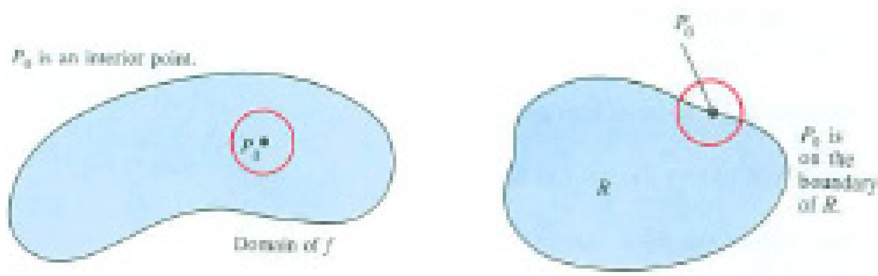


Figure 16.2.2

The domain of functions we look at will either be the entire xy-plane or a region bordered by curves or lines, or perhaps one with a few points omitted. Let P_0 be a point in the domain of a function f. If there is a disk with center



(a) (b)

Figure 16.2.3

P_0 that lies within the domain of f, we call P_0 an **interior** point of the domain. (See Figure 16.2.3(a).) When P_0 is an interior point of the domain of f, we know that $f(P)$ is defined for all points P sufficiently near P_0. A set R is called **open** if each point P of R is an interior point of R. The entire xy-plane is open. Any disk without its circumference is open. The set of points inside a closed curve but not on it forms an open set.

A point P_0 is on the **boundary** of a set if every disk centered at P_0, no matter how small, contains points in the set and points not in the set. (See Figure 16.2.3(b).) The boundary of the disk $x^2+y^2 \leq 1$ is the circle $x^2+y^2 = 1$.

A set R is called **closed** if it includes all of its boundary points. The entire xy-plane is both closed and open. Any disk with its circumference is closed. The set of points inside a closed curve, including the curve, forms a closed set.

The domain of f is a closed set and the domains of g and h are open sets.

The definition of the limit of $f(x, y)$ as (x, y) approaches $P_0 = (a, b)$ should not come as a surprise.

> **DEFINITION** *(Limit of $f(x, y)$ at $P_0 = (a, b)$)* Let f be a function defined at least at every point in some disk with center P_0, except perhaps at P_0. If there is a number L such that $f(P)$ approaches L whenever P approaches P_0 we call L the **limit of** $f(P)$ **as** P **approaches** P_0. We write
>
> $$\lim_{P \to P_0} f(P) = L$$
>
> or
>
> $$f(P) \to L \qquad \text{as} \qquad P \to P_0.$$
>
> We also write
>
> $$\lim_{(x,y)\to(a,b)} f(x, y) = L.$$

For most of our functions the limit will exist throughout its domain.

EXAMPLE 1 Let $f(x, y) = x^3/(x^2+y^2)$. Determine whether $\lim_{P\to(0,0)} f(P)$ exists.

SOLUTION The domain of f is the xy-plane without the origin. When P is near $(0, 0)$ both numerator and denominator approach 0, so we have an indeterminant limit.

Because of the presence of x^2+y^2, we introduce polar coordinates, replacing $x^2 + y^2$ by r^2 and x^3 by $r^3 \cos(\theta)^3$. The quotient now reads

$$\frac{r^3 \cos(\theta)^3}{r^2} = r \cos(\theta)^3.$$

Because $r \cos(\theta)^3$ approaches 0 as $r \to 0$, we conclude that $\lim_{P\to(0,0)} f(P) = 0$ ◇

EXAMPLE 2 Let $f(x, y) = \frac{x^2-y^2}{x^2+y^2}$. Determine whether $\lim_{(x,y)\to(0,0)} f(P)$ exists.

SOLUTION The function is not defined at $(0,0)$. When (x,y) is near $(0,0)$, both the numerator and denominator of $(x^2-y^2)/(x^2+y^2)$ are small numbers. As in Example 1 there are two opposing influences.

We try a few inputs near $(0,0)$. For instance, $(0.01,0)$ is near $(0,0)$ and

$$f(0.01,0)=\frac{(0.01)^2-0^2}{(0.01)^2+0^2}=1.$$

Also, $(0,0.01)$ is near $(0,0)$ and

$$f(0,0.01)=\frac{0^2-(0.01)^2}{0^2+(0.01)^2}=-1.$$

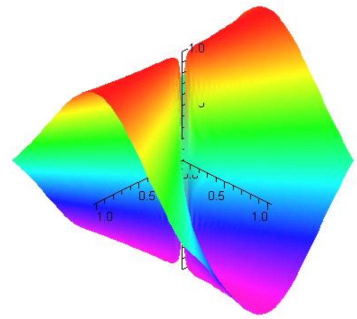


Figure 16.2.4

More generally, for $x\neq 0$,

$$f(x,0)=1$$

while for $y\neq 0$,

$$f(0,y)=-1$$

Since x can be as near 0 as we please and y can be as near 0 as we please, it is not the case that $\lim_{P\to(0,0)} f(P)$ exists. Figure 16.2.4 shows the graph of $z=\frac{x^2-y^2}{x^2+y^2}$. ◇

If P_0 is not an interior point of the domain of f, we modify the definition of limit slightly. Let P_0 be a point on the boundary of the domain of f. If $f(P)\to L$ as P approaches P_0 through points in the domain of f, we say that "L is the limit of $f(P)$ as $P\to P_0$." Example 2 is such a case.

Continuity of $f(x,y)$ at $P_0=(a,b)$

The definition of continuity for $f(x)$ in Section 2.4 generalizes to the definition of continuity for $f(x,y)$.

> **DEFINITION** *(Continuity of $f(x,y)$ at $P_0=(a,b)$).* Assume that $f(P)$ is defined throughout some disk with center P_0. Then f is **continuous** at P_0 if $\lim_{P\to P_0} f(P)=f(P_0)$.
>
> This means
>
> 1. $f(P_0)$ is defined (that is, P_0 is in the domain of f),
> 2. $\lim_{P\to P_0} f(P)$ exists, and
> 3. $\lim_{P\to P_0} f(P)=f(P_0)$.
>
> Continuity at a point on the boundary of the domain can be defined similarly. A function $f(P)$ is **continuous** if it is continuous at every point in its domain.

EXAMPLE 3 Determine whether $f(x,y) = \frac{x^2-y^2}{x^2+y^2}$ is continuous at $(1,1)$.
SOLUTION This is the function in Example 2. First, $f(1,1)$ is defined because it equals 0. Second, $\lim_{(x,y)\to(1,1)} \frac{x^2-y^2}{x^2+y^2}$ exists, because it is $\frac{0}{2} = 0$.) Third, $\lim_{(x,y)\to(1,1)} f(x,y)$ exists and has value $f(1,1)$. Hence, $f(x,y)$ is continuous at $(1,1)$. ◇

In fact, the function of Example 3 is continuous at every point (x,y) in its domain. We do not need to worry about the behavior of $f(x,y)$ when (x,y) is near $(0,0)$ because $(0,0)$ is not in the domain. Since $f(x,y)$ is continuous at every point in its domain, it is a continuous function.

Example 3 also illustrates that while the definition of continuity talks about $f(P)$ being defined throughout a disk, the domain does not have to be a disk.

The Two Partial Derivatives of $f(x,y)$

Let (a,b) be a point in the domain of $f(x,y)$. The trace on the surface $z = f(x,y)$ by a plane through (a,b) and parallel to the z-axis is a curve, as shown in Figure 16.2.5.

If f is well behaved, then, at the point $P = (a,b,f(a,b))$, the trace has a slope. It depends on the plane through (a,b). In this section we consider only the planes parallel to the coordinate planes $y = 0$ and $x = 0$. In the next section we treat the general case.

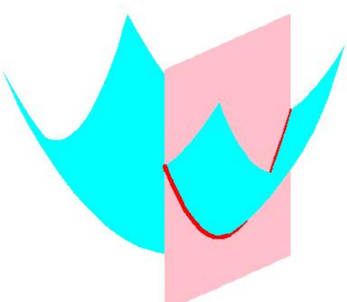


Figure 16.2.5

In $f(x,y) = x^2y^3$, if we hold y constant and differentiate with respect to x, we obtain $d(x^2y^3)/dx = 2xy^3$. This is called the partial derivative of x^2y^3 with respect to x. We could hold x fixed instead and find the derivative of x^2y^3 with respect to y, that is, $d(x^2y^3)/dy = 3x^2y^2$. This derivative is called the partial derivative of x^2y^3 with respect to y. Now define partial derivatives and then we will see what they mean in terms of slope and rate of change.

DEFINITION *(Partial derivatives.)* Assume that the domain of $f(x,y)$ includes the interior of some disk with center (a,b). If

$$\lim_{\Delta x\to 0} \frac{f(a+\Delta x, b) - f(a,b)}{\Delta x}$$

exists, it is called the **partial derivative of f with respect to x at** (a,b). Similarly, if

$$\lim_{\Delta x\to 0} \frac{f(a, b+\Delta y) - f(a,b)}{\Delta y}$$

exists, it is called the **partial derivative of f with respect to y at** (a,b).

There are several notations for the partial derivatives of $z = f(x, y)$ with respect to x:

Notations for partial derivatives.

$$\frac{\partial z}{\partial x}, \frac{\partial f}{\partial x}, f_x, f_1, \text{ or } z_x$$

and with respect to y:

$$\frac{\partial z}{\partial y}, \frac{\partial f}{\partial y}, f_y, f_2, \text{ or } z_y.$$

The symbol $\partial f/\partial x$ may be viewed as the rate at which the function $f(x, y)$ changes when x varies and y is kept fixed and $\partial f/\partial y$ records the rate at which the function $f(x, y)$ changes when y varies and x is kept fixed.

The value of $\partial f/\partial x$ at (a, b) is denoted

$$\frac{\partial f}{\partial x}(a, b) \quad \text{or} \quad \left.\frac{\partial f}{\partial x}\right|_{(a,b)}.$$

In the middle of a sentence, we will write it as $f_x(a, b)$ or $\partial f/\partial x(a, b)$.

Partial derivatives carry information about slopes. The graph of $z = f(x, b)$ is the trace of f in the plane $y = b$. The slope of this trace is given by $f_x(x, b)$. Likewise, $f_y(a, y)$ is the slope of the trace of f in the plane $x = a$.

SHERMAN: This is new. OK?

EXAMPLE 4 If $f(x, y) = \sin(x^2 y)$, find

(a) $\partial f/\partial x$

(b) $\partial f/\partial y$

(c) $\partial f/\partial y$ at $(1, \pi/4)$.

SOLUTION

1. To find $\frac{\partial}{\partial x}(\sin x^2 y)$, differentiate with respect to x, keeping y constant:

$$\begin{aligned} \tfrac{\partial}{\partial x}(\sin x^2 y) &= \cos(x^2 y)\tfrac{\partial}{\partial x}(x^2 y) && \text{(chain rule)} \\ &= \cos(x^2 y)(2xy) && (y \text{ is constant}) \\ &= 2xy\cos(x^2 y). \end{aligned}$$

2. To find $\frac{\partial}{\partial y}(\sin x^2 y)$, differentiate with respect to y, keeping x constant:

$$\begin{aligned} \tfrac{\partial}{\partial y}(\sin x^2 y) &= \cos(x^2 y)\tfrac{\partial}{\partial y}(x^2 y) && \text{(chain rule)} \\ &= \cos(x^2 y)(x^2) && (x \text{ is constant}) \\ &= x^2\cos(x^2 y). \end{aligned}$$

3. By (b)

$$\frac{\partial f}{\partial y}(1,\pi/4) \quad = \quad x^2\cos(x^2y)|_{(1,\pi/4)} = 1^2\cos\left(1^2\tfrac{\pi}{4}\right) = \tfrac{\sqrt{2}}{2}.$$

$\diamond$

As Example 4 shows, since partial derivatives are really ordinary derivatives, the procedures for computing derivatives of a function $f(x)$ of a single variable carry over to functions of two variables.

Higher-Order Partial Derivatives

Just as there are derivatives of derivatives, so are there partial derivatives of partial derivatives. For instance, if

$$z = 2x + 5x^4y^7,$$

we have

$$\frac{\partial z}{\partial x} = 2 + 20x^3y^7 \qquad \text{and} \qquad \frac{\partial z}{\partial y} = 35x^4y^6.$$

Then we may compute partial derivatives of $\partial z/\partial x$ and $\partial z/\partial y$:

$$\begin{array}{rclrcl} \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right) & = & 60x^2y^7 & \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial y}\right) & = & 140x^3y^6 \\ \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right) & = & 140x^3y^6 & \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial y}\right) & = & 210x^4y^5. \end{array}$$

There are four partial derivatives of the second order,

$$\frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right), \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right), \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial y}\right), \text{ and } \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right).$$

These are usually denoted, in the same order, as

$$\frac{\partial^2 z}{\partial x^2} = z_{xx}, \frac{\partial^2 z}{\partial y \partial x} = z_{xy}, \frac{\partial^2 z}{\partial y^2} = z_{yy}, \text{ and } \frac{\partial^2 z}{\partial x \partial y} = z_{yx}.$$

To compute $\partial^2 z/\partial x \partial y$, we first differentiate with respect to y, then with respect to x. To compute $\partial^2 z/\partial y \partial x$, we first differentiate with respect to x, then with respect to y. In both cases, we differentiate from right to left in the order that the variables occur.

The partial derivative $\frac{\partial f}{\partial x}$ is also denoted f_x and $\frac{\partial f}{\partial y}$ is denoted f_y. The second partial derivative $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial (f_y)}{\partial x} = (f_y)_x$ is denoted f_{yx}. In this case we differentiate from left to right, first f_y, then $(f_y)_x$. That is, $f_{yx} = (f_y)_x$, $f_{yy} = (f_y)_y$, and $f_{xy} = (f_x)_y$. In both notations the mixed partial is computed

The subscript notation, f_{yx}, is generally preferred in the midst of text.

in the order that resembles its definition with the parentheses removed. That is,

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) \qquad \text{and} \qquad f_{xy} = (f_x)_y$$

are the two mixed second partial derivatives of f.

In the computations just done, the two mixed partials z_{xy} and z_{yx} are equal. This is not a coincidence. For functions commonly encountered, the two mixed partials are equal.

Equality of the mixed partials

Exercise 54 has a function for which the two mixed particles are not equal.

EXAMPLE 5 Compute $\frac{\partial^2 z}{\partial x^2} = f_{xx}$, $\frac{\partial^2 z}{\partial y \partial x} = f_{xy}$, $\frac{\partial^2 z}{\partial x \partial y} = z_{yx}$, and $\frac{\partial^2 z}{\partial y^2} = z_{yy}$ for $z = y\cos(xy)$.

SOLUTION The first partial derivatives are

$$\frac{\partial z}{\partial x} = y\left(-\sin(xy)\frac{\partial}{\partial x}(xy)\right) = y(-\sin(xy)y) = -y^2\sin(xy)$$

and, using the chain rule,

$$\frac{\partial z}{\partial y} = \frac{\partial}{\partial y}(y)\cdot\cos(xy) + y\cdot\frac{\partial}{\partial y}\cos(xy) = \cos(xy) + y(-\sin(xy)x) = \cos(xy) - xy\sin(xy).$$

Now we can compute the four second derivatives.

$$\begin{aligned}
\frac{\partial^2 z}{\partial x^2} &= \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right) = \frac{\partial}{\partial x}(-y^2\sin(xy)) = -y^3\cos(xy),\\
\frac{\partial^2 z}{\partial y \partial x} &= \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right) = \frac{\partial}{\partial y}(-y^2\cos(xy)) = -2y\sin(xy) - xy^2\cos(xy),\\
\frac{\partial^2 z}{\partial x \partial y} &= \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right) = \frac{\partial}{\partial x}(-yx\sin(xy) + \cos(xy))\\
&= -y\frac{\partial}{\partial x}(x\sin(xy) + \frac{\partial}{\partial x}(\cos(xy)) = -y(xy\cos(xy) + \sin(xy)) - y\sin(xy)\\
&= -xy^2\cos(xy) - y\sin(xy) - y\sin(xy) = -2y\sin(xy) - xy^2\cos(xy),
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 z}{\partial y^2} &= \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial y}\right) = \frac{\partial}{\partial y}(-yx\sin(xy) + \cos(xy))\\
&= (-x\sin(xy) - yx(x\cos(xy))) + (-x\sin(xy))\\
&= -x\sin(xy) - xy^2\cos(xy) - x\sin(xy) = -2x\sin(xy) - x^2y\cos(xy).
\end{aligned}$$

While the computations of the two mixed partials are different, they are equal. ◇

In view of the importance of $f_{xy} = f_{yx}$, we state it as a theorem.

Theorem 16.2.1. *Equality of mixed partial derivatives Assume that $f(x, y)$ is defined in some disk centered at (a, b). If f_x and f_y exist in the disk and f_{xy} is continuous at (a, b), then $f_{yx}(a, b)$ exists and equals $f_{xy}(a, b)$.*

This is not obvious. Why should the rate at which the slope in the y-direction changes with respect to x be the same as the rate at which the slope in the x-direction changes with respect to y? A proof is outlined in Exercise 21 in the Chapter 16 Summary.

Differentiating Under the Integral Sign

Let $g(y) = \int_a^b f(x, y)\, dx$. The following theorem expresses the derivative of g in terms of a partial derivative of f. It gives conditions so we can differentiate under the integral sign.

Theorem 16.2.2. *Assume that $f(x, y)$ is defined in the square whose vertices are (a, a), (a, b), (b, b), and (b, a). Assume also that $\partial f/\partial y$ is continuous there. Then*

$$g'(y) = \int_a^b \frac{\partial f}{\partial y}\, dx.$$

EXAMPLE 6 Verify Theorem 16.2.2 when $g(y) = \int_1^2 e^{xy}\, dx$.
SOLUTION Evaluate $\int_1^2 \frac{\partial}{\partial y}\left(e^{xy}\right)\, dx$. We find

$$\begin{aligned} \int_1^2 \frac{\partial}{\partial y}\, dx &= \int_1^2 x e^{xy}\, dx = \left(\frac{x e^{xy}}{y} - \frac{e^{xy}}{y^2}\right)\Bigg|_{x=1}^{x=2} \\ &= \left(\frac{2e^{2y}}{y} - \frac{e^{2y}}{y}\right) - \left(\frac{e^y}{y} - \frac{e^y}{y^2}\right). \end{aligned} \tag{16.2.1}$$

On the other hand

$$g(y) = \int_1^2 e^{xy}\, dy = \frac{e^{xy}}{y}\Bigg|_{x=1}^{x=2} = \frac{e^{2y}}{y} - \frac{e^y}{y}.$$

Then

$$\begin{aligned} g'(y) &= \frac{2ye^{2y} - e^{2y}}{y^2} - \frac{ye^y - e^y}{y^2} \\ &= \frac{2e^{2y} - e^y}{y}. \end{aligned} \quad (16.2.2)$$

The final expressions in (16.2.1) and (16.2.2) are equal. ◇

Functions of More Than Two Variables

A quantity may depend on more than two variables. For instance, the chill factor depends on temperature, humidity, and wind velocity. The temperature T at a point in the atmosphere is a function of the three space coordinates, x, y, and z: $T = f(x, y, z)$.

The definitions and notations of partial derivatives carry over to functions of more than two variables. If $u = f(x, y, z, t)$, there are four first-order partial derivatives. The partial derivative of u with respect to x, holding y, z, and t fixed, is denoted

To differentiate, hold all variables constant except one.

$$\frac{\partial u}{\partial x}, \ \frac{\partial f}{\partial x}, \ \text{or} \ \ u_x.$$

Higher-ordered partial derivatives are defined and denoted similarly. Section 16.9 and the CIE on the Wave in a Rope at the end of Chapter 17 (page 1435) illustrate their use in physics.

Summary

We defined limits, continuity, and partial derivatives for functions of several variables. These are closely related to the one-variable versions.

A difference is that a partial derivative with respect to one variable, say x, is found by treating all other variables as constants and applying the standard differentiation rules with respect to x. Higher-order partial derivatives are defined much like higher-order derivatives. An important property of higher-order partial derivatives of functions usually met in applications is that the order in which the partial derivatives are calculated does not affect the results. For instance, $\dfrac{\partial}{\partial y}\left(\dfrac{\partial f}{\partial x}\right) = \dfrac{\partial}{\partial x}\left(\dfrac{\partial f}{\partial y}\right)$.

EXERCISES for Section 16.2

In Exercises 1 to 14 evaluate the limits if they exist.

1. $\lim_{(x,y)\to(2,3)} \frac{x+y}{x^2+y^2}$

2. $\lim_{(x,y)\to(1,1)} \frac{x^2}{x^2+y^2}$

3. $\lim_{(x,y)\to(0,0)} \frac{x^2}{x^2+y^2}$

4. $\lim_{(x,y)\to(0,0)} \frac{xy}{x^2+y^2}$

5. $\lim_{(x,y)\to(2,3)} x^y$

6. $\lim_{(x,y)\to(0,0)} (x^2)^y$

7. $\lim_{(x,y)\to(1,0)} \frac{e^{2xy}}{xy}$

8. $\lim_{(x,y)\to(1,2)} \frac{e^{2xy}}{xy}$

9. $\lim_{(x,y)\to(0,0)} \frac{\sin(x+y)}{x+x^2}$

10. $\lim_{(x,y)\to(0,0)} \frac{\sin(2x)}{\sin(3y)}$

11. $\lim_{(x,y)\to(2,2)} \frac{y^3-8}{x^2-4}$

12. $\lim_{(x,y)\to(3,2)} \frac{e^{2x}-e^{3y}}{4x^2-9y^2}$

13. $\lim_{(x,y)\to(0,0)} (1+xy)^{1/(xy)}$

14. $\lim_{(x,y)\to(0,0)} (1+x)^{1/y}$

In Exercises 15 to 22, (a) describe the domain of the function and (b) determine if it is continuous on its domain.

15. $f(x,y) = 1/(x+y)$

16. $f(x,y) = 1/(x^2+2y^2)$

17. $f(x,y) = 1/(9-x^2-y^2)$

18. $f(x,y) = 1/(9-x^2-y^2)^{1/3}$

19. $f(x,y) = \ln(x+2y)$

20. $f(x,y) = \ln(4-x^2+y^2)$

21. $f(x,y) = 1/\sqrt{x^2+y^2-25}$

22. $f(x,y) = \sqrt{49 - x^2 - y^2}$

In Exercises 23 to 28, R consists of all points (x, y) that satisfy the condition. Find the boundary of R.

23. $x^2 + y^2 \leq 1$

24. $x^2 + y^2 < 1$

25. $1/(x^2 + y^2)$ is defined

26. $1/(x + y)$ is defined

27. $y < x^2$

28. $y \leq x$

In Exercises 29 to 34 compute $\dfrac{\partial^2 f}{\partial y \partial x} = \dfrac{\partial}{\partial y}\left(\dfrac{\partial f}{\partial x}\right)$ and $\dfrac{\partial^2 f}{\partial x \partial y} = \dfrac{\partial}{\partial x}\left(\dfrac{\partial f}{\partial y}\right)$. (They should be equal.)

29. $f(x,y) = e^{3x^2 y}$

30. $f(x,y) = \frac{\sin(x+2y)}{x}$ (for $x \neq 0$)

31. $f(x,y) = \ln(2x + 3y)$ (for $2x + 3y > 0$)

32. $f(x,y) = \arctan(\sqrt{xy^3})$ (for $x \geq 0$, $y \geq 0$)

33. $f(x,y) = y/x$

34. $f(x,y) = \sqrt{x^2 + 3y^2}$

In Exercises 35 to 42 compute $\dfrac{\partial^2 f}{\partial x^2}$ and $\dfrac{\partial^2 f}{\partial y^2}$.

35. $f(x,y) = \ln(\sin^2(x)\cos^3(xy))$

36. $f(x,y) = \exp(x^3)$

37. $f(x,y) = \tan(3x^2 y^3)$

38. $f(x,y) = x^3/y^2$

39. $f(x,y) = 3x^2 y^3$

40. $f(x,y) = \arctan(y/x)$

41. $f(x,y) = e^{x^2+y^2}$

42. $f(x,y) = \ln(y^2 + y^4)$

43. Let $T(x,y,z) = 1/\sqrt{x^2 + y^2 + z^2}$, if (x, y, z) is not the origin $(0, 0, 0)$. Show that

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} = 0.$$

This equation arises in the theory of heat, as we will see in Section 16.4.

44. Solve Example 3 by using polar coordinates to express the function.

Check that differentiating under the integral sign gives correct results for the functions in Exercises 45 to 48.

45. $g(y) = \int_a^b x^m y^n \, dx$, $(m,\, n > 1)$

46. $g(y) = \int_a^b \sin(xy) \, dx$

47. $g(y) = \int_a^b x^y \, dx$, $(a,\, b,\, y > 1)$

48. $g(y) = \int_a^b \frac{dx}{xy}$

49. View $\int_a^b f(x,y) \, dx$ as a function of a, b, and y, say $g(a,b,y)$. Find

(a) $\partial g/\partial b$

(b) $\partial g/\partial a$

(c) $\partial g/\partial y$

50. Let $u = f(x,y)$. Assume that $u(1,2)$ is 3, $\partial u/\partial x$ at $(1,2)$ is 2, and $\partial u/\partial y$ at $(1,2)$ is 1.2.

(a) Estimate $u(1, 2.01)$.

(b) Estimate $u(0.98, 2)$.

(c) Estimate $u(1.02, 2.03)$.

Describe your reasoning.

51. Develop a convincing, but not necessarily rigorous, argument justifying differentiation under the integral sign. (Start with the definition of $g'(y)$.)

52. Assume that the domain of f is the xy-plane and $\frac{\partial f}{\partial x} = 0$ everywhere.

(a) Give an example of a non-constant function for which $\frac{\partial f}{\partial x} = 0$.

(b) What is the most general function for which $\frac{\partial f}{\partial x} = 0$?

53. Find all functions f defined throughout the xy-plane for which $\frac{\partial f}{\partial x} = 0$ and $\frac{\partial f}{\partial y} = 0$. Explain.

54. This exercise has a function $f(x,y)$ whose mixed partial derivatives at $(0,0)$ are not equal.

(a) Let $g(x,y) = \frac{x^2-y^2}{x^2+y^2}$ for (x,y) not $(0,0)$. Show that $\lim\limits_{k\to 0}(\lim\limits_{h\to 0} g(h,k)) = -1$ but $\lim\limits_{h\to 0}(\lim\limits_{k\to 0} g(h,k)) = 1$.

(b) Let $f(x, y) = xyg(x, y)$ for (x, y) not $(0, 0)$ and $f(0, 0) = 0$. Show that $f(x, y) = 0$ if x or y is 0 everywhere.

(c) Show that $f_{xy}(0, 0) = \lim_{k\to 0} \frac{f_x(0,k)-f_x(0,0)}{k}$.

(d) Show that $f_{xy}(0, 0) = \lim_{k\to 0} \left(\lim_{h\to 0} \frac{f(h,k)-f(0,k)-f(h,0)+f(0,0)}{hk}\right)$.

(e) Show that $f_{xy}(0, 0) = -1$.

(f) Similarly, show that $f_{yx}(0, 0) = 1$.

(g) Show that in polar coordinates the value of f at (r, θ) is $r^2 \sin(4\theta)/4$.

16.3 Change and the Chain Rule

For a function of one variable, $f(x)$, the change in its value as the input changes from a to $a + \Delta x$ is approximately $f'(a)\Delta x$. In this section we estimate the change in $f(x, y)$ as (x, y) moves from (a, b) to $(a + \Delta x, b + \Delta y)$.

It is used to obtain the chain rule for functions of several variables.

Estimating the Change Δf

Let $z = f(x, y)$ be a function of two variables with continuous partial derivatives at least throughout a disk centered at the point (a, b). We will express $\Delta f = f(a+\Delta x, b+\Delta y) - f(a, b)$ in terms of f_x and f_y. The change is shown in Figure 16.3.1. We can view it as obtained in two steps. First, there is the

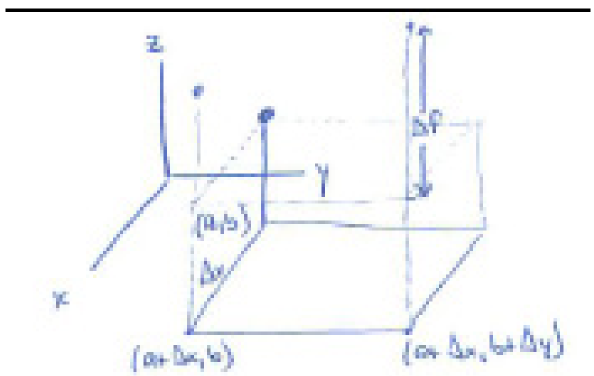


Figure 16.3.1

change as x goes from a to $a + \Delta x$, that is, $f(a + \Delta x, b) - f(a, b)$. Second, there is the change from $f(a+\Delta x, b)$ to $f(a+\Delta x, b+\Delta y)$, as y changes from b to $b + \Delta y$.

That is,

$$\Delta f = (f(a + \Delta x, b) - f(a, b)) + (f(a + \Delta x, b + \Delta y) - f(a + \Delta x, b)). \tag{16.3.1}$$

By the Mean-Value Theorem, there is a number c_1 between a and $a+\Delta x$ such that

$$f(a + \Delta x, b) - f(a, b) = \frac{\partial f}{\partial x}(c_1, b)\Delta x \tag{16.3.2}$$

Applying the Mmean-Value Theorem to the second bracketed expression in (16.3.1), we see that there is a number c_2 between b and $b + \Delta y$ such that

$$f(a + \Delta x, b + \Delta y) - f(a + \Delta x, b) = \frac{\partial f}{\partial y}(a + \Delta x, c_2)\Delta y. \tag{16.3.3}$$

. Combining (16.3.1), (16.3.2), and (16.3.3) we obtain

The Greek letter ϵ, pronounced ep-s$_{ə}$-l$_{ə}$n, corresponds to the Latin letter "e"

$$\Delta f = \frac{\partial f}{\partial x}(c_1, b)\Delta x + \frac{\partial f}{\partial y}(a + \Delta x, c_2)\Delta y. \qquad (16.3.4)$$

When both Δx and Δy are small, (c_1, b) and $(a + \Delta x, c_2)$ are near (a, b). If we assume that the partial derivatives f_x are continuous at (a, b), then

$$\frac{\partial f}{\partial x}(c_1, b) = \frac{\partial f}{\partial x}(a, b) + \epsilon_1 \quad \text{and} \quad \frac{\partial f}{\partial y}(a + \Delta x, c_2) = \frac{\partial f}{\partial y}(a, b) + \epsilon_2, \quad (16.3.5)$$

where ϵ_1 and ϵ_2 approach 0 as Δx and Δy approach 0.

Combining (16.3.4) and (16.3.5) gives the key to estimating the change in the function f. We state this important result as a theorem.

Theorem 16.3.1. *Let f have continuous partial derivatives f_x and f_y for all points within some disk with center at the point (a, b). Then Δf, the change $f(a + \Delta x, b + \Delta y) - f(a, b)$, can be written*

$$\Delta f = \frac{\partial f}{\partial x}(a, b)\Delta x + \frac{\partial f}{\partial y}(a, b)\Delta y + \epsilon_1 \Delta x + \epsilon_2 \Delta y, \qquad (16.3.6)$$

where ϵ_1 and ϵ_2 approach 0 as Δx and Δy approach 0. (Both ϵ_1 and ϵ_2 are functions of a, b, Δx, and Δy.)

Equation (16.3.6) is the core of this section. The term $f_x(a, b)\Delta x$ estimates the change due to the change in the x-coordinate and $f_y(a, b)\Delta y$ estimates the change due to the change in the y-coordinate.

We call $f(x, y)$ **differentiable** at (a, b) if (16.3.6) holds. If f_x and f_y exist in a disk around (a, b) and are continuous at (a, b), then f is differentiable at (a, b).

Since ϵ_1 and ϵ_2 in (16.3.6) approach 0 as Δx and Δy approach 0,

$$\Delta f \approx \frac{\partial f}{\partial x}(a, b)\Delta x + \frac{\partial f}{\partial y}(a, b)\Delta y, \qquad (16.3.7)$$

The approximation (16.3.7) gives us a way to estimate Δf when Δx and Δy are small.

EXAMPLE 1 Estimate $(2.1)^2(0.95)^3$.

SOLUTION Let $f(x, y) = x^2y^3$. We wish to estimate $f(2.1, 0.95)$. We know that $f(2,1)$ equals $2^2 1^3 = 4$. We use (16.3.7) to estimate $\Delta f = f(2.1, 0.95) - f(2,1)$. We have

$$\frac{\partial(x^2y^3)}{\partial x} = 2xy^3 \quad \text{and} \quad \frac{\partial(x^2y^3)}{\partial y} = 3x^2y^2.$$

Then

$$\frac{\partial f}{\partial x}(2,1) = 4 \quad \text{and} \quad \frac{\partial f}{\partial y}(2,1) = 12.$$

Since $\Delta x = 0.1$ and $\Delta y = -0.05$, we have

$$\Delta f \approx 4(0.1) + 12(-0.05) = 0.4 - 0.6 = -0.2.$$

The exact value is 3.78102375.

Thus $(2,1)^2(0.95)^3$ is approximately $4 + (-0.2) = 3.8$. ◇

The Chain Rule

We begin with two special cases of the chain rule for functions of more than one variable. Afterward we will state the chain rule for functions of any number of variables.

The first theorem considers $z = f(x,y)$ where x and y are functions of t. The second theorem is more general, where x and y are functions of two variables, t and u.

Theorem 16.3.2. *(Chain Rule, Special Case I) Let $z = f(x,y)$ have continuous partial derivatives f_x and f_y, and let $x = x(t)$ and $y = y(t)$ be differentiable functions of t. Then z is a differentiable function of t and*

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}. \tag{16.3.8}$$

Proof

A change Δt causes changes Δx and Δy, which cause a change Δz in z.

By definition,

$$\frac{dz}{dt} = \lim_{\Delta t \to 0} \frac{\Delta z}{\Delta t}.$$

According to Theorem 16.3.1,

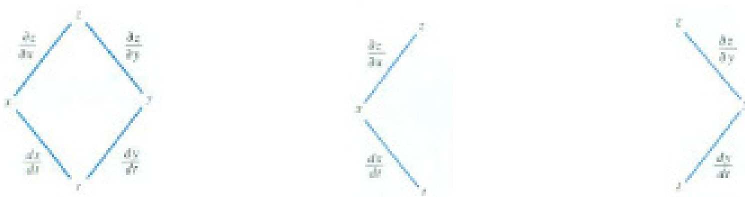
$$\Delta z = \frac{\partial f}{\partial x}(x,y)\Delta x + \frac{\partial f}{\partial y}(x,y)\Delta y + \epsilon_1 \Delta x + \epsilon_2 \Delta y,$$

where $\epsilon_1 \to 0$ and $\epsilon_2 \to 0$ as Δx and Δy approach 0. (x and y are fixed.) Thus

$$\frac{\Delta z}{\Delta t} = \frac{\partial f}{\partial x}(x,y)\frac{\Delta x}{\Delta t} + \frac{\partial f}{\partial y}(x,y)\frac{\Delta y}{\Delta t} + \epsilon_1\frac{\Delta x}{\Delta t} + \epsilon_2\frac{\Delta y}{\Delta t}.$$

and
$$\frac{dz}{dt} = \lim_{\Delta t \to 0} \frac{\Delta z}{\Delta t} = \frac{\partial f}{\partial x}(x,y)\frac{dx}{dt} + \frac{\partial f}{\partial y}(x,y)\frac{dy}{dt} + 0\frac{dx}{dt} + 0\frac{dy}{dt}.$$
This proves the theorem. •

The two summands on the right-hand side of (16.3.8) remind us of the chain rule for functions of one variable. Why is there a "+" in (16.3.8)? It first appears in (16.3.4) and you can trace it back to Figure 16.3.1.



(a) (b) (c)

Figure 16.3.2

The diagram in Figure 16.3.2(a) helps in using this case of the chain rule . There are two paths from the top variable z down to the bottom variable t. Label each edge with a partial derivative (or derivative). For each path there is a summand in the chain rule. The left-hand path (see Figure 16.3.2(b)) gives us the summand
$$\frac{\partial z}{\partial x}\frac{dx}{dt}.$$
The right-hand path (see Figure 16.3.2(c)) gives us the summand
$$\frac{\partial z}{\partial y}\frac{dy}{dt}.$$
Then dz/dt is their sum.

The total number of terms in a partial derivative is the number of paths from top to bottom. Each path from top to bottom produces one summand, the product of the factors that appear in the path. This simple fact provides a quick way to check if any terms have been omitted from a partial derivative.

SHERMAN: Converted margin note to this paragraph. OK?

EXAMPLE 2 Let $z = x^2y^3$, $x = 3t^2$, and $y = t/3$. Find dz/dt when $t = 1$.
SOLUTION To apply Theorem 16.3.2 compute z_x, z_y, dx/dt, and dy/dt:
$$\frac{\partial z}{\partial x} = 2xy^3, \quad \frac{\partial z}{\partial y} = 3x^2y^2, \quad \frac{dx}{dt} = 6t, \text{ and } \frac{dy}{dt} = \frac{1}{3}.$$

By Theorem 16.3.2,

$$\frac{dz}{dt} = 2xy^3 \cdot 6t + 3x^2y^2 \cdot \frac{1}{3}.$$

When $t = 1$, x is 3 and y is $\frac{1}{3}$, so

$$\frac{dz}{dt} = 2 \cdot 3 \left(\frac{1}{3}\right)^3 6 \cdot 1 + 3 \cdot 3^2 \left(\frac{1}{3}\right)^2 \frac{1}{3} = \frac{36}{27} + \frac{27}{27} = \frac{7}{3}.$$

◇

In Example 2, dz/dt can be found without using the theorem by writing z in terms of t:

$$z = x^2y^3 = (3t^2)^2 \left(\frac{t}{3}\right)^3 = \frac{t^7}{3}.$$

Then

$$\frac{dz}{dt} = \frac{7t^6}{3}.$$

When $t = 1$, this gives

$$\frac{dz}{dt} = \frac{7}{3},$$

as before.

EXAMPLE 3 The temperature at the point (x, y) on a window is $T(x, y)$. A bug wandering on the window is at the point $(x(t), y(t))$ at time t. How fast does the bug observe that the temperature of the glass changes as he crawls about?

SOLUTION The bug is asking us to find dT/dt. The chain rule (16.3.8) tells us that

$$\frac{dT}{dt} = \frac{\partial T}{\partial x}\frac{dx}{dt} + \frac{\partial T}{\partial y}\frac{dy}{dt}.$$

The bug can influence this rate by crawling faster or slower, changing $\frac{dx}{dt}$ and $\frac{dy}{dt}$. He may want to know the direction he should choose in order to cool off or warm up as quickly as possible. We will be able to tell him how to do this in the next section. ◇

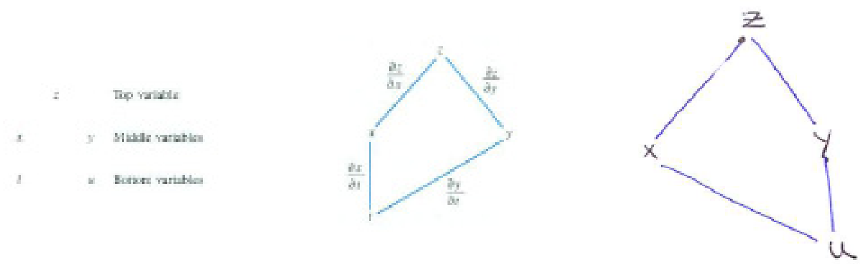
The proof of the next chain rule is almost identical to the proof of Theorem 16.3.2. (See Exercise 24.)

Theorem 16.3.3. *(Chain Rule, Special Case II) Let $z = f(x, y)$ have continuous partial derivatives, f_x and f_y. Let $x = x(t, u)$ and $y = y(t, u)$ have continuous partial derivatives*

$$\frac{\partial x}{\partial t}, \qquad \frac{\partial x}{\partial u}, \qquad \frac{\partial y}{\partial t}, \qquad \frac{\partial y}{\partial u}.$$

Then
$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t} \qquad \text{and} \qquad \frac{\partial z}{\partial u} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial u}.$$

The variables are listed in Figure 16.3.3(a).



(a) (b) (c)

Figure 16.3.3

To find z_t, draw paths from z down to t. Label their edges by the appropriate partial derivative, as shown in Figure 16.3.3(b). Finding z_u is shown in Figure 16.3.3(c).

Each path from the top variable to the bottom variable contributes a summand in the chain rule. The only difference between Figure 16.3.2 and Figure 16.3.3(b) is that ordinary derivatives dx/dt and dy/dt appear in Figure 16.3.2, while partial derivatives x_t and y_t appear in Figure 16.3.3(b).

In Theorem 16.3.2 there are two middle variables and one bottom variable. In Theorem 16.3.3 there are two middle variables and two bottom variables. The chain rule holds for any number of middle variables and any number of bottom variables. There may be three middle variables and, say, four bottom variables. Then there are three summands for each of four partial derivatives.

In the next example there is one middle variable and two bottom variables.

EXAMPLE 4 Let $z = f(u)$ be a function of a single variable. Let $u = 2x + 3y$. Then z is a composite function of x and y. Show that

$$2\frac{\partial z}{\partial y} = 3\frac{\partial z}{\partial x}. \tag{16.3.9}$$

SOLUTION We will evaluate z_x and z_y by the chain rule and then show that (16.3.9) is true.

To find z_x we consider paths from z down to x. (See Figure 16.3.4.) There is only one middle variable so there is only one path. Since $u = 2x+3y$, $u_x = 2$ and

$$\frac{\partial z}{\partial x} = \frac{dz}{du}\frac{\partial u}{\partial x} = \frac{dz}{du} \cdot 2 = 2\frac{dz}{du}$$

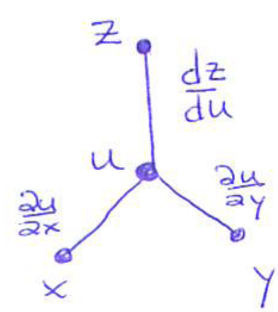


Figure 16.3.4

(One derivative is ordinary, while the other is partial.)

Next we find z_y. There is only one summand. Since $u = 2x + 3y$, $u_y = 3$. Thus

$$\frac{\partial z}{\partial y} = \frac{dz}{du}\frac{\partial u}{\partial y} = \frac{dz}{du} \cdot 3 = 3\frac{dz}{du}.$$

Thus $z_x = 2dz/du$ and $z_y = 3dz/du$. Substitute these into

$$2\frac{\partial z}{\partial y} = 3\frac{\partial z}{\partial x}$$

andr we have a true equation:

$$2\left(3\frac{dz}{du}\right) = 3\left(2\frac{dz}{du}\right). \tag{16.3.10}$$

Since (16.3.10) is true, we have verified (16.3.9). ◇

An Important Use of the Chain Rule

There is a difference between Example 2 and Example 4. In the first, we were dealing with explicitly given functions. We did not need to use the chain rule to find the derivative, dz/dt. We could have shown that $z = t^7/3$ and easily found that $dz/dt = 7t^6/3$. In Example 4, we were dealing with a general type of function formed in a certain way: We showed that (16.3.9) holds for every differentiable function $f(u)$. No matter what $f(u)$ we choose, if $u = 2x + 3y$, we know that $2z_y = 3z_x$.

Example 4 shows why the chain rule is important. It enables us to make general statements about the partial derivatives of an infinite number of functions, all of which are formed the same way. The next example illustrates this use again.

The wave equation also appears in the study of sound or light.

We begin with a brief introduction to that example. D'Alembert in 1746 obtained the partial differential equation for a vibrating string:

$$\frac{\partial^2 y}{\partial t^2} = k^2\frac{\partial^2 y}{\partial x^2}. \tag{16.3.11}$$

(See Figure C.23.3 in the CIE about the Wave in a Rope at the end of this chapter.) This wave equation created a great deal of excitement, especially since d'Alembert showed that any differentiable function of the form

$$g(x + kt) + h(x - kt)$$

is a solution. Here k is a constant, which may be positive, negative, or zero.

Before we show that d'Alembert is right, we note that it is enough to check his claim for $g(x+kt)$. If you replace k by $-k$ in it, you will also have a solution since replacing k by $-k$ in (16.3.11) doesn't change the equation.

EXAMPLE 5 Show that any function $y = g(x+kt)$ satisfies the partial differential equation (16.3.11).
SOLUTION To find the partial derivatives y_{xx} and y_{tt} we express $y = g(x+kt)$ as a composition of functions:

$$y = g(u) \qquad \text{where} \qquad u = x + kt.$$

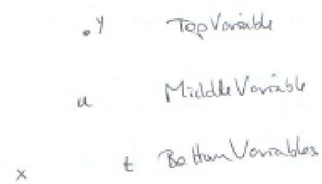


Figure 16.3.5

Note that g is a function of one variable. Figure 16.3.5 lists the variables.

We will compute y_{xx} and y_{tt} in terms of derivatives of g and then check whether (16.3.11) holds. We first compute y_{xx}.

Recall that $u = x + kt$.

$$\frac{\partial y}{\partial x} = \frac{dy}{du}\frac{\partial u}{\partial x} = \frac{dy}{du}\cdot 1 = \frac{dy}{du}. \tag{16.3.12}$$

(There is only one path from y down to x. See Figure 16.3.5.) In (16.3.12) dy/du is viewed as a function of x and t; that is, u is replaced by $x+kt$. Next,

$$\frac{\partial^2 y}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial y}{\partial x}\right) = \frac{\partial}{\partial x}\left(\frac{dy}{du}\right).$$

Now, dy/du, viewed as a function of x and t, may be expressed as a composite function. Letting $w = dy/du$, we have

$$w = f(u), \qquad \text{where} \qquad u = x + kt.$$

Therefore

$$\begin{aligned} \frac{\partial^2 y}{\partial x^2} &= \frac{\partial}{\partial x}\left(\frac{\partial y}{\partial x}\right) = \frac{\partial w}{\partial x} \\ &= \frac{dw}{du}\cdot\frac{\partial u}{\partial x} && \text{(only one path down to } x\text{)} \\ &= \frac{d}{du}\left(\frac{dy}{du}\right)\frac{\partial u}{\partial x} = \frac{d^2 y}{du^2}\cdot 1 \end{aligned}$$

so

$$\frac{\partial^2 y}{\partial x^2} = \frac{d^2 y}{du^2}. \tag{16.3.13}$$

Figure 16.3.6

Then we express y_{tt} in terms of d^2y/du^2, as follows. For

$$\frac{\partial y}{\partial t} = \frac{dy}{du}\frac{\partial u}{\partial t} = \frac{dy}{du}\cdot k = k\frac{dy}{du}$$

(see Figure 16.3.6).

We get

$$\begin{array}{rcll} \frac{\partial^2 y}{\partial t^2} & = & \frac{\partial}{\partial t}\left(\frac{\partial y}{\partial t}\right) = \frac{\partial}{\partial t}\left(k\frac{dy}{du}\right) & \\ & = & k\frac{d}{dy}\left(\frac{dy}{du}\right)\cdot\frac{\partial u}{\partial t} & \text{(only one path down to } t\text{)} \\ & = & k\frac{d^2y}{du^2}\cdot k & \end{array}$$

so

$$\frac{\partial^2 y}{\partial t^2} = k^2\frac{d^2 y}{du^2} \tag{16.3.14}$$

Comparing (16.3.13) and (16.3.14) shows that

$$\frac{\partial^2 y}{\partial t^2} = k^2\frac{\partial^2 y}{\partial x^2}$$

◇

Summary

The section opened by showing that, under suitable assumptions on $f(x, y)$, $\Delta f = f(a + \Delta x, b + \Delta y) - f(a, b)$ equals

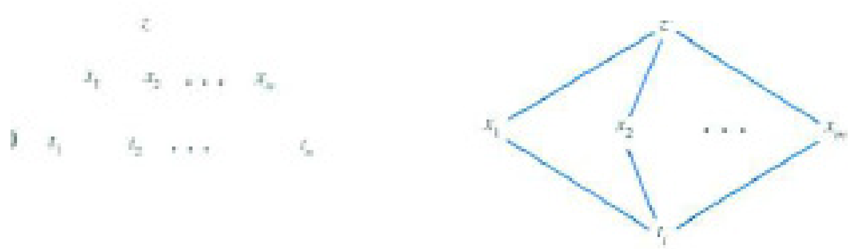
$$\frac{\partial f}{\partial x}(a, b)\Delta x + \frac{\partial f}{\partial y}(a, b) + \epsilon_1\Delta x + \epsilon_2\Delta y, \tag{16.3.15}$$

where ϵ_1 and ϵ_2 approach 0 as Δx and Δy approach 0. This gave us a way to estimate Δf, namely

$$\Delta f \approx \frac{\partial f}{\partial x}(a, b)\Delta x + \frac{\partial f}{\partial y}(a, b)\Delta y$$

The change is due to both the change in x and the change in y. Equation (16.3.15) generalizes to any number of variables and also is the basis for the various chain rules for partial derivatives. This is the general case:

If z is a function of $x_1, x_2, \ldots, x_m$ and each x_i is a function of $t_1, t_2, \ldots, t_n$, then there are n partial derivatives $\partial z/\partial t_j$, $j = 1, 2, \ldots, n$. Each is a sum of m products of the form $(\partial z/\partial x_i)(\partial x_i/\partial t_j)$. To organize the calculation, first make a roster as shown in Figure 16.3.7(a). To compute $\partial z/\partial t_j$, list all paths from z down to t_j, as shown in Figure 16.3.7(b). Each path that starts at z and goes to t_j contributes a product.



(a) (b)

Figure 16.3.7

EXERCISES for Section 16.3

In Exercises 1 to 4 verify the Chain Rule, Special Case I (Theorem 16.3.2, by computing dz/dt two ways: (a) with the chain rule, (b) without the chain rule, by writing z as a function of t.

1. $z = x^2y^3$, $x = t^2$, $y = t^3$

2. $z = xe^y$, $x = t$, $y = 1 + 3t$

3. $z = \cos(xy^2)$, $x = e^{2t}$, $y = \sec(3t)$

4. $z = \ln(x + 3y)$, $x = t^2$, $y = \tan(3t)$.

In Exercises 5 and 6 verify the Chain Rule, Special Case II (Theorem 16.3.3, by computing dz/dt two ways: (a) with the chain rule, (b) without the chain rule, by writing z as a function of t and u.

5. $z = x^2y$, $x - 3t + 4u$, $y = 5t - u$

6. $z = \sin(x + 3y)$, $x = \sqrt{t/u}$, $y = \sqrt{t} + \sqrt{u}$

7. Assume that $z = f(x_1, x_2, x_3, x_4, x_5)$ and that x_i is a function of t_1, t_2, t_3.

(a) List all variables, showing top, middle, and bottom variables.

(b) Draw the paths involved in expressing $\partial z/\partial t_3$ in terms of the chain rule.

(c) Express $\partial z/\partial t_3$ in terms of the sum of products of partial derivatives.

(d) When computing $\partial z/\partial t_3$, which variables are constant?

(e) When computing $\partial z/\partial t_2$, which variables are constant?

8. If $z = f(g(t_1, t_2, t_3), h(t_1, t_2, t_3))$

(a) How many middle variables are there?

(b) How many bottom variables?

(c) What does the chain rule say about $\partial z/\partial t_3$? Include a diagram showing the paths.

9. Find dz/dt if $z_x = 4$, $z_y = 3$, $dx/dt = 4$, and $dy/dt = 1$.

10. Find dz/dt if $z_x = 3$, $z_y = 2$, $dx/dt = 4$, and $dy/dt = -3$.

11. Let $z = f(x, y)$, $x = u + v$, and $y = u - v$.

(a) Show that $(z_x)^2 - (z_y)^2 = (z_u)(z_v)$. Include diagrams.

(b) Verify (a) when $f(x, y) = x^2 + 2y^3$.

12. Let $z = f(x, y)$, $x = u^2 - v^2$, and $y = v^2 - u^2$.

(a) Show that $u\frac{\partial z}{\partial v} + v\frac{\partial z}{\partial u} = 0$. Include diagrams.

(b) Verify (a) when $f(x, y) = \sin(x + 2y)$.

13. Let $z = f(t - u, -t + u)$.

(a) Show that $\frac{\partial z}{\partial t} + \frac{\partial z}{\partial u} = 0$ (Include diagrams.)

(b) Verify (a) when $f(x, y) = x^2 y$

14. Let $w = f(x - y, y - z, z - x)$.

(a) Show that $\frac{\partial w}{\partial x} + \frac{\partial w}{\partial y} + \frac{\partial w}{\partial z} = 0$. Include diagrams.

(b) Verify (a) when $f(s, t, u) = s^2 + t^2 - u$.

15. Let $z = f(u, v)$, where $u = ax + by$, $v = cx + dy$, and a, b, c, d are constants. Show that

(a) $\dfrac{\partial^2 z}{\partial x^2} = a^2\dfrac{\partial^2 f}{\partial u^2} + 2ac\dfrac{\partial^2 f}{\partial u \partial v} + c^2\dfrac{\partial^2 f}{\partial v^2}$

(b) $\dfrac{\partial^2 z}{\partial y^2} = b^2\dfrac{\partial^2 f}{\partial u^2} + 2bd\dfrac{\partial^2 f}{\partial u \partial v} + d^2\dfrac{\partial^2 f}{\partial v^2}$

(c) $\dfrac{\partial^2 z}{\partial x \partial y} = ab\dfrac{\partial^2 f}{\partial u^2} + (ad + bc)\dfrac{\partial^2 f}{\partial u \partial v} + cd\dfrac{\partial^2 f}{\partial v^2}$.

16. Let a, b, and c be constants and consider the partial differential equation

$$a\frac{\partial^2 z}{\partial x^2} + b\frac{\partial^2 z}{\partial x \partial y} + c\frac{\partial^2 z}{\partial y^2} = 0.$$

Suppose it has a solution $z = f(y + mx)$, where m is a constant. Show that $am^2 + bm + c$ must be 0.

17.

(a) Show that any function of the form $z = f(x+y)$ satisfies

$$\frac{\partial^2 z}{\partial x^2} - 2\frac{\partial^2 z}{\partial x \partial y} + \frac{\partial^2 z}{\partial y^2} = 0.$$

(b) Verify (a) when $z = (x+y)^3$.

18.

(a) Show that any function of the form $z = f(x+y) + e^y f(x-y)$ satisfies

$$\frac{\partial^2 z}{\partial x^2} - \frac{\partial^2 z}{\partial y^2} - \frac{\partial z}{\partial x} + \frac{\partial z}{\partial y} = 0.$$

(b) Check (a) for $z = (x+y)^2 + e^y \sin(x-y)$.

19. Let $z = f(x,y)$ denote the temperature at the point (x,y) in the first quadrant. If polar coordinates are used, then we would write $z = f(r,\theta)$.

(a) Express z_r in terms of z_x and x_y. (Use the relation between rectangular coordinates (x,y) and polar coordinates (r,θ).)

(b) Express z_θ in terms of z_x and z_y.

(c) Show that

$$\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2 = \left(\frac{\partial z}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial z}{\partial \theta}\right)^2.$$

(d) What variable is held constant in $\frac{\partial z}{\partial \theta}$?

(e) What variable is held constant in $\frac{\partial z}{\partial x}$?

20. Let $u = f(r)$ and $r = (x^2+y^2+z^2)^{1/2}$. Show that

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = \frac{d^2 u}{dr^2} + \frac{2}{r}\frac{du}{dr}.$$

21. At what rate is the volume of a rectangular box changing when its width is 3 feet and increasing at the rate of 2 feet per second, its length is 8 feet and decreasing at the rate of 5 feet per second, and its height is 4 feet and increasing at the the rate of 2 feet per second?

22. The temperature T at (x, y, z) in space is $f(x, y, z)$. An astronaut is traveling so that his x and y coordinates increase at the rate of 4 miles per second and his z coordinate decreases at the rate of 3 miles per second. Compute the rate dT/dt at which the temperature changes at a point where

$$\frac{\partial T}{\partial x} = 4, \quad \frac{\partial T}{\partial y} = 7, \quad \text{and} \quad \frac{\partial T}{\partial z} = 9.$$

23. Let $u(x, t)$ be the temperature at point x along a rod at time t. The function u satisfies the one-dimensional heat equation for a constant k:

$$\frac{\partial u}{\partial t} = k\frac{\partial^2 u}{\partial x^2}.$$

(a) Show that $u(x, t) = e^{kt}g(x)$ satisfies the heat equation if $g(x)$ is any function such that $g''(x) = g(x)$.

(b) Show that if $g(x) = 3e^{-x} + 4e^{x}$, then $g''(x) = g(x)$.

24. We proved Theorem 16.3.2 when there are two middle variables and one bottom variable. Prove Theorem 16.3.3 when there are two middle variables and two bottom variables.

25. To prove the general chain rule when there are three middle variables, we need an analog of Theorem 16.3.1 concerning Δf when f is a function of three variables.

(a) Let $y = f(x, y, x)$ be a function of three variables. Show that

$$\begin{aligned} \Delta f &= f(x + \Delta x, y + \Delta y, z + \Delta z) - f(x, y, z) \\ &= (f(x + \Delta x, y + \Delta y, z + \Delta z) - f(x + \Delta x, y + \Delta y, z) + (f(x + \Delta x, y + \Delta y, z) - f(x + \Delta x, y, z)) \\ &\quad +(f(x + \Delta x, y, z) - f(x, y, z)). \end{aligned}$$

(b) Using (a) show that

$$\Delta f = \frac{\partial f}{\partial x}(x, y, z)\Delta x + \frac{\partial f}{\partial y}(x, y, z)\Delta y + \frac{\partial f}{\partial z}(x, y, z)\Delta z + \epsilon_1\Delta x + \epsilon_2\Delta y + \epsilon_3\Delta z,$$

where ϵ_1, ϵ_2, $\epsilon_3 \to 0$ as Δx, Δy, $\Delta z \to 0$.

(c) Obtain the general chain rule for three middle variables and any number of bottom variables.

26. Let $z = f(x,y)$, where $x = r\cos(\theta)$ and $y = r\sin(\theta)$. Show that

$$\frac{\partial^2 z}{\partial r^2} = \cos^2(\theta)\frac{\partial^2 f}{\partial x^2} + 2\cos(\theta)\sin(\theta)\frac{\partial^2 f}{\partial x \partial y} + \sin^2(\theta)\frac{\partial^2 f}{\partial y^2}.$$

27. Let $u = f(x,y)$, where $x = r\cos(\theta)$ and $y = r\sin(\theta)$. Verify the following equation, which appears in electromagnetic theory,

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}.$$

28. Let u be a function of x and y, where x and y are both functions of s and t. Show that

$$\frac{\partial^2 u}{\partial s^2} = \frac{\partial^2 u}{\partial x^2}\left(\frac{\partial x}{\partial s}\right)^2 + 2\frac{\partial^2 u}{\partial x \partial y}\frac{\partial x}{\partial s}\frac{\partial y}{\partial x} + \frac{\partial^2 u}{\partial y^2}\left(\frac{\partial y}{\partial x}\right)^2 + \frac{\partial u}{\partial x}\frac{\partial^2 x}{\partial s^2} + \frac{\partial u}{\partial y}\frac{\partial^2 y}{\partial s^2}.$$

29. Let (r,θ) be polar coordinates for the point (x,y) given in rectangular coordinates.

(a) From $r = \sqrt{x^2+y^2}$, show that $\partial r/\partial x = \cos(\theta)$.

(b) From $r = x/\cos\theta$, show that $\partial r/\partial x = 1/\cos(\theta)$.

(c) Explain why (a) and (b) are not contradictory.

30. In developing (16.3.6), we used the path that started at (x,y), went to $(x+\Delta x, y)$, and ended at $(x+\Delta x, y+\Delta y)$. Could we have used the path from (x,y), through $(x, y+\Delta y)$, to $(x+\Delta x, y+\Delta y)$ instead? If not, explain why. If so write out the argument, using the path.

In Exercises 31 to 35 concern **homogeneous functions**. A function $f(x,y)$ is homogeneous of degree r if $f(kx,ky) = k^r f(x,y)$ for all $k > 0$.

31. Show that the following functions are homogeneous, and find the degree r.

(a) $f(x,y) = x^2(\ln x - \ln y)$

(b) $f(x,y) = 1/\sqrt{x^2+y^2}$

(c) $f(x,y) = \sin\left(\frac{y}{x}\right)$

32. Show that if f is homogeneous of degree r, then $xf_x + yf_y = rf$. This is **Euler's theorem**.

33. Verify that the following functions are homogeneous of degree 1 and that they satisfy the conclusion of Euler's theorem (with $r = 1$):

$$f(x,y) = x\frac{\partial f}{\partial x} + y\frac{\partial f}{\partial y}.$$

(a) $f(x,y) = 3x + 4y$

(b) $f(x,y) = x^3 y^{-2}$

(c) $f(x,y) = xe^{x/y}$

34. (See Exercise 32.) Verify Euler's theorem for the functions in Exercise 31.

35. (See Exercise 31.) Show that if f is homogeneous of degree r, then $\partial f/\partial x$ is homogeneous of degree $r - 1$.

16.4 Directional Derivatives and the Gradient

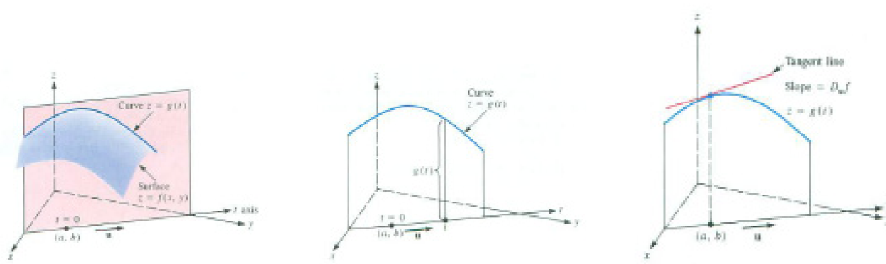
In this section we generalize the notion of a partial derivative to that of a directional derivative. Then we introduce a vector, called the gradient, to provide a formula for the directional derivative. The gradient will have other uses later in this chapter and in Chapter 18.

Directional Derivatives

If $z = f(x, y)$, the partial derivative $\partial f/\partial x$ tells us how rapidly z changes as we move (x, y) in a direction parallel to the x-axis with increasing x. Similarly, f_y tells how fast z changes as we move parallel to the y-axis with increasing y. How rapidly does z change when we move the input point (x, y) in any fixed direction in the xy plane? The answer is given by the directional derivative.

It is important to remember that $|\mathbf{u}| = 1$.

Let $z = f(x, y)$, (a, b) be a point, and $\mathbf{u}$ be a unit vector in the xy-plane. Draw a line through (a, b) parallel to $\mathbf{u}$. Call it the t-axis and let its positive part point in the direction of $\mathbf{u}$. Place the 0 of the t-axis at (a, b). (See Figure 16.4.1.) A value of t determines a point (x, y) on the t-axis and thus a value of z. Along the t-axis, z is a function of t, $z = g(t)$. The derivative dg/dt, evaluated at $t = 0$, is called the **directional derivative** of $z = f(x, y)$ at (a, b) in the direction $\mathbf{u}$. It is denoted $D_{\mathbf{u}}f$. The directional derivative is the slope of the tangent line to the curve $z = g(t)$ at $t = 0$. (See Figure 16.4.1(c).)



(a) (b) (c)

Figure 16.4.1

When $\mathbf{u} = \mathbf{i}$, we obtain the directional derivative $D_{\mathbf{i}}f$, which is simply f_x. When $\mathbf{u} = \mathbf{j}$, we obtain $D_{\mathbf{j}}f$, which is f_y.

The directional derivative generalizes the partial derivatives f_x and f_y, giving the rate of change of $z = f(x, y)$ in any direction in the xy-plane, not just in the directions indicated by $\mathbf{i}$ and $\mathbf{j}$.

The following theorem shows how to compute a directional derivative.

Theorem 16.4.1. *(Directional Derivatives) If $f(x,y)$ has continuous partial derivatives f_x and f_y, then the directional derivative of f at (a,b) in the direction of $\mathbf{u} = \cos(\theta)\mathbf{i} + \sin(\theta)\mathbf{j}$, where θ is the angle between $\mathbf{u}$ and $\mathbf{i}$, is*

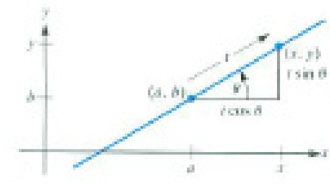
$$\frac{\partial f}{\partial x}(a,b)\cos(\theta) + \frac{\partial f}{\partial y}(a,b)\sin(\theta). \tag{16.4.1}$$

Proof

The directional derivative of f at (a,b) in the direction $\mathbf{u}$ is the derivative of

$$g(t) = f(a + t\cos(\theta), b + t\sin(\theta))$$

when $t = 0$. (See Figure 16.4.2 and Figure 16.4.3.)



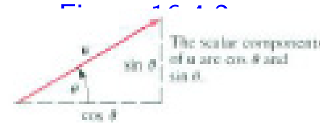


Figure 16.4.3

Because g is a composite function $g(t) = f(x,y)$ where $x = a + t\cos(\theta)$ and $y = b + t\sin(\theta)$, the chain rule tells us that

$$g'(t) = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

Because

$$\frac{dx}{dt} = \cos(\theta) \qquad \text{and} \qquad \frac{dy}{dt} = \sin(\theta)$$

it follows that

$$g'(0) = \frac{\partial f}{\partial x}(a,b)\cos\theta + \frac{\partial f}{\partial y}(a,b)\sin\theta,$$

and the theorem is proved. •

When $\theta = 0$, that is, $\mathbf{u} = \mathbf{i}$, (16.4.1) becomes

$$\frac{\partial f}{\partial x}(a,b)\cos(0) + \frac{\partial f}{\partial y}(a,b)\sin(0) \quad = \quad \frac{\partial f}{\partial x}(a,b)(1) + \frac{\partial f}{\partial y}(a,b)(0) = \frac{\partial f}{\partial x}(a,b).$$

When $\theta = \pi$, that is when $\mathbf{u} = -\mathbf{i}$, (16.4.1) becomes

$$\frac{\partial f}{\partial x}(a,b)\cos(\pi) + \frac{\partial f}{\partial y}(a,b)\sin(\pi) \quad = \quad \frac{\partial f}{\partial x}(a,b)(-1) + \frac{\partial f}{\partial y}(a,b)(0) = -\frac{\partial f}{\partial x}(a,b).$$

(It makes sense that if g increases in one direction then it decreases in the opposite direction.)

SHERMAN: You ask if I agree with Woody. Here, I do. I think in other places he goes too far.

When $\theta = \frac{\pi}{2}$, that is when $\mathbf{u} = \mathbf{j}$, (16.4.1) asserts that the directional derivative is

$$\frac{\partial f}{\partial x}(a,b)\cos(\frac{\pi}{2}) + \frac{\partial f}{\partial y}(a,b)\sin(\frac{\pi}{2}) \quad = \quad \frac{\partial f}{\partial x}(a,b)(0) + \frac{\partial f}{\partial y}(a,b)(1) = \frac{\partial f}{\partial y}(a,b),$$

also as expected.

EXAMPLE 1 Compute the derivative of $f(x,y) = x^2y^3$ at $(1,2)$ in the direction given by the angle $\pi/3$. (That is, $\mathbf{u} = \cos(\pi/3)\mathbf{i} + \sin(\pi/3)\mathbf{j}$.) Interpret the result if f describes a temperature distribution.

SOLUTION Because $\frac{\partial f}{\partial x} = 2xy^3$ and $\frac{\partial f}{\partial y} = 3x^2y^2$, $\frac{\partial f}{\partial x}(1,2) = 16$ and $\frac{\partial f}{\partial y}(1,2) = 12$. Because $\cos\left(\frac{\pi}{3}\right) = \frac{1}{2}$ and $\sin\left(\frac{\pi}{3}\right) = \frac{\sqrt{3}}{2}$, the derivative of f in the direction given by $\theta = \pi/3$ is

$$16\left(\frac{1}{2}\right) + 12\left(\frac{\sqrt{3}}{2}\right) = 8 + 6\sqrt{3} \approx 18.3923.$$

If x^2y^3 is the temperature in degrees at (x,y), where x and y are measured in centimeters, then the rate at which the temperature changes at $(1,2)$ in the direction given by $\theta = \pi/3$ is approximately 18.4 degrees per centimeter. ◇

The Gradient

Equation (16.4.1) resembles a dot product. To exploit this similarity, we introduce the vector whose scalar components are $f_x(a,b)$ and $f_y(a,b)$.

DEFINITION *(The gradient of $f(x,y)$)* The vector

$$\frac{\partial f}{\partial x}(a,b)\mathbf{i} + \frac{\partial f}{\partial y}(a,b)\mathbf{j}$$

is the **gradient** of f at (a,b) and is denoted ∇f. (It is also called del f, because of the upside-down delta ∇.)

The del symbol is in boldface because the gradient of f is a vector. Let $f(x,y) = x^2 + y^2$. We compute and draw ∇f at a few points, listed in Table 16.4.1:

(x,y)	$\frac{\partial f}{\partial x} = 2x$	$\frac{\partial f}{\partial y} = 2y$	∇f
$(1,2)$	2	4	$2\mathbf{i} + 4\mathbf{j}$
$(3,0)$	6	0	$6\mathbf{i}$
$(2,-1)$	4	-2	$4\mathbf{i} - 2\mathbf{j}$

Table 16.4.1

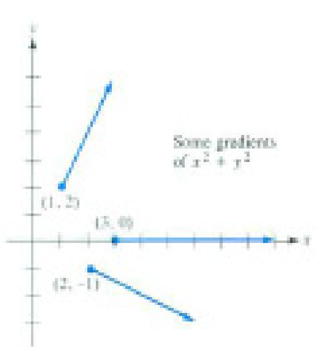


Figure 16.4.4

Figure 16.4.4 shows ∇f, with its tail placed where ∇f is evaluated.
In vector notation, Theorem 16.4.1 reads as follows:

Theorem 16.4.2. *(Directional Derivative, Rephrased) If $z = f(x, y)$ has continuous partial derivatives f_x and f_y, then at (a, b)*

Shorthand: $D_{\mathbf{u}}f = \nabla f \cdot \mathbf{u}$.

$$D_{\mathbf{u}}f = \nabla f(a, b) \cdot \mathbf{u} = (f_x(a, b)\mathbf{i} + f_y(a, b)\mathbf{j}) \cdot \mathbf{u}.$$

The gradient is introduced not merely to provide a short notation for directional derivatives. Its importance is made clear in the next theorem.

A Different View of the Gradient

The gradient vector $\nabla f(a, b)$ provides two pieces of geometric information about a function. First, it points in the direction in the xy-plane in which the function increases most rapidly from the point (a, b). Second, its length, $|\nabla f(a, b)|$, is the largest directional derivative of f at (a, b).

The meaning of $|\nabla f|$ and the direction of ∇f

Theorem 16.4.3. *(Significance of ∇f) Let $z = f(x, y)$ have continuous partial derivatives f_x and f_y. Let (a, b) be a point where ∇f is not $\mathbf{0}$. Then the length of ∇f at (a, b) is the largest directional derivative of f at (a, b). The direction of ∇f is the direction of the largest directional derivative at (a, b).*

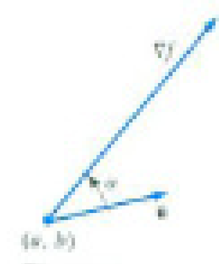


Figure 16.4.5

Proof

If $\mathbf{u}$ is a unit vector, then at (a, b)

$$D_{\mathbf{u}}f = \nabla f \cdot \mathbf{u}.$$

By the definition of the dot product

$$\nabla f \cdot \mathbf{u} = |\nabla f||\mathbf{u}| \cos((\nabla f, \mathbf{u}),$$

(In Figure 16.4.5 the angle between ∇f and $\mathbf{u}$ is labeled α.) Since $|\mathbf{u}| = 1$,

$$D_{\mathbf{u}}f = |\nabla f| \cos(\nabla f, \mathbf{u}). \tag{16.4.2}$$

The largest value of $\cos(\nabla f, \mathbf{u})$ occurs when the angle is 0. That is, when $\mathbf{u}$ points in the direction of ∇f. Thus, by (16.4.2), the largest directional derivative of $f(x, y)$ at (a, b) occurs when the direction is that of ∇f at (a, b). For that $\mathbf{u}$, $D_{\mathbf{u}}f = |\nabla f|$. This proves the theorem. •

EXAMPLE 2 What is the largest directional derivative of $f(x, y) = x^2y^3$ at $(2, 3)$? In what direction does the maximum directional derivative point?
SOLUTION At (x, y),

$$\nabla f = 2xy^3\mathbf{i} + 3x^2y^2\mathbf{j}.$$

At $(2, 3)$,

$$\nabla f = 108\mathbf{i} + 108\mathbf{j},$$

which is sketched in Figure 16.4.6. Its angle θ is $\pi/4$. The maximal directional derivative of x^2y^3 at $(2, 3)$ is $|\nabla f| = 108\sqrt{2} \approx 152.735$. This occurs at the angle $\theta = \pi/4$ relative to the x-axis, that is, for

$$\mathbf{u} = \cos\left(\frac{\pi}{4}\right)\mathbf{i} + \sin\left(\frac{\pi}{4}\right)\mathbf{j} = \frac{\sqrt{2}}{2}\mathbf{i} + \frac{\sqrt{2}}{2}\mathbf{j}.$$

◇

Direction of fastest decrease is $-\nabla f$

If $f(x, y)$ denotes the temperature at (x, y), the gradient ∇f helps indicate the direction in which heat flows. It tends to flow toward the cold, which is the mathematical assertion, that heat tends to flow in the direction of $-\nabla f$.

The gradient and directional derivative have been interpreted in terms of a temperature distribution in the plane. It is also instructive to interpret them for a hiker on a mountain.

The elevation of a point on the surface of a mountain above the point (x, y) will be denoted by $f(x, y)$. The directional derivative $D_{\mathbf{u}}f$ indicates the rate at which elevation changes per unit change in horizontal distance in the direction of $\mathbf{u}$. The gradient ∇f at (a, b) points in the direction of steepest ascent. The length of ∇f tells the hiker how steep the slope is. (See Figure 16.4.7.)

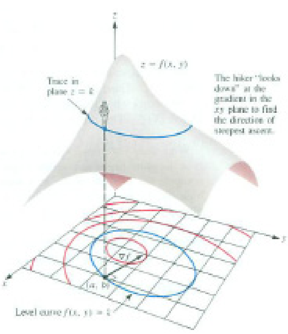


Figure 16.4.7

Generalization to $f(x, y, z)$

Directional derivatives and gradients can be generalized to functions of three or more variables. The directional derivative of $f(x, y, z)$ in a direction in space indicates the rate of change in that direction.

Let $\mathbf{u}$ be a unit vector in space with direction angles α, β, and γ. Then $\mathbf{u} = \cos\alpha\mathbf{i} + \cos\beta\mathbf{j} + \cos\gamma\mathbf{k}$. We now define the derivative of $f(x, y, z)$ in the direction $\mathbf{u}$.

> **DEFINITION** *(Directional Derivative of $f(x, y, z)$.)* The **directional derivative** of f at (a, b, c) in the direction of the unit vector $\mathbf{u} = \cos(\alpha)\mathbf{i} + \cos(\beta)\mathbf{j} + \cos(\gamma)\mathbf{k}$ is $g'(0)$, where g is defined by
>
> $$g(t) = f(a + t\cos(\alpha), b + t\cos(\beta), c + t\cos(\gamma)).$$
>
> It is denoted $D_{\mathbf{u}}f$.

We see that t is the measure of length along the line through (a, b, c) with direction angles α, β, and γ. Therefore $D_{\mathbf{u}}f$ is a derivative along the t-axis.

The proof of Theorem 16.4.4 for a function of three variables is like those for functions of two variables.

Theorem 16.4.4. *(Directional Derivative of $f(x, y, z)$) If $f(x, y, z)$ has continuous partial derivatives f_x, f_y, and f_z, then the directional derivative of f at (a, b, c) in the direction of the unit vector $\mathbf{u} = \cos(\alpha)\mathbf{i} + \cos(\beta)\mathbf{j} + \cos(\gamma)\mathbf{k}$ is*

$$\frac{\partial f}{\partial x}(a, b, c)\cos(\alpha) + \frac{\partial f}{\partial y}(a, b, c)\cos(\beta) + \frac{\partial f}{\partial z}(a, b, c)\cos(\gamma).$$

To write the directional derivative in space as a dot product brings us to the definition of the gradient of a function of three variables.

SHERMAN: This paragraph is new. I did not like the theorem followed immediately by the definition with no intervening text. Your thoughts?

DEFINITION *(The gradient of $f(x, y, z)$.)* The vector

$$\frac{\partial f}{\partial x}(a, b, c)\mathbf{i} + \frac{\partial f}{\partial y}(a, b, c)\mathbf{j} + \frac{\partial f}{\partial z}(a, b, c)\mathbf{k}$$

is the **gradient** of f at (a, b, c) and is denoted ∇f.

Theorem 16.4.4 asserts that

The derivative of $f(x, y, z)$ in the direction of the unit vector $\mathbf{u}$ equals the dot product of the gradient of f and $\mathbf{u}$:

$$D_{\mathbf{u}} f = \nabla f \cdot \mathbf{u}.$$

As for a function of two variables, ∇f evaluated at (a, b, c), points in the direction $\mathbf{u}$ that produces the largest directional derivative there and $|\nabla f|$ is the largest directional derivative. As for two variables, the key steps in the proof of this theorem are writing $\nabla f \cdot \mathbf{u} = |\nabla f||\mathbf{u}|\cos(\nabla f, \mathbf{u})$ and recalling that $\mathbf{u}$ is a unit vector.

EXAMPLE 3 The temperature at the point (x, y, z) in a solid piece of metal is given by $f(x, y, z) = e^{2x+y+3z}$ degrees. In what direction at the point $(0, 0, 0)$ does the temperature increase most rapidly?

SOLUTION Because

$$\frac{\partial f}{\partial x} = 2e^{2x+y+3z}, \quad \frac{\partial f}{\partial y} = e^{2x+y+3z}, \quad \frac{\partial f}{\partial z} = 3e^{2x+y+3z},$$

the gradient vector is

$$\nabla f = 2e^{2x+y+3z}\mathbf{i} + e^{2x+y+3z}\mathbf{j} + 3e^{2x+y+3z}\mathbf{k}.$$

At $(0,0,0)$,

$$\nabla f = 2\mathbf{i} + \mathbf{j} + 3\mathbf{k}.$$

The direction of most rapid increase in temperature is that given by the vector $2\mathbf{i} + \mathbf{j} + 3\mathbf{k}$, that is, $\mathbf{u} = (2\mathbf{i} + \mathbf{j} + 3\mathbf{k})/\sqrt{14}$. The rate of increase is

$$|2\mathbf{i} + \mathbf{j} + 3\mathbf{k}| = \sqrt{14} \text{ degrees per unit length.}$$

If the line through $(0,0,0)$ parallel to $2\mathbf{i} + \mathbf{j} + 3\mathbf{k}$ is given a coordinate system so that it becomes the t-axis, with $t = 0$ at the origin and the positive part in the direction of $2\mathbf{i} + \mathbf{j} + 3\mathbf{k}$, then $df/dt = \sqrt{14}$ at 0. ◇

The gradient was denoted Δ by Hamilton in 1846. By 1870 it was denoted ∇, an upside-down delta, and therefore called "atled." In 1871 Maxwell wrote, "The quantity ∇P is a vector. I venture, with much diffidence, to call it the *slope* of P. Slope is no longer used in this context, having been replaced by gradient which comes from grade, the slope of a road or surface. The name "del" first appeared in print in 1901, in *Vector Analysis, A text-book for the use of students of mathematics and physics founded upon the lectures of J. Willard Gibbs*, by E.B. Wilson.

Summary

We defined the derivative of $f(x,y)$ at (a,b) in the direction of the unit vector $\mathbf{u}$ in the xy plane and the derivative of $f(x,y,z)$ at (a,b,c) in the direction of the unit vector $\mathbf{u}$ in space. Both are denoted $D_{\mathbf{u}}f$. Then we introduced the gradient vector ∇f in terms of its components and obtained the formula

$$D_{\mathbf{u}}f = \nabla f \cdot \mathbf{u}.$$

By examining this formula we saw that the length and direction of ∇f at a given point are significant: ∇f points in the direction $\mathbf{u}$ that maximizes $D_{\mathbf{u}}f$ at the given point and $|\nabla f|$ is the maximum directional derivative of f at the given point.

EXERCISES for Section 16.4

We assume that all functions mentioned in these Exercises have continuous partial derivatives.

SHERMAN: Woody suggests shortening "all functions mentioned" to "functions". I have a problem with this: it's a false assumption (so everything follows from it). Don't we have to have some qualifiers?

1. In what direction from (a, b) does a function decrease most rapidly? What is the maximum rate of decrease?

2. Explain in words, using no symbols, the meaning of $D_{\mathbf{u}}f$.

In Exercises 3 and 4 compute the directional derivative of x^4y^5 at $(1, 1)$ in the given direction.

3. (a) $\mathbf{i}$, (b) $-\mathbf{i}$, (c) $\cos(\pi/4)\mathbf{i} + \sin(\pi/4)\mathbf{j}$

4. (a) $\mathbf{j}$, (b) $-\mathbf{j}$, (c) $\cos(\pi/3)\mathbf{i} + \sin(\pi/3)\mathbf{j}$

In Exercises 5 and 6 compute the directional derivative of x^2y^3 in the given direction.

5. (a) $\mathbf{j}$, (b) $\mathbf{k}$, (c) $-\mathbf{i}$

6. (a) $\mathbf{i}+\mathbf{j}+\mathbf{k}$, (b) $2\mathbf{i}-\mathbf{j}+2\mathbf{k}$, (c) $\mathbf{i}+\mathbf{k}$ These are not unit vectors. First construct a unit vector with the same direction.

7. Assume that at the point $(2, 3), \partial f/\partial x = 4$ and $\partial f/\partial y = 5$.

(a) Draw ∇f at $(2, 3)$.

(b) What is the maximal directional derivative of f at $(2, 3)$?

(c) For which $\mathbf{u}$ is $D_{\mathbf{u}}f$ at $(2, 3)$ maximal? (Write $\mathbf{u}$ as $x\mathbf{i} + y\mathbf{j}$.)

8. Assume that at the point $(1, 1), \partial f/\partial x = 3$ and $\partial f/\partial y = -3$.

(a) Draw ∇f at $(1, 1)$.

(b) What is the maximal directional derivative of f at $(1, 1)$?

(c) For which $\mathbf{u}$ is $D_{\mathbf{u}}f$ at $(1, 1)$ maximal? (Write $\mathbf{u}$ in the form $x\mathbf{i} + y\mathbf{j}$.)

In Exercises 9 and 10 compute and draw ∇f at the given point.

9. $f(x, y) = x^2y$ at (a) $(2, 5)$, (b) $(3, 1)$

10. $f(x, y) = 1/\sqrt{x^2 + y^2}$ at (a) $(1, 2)$, (b) $(3, 0)$

11. If the maximal directional derivative of f at (a, b) is 5, what is the minimal directional derivative there? Explain.

12. For a function $f(x, y)$ at (a, b) is there always a direction in which the directional derivative is 0? Explain.

13. If $(\partial f/\partial x)(a, b) = 2$ and $(\partial f/\partial y)(a, b) = 3$, in what direction should a directional derivative at (a, b) be computed in order that it be

(a) 0?

(b) as large as possible?

(c) as small as possible?

14. If, at the point (a, b, c), $\partial f/\partial x = 2$, $\partial f/\partial y = 3$, $\partial f/\partial z = 4$, what is the largest directional derivative of f at (a, b, c)?

15. Assume that $f(1, 2) = 2$ and $f(0.99, 2.01) = 1.98$.

(a) Which directional derivatives $D_{\mathbf{u}}f$ at $(1, 2)$ can be estimated? (Give $\mathbf{u}$.)

(b) Estimate the directional derivatives in (a).

16. Assume that $f(1, 1, 1) = 3$ and $f(1.1, 1.2, 1.1) = 3.1$.

(a) Which directional derivatives $D_{\mathbf{u}}f$ at $(1, 1, 1)$ can be estimated? (Give $\mathbf{u}$.)

(b) Estimate the directional derivatives in (a).

17. When moving east on the xy-plane, the temperature increases at the rate of $0.02°$ per centimeter. When moving north, the temperature decreases at the rate of $-0.03°$ per centimeter.

(a) At what rate does the temperature change when going south?

(b) At what rate does the temperature change when moving $30°$ north of east?

(c) In what direction should on move to keep the temperature the same?

18. The temperature increases at the rate of $2°$ per kilometer towards the east and decreases at the rate of $2°$ per kilometer towards the north. In what direction does the temperature

(a) increase most rapidly?

(b) decrease most rapidly?

(c) change as little as possible?

19. In the direction **i**, the temperature increases at the rate of 0.03° per centimeter. In the direction **j**, the temperature decreases at the rate of 0.02° per centimeter. In the direction **k** the temperature increases at the rate of 0.05° per centimeter. Does the temperature increase, decrease, or stay the same in the direction $\langle 2,5,1\rangle$?

20. Assume that $f(1,2) = 3$ and that the directional derivative of f at $(1,2)$ in the direction of the (non-unit) vector $\mathbf{i}+\mathbf{j}$ is 0.7. Estimate $f(1.1, 2.1)$.

21. Assume that $f(1,1,2) = 4$ and that the directional derivative of f at $(1,1,2)$ in the direction of the vector from $(1,1,2)$ to $(1.01, 1.02, 1.99)$ is 3. Estimate $f(0.99, 0.98, 2.01)$.

In Exercises 22 to 27 find the directional derivative of the function in the given direction and the maximum directional derivative.

22. xyz^2 at $(1,0,1)$, $\mathbf{i}+\mathbf{j}+\mathbf{k}$

23. x^3yz at $(2,1,-1)$, $2\mathbf{i}-\mathbf{k}$

24. $e^{xy\sin(z)}$ at $(1,1,\pi/4)$, $\mathbf{i}+\mathbf{j}++3\mathbf{k}$

25. $\arctan(\sqrt{x^2+y+z})$ at $(1,1,1)$, $-\mathbf{i}$

26. $\ln(1+xyz)$ at $(2,3,1)$, $-\mathbf{i}+\mathbf{j}$

27. $x^xye^{z^2}$ at $(1,1,0)$, $\mathbf{i}-\mathbf{j}+\mathbf{k}$

28. Let $f(x,y,z) = 2x+3y+z$.

(a) Compute ∇f at (i) $(0,0,0)$ and (ii) $(1,1,1)$.

(b) Draw ∇f for the points in (a), putting the tail at the point.

29. Let $f(x,y,z) = x^2+y^2+z^2$.

(a) Compute ∇f at (i) $(2,0,0)$, (ii) $(0,2,0)$, and (iii) $(0,0,2)$.

(b) Draw ∇f for the points in (a), putting the tail at the point.

30. Let $T(x,y,z)$ be the temperature at the point (x,y,z). Assume that ∇T at $(1,1,1)$ is $2\mathbf{i}+3\mathbf{j}+4\mathbf{k}$.

(a) Find $D_{\mathbf{u}}T$ at $(1,1,1)$ if $\mathbf{u}$ is in the direction of $\mathbf{i}-\mathbf{j}+2\mathbf{k}$.

(b) Estimate the change in temperature as you move from the point $(1,1,1)$ a distance 0.2 in the direction of $\mathbf{i}-\mathbf{j}+2\mathbf{k}$.

(c) Find three unit vectors $\mathbf{u}$ such that $D_{\mathbf{u}}T = 0$ at $(1,1,1)$.

31. Let $f(x,y) = 1/\sqrt{x^2+y^2}$ and $\mathbf{r} = \langle x, y\rangle$. (Note that f is not defined at $(0,0)$.)

(a) Show that $\nabla f = -\mathbf{r}/|\mathbf{r}|^3$.

(b) Show that $|\nabla f| = -1/|\mathbf{r}|^2$.

32. Let $f(x,y,z) = 1/\sqrt{x^2+y^2+z^2}$ and $\mathbf{r} = x\mathbf{i}+y\mathbf{j}+z\mathbf{k}$. (Note that f is not defined at $(0,0,0)$.)

(a) Express ∇f in terms of $\mathbf{r}$.

(b) Express $|\nabla f|$ in terms of $\mathbf{r}$.

33. Let $f(x,y) = x^2+y^2$. Prove that if (a,b) is a point on the curve $x^2+y^2=9$, then ∇f computed at (a,b) is perpendicular to the tangent line to the curve at (a,b).

34. Let $f(x,y,z)$ be the temperature at (x,y,z). Let $P=(a,b,c)$ and Q be a point near (a,b,c). Show that $\nabla f \cdot \overrightarrow{PQ}$ is a good estimate of the change in temperature from P to Q.

35. If $f(P)$ is the electric potential at the point P, then the electric field $\mathbf{E}$ at P is given by $(-1/c^2)\nabla f$. Calculate $\mathbf{E}$ if $f(x,y) = \sin(\alpha x)\cos(\beta y)$, where α and β are constants, where c is a constant.

36. The equality $\partial^2 f/\partial x\partial y = \partial^2 f/\partial y\partial x$ can be written as $D_{\mathbf{i}}(D_{\mathbf{j}}f) = D_{\mathbf{j}}(D_{\mathbf{i}}f)$. Show that $D_{\mathbf{u}_2}(D_{\mathbf{u}_1}f) = D_{\mathbf{u}_1}(D_{\mathbf{u}_2}f)$ for any unit vectors $\mathbf{u}_1$ and $\mathbf{u}_2$. (Assume that partial derivatives of f of all orders are continuous.)

37. Figure 16.4.8 shows two level curves of $f(x,y)$ near $(1,2)$, namely $f(x,y)=3$ and $f(x,y)=3.02$. Use it to estimate

(a) $D_{\mathbf{i}}f$ at $(1,2)$

(b) $D_{\mathbf{j}}f$ at $(1,2)$

(c) Draw ∇f at $(1,2)$.

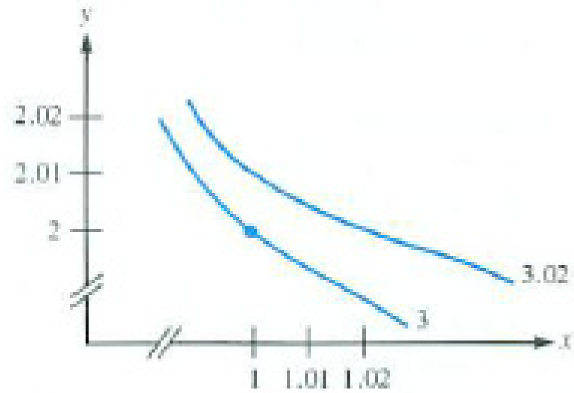


Figure 16.4.8

38. Assume that ∇f at (a,b) is not $\mathbf{0}$. Show that there are unit vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ such that the directional derivatives of f at (a,b) in the direction of $\mathbf{u}_1$ and $\mathbf{u}_2$ are 0.

39. Assume that ∇f at (a,b,c) is not $\mathbf{0}$. How many unit vectors $\mathbf{u}$ are there such that $D_{\mathbf{u}}f(a,b,c)=0$? Explain.

40. Let $f(x,y)$ be the temperature at (x,y). Assume that ∇f at $(1,1)$ is $2\mathbf{i}+3\mathbf{j}$. A particle moves northwest at the rate of 3 centimeters per second. Let $g(t)$ be the temperature at the point where the particle is at time t seconds. Then dg/dt is the rate at which temperature changes on the particle's journey (degrees per second.) Find dg/dt when the bug is at $(1,1)$.

41. Assume that f is defined throughout the xy-plane. If $f(x,y)=g(y)$, then $\partial f/\partial x=0$. Is the converse true? That is, if $\partial f/\partial x=0$, is there a function g of one variable such that $f(x,y)=g(y)$?

42. Let f and g be scalar functions defined throughout the xy-plane. Assume they have the same gradient, $\nabla f=\nabla g$ at all points. Must $f=g$? Is there any relation between f and g?

43. Let $f(x,y)=xy$.

(a) Draw the level curve $xy=4$ carefully.

(b) Compute ∇f at three points on the level curve and draw it with its tail at the point where it is evaluated.

(c) What angle does ∇f seem to make with the curve at the point where it is evaluated?

(d) Prove that the angle is what you think it is.

44. Without the aid of vectors, prove that the maximum value of

$$g(\theta) = \frac{\partial f}{\partial x}(a,b)\cos(\theta) + \frac{\partial f}{\partial y}(a,b)\sin(\theta)$$

is $\sqrt{(\partial f/\partial x(a,b))^2 + (\partial f/\partial y(a,b))^2}$. This is the first part of Theorem 16.4.3.

16.5 Normals and Tangent Planes

In this section we find a normal vector to a curve whose equation is $f(x, y) = k$. Then we find a normal to a surface whose equation is $f(x, y, z) = k$. We can then find the tangent plane at a point on a surface.

Normals to a Curve in the xy-Plane

We saw in Section 14.4 how to find a normal vector to a curve when the curve is given parametrically, $\mathbf{r} = \mathbf{G}(t)$. Now we will see how to find a normal when $f(x, y) = k$. We assume that functions are well behaved, meaning curves have continuous tangent vectors and functions have continuous partial derivatives.

Theorem 16.5.1. *The gradient ∇f at (a, b) is a normal to the level curve of f passing through (a, b).*

Proof

Let $\mathbf{G}(t) = x(t)\mathbf{i} + y(t)\mathbf{j}$ be a parameterization of the level curve of f that passes through (a, b). On it, $f(x, y)$ is a constant and has value $f(a, b)$. Let $\mathbf{G}'(t_0) = \frac{dx}{dt}(t_0)\mathbf{i} + \frac{dy}{dt}(t_0)\mathbf{j}$ be the tangent vector to the curve at (a, b) and let the gradient of f at (a, b) be $\nabla f(a, b) = f_x(a, b,)\mathbf{i} + f_y(a, b)\mathbf{j}$. We wish to show that

$$\nabla f \cdot \mathbf{G}'(t_0) = 0$$

or

$$\frac{\partial f}{\partial x}(a, b)\frac{dx}{dt}(t_0) + \frac{\partial f}{\partial y}(a, b)\frac{dy}{dt}(t_0) = 0. \tag{16.5.1}$$

The left side of (16.5.1) has the form of a chain rule. Let

$$u(t) = f(x(t), y(t)).$$

Because f has the same value at every point on a level curve, $u(t) = f(a, b)$. Thus $u(t)$ is a constant function, and so $du/dt = 0$.

Because u can be viewed as a function of x and y, where x and y are functions of t, the chain rule gives

$$\frac{du}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

Since $du/dt = 0$, (16.5.1) follows. Hence ∇f evaluated at (a, b) is a normal vector to the level curve of f that passes through (a, b). •

What does the theorem say about a weather map that shows the barometric pressure? On level curves the pressures are equal. The gradient ∇f points in the direction in which the pressure increases most rapidly. So $-\nabla f$ points to where the pressure is decreasing most rapidly. Since the wind tends to go from high pressure to low pressure, we can think of $-\nabla f$ as representing the wind. (Earth's rotation also affects the wind.)

Figure 16.5.1 shows a level curve and gradient. The gradient is perpendicular to the level curve. As we saw in Section 16.4, the gradient points in the direction in which the function increases most rapidly.

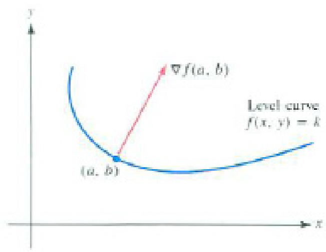


Figure 16.5.1

EXAMPLE 1 Find and draw a normal vector to the hyperbola $xy = 6$ at the point $(2, 3)$.
SOLUTION Let $f(x, y) = xy$. Then $f_x = y$ and $f_y = x$. Hence,

$$\nabla f = y\mathbf{i} + x\mathbf{j}$$

so

$$\nabla f(2, 3) = 3\mathbf{i} + 2\mathbf{j}.$$

This gradient and level curve $xy = 6$ are shown in Figure 16.5.2. ◇

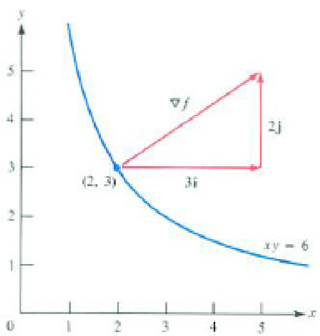


Figure 16.5.2

EXAMPLE 2 Find an equation of the tangent line to the ellipse $x^2 + 3y^2 = 7$ at $(2, 1)$.
SOLUTION As we saw in Section 14.4, we may write the equation of a line in the plane if we know a point on it and a vector normal to it. We know that $(2, 1)$ lies on the line. We use a gradient to produce a normal.

The ellipse $x^2 + 3y^2 = 7$ is a level curve of $f(x, y) = x^2 + 3y^2$. Since $f_x = 2x$ and $f_y = 6y$, $\nabla f = 2x\mathbf{i} + 6y\mathbf{j}$ so

$$\nabla f(2, 1) = 4\mathbf{i} + 6\mathbf{j}.$$

Hence the tangent line at $(2, 1)$ has equation

$$4(x - 2) + 6(y - 1) = 0 \qquad \text{or} \qquad 4x + 6y = 14.$$

The level curve, normal vector, and tangent line are shown in Figure 16.5.3. ◇

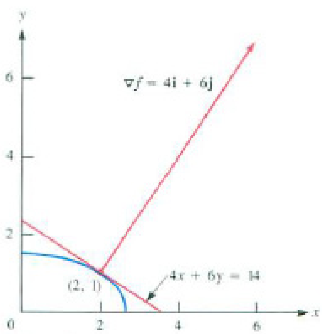


Figure 16.5.3

Normals to a Surface

We can construct a vector perpendicular to a surface $f(x, y, z) = k$ at a point $P = (a, b, c)$ as we constructed a vector perpendicular to a curve. The gradient vector ∇f evaluated at (a, b, c) is perpendicular to the surface $f(x, y, z) = k$. The proof of this is similar to the proof of Theorem 16.5.2 for normal vectors to a level curve.

We define what is meant by a vector being perpendicular to a surface.

DEFINITION *(Normal vector to a surface)* A vector is perpendicular to a surface at (a, b, c) on the surface if it is perpendicular to each curve on the surface through (a, b, c). It is called a **normal vector**.

The next theorem provides a way to find normal vectors to a level surface $f(x, y, z) = k$.

Theorem 16.5.2. *(Normal vectors to a level surface.) The gradient ∇f at (a, b, c) is a normal to the level surface of f passing through (a, b, c).*

The proof consists of showing that ∇f at (a, b, c) is perpendicular to each curve in the level surface of f at (a, b, c). It differs from the proof of Theorem 16.5.1 in that the vectors have three components instead of two.

A check of this theorem is to see whether it is correct when the level surfaces are planes. Consider $f(x, y, z) = Ax + By + Cz + D$. The plane $Ax + By + Cz + D = 0$ is the level surface $f(x, y, z) = 0$. According to the theorem, ∇f is perpendicular to it. Because $f_x = A$, $f_y = B$, and $f_z = C$,

$$\nabla f \quad = \quad \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} + \frac{\partial f}{\partial z}\mathbf{k} = A\mathbf{i} + B\mathbf{j} + C\mathbf{k}.$$

This agrees with the fact that $A\mathbf{i} + B\mathbf{j} + C\mathbf{k}$ is normal to the plane, as we saw by vector algebra in Section 14.4.

EXAMPLE 3 Find a normal vector to the ellipsoid $x^2 + y^2/4 + z^2/9 = 3$ at $(1, 2, 3)$.

SOLUTION The ellipsoid is a level surface of

$$f(x, y, z) = x^2 + \frac{y^2}{4} + \frac{z^2}{9}.$$

The gradient of f at (x, y, z) is

$$\nabla f = 2x\mathbf{i} + \frac{y}{2}\mathbf{j} + \frac{2z}{9}\mathbf{k}.$$

At $(1, 2, 3)$

$$\nabla f = 2\mathbf{i} + \mathbf{j} + \frac{2}{3}\mathbf{k}.,$$

which is normal to the ellipsoid at $(1, 2, 3)$. ◇

Tangent Planes to a Surface

Now that we can find a normal to a surface we can define a tangent plane at a point on the surface.

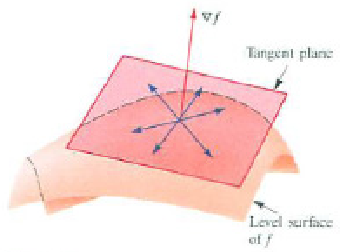


Figure 16.5.4

> **DEFINITION** *(Tangent plane to a surface)* Consider a surface that is a level surface of a function $u = f(x, y, z)$. Let (a, b, c) be a point on it where ∇f is not $\mathbf{0}$. The tangent plane to the surface at (a, b, c) is the plane through (a, b, c) that is perpendicular to the vector ∇f evaluated at (a, b, c).

The tangent plane at (a, b, c) is the plane that best approximates the surface near (a, b, c). It consists of all the tangent lines to curves in the surface that pass through (a, b, c). See Figure 16.5.4.

An equation of the plane through (x_0, y_0, z_0) perpendicular to $\langle A, B, C\rangle$ is $A(x - x_0) + B(y - y_0) + C(z - z_0) = 0$.

An equation of the tangent plane to the surface $f(x, y, z) = k$ at (a, b, c) is

$$\frac{\partial f}{\partial x}(a, b, c)(x - a) + \frac{\partial f}{\partial y}(a, b, c)(y - b) + \frac{\partial f}{\partial z}(a, b, c)(z - c) = 0.$$

EXAMPLE 4 Find an equation of the tangent plane to the ellipsoid $x^2 + y^2/4 + z^2/9 = 3$ at $(1, 2, 3)$.

SOLUTION By Example 3, $2\mathbf{i} + \mathbf{j} + \frac{2}{3}\mathbf{k}$ is normal to the surface at the point $(1, 2, 3)$. The tangent plane consequently has an equation

$$2(x - 1) + 1(y - 2) + 2/3(z - 3) = 0.$$

◇

Normals and Tangent Planes to $z = f(x, y)$

A surface may be described explicitly in the form $z = f(x, y)$ rather than implicitly in the form $f(x, y, z) = k$. The techniques we developed enable us to find the normal and tangent plane for $z = f(x, y)$ as well.

Finding a normal to the surface $z = f(x, y)$

If we write $z = f(x, y)$ as $z - f(x, y) = 0$ and let $g(x, y, z)$ be $z - f(x, y)$, then the surface $z - f(x, y) = 0$ is the level surface $g(x, y, z) = 0$. There is no need to memorize a formula for a vector normal to the surface $z = f(x, y)$. The next example illustrates this.

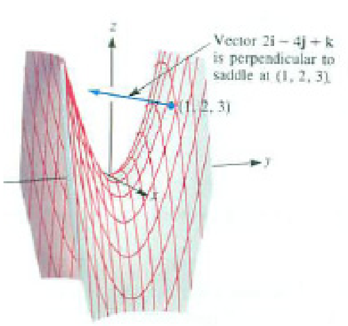


Figure 16.5.5

EXAMPLE 5 Find a vector perpendicular to $z = y^2 - x^2$ at $(1, 2, 3)$.

SOLUTION Write $z = y^2 - x^2$ as $z + x^2 - y^2 = 0$. The surface is a level surface of $g(x, y, z) = z + x^2 - y^2$. Hence $\nabla g = 2x\mathbf{i} - 2y\mathbf{j} + \mathbf{k}$ and $2\mathbf{i} - 4\mathbf{j} + \mathbf{k}$ is perpendicular to the surface at $(1, 2, 3)$.

The surface looks like a saddle near the origin. The surface and the normal vector are shown in Figure 16.5.5. ◇

Estimates by Tangent Planes

For a function of one variable, $y = f(x)$, the tangent line at $(a, f(a))$ closely approximates the graph of $y = f(x)$ near $(a, f(a))$. The equation of the tangent line, $y = f(a) + f'(a)(x - a)$, gives us a linear approximation of $f(x)$. (See Section 5.4.)

We can use the tangent plane to the surface $z = f(x, y)$ similarly. To find the equation of the plane tangent at $(a, b, f(a, b))$, we first write the equation of the surface as

$$g(x, y, z) = f(x, y) - z = 0.$$

Then ∇g is a normal to the surface at $(a, b, f(a, b))$. Because

$$\nabla g = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} - \mathbf{k},$$

(where the partial derivatives are evaluated at (a, b)), the equation of the tangent plane at $(a, b, f(a, b))$ is

$$\frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b) - (z - f(a, b)) = 0.$$

We can write this as

$$z = f(a, b) + \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b). \tag{16.5.2}$$

Letting $\Delta x = x - a$ and $\Delta y = y - b$, (16.5.2) becomes

$$z = f(a, b) + \frac{\partial f}{\partial x}(a, b)\Delta x + \frac{\partial f}{\partial y}(a, b)\Delta y.$$

This tells us that the change of the z coordinate on the tangent plane, as the x coordinate changes from a to $a + \Delta x$ and the y coordinate changes from b to $b + \Delta y$, is exactly

$$\frac{\partial f}{\partial x}(a, b)\Delta x + \frac{\partial f}{\partial y}(a, b)\Delta y.$$

By (16.3.1) in Section 16.3, this is an estimate of the change Δf in the function f as its argument changes from (a, b) to $(a + \Delta x, b + \Delta y)$. This is another way of saying that the tangent plane to $z = f(x, y)$ at $(a, b, f(a, b))$ is close to the surface. See Figure 16.5.6.

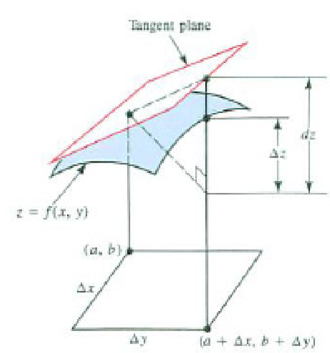


Figure 16.5.6

The expression $f_x(a, b)\ dx + f_y(a, b)\ dy$ is called the **differential** of f at (a, b). For small values of dx and dy it is a good estimate of $f(a + dx, b + dy) - f(a, b)$.

EXAMPLE 6 Let $z = f(x, y) = x^2 y$. Let $\Delta z = f(1.01, 2.02) - f(1, 2)$ and let

$$dz = \frac{\partial f}{\partial x}(1, 2) \cdot 0.01 + \frac{\partial f}{\partial y}(1, 2) \cdot 0.02.$$

Compute Δz and dz.
SOLUTION

$$\Delta z = (1.01)^2(2.02) - 1^2 2 = 2.060602 - 2 = 0.060602$$

Since $f_x = 2xy$ and $f_y = x^2$, we have $f_x = 4$ and $f_y = 1$ at $(1, 2)$. Hence,

$$dz = (4)(0.01) + (1)(0.02) = 0.06,$$

so dz is a good approximation of Δz. $\diamond$

Summary

Table 16.5.1 summarizes normals and tangents.

Function	Level Curve/Surface	Normal	Tangent
$f(x, y)$	level curve: $f(x, y) = k$	$\nabla f = f_x\mathbf{i} + f_y\mathbf{j}$	Tangent line at (a, b) is $f_x(a, b)(x - a) + f_y(a, b)(y - b) = f(a, b)$
$f(x, y, z)$	level surface: $f(x, y, z) = k$	$\nabla f = f_x\mathbf{i} + f_y\mathbf{j} + f_z\mathbf{k}$	Tangent plane at (a, b, c) is $f_x(a, b, c)(x - a)$ $+ f_y(a, b, c)(y - b)$ $+ f_z(a, b, c)(z - c) = f(a, b, c)$

Table 16.5.1

To find a normal and tangent plane to a surface given as $z = f(x, y)$, treat the surface as a level surface of $z - f(x, y)$, namely $z - f(x, y) = 0$.

We also showed that the differential approximation of Δf in Section 16.3 is the change along the tangent plane.

EXERCISES for Section 16.5

The angle between two surfaces that pass through (a, b, c) is defined as the angle between the lines through (a, b, c) that are perpendicular to the surfaces at the point (a, b, c). This angle is taken to be acute. Use this definition in Exercises 1 to 3.

1.

(a) Show that $(1, 1, 2)$ lies on the surfaces $xyz = 2$ and $x^3yz^2 = 4$.

(b) Find the angle between the surfaces at $(1, 1, 2)$.

2.

(a) Show that $(1, 2, 3)$ lies on the plane $2x+3y-z = 5$ and the sphere $x^2+y^2+z^2 = 14$.

(b) Find the angle between them at the point $(1, 2, 3)$.

3.

(a) Show that the surfaces $z = x^2y^3$ and $z = 2xy$ pass through $(2, 1, 4)$.

(b) At what angle do they cross at that point?

4. Let $z = f(x, y)$ describe a surface. Assume that at $(3, 5)$, $z = 7$, $\partial z/\partial x = 2$, and $\partial z/\partial y = 3$.

(a) Find a normal to the surface at $(3, 5, 7)$.

(b) Find two vectors that are tangent to it at $(3, 5, 7)$.

(c) Estimate $f(3.02, 4.99)$.

5. Let $T(x, y, z)$ be the temperature at the point (x, y, z), where ∇T is not $\mathbf{0}$. A level surface $T(x, y, z) = k$ is called an **isotherm**. Show that if you are at the point (a, b, c) and wish to move in the direction in which the temperature changes most rapidly, you would move in a direction perpendicular to the isotherm that passes through (a, b, c).

6. Write a short essay on the chain rule. Include a description of how it was used to show that $D_{\mathbf{u}}f = \nabla f \cdot \mathbf{u}$ and in showing that ∇f is a normal to the level surface of f at the point where it is evaluated.

7. Suppose you are at (a, b, c) on the level surface $f(x, y, z) = k$. There $\nabla F = 2\mathbf{i} + 3\mathbf{j} - 4\mathbf{k}$.

(a) If $\mathbf{u}$ is tangent to the surface at (a, b, c), what would $D_{\mathbf{u}}f$ equal? (There are infinitely many $\mathbf{u}$s.)

(b) If $\mathbf{u}$ is normal to the level surface at (a, b, c), what would $D_{\mathbf{u}}f$ equal? (There are two $\mathbf{u}$s.)

8. We have found a way to find a normal and a tangent plane to a surface. How would you find a tangent line to a surface? Illustrate your method by finding a line that is tangent to $z = xy$ at $(2, 3, 6)$.

9.

(a) Draw three level curves of the function f defined by $f(x, y) = xy$. Include the curve through $(1, 1)$ as one of them.

(b) Draw three level curves of the function g defined by $g(x, y) = x^2 - y^2$. Include the curve through $(1, 1)$ as one of them.

(c) Prove that a level curve of f intersects a level curve of g at a right angle.

(d) If we think of f as air pressure, how may we interpret the level curves of g?

10.

(a) Draw a level curve for $2x^2 + y^2$.

(b) Draw a level curve for y^2/x.

(c) Prove that a level curve of $2x^2 + y^2$ crosses a level curve of y^2/x at a right angle.

11. Two surfaces $f(x, y, z) = 0$ and $g(x, y, z) = 0$ pass through (a, b, c). Their intersection is a curve. How would you find a tangent vector to it at (a, b, c)?

12. How far is the point $(2, 1, 3)$ from the tangent plane to $z = xy$ at $(3, 4, 12)$?

13. The map in Figure 16.5.7 shows isobars, level curves of the pressure $p(x, y)$. At which of the labeled points on the map is the gradient of p, ∇p, the longest? In what direction does it point? In which direction (approximately) would the wind vector point?

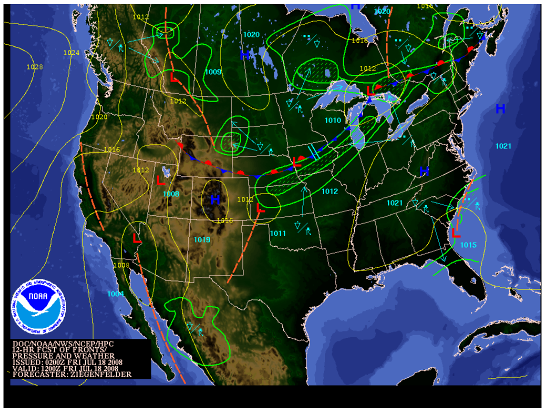


Figure 16.5.7

14. Prove Theorem 16.5.2. (The proof is a slight modification of the proof of Theorem 16.5.1.)

15. The surfaces $x^2yz = 1$ and $xy + yz + zx = 3$ pass through $(1, 1, 1)$. Their tangent planes meet in a line. Find parametric equations for it.

16. The surface $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$ is called an ellipsoid. If $a^2 = b^2 = c^2$ it is a sphere. Show that if a^2, b^2, and c^2 are distinct, then there are exactly six normals on the ellipsoid that pass through the origin.

17. If $f(x)$ is defined for all x and its derivative is always 0, it is constant. Assume $f(x, y)$ is defined at all points (x, y) and its gradient is always $\mathbf{0}$. Must $f(x, y) = C$ for some constant C?

18. Let $\mathcal{S}$ be a surface with equation $f(x, y, z) = 0$, such that each ray from the origin O meets $\mathcal{S}$ at exactly one point. Assume that at each point P in $\mathcal{S}$, the tangent plane at P is perpendicular to the radial vector $\mathbf{r} = \overrightarrow{OP}$. Show that $\mathcal{S}$ is a sphere. (Let P_0 and P_1 be two points on $\mathcal{S}$. Take a curve with parameterization $(x(t), y(t), z(t))$ and that on it $|\mathbf{r}|$ is constant.)

19. Assume that $f(x, y, z)$ and $g(x, y, z)$ have the property that ∇f and ∇g, evaluated at the same point, are always parallel. Show that they have the same level surfaces.

20. Prove Theorem 16.5.2.

16.6 Critical Points and Extrema

For a function of one variable, $y = f(x)$, the first and second derivatives were of use in searching for relative extrema. We looked for critical numbers, that is, solutions of the equation $f'(x) = 0$ and we checked the value of $f''(x)$ there. If $f''(x)$ was positive, the critical number gave a relative minimum. If $f''(x)$ was negative, the critical number gave a relative maximum. If $f''(x)$ was 0, then anything might happen: a minimum, maximum, or neither. To tell we have to resort to other tests.

If $f''(x)$ is positive, the graph of f is concave up; if $f''(x)$ is negative, the graph of f is concave down.

This section extends the idea of a critical point to functions $f(x, y)$ of two variables and shows how to use the second-order partial derivatives f_{xx}, f_{yy}, and f_{xy} to see whether the critical point provides a relative maximum, relative minimum, or neither.

The subscript notation $\frac{\partial f}{\partial x} = f_x$. is used in text to save space.

Extrema of $f(x, y)$

The number M is called the **maximum** (or **global maximum**) of f over a set R in the plane if it is the largest value of $f(x, y)$ for (x, y) in R. A **relative maximum** (or **local maximum**) of f occurs at a point (a, b) in R if there is a disk around (a, b) such that $f(a, b) \geq f(x, y)$ for all points (x, y) in the disk. **Minimum** and **relative** (or **local) minimum** are defined similarly.

Let us look at the surface above a point (a, b) where a relative maximum of f occurs. Assume that f is defined for all points in some disk around (a, b) and possesses partial derivatives at (a, b). Assume, for convenience, that the values of f are positive. Let L_1 be the line $y = b$ in the xy plane; let L_2 be the line $x = a$ in the xy plane. (See Figure 16.6.1.)

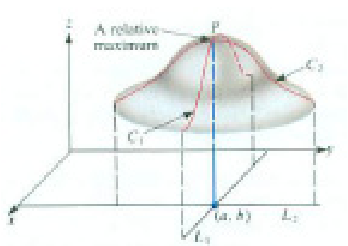


Figure 16.6.1

Let C_1 be the curve in the surface above the line L_1, let C_2 be the curve in the surface above the line L_2, and let P be the point on the surface above (a, b).

Since f has a relative maximum at (a, b), no point on the surface near P is higher than P. Thus P is a highest point on C_1 and on C_2 (for points near P). So both curves have horizontal tangents at P. That is, at (a, b) both partial derivatives of f must be 0:

$$\frac{\partial f}{\partial x}(a, b) = 0 \qquad \text{and} \qquad \frac{\partial f}{\partial y}(a, b) = 0.$$

This conclusion is summarized in the following theorem.

Theorem 16.6.1. *(Relative Extremum of $f(x, y)$) Let f be defined on a domain that includes the point (a, b) and all points in some disk whose center is (a, b). If f has a relative maximum (or relative minimum) at (a, b) and f_x and f_y exist at (a, b), then*

$$\frac{\partial f}{\partial x}(a, b) = 0 = \frac{\partial f}{\partial y}(a, b),$$

So, the gradient of f, ∇f is $\mathbf{0}$ at a relative extremum.

Points (a, b) where both partial derivatives f_x and f_y are 0 is clearly of importance. They are analogous to the critical points of a function of one variable.

> **DEFINITION** *(Critical point)* If $f_x(a, b) = 0$ and $f_y(a, b) = 0$, the point (a, b) is a **critical point** of the function $f(x, y)$.

You might expect that if (a, b) is a critical point of f and the two second partial derivatives f_{xx} and f_{yy} are both positive at (a, b), then f necessarily has a relative minimum at (a, b). The next example shows that this is false.

EXAMPLE 1 Find the critical points of $f(x, y) = x^2 + 3xy + y^2$ and determine whether there are extrema there.

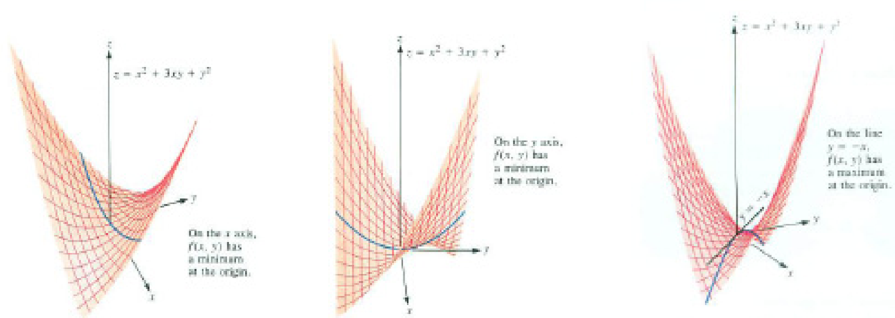
SOLUTION Find critical points by setting both f_x and f_y equal to 0. This gives

$$2x + 3y = 0 \qquad \text{and} \qquad 3x + 2y = 0,$$

whose only solution is $(x, y) = (0, 0)$. The function has one critical point, $(0, 0)$.

Now look at the graph of f for (x, y) near $(0, 0)$. Because $f(x, 0) = x^2 + 3 \cdot x \cdot 0 + 0^2 = x^2$, considered as a function of x, the function has a minimum at the origin. (See Figure 16.6.2(a).)

On the y-axis, the function reduces to $f(0, y) = y^2$, whose graph is another parabola with a minimum at the origin. (See Figure 16.6.2(b).) Also, $f_{xx} = 2$ and $f_{yy} = 2$, so both are positive at $(0, 0)$.



(a) (b) (c)

Figure 16.6.2

We might think that f has a relative minimum at $(0, 0)$. On the line $y = -x$,

$$f(x, y) = f(x, -x) = x^2 + 3x(-x) + (-x)^2 = -x^2.$$

On the line the function assumes negative values, and its graph is a parabola opening downward, as shown in Figure 16.6.2(c).

Thus $f(x, y)$ has neither a relative maximum nor minimum at the origin. Its graph resembles a saddle. $\diamond$

Example 1 shows that to determine whether a critical point of $f(x, y)$ provides an extremum, it is not enough to look at f_{xx} and f_{yy} The criteria are more complicated and involve the mixed partial derivative f_{xy} as well. Exercise 56 outlines a proof of Theorem 16.6.2. At the end of this section we give a proof when $f(x, y)$ is a polynomial of the form $Ax^2 + Bxy + Cy^2$, where A, B and C are constants.

In subscript notation, $D = f_{xx}f_{yy} - (f_{xy})^2$.

Theorem 16.6.2. *(Second-partial-derivative test for $f(x, y)$) Let (a, b) be a critical point of the function $f(x, y)$. Assume that the partial derivatives f_x, f_y, f_{xx}, f_{xy}, and f_{yy} are continuous at and near (a, b). Let*

$$D = \frac{\partial^2 f}{\partial x^2}(a, b)\frac{\partial^2 f}{\partial y^2}(a, b) - \left(\frac{\partial^2 f}{\partial x \partial y}(a, b)\right)^2.$$

Case 1: If $D > 0$ and $f_{xx}(a, b) > 0$, then f has a relative minimum at (a, b).

Case 2: If $D > 0$ and $f_{xx}(a, b) < 0$, then f has a relative maximum at (a, b).

*Case 3: If $D < 0$, then f has neither a relative minimum nor a relative maximum at (a, b). (There is a **saddle point** at (a, b).)*

If $D = 0$, then anything can happen: there may be a relative minimum, a relative maximum, or a saddle point as illustrated in Exercise 43.

The expression $f_{xx}f_{yy} - (f_{xy})^2$ is called the **discriminant** of f.

To see what the theorem says, consider case 1, the test for a relative minimum. It says that $f_{xx}(a, b) > 0$ (which is to be expected) and that

$$\frac{\partial^2 f}{\partial x^2}(a, b)\frac{\partial^2 f}{\partial y^2}(a, b) - \left(\frac{\partial^2 f}{\partial x \partial y}(a, b)\right)^2 > 0,$$

Or equivalently,

$$\left(\frac{\partial^2 f}{\partial x \partial y}(a, b)\right)^2 < \frac{\partial^2 f}{\partial x^2}(a, b)\frac{\partial^2 f}{\partial y^2}(a, b). \qquad (16.6.1)$$

Memory aid regarding size of f_{xy}

Since the square of a real number is never negative and $f_{xx}(a, b)$ is positive, it follows that $f_{yy}(a, b) > 0$, which was to be expected. But (16.6.1) says more. It says that the mixed partial $f_{xy}(a, b)$ must not be too large. For a relative maximum or minimum, (16.6.1) must hold. This may be easier to remember than $D > 0$.

EXAMPLE 2 Find all relative extrema of:

1. $f(x,y) = x^2 + 3xy + y^2$
2. $g(x,y) = x^2 + 2xy + y^2$
3. $h(x,y) = x^2 + xy + y^2$

SOLUTION

1. The case $f(x,y) = x^2 + 3xy + y^2$ is Example 1. The origin is the only critical point, and it provides neither a relative maximum nor a relative minimum. We can check this by the discriminant. We have

$$\frac{\partial^2 f}{\partial x^2}(0,0) = 2, \qquad \frac{\partial^2 f}{\partial x \partial y}(0,0) = 3, \qquad \text{and} \quad \frac{\partial^2 f}{\partial y^2}(0,0) = 2.$$

Hence $D = 2 \cdot 2 - 3^2 = -5$ is negative. By Theorem 16.6.1, there is neither a relative maximum nor a relative minimum at the origin. There is a saddle point there.

2. It is straightforward to show that all the points on the line $x + y = 0$ are critical points of $g(x,y) = x^2 + 2xy + y^2$. Because

$$\frac{\partial^2 g}{\partial x^2}(x,y) = 2, \qquad \frac{\partial^2 g}{\partial x \partial y}(x,y) = 2, \qquad \text{and} \quad \frac{\partial^2 g}{\partial y^2}(x,y) = 2,$$

the discriminant $D = 2 \cdot 2 - 2^2 = 0$. Since $D = 0$, it gives no information.

However, if $x^2 + 2xy + y^2 = (x+y)^2$ and so, being the square of a real number, is always greater than or equal to 0. Hence the origin provides a relative minimum of $x^2 + 2xy + y^2$. (So does any point on $x + y = 0$.) Since $g(x,y) = (x+y)^2$, it is constatt on a lne $x + y = c$, it is constant on a line $x + y = c$. See Figure 16.6.3.)

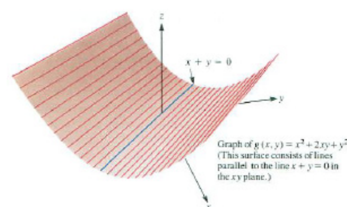


Figure 16.6.3

3. For $h(x,y) = x^2 + xy + y^2$, it is easy to check that the origin is the only critical point and we have

$$\frac{\partial^2 h}{\partial x^2}(0,0) = 2, \qquad \frac{\partial^2 h}{\partial x \partial y}(0,0) = 1, \qquad \text{and} \quad \frac{\partial^2 h}{\partial y^2}(0,0) = 2.$$

Then, $D = 2 \cdot 2 - 1^2 = 3$ is positive and $h_{xx}(0,0) > 0$. Hence $x^2 + xy + y^2$ has a relative minimum at the origin.

The graph of h is shown in Figure 16.6.4

Figure 16.6.4

$\diamond$

EXAMPLE 3 Examine $f(x, y) = x + y + 1/(xy)$ for global and relative extrema.

SOLUTION When x and y are both large positive numbers or small positive numbers, then $f(x, y)$ may be arbitrarily large. There is therefore no global maximum. By allowing x and y to be negative numbers of large absolute values, we see that there is no global minimum.

The function has no global extrema.

Local extrema will occur at critical points. We have

$$\frac{\partial f}{\partial x} = 1 - \frac{1}{x^2y} \qquad \text{and} \qquad \frac{\partial f}{\partial y} = 1 - \frac{1}{xy^2}.$$

Setting these partial derivatives equal to 0 gives

$$\frac{1}{x^2y} = 1 \qquad \text{and} \qquad \frac{1}{xy^2} = 1. \tag{16.6.2}$$

Hence $x^2y = xy^2$. Since f is not defined when x or y is 0, we may assume $xy \neq 0$. Dividing both sides of $x^2y = xy^2$ by xy gives $x = y$. By (16.6.2), $1/x^3 = 1$ so $x = 1$, and $y = 1$. Thus there is only one critical point, $(1, 1)$.

To find whether it is a relative extremum, use Theorem 16.6.2. We have

$$\frac{\partial^2 f}{\partial x^2} = \frac{2}{x^3y}, \qquad \frac{\partial^2 f}{\partial x \partial y} = \frac{1}{x^2y^2}, \qquad \text{and} \qquad \frac{\partial^2 f}{\partial y^2} = \frac{2}{xy^3}.$$

Thus at $(1, 1)$,

$$\frac{\partial^2 f}{\partial x^2} = 2, \qquad \frac{\partial^2 f}{\partial x \partial y} = 1, \qquad \text{and} \qquad \frac{\partial^2 f}{\partial y^2} = 2.$$

Therefore,

$$D = 2 \cdot 2 - 1^2 = 3 > 0.$$

Since the discriminant is positive and $f_{xx}(1, 1) > 0$, the point $(1, 1)$ provides a relative minimum. $\diamond$

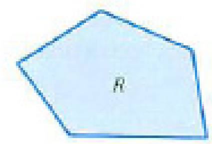
Extrema on a Bounded Region

In Section 4.3, we saw how to find a maximum of a differentiable function, $y = f(x)$, on an interval $[a, b]$. The procedure was:

1. Find x in $[a, b]$ (other than a or b) where $f'(x) = 0$. It is called a critical number. If there are no critical numbers, the maximum occurs at a or b.

2. If there are critical numbers, evaluate f at them. Also find the values of $f(a)$ and $f(b)$. The maximum of f in $[a, b]$ is the largest of $f(a)$, $f(b)$, and the values of f at critical numbers.

We can similarly find the maximum of $f(x, y)$ in a region R in the plane bounded by some polygon or curve. (See Figure 16.5.6.) It is assumed that R includes its border and is a finite region in the sense that it lies within some disk. (In advanced courses it is proved that a continuous function defined on such a domain has a maximum – and a minimum – value.) If f has continuous partial derivatives, the procedure for finding a maximum is similar to that for maximizing a function on a closed interval.



A continuous function on R (which includes the border) has a maximum value at some point in R.

Figure 16.6.5

1. First find points that are in R but not on the boundary of R where both f_x and f_y are 0. They are called **critical points**. If there are no critical points in R, the maximum occurs on the boundary.

2. If there are critical points, evaluate f at them. Also find the maximum of f on the boundary. The maximum of f on R is the largest value of f on the boundary and at critical points.

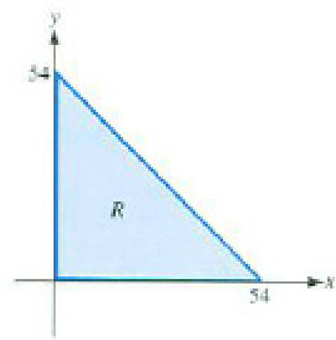


Figure 16.6.6

A similar procedure finds the minimum value on a bounded region.

EXAMPLE 4 Maximize the function

$$f(x, y) = xy(108 - 2x - 2y) = 108xy - 2x^2y - 2xy^2$$

on the triangle R bounded by the x-axis, the y-axis, and the line $x + y = 54$. (See Figure 16.6.6.)

SOLUTION Find any critical points not on the boundary of R. We have

Find critical points.

$$\begin{aligned}\frac{\partial f}{\partial x}(x, y) &= 108y - 4xy - 2y^2 = 0\\ \frac{\partial f}{\partial y}(x, y) &= 108x - 2x^2 - 4xy = 0\end{aligned}$$

which give the simultaneous equations

$$\begin{aligned}2y(54 - 2x - y) &= 0,\\ 2x(54 - x - 2y) &= 0.\end{aligned}$$

Because y is not zero, by the first equation, $y = 54 - 2x$. Substituting into the second equation gives $54 - x - 2(54 - 2x) = 0$, or $-54 + 3x = 0$. Hence $x = 18$ and therefore $y = 54 - 2 \cdot 18 = 18$.

Evaluate f at critical points.

The point $(18, 18)$ lies in the interior of R, since it lies above the x-axis, to the right of the y-axis, and below the line $x + y = 54$, and $f(18, 18) = 18 \cdot 18(108 - 2 \cdot 18 - 2 \cdot 18) = 11,664$.

Evaluate f on boundary.

Next we examine the function $f(x, y) = xy(108-2x-2y)$ on the boundary of R. On its base, $y = 0$, so $f(x, y) = 0$. On its left edge, $x = 0$, so again $f(x, y) = 0$. On the slanted edge, which lies on $x + y = 54$, we have $108 - 2x - 2y = 0$, so $f(x, y) = 0$ on this edge also. Thus $f(x, y) = 0$ on the entire boundary.

Therefore, the local maximum occurs at the critical point $(18, 18)$ and has the value $11,664$. ◇

EXAMPLE 5 The combined length and girth (distance around) of a package sent through the mail cannot exceed 108 inches. If the package is a rectangular box, how large can its volume be?

SOLUTION We label the length of the bar (a longest side) z and the other sides x and y, as in Figure 16.6.7. The volume $V = xyz$ is to be maximized, subject to girth plus length being at most 108, that is,

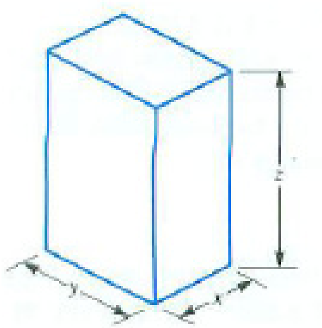


Figure 16.6.7

$$2x + 2y + z \leq 108.$$

Since we want the largest box, wewill restrict our attention to boxes for which

$$2x + 2y + z = 108. \tag{16.6.3}$$

By (16.6.3), $z = 108 - 2x - 2y$. Thus $V = xyz$ can be expressed as a function of two variables:

$$V = f(x, y) = xy(108 - 2x - 2y).$$

Why is $2x + 2y \leq 108$?

This is to be maximized on the triangle described by $x \geq 0$, $y \geq 0$, $2x + 2y \leq 108$, that is, $x + y \leq 54$.

These are the same function and region as in the previous example. Hence, the largest box has $x = y = 18$ and $z = 108-2x-2y = 108-2\cdot18-2\cdot18 = 36$. Its dimensions are 18 inches by 18 inches by 36 inches and its volume is $11,664$ cubic inches. ◇

In Example 5 we let z be the length of a longest side, an assumption that was never used. So if the Postal Service regulations read "The length of one edge plus the girth around the other edges shall not exceed 108 inches," the effect would be the same. You would not be able to send a larger box by measuring the girth around the base formed by its largest edges.

EXAMPLE 6 Let $f(x, y) = x^2 + y^2 - 2x - 4y$. Find the maximum and minimum values of $f(x, y)$ on the disk R of radius 3 and center $(0, 0)$.

SOLUTION Find critical points. We have

$$\frac{\partial f}{\partial x} = 2x - 2 \qquad \text{and} \qquad \frac{\partial f}{\partial y} = 2y - 4.$$

The equations

$$\begin{aligned} 2x - 2 &= 0 \\ 2y - 4 &= 0 \end{aligned}$$

have the solution $x = 1$ and $y = 2$. The point $(1, 2)$ lies in R since its distance from the origin is $\sqrt{1^2 + 2^2} = \sqrt{5}$, which is less than 3. At $(1, 2)$ the value of the function is $1^2 + 2^2 - 2(1) - 4(2) = 5 - 2 - 8 = -5$.

Now we find the behavior of f on the boundary, which is a circle of radius 3. We parameterize this circle as

$$x = 3\cos(\theta), \qquad y = 3\sin(\theta) \qquad \text{for } 0 \le \theta \le 2\pi.$$

On it

$$\begin{aligned} f(x, y) &= x^2 + y^2 - 2x - 4y \\ &= (3\cos(\theta))^2 + (3\sin(\theta))^2 - 2(3\cos(\theta)) - 4(3\sin(\theta)) \\ &= 9\cos^2(\theta) + 9\sin^2(\theta) - 6\cos(\theta) - 12\sin(\theta) \\ &= 9 - 6\cos(\theta) - 12\sin(\theta). \end{aligned}$$

We now find the maximum and minimum of the single-variable function $g(\theta) = 9 - 6\cos(\theta) - 12\sin(\theta)$ for θ in $[0, 2\pi]$.

To do this, find $g'(\theta)$:

$$g'(\theta) = 6\sin\theta - 12\cos\theta.$$

Setting $g'(\theta) = 0$ gives

$$0 = 6\sin(\theta) - 12\cos(\theta)$$

or

$$\sin(\theta) = 2\cos(\theta). \tag{16.6.4}$$

To solve (16.6.4), divide by $\cos(\theta)$ (which will not be 0), getting

Why is $\cos(\theta)$ not 0?

$$\frac{\sin(\theta)}{\cos(\theta)} = 2$$

or

$$\tan(\theta) = 2.$$

There are two angles θ in $[0, 2\pi]$ such that $\tan(\theta) = 2$. One is in the first quadrant, $\theta = \arctan(2)$, and the other is in the third quadrant, $\theta = \pi + \arctan(2)$. To evaluate $g(\theta) = 9 - 6\cos(\theta) - 12\sin(\theta)$ there we must compute their cosine and sine. The right triangle in Figure 16.6.8 helps us do this.

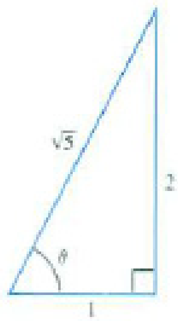


Figure 16.6.8

Figure 16.6.8 shows that for $\theta = \arctan(2)$,

$$\cos(\theta) = \frac{1}{\sqrt{5}} \qquad \text{and} \qquad \sin(\theta) = \frac{2}{\sqrt{5}}.$$

Then

$$g(\arctan(2)) \;=\; 9 - 6\left(\frac{1}{\sqrt{5}}\right) - 12\left(\frac{2}{\sqrt{5}}\right) = 9 - \frac{30}{\sqrt{5}} \approx -4.41641.$$

When $\theta = \pi + \arctan(2)$,

$$\cos(\theta) = \frac{-1}{\sqrt{5}} \qquad \text{and} \qquad \sin(\theta) = \frac{-2}{\sqrt{5}}.$$

so

$$g(\pi + \arctan(2)) \;=\; 9 - 6\left(\frac{-1}{\sqrt{5}}\right) - 12\left(\frac{-2}{\sqrt{5}}\right)$$
$$= 9 + \frac{30}{\sqrt{5}} \approx 22.41641.$$

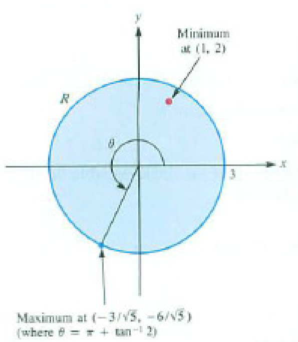


Figure 16.6.9

At the ends of $[0, 2\pi]$, $g(2\pi) = g(0) = 9 - 6(1) - 12(0) = 3$. Thus the maximum of f on the boundary of R is about 22.41641 and the minimum is about -4.41641.

At the critical point the value of f is -5. We conclude that the maximum value of f on R is about 22.41641 and the minimum value is -5. See Figure 16.6.9. $\diamond$

Proof of Theorem 16.6.2 in a Special Case

We will prove Theorem 16.6.2 when $f(x, y)$ is a second-degree polynomial.

Theorem 16.6.3. *Let $f(x, y) = Ax^2 + Bxy + Cy^2$, where A, B, and C are constants. Then $(0, 0)$ is a critical point. Let*

$$D = \frac{\partial^2 f}{\partial x^2}(0,0)\frac{\partial^2 f}{\partial y^2}(0,0) - \left(\frac{\partial^2 f}{\partial x \partial y}(0,0)\right)^2.$$

Case 1: If $D > 0$ *and* $f_{xx}(0,0) > 0$*, then* f *has a relative minimum at* $(0,0)$*.*

Case 2: If $D > 0$ *and* $f_{xx}(0,0) < 0$*, then* f *has a relative maximum at* $(0,0)$*.*

Case 3: If $D < 0$*, then* f *has neither a relative minimum nor a relative maximum at* $(0,0)$*.*

The case when $D = 0$ is addressed in Exercise 60.

Proof

We prove Case 1, leaving Cases 2 and 3 for Exercises 58 and 59.

Compute the first- and second-order partial derivatives of f:

$$\frac{\partial f}{\partial x} = 2Ax + By, \qquad \frac{\partial f}{\partial y} = Bx + 2Cy,$$

$$\frac{\partial^2 f}{\partial x^2} = 2A, \quad \frac{\partial^2 f}{\partial x \partial y} = B, \quad \frac{\partial^2 f}{\partial y^2} = 2C.$$

Both f_x and f_y are 0 at $(0,0)$. Hence $(0,0)$ is a critical point; the function has the value 0 at that input. We must show that $f(x,y) \geq 0$ for (x,y) near $(0,0)$. (In fact we will show that $f(x,y) \geq 0$ for all (x,y).)

In terms of A, B, and C, we have

$$D = f_{xx}(0,0)f_{yy}(0,0) - f_{xy}^2(0,0) = (2A)(2C) - B^2 = 4AC - B^2 > 0,$$

and $f_{xx}(0,0) = 2A > 0$. So, we are assuming that $4AC - B^2 > 0$ and $A > 0$ and want to deduce that $f(x,y) = Ax^2 + Bxy + Cy^2 \geq 0$, for (x,y) near $(0,0)$.

Since A is positive, this amounts to showing that

$$A(Ax^2 + Bxy + Cy^2) \geq 0. \tag{16.6.5}$$

We multiply by A to simplify completing the square:

$$\begin{aligned}
A(Ax^2 + Bxy + Cy^2) &= A^2x^2 + ABxy + ACy^2 \\
&= A^2x^2 + ABxy + \frac{B^2}{4}y^2 - \frac{B^2}{4}y^2 + ACy^2 \\
&= (Ax + \frac{B}{2}y)^2 + (AC - \frac{B^2}{4})y^2 \\
&= (Ax + \frac{B}{2}y)^2 + (\frac{4AC - B^2}{4})y^2.
\end{aligned}$$

We know that $\left(Ax + \frac{B}{2}y\right)^2 \geq 0$ and $y^2 \geq 0$, since they are squares of real numbers. By our assumption on D, $4AC - B^2$ is positive. Thus (16.6.5) holds for all (x,y), not only for (x,y) near $(0,0)$. Case 1 of the theorem is proved. •

Summary

We defined a critical point of $f(x, y)$ as a point where both partial derivatives f_x and f_y are 0. Even if f_{xx} and f_{yy} are negative there, such a point need not provide a relative maximum. We must also know that $|f_{xy}|$ is not too large.

If $f_{xx} < 0$ and $f_{xy}^2 < f_{xx}f_{yy}$, then there is a relative maximum at the critical point. (The two inequalities imply $f_{yy} < 0$.)

Similar criteria hold for a relative minimum: if $f_{xx} > 0$ and $f_{xy}^2 < f_{xx}f_{yy}$, then this critical point is a relative minimum.

The critical point is a saddle point when $f_{xy}^2 > f_{xx}f_{yy}$.

When $f_{xy}^2 = f_{xx}f_{yy}$, the critical point may be a relative maximum, relative minimum, or neither.

We also saw how to find extrema of a function defined on a bounded region.

EXERCISES for Section 16.6

In Exercises 1 to 10 use Theorems 16.6.1 and 16.6.2 to determine relative maxima or minima.

1. $x^2 + 3xy + y^2$

2. $x^2 - y^2$

3. $x^2 - 2xy + 2y^2 + 4x$

4. $x^4 + 8x^2 + y^2 - 4y$

5. $x^2 - xy + y^2$

6. $x^2 + 2xy + 2y^2 + 4x$

7. $2x^2 + 2xy + 5y^2 + 4x$

8. $-4x^2 - xy - 3y^2$

9. $4/x + 2/y + xy$

10. $x^3 - y^3 + 3xy$

In each of Exercises 11 to 16 let f by a function of x and y such that at (a, b) both f_x and f_y equal 0. Values are specified for f_{xx}, f_{xy}, and f_{yy} at (a, b). Assume that the partial derivatives are continuous. Decide whether, at (a, b), f has a relative maximum, a relative minimum, a saddle point, or there is not enough information to classify the critical point.

11. $f_{xy} = 4$, $f_{xx} = 2$, $f_{yy} = 8$

12. $f_{xy} = -3$, $f_{xx} = 2$, $f_{yy} = 4$

13. $f_{xy} = 3$, $f_{xx} = 2$, $f_{yy} = 4$

14. $f_{xy} = 2$, $f_{xx} = 3$, $f_{yy} = 4$

15. $f_{xy} = -2$, $f_{xx} = -3$, $f_{yy} = -4$

16. $f_{xy} = -2$, $f_{xx} = 3$, $f_{yy} = -4$

In Exercises 17 to 24 find the critical points and the relative extrema.

17. $x + y - \frac{1}{xy}$

18. $3xy - x^3 - y^3$

19. $12xy - x^3 - y^3$

20. $6xy - x^2y - xy^2$

21. $\exp(x^3 + y^3)$

22. 2^{xy}

23. $3x + xy + x^2y - 2y$

24. $x + y + \frac{8}{xy}$

25. Find the dimensions of the open rectangular box of volume 1 that has the smallest surface area. Use Theorem 16.6.2 as a check that the critical point provides a minimum.

26. The material for the top and bottom of a rectangular box costs 3 cents per square foot, and that for the sides 2 cents per square foot. What is the least expensive box that has a volume of 1 cubic foot? Use Theorem 16.6.2 as a check that the critical point provides a minimum.

27. UPS ships packages whose combined length and girth is at most 165 inches (and weigh at most 150 pounds).

(a) What are the dimensions of the package with the largest volume that it ships?

(b) What are the dimensions of the package with maximum surface area that UPS will ship?

28. Let $P_1 = (x_1, y_1)$, $P_2 = (x_2, y_2)$, $P_3 = (x_3, y_3)$, and $P_4 = (x_4, y_4)$. Find the coordinates of the point P that minimizes the sum of the squares of the distances from P to the four points.

29. Find the dimensions of the rectangular box of largest volume whose total surface area is 12 square meters.

30. Three nonnegative numbers x, y, and z have sum 1.

(a) How small can $x^2 + y^2 + z^2$ be?

(b) How large can it be?

31. Each year a firm can produce r radios and t television sets at a cost of $2r^2 + rt + 2t^2$ dollars. It sells a radio for \$600 and a television set for \$900.

(a) What is the profit from the sale of r radios and t television sets? Profit is revenue less cost.

(b) Find the combination of r and t that maximizes profit. Use the discriminant as a check.

32. Find the dimensions of the rectangular box of largest volume that can be inscribed in a sphere of radius 1.

33. For what values of k does $x^2 + kxy + 3y^2$ have a relative minimum at $(0, 0)$?

34. For what values of k does the function $kx^2+5xy+4y^2$ have a relative minimum at $(0, 0)$?

35. Let $f(x, y) = (2x^2 + y^2)e^{-x^2-y^2}$.

(a) Find all critical points of f.

(b) Examine the behavior of f when (x, y) is far from the origin.

(c) What is the minimum value of f?

(d) What is the maximum value of f?

36. Find the maximum and minimum values of the function in Exercise 35 on the circle

(a) $x^2 + y^2 = 1$,

(b) $x^2 + y^2 = 4$.

(Express f in terms of θ.)

37. Find the maximum value of $f(x, y) = 3x^2 - 4y^2 + 2xy$ for points (x, y) in the square region whose vertices are $(0, 0)$, $(0, 1)$, $(1, 0)$, and $(1, 1)$.

38. Find the maximum value of $f(x, y) = xy$ for points (x, y) in the triangular region whose vertices are $(0, 0)$, $(1, 0)$, and $(0, 1)$.

39. Maximize the function $-x+3y+6$ on the region bounded by the quadrilateral whose vertices are $(1, 1)$, $(4, 2)$, $(0, 3)$, and $(5, 6)$.

40.

(a) Show that $z = x^2 - y^2 + 2xy$ has no maximum and no minimum.

(b) Find the minimum and maximum of z if we consider only (x, y) on the circle of radius 1 and center $(0, 0)$. That is all (x, y) such that $x^2 + y^2 = 1$.

(c) Find the minimum and maximum of z if we consider all (x, y) in the disk of radius 1 and center $(0, 0)$. That is, all (x, y) such that $x^2 + y^2 \leq 1$.

41. Suppose z is a function of x and y with continuous second partial derivatives. If, at (x_0, y_0), $z_x = 0 = z_y$, $z_{xx} = 3$, and $z_{yy} = 12$, for what values of z_{xy} is it certain that z has a relative minimum at (x_0, y_0)?

42. Let $U(x, y, z) = x^{1/2}y^{1/3}z^{1/6}$ be the utility to a consumer of the amounts x, y, and z of three commodities. Their prices are, respectively, 2 dollars, 1 dollar, and 5 dollars, and the consumer has 60 dollars to spend. How much of each product should he buy to maximize the utility?

43. This exercise shows that if the discriminant D is 0, then any of the three outcomes mentioned in Theorem 16.6.2 are possible.

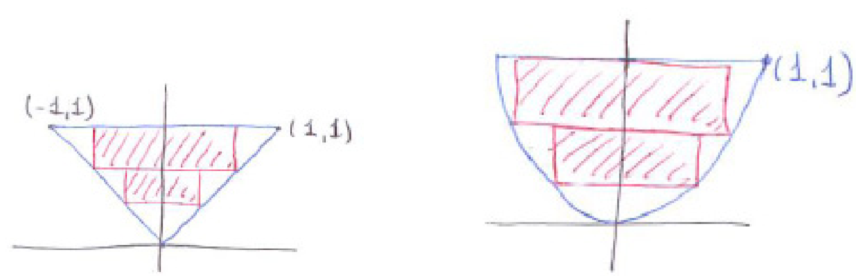
(a) Let $f(x, y) = x^2 + 2xy + y^2$. Show that at $(0, 0)$ both f_x and f_y are 0, f_{xx} and f_{yy} are positive, $D = 0$, and f has a relative minimum.

(b) Let $f(x, y) = x^2 + 2xy + y^2 - x^4$. Show that at $(0, 0)$ both f_x and f_y are 0, f_{xx} and f_{yy} are positive, $D = 0$, and f has neither a relative maximum nor a relative minimum.

(c) Give an example of a function $f(x, y)$ for which $(0, 0)$ is a critical point and $D = 0$ there, but f has a relative maximum at $(0, 0)$.

44. Let $f(x, y) = ax + by + c$ for constants a, b, and c. Let R be a polygon in the xy plane. Show that the maximum and minimum values of $f(x, y)$ on R are assumed only at vertices of the polygon.

45. Two rectangles are placed in the triangle whose vertices are $(0, 0)$, $(1, 1)$, and $(-1, 1)$ as shown in Figure 16.6.10(a).



(a) (b)

Figure 16.6.10

Show that they can fill as much as 2/3 of the area of the triangle.

46. Two rectangles are placed in the region bounded by the line $y = 1$ and the parabola $y = x^2$ as shown in Figure 16.6.10(b). How large can their total area be?

47. Let $P_0 = (a, b, c)$ be a point not on the surface $f(x, y, z) = 0$. Let P be the point on the surface nearest P_0. Show that $\overrightarrow{PP_0}$ is perpendicular to the surface at P. (Show it is perpendicular to each curve on the surface that passes through P.)

48. Let (x_1, y_1), (x_2, y_2), $\dots$, (x_n, y_n) be n points in the plane. Statisticians define the **line of regression** as the line that minimizes the sum of the squares of the differences between y_i and the ordinates of the line at x_i. (See Figure 16.6.11.) Let a line in the plane have the equation $y = mx + b$.

(a) Show that the line of regression minimizes $\sum_{i=1}^{n} (y_i - (mx_i + b))^2$ considered as a function of m and b.

(b) Let $f(m, b) = \sum_{i=1}^{n} (y_i - (mx_i + b))^2$. Compute f_m and f_b.

(c) Show that when $f_m = 0 = f_b$, we have

$$m \sum_{i=1}^{n} x_i^2 + b \sum_{i=1}^{n} x_1 = \sum_{i=1}^{n} x_i y_i$$

and

$$m \sum_{i=1}^{n} x_i + nb = \sum_{i=1}^{n} y_i.$$

(d) When do the simultaneous equations in (c) have a unique solution for m and b?

(e) Find the regression line for the points $(1, 1)$, $(2, 3)$, and $(3, 5)$.

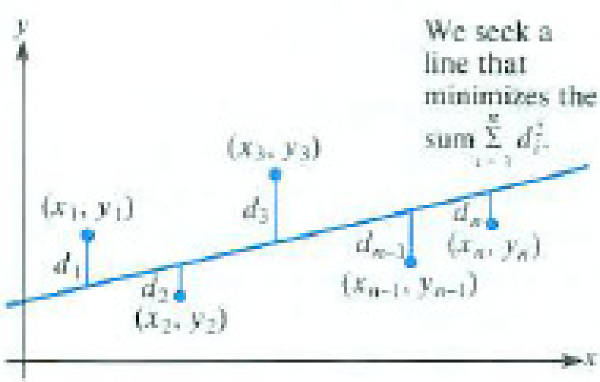


Figure 16.6.11

49. If your calculator is programmed to compute lines of regression, find and draw the line of regression for the points $(1, 1)$, $(2, 1.5)$, $(3, 3)$, $(4, 2)$ and $(5, 3.5)$.

50. Let $f(x, y) = (y - x^2)(y - 2x^2)$.

(a) Show that f has neither a local minimum nor a local maximum at $(0, 0)$.

(b) Show that f has a local minimum at $(0, 0)$ when considered only on any line through $(0, 0)$. (Graph $y = x^2$ and $y = 2x^2$ and show where $f(x, y)$ is positive and where it is negative.

51. Find (a) the minimum value of xyz, and (b) the maximum value of xyz, for nonnegative real numbers x, y, z such that $x + y + z = 1$.

52.

(a) Deduce from Exercise 51 that for three nonnegative numbers a, b, and c, $\sqrt[3]{abc} \leq (a + b + c)/3$. This shows that the **geometric mean** of three numbers is not larger than their **arithmetic mean**.

(b) Obtain a corresponding result for four numbers.

53. The dimensions of a box are x, y, and z. Its girth plus length is at most 165 inches. If you are free to choose which dimension is the length, which would you choose if you wanted to maximize the volume of the box? Assume $x < y < z$.

54. A surface is called **closed** when it is the boundary of a region R, as a balloon is the boundary of the air within it. A surface is called **smooth** when it has a continuous outward unit normal vector at each point of the surface. Let S be a smooth closed surface bounding a region R. Show that for a point P_0 in R, there are at least two points on S such that $\overrightarrow{P_0P}$ is normal to S. It is conjectured that if P_0 is the centroid of R, then there are at least four points on S such that P_0P is normal to S. The **centroid** of a triangle is the point where its three medians meet.

55. Find the point P on the plane $Ax + By + Cz + D = 0$ nearest the point $P_0 = (x_0, y_0, z_0)$, which is not on that plane.

(a) Find P by calculus.

(b) Find P by using the algebra of vectors. (Why is $\overrightarrow{P_0P}$ perpendicular to the plane?)

56. This exercise outlines the proof of Theorem 16.6.3 when $f_{xx}(a,b) > 0$ and $f_{xx}(a,b)f_{yy}(a,b) - (f_{xy})^2(a,b) > 0$. Assuming that f_{xx}, f_{yy}, and f_{xy} are continuous, we know by the permanence principle that f_{xx} and $f_{xx}f_{yy} - (f_{xy})^2$ remain positive on a some disk R whose center is (a,b). The following steps show that f has a minimum (a,b) on each line L through (a,b). Let $\mathbf{u} = \cos(\theta)\mathbf{i} + \sin(\theta)\mathbf{j}$ be a unit vector. Show that $D_{\mathbf{u}}(D_{\mathbf{u}}f)$ is positive on the part of L that lies in the disk R.

(a) Show that $D_{\mathbf{u}}f(a,b) = 0$.

(b) Show that $D_{\mathbf{u}}(D_{\mathbf{u}}f) = f_{xx}\cos^2(\theta) + 2f_{xy}\sin(\theta)\cos(\theta) + f_{yy}\sin^2(\theta)$.

(c) Show that $(f_{xx})D_u(D_uf) = (f_{xx}\cos(\theta) + f_{xy}\sin(\theta))^2 + \left(f_{xx}f_{yy} - (f_{xy})^2\right)\sin^2(\theta)$.

(d) Deduce from (b) that f is concave up on the part of each line through (a,b) inside the disk R.

(e) Deduce that the graph of f lies above its tangent lines at $(a,b,f(a,b))$, so f has a relative minimum at (a,b).

Exercise 57 provides another motivation for the definition of the Fourier series of a function f defined on the interval $[0,2\pi]$.

57. For an integer n let

$$S(x) = \frac{a_0}{2} + \sum_{k=1}^{n} \left(a_k\cos(kx) + b_k\sin(kx)\right).$$

Let $f(x)$ be a continuous function defined on $[0,2\pi]$. The definite integral

$$\int_0^{2\pi} \left(f(x) - S(x)\right)^2 \, dx$$

is a measure of how close $S(x)$ is to $f(x)$ on the interval $[0,2\pi]$. The integral can never be negative. (Why?) The smaller the integral, the better S approximates f on $[0,2\pi]$. Show that the S that minimizes the integral is a partial sum of the Fourier series associated with $f(x)$.

58. Prove Case 2 of Theorem 16.6.3.

59. Prove Case 3 of Theorem 16.6.3.

60. Show that if $B^2 - 4AC = 0$, then C/A is positive and $Ax^2 + Bxy + Cy^2 = A\left(x + \sqrt{C/A}\right)^2$. What does this imply about the critical point $(0,0)$?

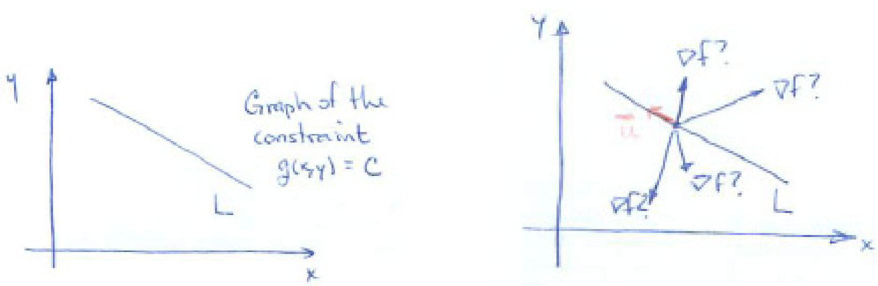
16.7 Lagrange Multipliers

Another method of finding maxima or minima of a function is due to Joseph Louis Lagrange (1736–1813). It makes use of the fact that a gradient of a function is perpendicular to its level curves (or level surfaces).

The Essence of the Method

We first consider a simple case. Suppose we want to find a maximum or a minimum of $f(x,y)$ for points (x,y) on the line L that has the equation $g(x,y)=C$. See Figure 16.7.1(a).)

Assume the extremum occurs at (a,b). Let ∇f be the gradient of f evaluated at (a,b). What can we say about its direction? (See Figure 16.7.1(b).)



(a) (b)

Figure 16.7.1

Suppose that ∇f is not perpendicular to L. Let $\mathbf{u}$ be a unit vector parallel to L. Then $D_{\mathbf{u}}f = (\nabla f)\cdot\mathbf{u}$ is not 0. If $D_{\mathbf{u}}f$ is positive then $f(x,y)$ is increasing in the direction $\mathbf{u}$, which is along L. In the direction $-\mathbf{u}$, $f(x,y)$ is decreasing. Therefore the point (a,b) could not provide either a maximum or a minimum. That means ∇f must be perpendicular to L. Since $g(x,y)=C$ is a level curve of g, ∇g is perpendicular to L. So ∇f and ∇g are parallel and there is a scalar λ such that

$$\nabla f = \lambda \nabla g. \tag{16.7.1}$$

The scalar λ is called a **Lagrange multiplier**.

λ, pronounced *lambda*, is the Greek letter corresponding to the lowercase letter l.

EXAMPLE 1 Find the minimum of x^2+2y^2 on the line $x+y=2$.

SOLUTION Since x^2+2y^2 increases without bound in both directions along the line it must have a minimum somewhere.

Here $f(x,y)=x^2+2y^2$ and $g(x,y)=x+y$; so

$$\nabla f = 2x\mathbf{i}+4y\mathbf{j} \qquad \text{and} \qquad \nabla g = \mathbf{i}+\mathbf{j}.$$

At the minimum, the gradients of f and g are parallel. That is, there is a scalar λ such that

$$\nabla f = \lambda \nabla g.$$

This implies that

$$2x\mathbf{i} + 4y\mathbf{j} = \lambda(\mathbf{i} + \mathbf{j}).$$

This single vector equation leads to the two scalar equations

$$\begin{array}{rcll} 2x & = & \lambda & \qquad \text{(equating } \mathbf{i} \text{ components)} \\ 4y & = & \lambda & \qquad \text{(equating } \mathbf{j} \text{ components)} \end{array} \tag{16.7.2}$$

But we also know that

$$x + y = 2. \tag{16.7.3}$$

From (16.7.2), $2x = 4y$ or $x = 2y$. Substituting into (16.7.3) gives $2y + y = 2$ or $y = 2/3$, hence $x = 2y = 4/3$. The minimum is $f\left(\frac{4}{3}, \frac{2}{3}\right) = \left(\frac{4}{3}\right)^2 + 2\left(\frac{2}{3}\right)^2 = \frac{24}{9} = \frac{8}{3}$. There is no need to find λ. ◇

The General Method

Let us see why Lagrange's method works when the constraint is not a line, but a curve. The problem is

Maximize or minimize $u = f(x, y)$, given the constraint $g(x, y) = k$.

The graph of $g(x, y) = k$ is in general a curve C, as shown in Figure 16.7.2. Assume that f, considered only on points of C, takes a maximum (or minimum) value at P_0. Let C be parameterized by $\mathbf{G}(t) = x(t)\mathbf{i} + y(t)\mathbf{j}$. Let $\mathbf{G}(t_0) = \overrightarrow{OP_0}$. Then u is a function of t:

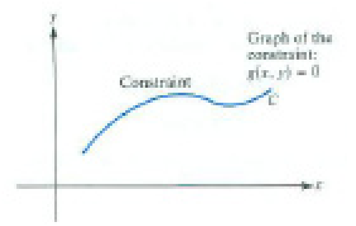


Figure 16.7.2

$$u = f(x(t), y(t)),$$

and, as shown in the proof of Theorem 16.5.1 of Section 16.5,

$$\frac{du}{dt} = \nabla f \cdot \mathbf{G}'(t_0). \tag{16.7.4}$$

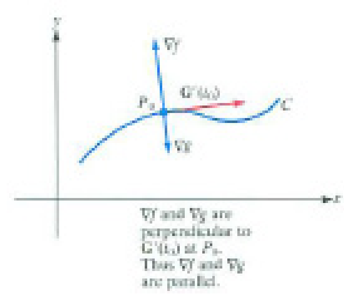


Figure 16.7.3

Since f, considered only on C, has an extremum at $\mathbf{G}(t_0)$,

$$\frac{du}{dt} = 0 \qquad \text{at } t = 0.$$

Thus, by (16.7.4),

$$\nabla f \cdot \mathbf{G}'(t_0) = 0.$$

This means that ∇f is perpendicular to $\mathbf{G}'(t_0)$ at P_0. But ∇g, evaluated at P_0, is also perpendicular to $\mathbf{G}'(t_0)$, since the gradient ∇g is perpendicular to the level curve $g(x, y) = 0$. (We assume that ∇g is not $\mathbf{0}$.) See Figure 16.7.3. Thus

$\boxed{\nabla f \text{ is parallel to } \nabla g \text{ so there is a scalar } \lambda \text{ such that } \nabla f = \lambda \nabla g.}$

EXAMPLE 2 Maximize the function x^2y for points (x, y) on the unit circle $x^2 + y^2 = 1$.

SOLUTION Let $g(x, y) = x^2 + y^2$. We wish to maximize $f(x, y) = x^2y$ for points on the circle $g(x, y) = 1$. Then

$$\nabla f = \nabla(x^2y) = 2xy\mathbf{i} + x^2\mathbf{j}$$

and

$$\nabla g = \nabla(x^2 + y^2) = 2x\mathbf{i} + 2y\mathbf{j}.$$

At an extreme point of f, $\nabla f = \lambda \nabla g$ for some scalar λ. This gives us two scalar equations:

$$\begin{aligned} 2xy &= \lambda(2x) \qquad (\mathbf{i} \text{ component}) \qquad (16.7.5)\\ x^2 &= \lambda(2y) \qquad (\mathbf{j} \text{ component}) \qquad (16.7.6) \end{aligned}$$

The third equation is the constraint,

$$x^2 + y^2 = 1. \qquad (16.7.7)$$

Since the maximum does not occur when $x = 0$, we may assume x is not 0. Dividing both sides of (16.7.5) by x, we get $2y = 2\lambda$ or $y = \lambda$. Thus (16.7.6) becomes

$$x^2 = 2y^2. \qquad (16.7.8)$$

Combining this with (16.7.7), we have

$$2y^2 + y^2 = 1$$

or

$$y^2 = \frac{1}{3}.$$

Thus

$$y = \frac{\sqrt{3}}{3} \qquad \text{or} \qquad y = -\frac{\sqrt{3}}{3}.$$

By (16.7.8),

$$x = \sqrt{2}y \qquad \text{or} \qquad x = -\sqrt{2}y.$$

There are four points to be considered on the circle:

$$\left(\frac{\sqrt{6}}{3}, \frac{\sqrt{3}}{3}\right), \left(\frac{-\sqrt{6}}{3}, \frac{\sqrt{3}}{3}\right), \left(\frac{-\sqrt{6}}{3}, \frac{-\sqrt{3}}{3}\right), \left(\frac{\sqrt{6}}{3}, \frac{-\sqrt{3}}{3}\right).$$

At the first and second x^2y is positive, while at the third and fourth x^2y is negative. The first two points provide the maximum value of x^2y on the circle $x^2 + y^2 = 1$,

$$\left(\frac{\sqrt{6}}{3}\right)^2 \frac{\sqrt{3}}{3} = \frac{2\sqrt{3}}{9}.$$

The third and fourth points provide the minimum value of x^2y,

$$\frac{-2\sqrt{3}}{9}.$$

◇

More Variables

In the preceding examples we examined the maximum and minimum of $f(x, y)$ on a curve $g(x, y) = k$. The same method works for finding extreme values of $f(x, y, z)$ on a surface $g(x, y, z) = k$. If f has, say, a minimum at (a, b, c), then it does on any level curve on the surface $g(x, y, z) = k$. Thus ∇f is perpendicular to any curve on the surface through P. But so is ∇g. Thus ∇f and ∇g are parallel, and there is a scalar λ such that the $\nabla f = \lambda \nabla g$. So we will have four scalar equations: three from the vector equation $\nabla f = \lambda \nabla g$ and one from the constraint $g(x, y, z) = k$. That gives four equations in four unknowns, x, y, z and λ, but it is not necessary to find λ.

EXAMPLE 3 Find the rectangular box with the largest volume, if its surface area is 96 square feet.

SOLUTION Let the dimensions be x, y and z and the volume be V, which equals xyz. The surface area is $2xy + 2xz + 2yz$. See Figure 16.7.4.

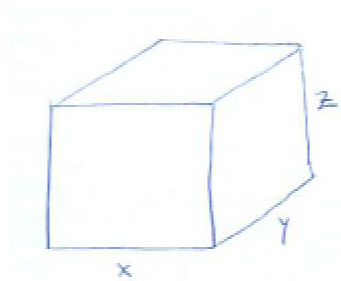


Figure 16.7.4

We wish to maximize $V(x, y, z) = xyz$ subject to the constraint

$$g(x, y, z) = 2xy + 2xz + 2yz = 96. \qquad (16.7.9)$$

We have

$$\nabla V = yz\mathbf{i} + xz\mathbf{j} + xy\mathbf{k}$$

and

$$\nabla g = (2y + 2z)\mathbf{i} + (2x + 2z)\mathbf{j} + (2x + 2y)\mathbf{k}.$$

The vector equation $\nabla V = \lambda \nabla g$ provides three scalar equations

$$\begin{aligned} yz &= \lambda(2y + 2z) && (16.7.10) \\ xz &= \lambda(2x + 2z) && (16.7.11) \\ xy &= \lambda(2x + 2y) \end{aligned}$$

The fourth equation is the constraint,

$$2xy + 2xz + 2yz = 96.$$

Solving for λ in (16.7.10) and in (16.7.11), and equating the results gives

$$\frac{yz}{2y+2z} = \frac{xz}{2x+2z}.$$

Since $z \neq 0$, we have

$$\frac{y}{2y+2z} = \frac{x}{2x+2z}.$$

Clearing denominators gives

$$\begin{array}{rrcl} & 2xy + 2yz & = & 2xy + 2xz, \\ \text{hence,} & 2yz & = & 2xz. \end{array}$$

Since $z \neq 0$, we conclude that

$$x = y.$$

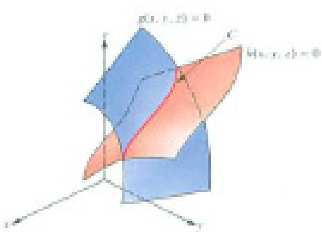


Figure 16.7.5

Since x, y and z play the same roles in both the volume xyz and in the surface area, $2(xy + xz + yz)$, we conclude also that

$$x = z.$$

Then $x = y = z$. The box of maximum volume is a cube.

To find its dimensions use the constraint, which tells us that $6x^2 = 96$ or $x = 4$. Hence y and z are 4 also. $\diamond$

More Constraints

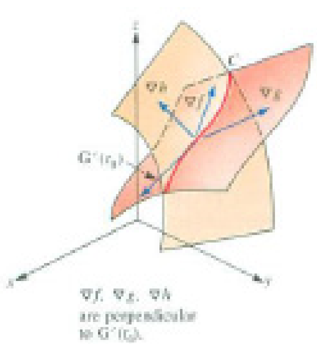


Figure 16.7.6

Lagrange multipliers can also be used to maximize $f(x, y, z)$ subject to more than one constraint. The constraints could be

$$g(x, y, z) = k_1 \qquad \text{and} \qquad h(x, y, z) = k_2. \tag{16.7.12}$$

The two surfaces (16.7.12) in general meet in a curve C, as shown in Figure 16.7.5. Assume that C is parameterized by the function $\mathbf{G}$. Then at a maximum (or minimum) of f at a point $P_0(x_0, y_0, z_0)$ on C,

$$\nabla f \cdot \mathbf{G}'(t_0) = 0.$$

Thus ∇f, evaluated at P_0, is perpendicular to $\mathbf{G}'(t_0)$. But ∇g and ∇h, being normal vectors at P_0 to the level surfaces $g(x, y, z) = k_1$ and $h(x, y, z) = k_2$, respectively, are both perpendicular to $\mathbf{G}'(t_0)$. Thus

∇f, ∇g, and ∇h are all perpendicular to $\mathbf{G}'(t_0)$ at (x_0, y_0, z_0).

(See Figure 16.7.6.) Consequently, ∇f lies in the plane determined by the vectors ∇g and ∇h which we assume are not parallel. Hence there are scalars λ and μ such that

$$\nabla f = \lambda \nabla g + \mu \nabla h.$$

mu, mew, is Greek for the letter M.

This vector equation provides three scalar equations in λ, μ, x, y, z. The two constraints give two more equations. So we have five equations in five unknowns. We find λ and μ only if they assist the algebra.

If a maximum occurs at an endpoint of the curves or if the two surfaces do not meet in a curve or if ∇g and ∇h are parallel, this method does not apply. We will content ourselves by illustrating the method with an example in which there are two constraints. A rigorous development of the material in this section belongs in an advanced calculus course.

EXAMPLE 4 Minimize the quantity $x^2+y^2+z^2$ subject to the constraints $x+2y+3z=6$ and $x+3y+9z=9$.

SOLUTION There are three variables and two constraints. Each constraint is a plane. Together they give a line. The function $x^2+y^2+z^2$ is the square of the distance from (x,y,z) to the origin. So the problem can be rephrased as how far is the origin from a certain line? (It could be solved by vector algebra. See Exercises 19 and 20.) When viewed this way, the problem certainly has a solution so we know there is a minimum.

We have

$$\begin{aligned} f(x,y,z) &= x^2+y^2+z^2 \\ g(x,y,z) &= x+2y+3z \\ h(x,y,z) &= x+3y+9z. \end{aligned}$$

Thus

$$\begin{aligned} \nabla f &= 2x\mathbf{i}+2y\mathbf{j}+2z\mathbf{k} \\ \nabla g &= \mathbf{i}+2\mathbf{j}+3\mathbf{k} \\ \nabla h &= \mathbf{i}+3\mathbf{j}+9\mathbf{k}. \end{aligned}$$

There are constants λ and μ such that

$$\nabla f = \lambda \nabla g + \mu \nabla h.$$

Therefore, the five equations for x, y, z, λ, and μ are

$$2x = \lambda + \mu \qquad (16.7.13)$$

$$2y = 2\lambda + 3\mu \qquad (16.7.14)$$

$$2z = 3\lambda + 9\mu \qquad (16.7.15)$$

$$x+2y+3z = 6 \qquad (16.7.16)$$

$$x+3y+9z = 9 \qquad (16.7.17)$$

Another way is to use software programs that solve simultaneous linear equations.

There are several ways to solve them.

One way is to use the first three equations to express x, y, and z in terms of λ and μ. Then substitute in the last two equations, getting two simultaneous equations in two unknowns.

By (16.7.13), (16.7.14), and (16.7.15),

$$x = \frac{\lambda + \mu}{2}, \qquad y = \frac{2\lambda + 3\mu}{2}, \qquad z = \frac{3\lambda + 9\mu}{2}.$$

Equations (16.7.16) and (16.7.17) then become

$$\frac{\lambda + \mu}{2} + \frac{2(2\lambda + 3\mu)}{2} + \frac{3(3\lambda + 9\mu)}{2} = 6$$

and

$$\frac{\lambda + \mu}{2} + \frac{3(2\lambda + 3\mu)}{2} + \frac{9(3\lambda + 9\mu)}{2} = 9,$$

which simplify to

$$14\lambda + 34\mu = 12 \qquad (16.7.18)$$

and

$$34\lambda + 91\mu = 18. \qquad (16.7.19)$$

Solving (16.7.18) and (16.7.19) gives

$$\lambda = \frac{240}{59} \qquad \mu = -\frac{78}{59}.$$

Thus

$$\begin{aligned} x &= \frac{\lambda + \mu}{2} = \frac{81}{59} \approx 1.37288 \\ y &= \frac{2\lambda + 3\mu}{2} = \frac{123}{59} \approx 2.08475 \\ z &= \frac{3\lambda + 9\mu}{2} = \frac{9}{59} \approx 0.15254. \end{aligned}$$

Since there is no maximum, this must be a minimum. Why?

The minimum of $x^2 + y^2 + z^2$ is thus

$$\left(\frac{81}{59}\right)^2 + \left(\frac{123}{59}\right)^2 + \left(\frac{9}{59}\right)^2 = \frac{21,771}{3,481} = \frac{369}{59} \approx 6.24542.$$

◇

In Example 4 there were three variables, x, y, and z, and two constraints. There may be many variables, x_1, x_2, $\ldots x_n$, and many constraints. If there are m constraints, g_1, $g_2 \ldots g_m$, introduce Lagrange multipliers λ_1, λ_2, $\ldots \lambda_m$, one for each constraint. So there would be $m + n$ equations, n from

$$\nabla f = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2 + \cdots + \lambda_m \nabla g_m$$

and m more from the m constraints. There would be $m+n$ unknowns, λ_1, λ_2, $\ldots$, λ_m, x_1, x_2, $\ldots$, x_n.

Summary

The basic idea of Lagrange multipliers is that if $f(x, y, z)$ has an extreme value on the surface $g(x, y, z) = C$, then ∇f and ∇g are parallel where the extreme value occurs. If there are two constraints $g(x, y, z) = k_1$ and $h(x, y, z) = k_2$, then ∇f lies in the plane of ∇g and ∇h. In the first case there is a scalar λ such that $\nabla f = \lambda \nabla g$. In the second case, there are scalars λ and μ such that $\nabla f = \lambda \nabla g + \mu \nabla h$. These vector equations, together with the constraints, provide simultaneous scalar equations, which must then be solved. Similarly, if $f(x, y)$ has an extremum at P_0 on the curve $g(x, y) = k$, then ∇f and ∇g are parallel.

EXERCISES for Section 16.7

In the exercises use Lagrange multipliers unless told otherwise.

1. Maximize xy for points on the circle $x^2 + y^2 = 4$.

2. Minimize $x^2 + y^2$ for points on the line $2x + 3y = 6$.

3. Minimize $2x+3y$ on the portion of the hyperbola $xy = 1$ in the first quadrant.

4. Maximize $x + 2y$ on the ellipse $x^2 + y^2 = 8$.

5. Find the largest area of rectangles whose perimeters are 12 centimeters.

6. A rectangular box is to have a volume of 1 cubic meter. Find its dimensions if its surface area is minimal.

7. Find the point on the plane $x + 2y + 3z = 6$ that is closest to the origin. (Minimize the square of the distance to avoid square roots.)

8. Maximize $x + y + 2z$ on the sphere $x^2 + y^2 + z^2 = 9$.

9. Minimize the distance from (x, y, z) to $(1, 3, 2)$ for points on the plane $2x+y+z = 5$.

10. Find the dimensions of the box of largest volume whose surface area is 6 square inches.

11. Maximize $x^2y^2z^2$ subject to $x^2 + y^2 + z^2 = 1$.

12. Find the points on the surface $xyz = 1$ closest to the origin.

13. Minimize $x^2 + y^2 + z^2$ on the line common to the two planes $x + 2y + 3z = 0$ and $2x + 3y + z = 4$.

14. The plane $2y + 4z - 5 = 0$ meets the cone $z^2 = 4(x^2 + y^2)$ in a curve. Find the point on it nearest the origin.

In Exercises 15 to 18 solve the exercise in Section 16.5 by Lagrange multipliers.

15. Exercise 25

16. Exercise 26

17. Exercise 29

18. Exercise 30

19. Solve Example 4 by vector algebra.

20. Solve Exercise 13 by vector algebra.

21.

(a) Sketch the elliptical paraboloid $z = x^2 + 2y^2$.

(b) Sketch the plane $x + y + z = 1$.

(c) Sketch the intersection of the surfaces in (a) and (b).

(d) Find the highest point on the intersection in (c).

22.

(a) Sketch the ellipsoid $x^2 + y^2/4 + z^2/9 = 1$ and the point $P(2, 1, 3)$.

(b) Find the point Q on the ellipsoid that is nearest P.

(c) What is the angle between PQ and the tangent plane at Q?

23.

(a) Sketch the hyperboloid $x^2 - y^2/4 - z^4/9 = 1$. How many sheets does it have?

(b) Plot the point $(1, 1, 1)$. Is it inside or outside the hyperboloid?

(c) Find the point on the hyperboloid nearest P.

24. Maximize $x^3 + y^3 + 2z^3$ on the intersection of the spheres $x^2 + y^2 + z^2 = 4$ and $(x - 3)^2 + y^2 + z^2 = 4$.

25. Show that a triangle in which the product of the sines of the three angles is maximized is equilateral.

26. Solve Exercise 25 by labeling the angles x,y, and $\pi - x - y$ and minimizing a function of x and y by the method of Section 16.6.

27. Maximize $x + 2y + 3z$ subject to the constraints $x^2 + y^2 + z^2 = 2$ and $x + y + z = 0$.

28.

(a) Maximize $x_1 x_2 \cdots x_n$ subject to the constraints that $\sum_{i=1}^n x_i = 1$ and all $x_i \geq 0$.

(b) Deduce that for nonnegative numbers a_1, a_2, ..., a_n, $\sqrt[n]{a_1 a_2 \cdots a_n} \leq (a_1 + a_2 + \cdots + a_n)/n$. The **geometric mean** is less than or equal to the **arithmetic mean**.

29.

(a) Maximize $\sum_{i=1}^n x_i y_i$ subject to the constraints $\sum_{i=1}^n x_i^2 = 1$ and $\sum_{i=1}^n y_i^2 = 1$.

(b) Deduce the **Cauchy-Schwarz inequality**: for numbers a_1, a_2, ..., a_n and $b_1, b_2, \dots b_n$,

$$\sum_{i=1}^n a_i b_i \leq \left(\sum_{i=1}^n a_i^2\right)^{1/2} \left(\sum_{i=1}^n b_i^2\right)^{1/2}.$$

(Let $x_i = \frac{a_i}{(\sum_{i=1}^n a_i^2)^{1/2}}$ and $y_1 = \frac{b_i}{(\sum_{i=1}^n b_i^2)^{1/2}}$. See also the Average Speed CIE at the end of Chapter 7.)

(c) How would you justify the inequality in (b), for $n = 3$, using vectors?

30. Let $a_1, a_2, \dots, a_n$ be fixed nonzero numbers. Maximize $\sum_{i=1}^n a_i x_i$ subject to $\sum_{i=1}^n x_i^2 = 1$.

31. Let p and q be positive numbers that satisfy the equation $1/p + 1/q = 1$. Obtain Hölder's inequality for nonnegative numbers a_i and b_i,

$$\sum_{i=1}^n a_i b_i \leq \left(\sum_{i=1}^n a_i^p\right)^{1/p} \left(\sum_{i=1}^n b_i^q\right)^{1/q},$$

as follows.

(a) Maximize $\sum_{i=1}^n x_i y_i$ subject to $\sum_{i=1}^n x_i^p = 1$ and $\sum_{i=1}^n y_i^q = 1$.

(b) By letting $x_i = \frac{a_i}{(\sum_{i=1}^n a_i^p)^{1/p}}$ and $y_i = \frac{b_i}{(\sum_{i=1}^n b_i^q)^{1/q}}$, obtain Hölder's inequality.

Hölder's inequality, with $p = 2$ and $q = 2$, reduces to the Cauchy-Schwarz inequality in Exercise 29.

16.8 Mappings and Their Magnification

This section explores mappings from one set to another. In the next chapter mappings will be used to evaluate integrals over surfaces and solid regions.

Mappings

A **mapping** or **transformation** is a one-to-one function. For instance, the function that assigns to each number θ in $[0, 2\pi)$ the point $(x, y) = (\cos(\theta), \sin(\theta))$ is a mapping from the interval $[0, 2\pi)$ to the unit circle with center at $(0, 0)$.

We have met other examples, including the function that assigns to points in the xy-plane points on a surface. To be specific, let $\mathcal{R}$ be a region in the plane and $f(x, y)$ be a scalar function defined on $\mathcal{R}$. Let $F(x, y) = (x, y, f(x, y))$, a point above (or below) (x, y) on the surfce $\mathcal{S}$ whose equation is $z = f(x, y)$.

The projection of a slide on a screen is a mapping. It projects a point on the slide to a point on the screen.

If the mapping F pairs points in $\mathcal{R}$ with points in $\mathcal{S}$, $\mathcal{S}$ is called the **image** of $\mathcal{R}$ and we write $F(\mathcal{R}) = \mathcal{S}$. $\mathcal{S}$ consists of all points $F(P)$ where P is in $\mathcal{R}$. F has an inverse, from $\mathcal{S}$ to $\mathcal{R}$, denoted $\mathsf{inv}F$ or F^{-1}.

EXAMPLE 1 Let F be the mapping that assigns to the point (u, v) in the uv-plane the point $(2u, 3v)$ in the xy-plane, so $F(u, v) = (2u, 3v)$.

(a) Describe the mapping geometrically.

(b) Find the formula for $\mathsf{inv}F$.

(c) Show that the image of a line is a line.

(d) Find the image of the square whose vertices are $(0, 0)$, $(1, 0)$, $(1, 1)$, and $(0, 1)$.

(e) Find the image of the disk, $u^2 + v^2 \leq 1$.

SOLUTION

(a) If $F(u, v) = (x, y)$, we have $x = 2u$ and $y = 3v$. That implies that F magnifies horizontal distances by a factor 2 and vertical distances by a factor 3.

(b) To find a formula for $\mathsf{inv}F$, solve for u and v in terms of x and y. Because $x = 2u$ and $y = 3v$, we have

$$u = \frac{x}{2} \qquad \text{and} \qquad v = \frac{y}{3}.$$

Thus $\mathsf{inv}F$ maps (x, y) to $(x/2, y/3)$.

(c) A line L in the uv-plane has an equation of the form $au + bv + c = 0$ where not both a and b are zero. If (x, y) is in the image of the line $(\mathsf{inv}F)(x, y)$ lies on L. Thus $(x/2, y/3)$ lies on L in the uv-plane, which implies that

$$a\frac{x}{2} + b\frac{y}{3} + c = 0.$$

Clearing denominators, we conclude that the image of L, $F(L)$, is described by the equation $3ax + 2by + 6c = 0$ Thus, the image of L is another line.

(d) The square $\mathcal{R}$ in Figure 16.8.1 is bounded by four lines. So its image is also bounded by four lines. F takes the four corners of the square to $F(0, 0) = (0, 0)$, $F(1, 0) = (2, 0)$, $F(1, 1) = (2, 3)$, and $F(0, 1) = (0, 3)$. So the image is the rectangle $\mathcal{S}$ As expected, the area of $\mathcal{S}$ is six times the area of the square $\mathcal{R}$.

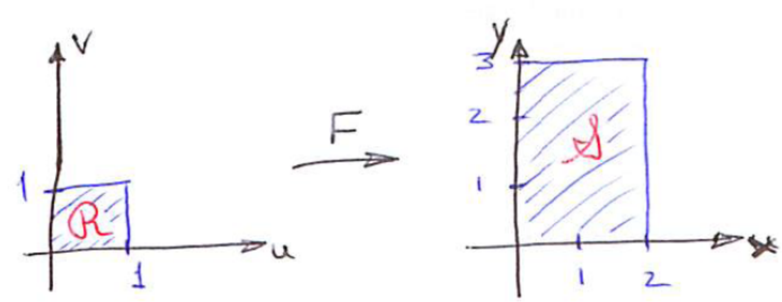


Figure 16.8.1

(e) If $u^2 + v^2 \leq 1$ and $(x, y) = F(u, v)$, then $(x/2)^2 + (y/3)^2 \leq 1$ for a point (x, y) in the image of the disk. So points in the image satisfy the inequality

$$\frac{x^2}{2^2} + \frac{y^2}{3^2} \leq 1.$$

It is the set bounded by the ellipse $x^2/2^2 + y^2/3^2 = 1$, shown in Figure 16.8.2.

◇

EXAMPLE 2 A parallelogram $\mathcal{S}$ in the xy-plane is bounded by two level curves of $x + y$, namely $x + y = 1$ and $x + y = 4$, and by two level curves of $y - 2x$, namely $y - 2x = 2$ and $y - 2x = 3$. Find a rectangle $\mathcal{R}$ in the uv-plane and a mapping F such that $F(\mathcal{R}) = \mathcal{S}$. (See Figure 16.8.3.)

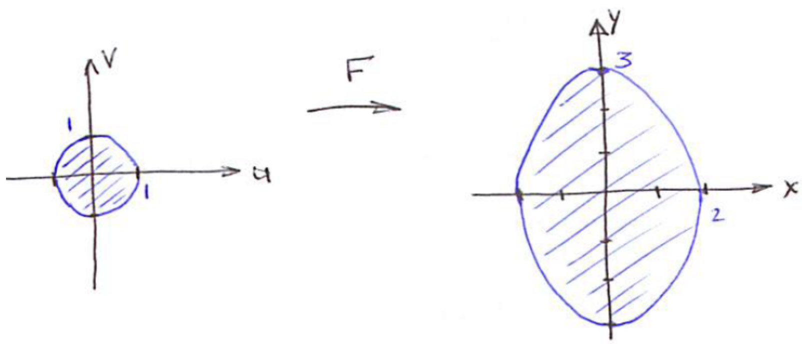


Figure 16.8.2 Because F magnifies areas by a factor 6, the area of the ellipse is 6π.

SOLUTION On $\mathcal{S}$, $x+y$ is in the interval from 1 to 4. So let $u = x+y$, which is in the interval $[1, 4]$. Similarly, $y - 2x$ goes from 2 to 3. So let $v = y - 2x$. The equations

$$u = x + y, \qquad v = y - 2x \tag{16.8.1}$$

describe a mapping that maps the parallelogram $\mathcal{S}$ to the rectangle $\mathcal{R}$, described by $1 \leq u \leq 4$, $2 \leq v \leq 3$. Thus $\mathsf{inv}F$ maps $\mathcal{R}$ to $\mathcal{S}$. Solving (16.8.1) for x and y in terms of u and v, we find that

$$x = \frac{u}{3} - \frac{v}{3} \qquad y = \frac{2u}{3} + \frac{v}{3}.$$

So the mapping F given by $F(u, v) = \left(\frac{u}{3} - \frac{v}{3}, \frac{2u}{3} + \frac{v}{3}\right)$ maps the rectangle $\mathcal{R}$ onto the parallelogram $\mathcal{S}$, as shown in Figure 16.8.3. ◇

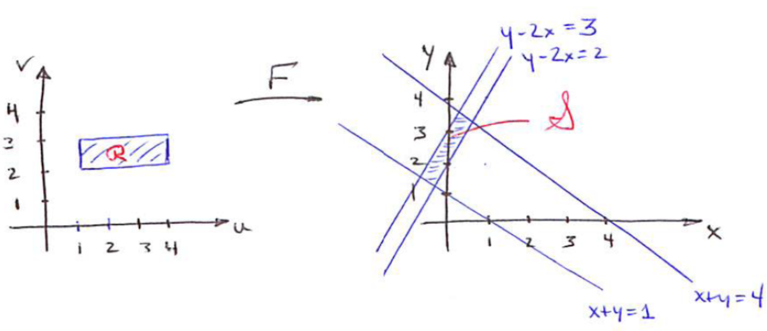


Figure 16.8.3

A function of the form $F(x, y) = (au + bv, cu + dv)$, where $ad - bc$ is not 0, is called a **linear mapping**. It takes lines to lines and the origin to the origin.

The next example provides a fresh perspective on polar coordinates.

EXAMPLE 3 Let $F(u,v) = (u\cos(v), u\sin(v))$ and let $\mathcal{R}$ be the rectangle $1 \le u \le 2$, $\pi/6 \le v \le \pi/4$. Sketch the image of $\mathcal{R}$, $F(\mathcal{R})$.

SOLUTION We have $x = u\cos(v)$, $y = \sin(v)$. Then $x^2 + y^2 = u^2\cos^2(v) + u^2\sin^2(v) = u^2$. We see that $u^2 = r^2$,where r is part of the polar coordinates (r,θ) for (x,y). Thus, $u = r$. Similarly v is the angle θ of polar coordinates.

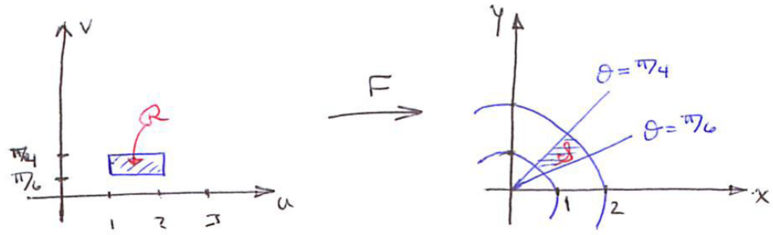


Figure 16.8.4

Thus the image of $\mathcal{R}$ is bounded by circles of radii 1 and 2 and by rays at angles $\pi/6$ and $\pi/4$, shown in Figure 16.8.4. ◇

If, in Example 3, we had used the letters r and θ instead of u and v, we would see that r and θ are rectangular coordinates in the $r\theta$-plane. The mapping F makes them the polar coordinates in the xy-plane.

This is typical of a mapping. That is why a mapping is also called a **coordinate transformation**. The point (x,y) is assigned the coordinates in the uv-plane of $(\mathsf{inv}F)(x,y)$.

To put it another way, a mapping lifts the coordinates from the uv-plane, where they are ordinary rectangular coordinates, and places them like tags on the xy-plane, where they no longer appear as rectangular coordinates.

Magnification of a Mapping

The mapping F given by $F(u,v) = (2u, 3v)$ magnifies areas by a factor 6. The magnification of other mappings may vary from point to point, and can be thought of as the local magnification. This leads to the question, If $F(u,v) = (f(u,v), g(u,v))$, by what factor does it magnify or shrink the area of a small patch near a point(u_0, v_0)

We will determine by what factor the mapping F magnifies the area of a small rectangle near (u_0, v_0).

For positive changes Δu in u and Δv in v consider the rectangle B in the uv-plane whose vertices are (u_0, v_0), $(u_0 + \Delta u, v_0)$, $(u_0 + \Delta u, v_0 + \Delta v)$, and $(u_0, v_0 + \Delta v)$ shown in Figure 16.8.5. Its area is $\Delta u \Delta v$. The image $F(B)$,

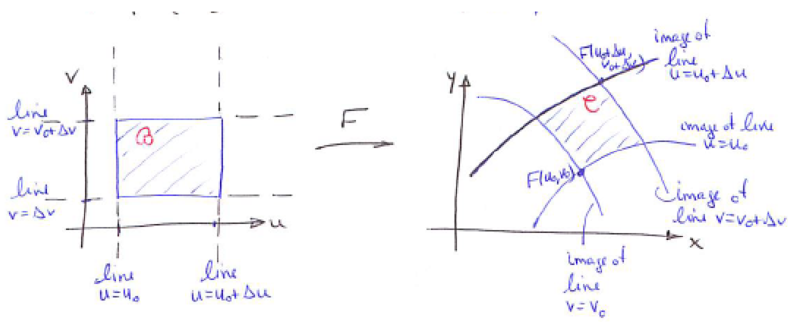


Figure 16.8.5

which we call C, is bounded by curves that are the images of the edges of B. They need not be straight lines.

The magnification of F at (u_0, v_0) is defined as a limit:

$$\lim_{\Delta u, \Delta v \to 0} \frac{\text{Area of } C}{\text{Area of } B} = \lim_{\Delta u, \Delta v \to 0} \frac{\text{Area of } C}{\Delta u \Delta v}.$$

To express the limit in terms of the components $f(u, v)$ and $g(u, v)$, we will estimate the area of C. We will use vectors and the fact that $|\mathbf{A} \times \mathbf{B}|$ is the area of the parallelogram spanned by the vectors $\mathbf{A}$ and $\mathbf{B}$.

Let O be the origin of the xy-coordinates and $\mathbf{r}(u, v) = \overrightarrow{OF(u, v)}$. Let $\Delta \mathbf{r} = \mathbf{r}(u_0 + \Delta u, v_0) - \mathbf{r}(u_0, v_0)$ be the change in $\mathbf{r}$ due to a change Δu in u. Then $\Delta \mathbf{r}/\Delta u$ approximates $\partial \mathbf{r}/\partial u$ evaluated at (u_0, v_0). If follows that $\partial \mathbf{r}/\partial u$ approximates $\Delta \mathbf{r}/\Delta u$. Thus

$$\Delta \mathbf{r} \approx \left(\frac{\partial \mathbf{r}}{\partial u} \right) \Delta u.$$

Similarly, $(\partial \mathbf{r}/\partial v)\Delta v$ approximates the change in $\mathbf{r}$ due to the change Δv in v.

With these observations, we are ready to estimate the area of C.

The vector $(\partial \mathbf{r}/\partial u)\Delta u$ is tangent to the curve between $F(u_0, v_0)$ and $F(u_0 + \Delta u, v_0)$ and approximates the vector $\mathbf{r}(u_0 + \Delta u, v_0) - \mathbf{r}(u_0, v_0)$. Similarly, $(\partial \mathbf{r}/\partial v)\Delta v$ approximates the vector from $F(u_0, v_0)$ to $F(u_0, v_0 + \Delta v)$. Then the area of the parallelogram spanned by the vectors $(\partial \mathbf{r}/\partial u)\Delta u$ and $(\partial \mathbf{r}/\partial v)\Delta v$

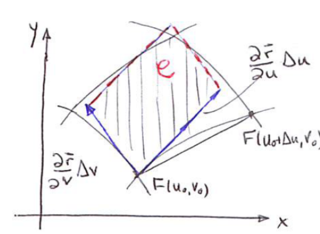


Figure 16.8.6

approximates the area of C. For this reason we expect that

$$\begin{aligned}
\text{Magnification of } F \text{ at } (u_0, v_0) &= \lim_{\Delta u, \Delta v \to 0} \frac{\text{Area of parallelogram}}{\Delta u \Delta v} \\
&= \lim_{\Delta u, \Delta v \to 0} \frac{\left|\frac{\partial \mathbf{r}}{\partial u}\Delta u \times \frac{\partial \mathbf{r}}{\partial v}\Delta v\right|}{\Delta u \Delta v} \\
&= \lim_{\Delta u, \Delta v \to 0} \frac{\left|\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}\right| \Delta u \Delta v}{\Delta u \Delta v} \\
&= \left|\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}\right|.
\end{aligned}$$

Because $\partial\mathbf{r}/\partial u$ and $\partial\mathbf{r}/\partial v$ are evaluated at (u_0, v_0) we have

$$\boxed{\text{Magnification of } F \text{ at } (u_0, v_0) = \left|\frac{\partial \mathbf{r}}{\partial u}(u_0, v_0) \times \frac{\partial \mathbf{r}}{\partial v}(u_0, v_0)\right|.}$$

To express this in terms of $f(u, v)$ and $g(u, v)$, we use $\mathbf{r}(u, v) = \overrightarrow{OF}(u, v) = f(u, v)\mathbf{i} + g(u, v)\mathbf{j}$. So

$$\begin{aligned}
\text{Magnification of } F \text{ at } (u_0, v_0) &= \left|\det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial f}{\partial u} & \frac{\partial g}{\partial u} & 0 \\ \frac{\partial f}{\partial v} & \frac{\partial g}{\partial v} & 0 \end{pmatrix}\right| \\
&= \left|0\mathbf{i} + 0\mathbf{j} + \begin{vmatrix} \frac{\partial f}{\partial u} & \frac{\partial g}{\partial u} \\ \frac{\partial f}{\partial v} & \frac{\partial g}{\partial v} \end{vmatrix} \mathbf{k}\right| \\
&= \left|\left(\frac{\partial f}{\partial u}\frac{\partial g}{\partial v} - \frac{\partial f}{\partial v}\frac{\partial g}{\partial u}\right)\mathbf{k}\right| \\
&= \left|\frac{\partial f}{\partial u}\frac{\partial g}{\partial v} - \frac{\partial f}{\partial v}\frac{\partial g}{\partial u}\right|.
\end{aligned}$$

The magnification is the absolute value of the determinant of the 2×2 matrix

$$\begin{pmatrix} \frac{\partial f}{\partial u} & \frac{\partial g}{\partial u} \\ \frac{\partial f}{\partial v} & \frac{\partial g}{\partial v} \end{pmatrix}$$

Both the matrix and its determinant are called the **Jacobian** of the mapping F. The most descriptive notation is $J[F](Q)$, indicating that the determinant depends on the transformation F, and the mapping is evaluated at Q. In most situations we will abbreviate it by J, $J(u, v)$, or $\frac{\partial(x,y)}{\partial(u,v)}$. The absolute value of the determinant is the magnification.

Where the local magnification is greater than 1, F increases area.

If the magnification is less than 1, it shrinks area. It may sound odd to treat shrinking area as a form of magnification, but we don't complain when we go down in an elevator or when a football team gains -3 yards on a play.

Examples of Magnification

We conclude this introduction to magnification with two examples that will re-appear in the next chapter.

EXAMPLE 4 Find the Jacobian of the mapping F in Example 3, $x = u\cos(v)$, $y = u\sin(v)$.

SOLUTION Because $f(u,v) = u\cos(v)$ and $g(u,v) = u\sin(v)$, the Jacobian is

$$\begin{vmatrix} \frac{\partial(u\cos(v))}{\partial u} & \frac{\partial(u\sin(v))}{\partial u} \\ \frac{\partial(u\cos(v))}{\partial v} & \frac{\partial(u\sin(v))}{\partial v} \end{vmatrix} = \begin{vmatrix} \cos(v) & \sin(v) \\ -u\sin(v) & \cos(v) \end{vmatrix} = u\cos^2(v) + u\sin^2(v) = u.$$

◇

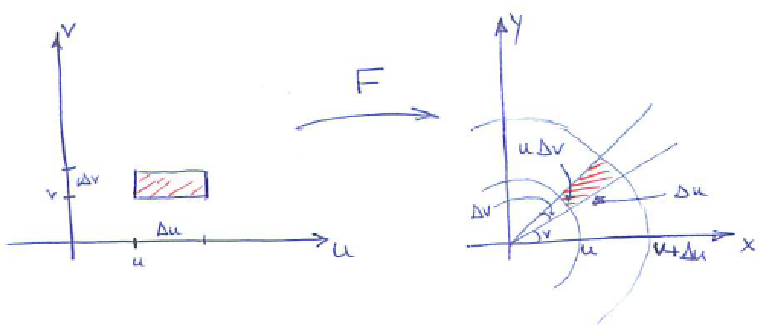


Figure 16.8.7

Figure 16.8.7 shows that the magnification in Example 4 is to be expected. In the uv-plane, u and v serve as rectangular coordinates. A small patch in the uv-plane determined by Δu and Δv is a rectangle of area $\Delta u\ \Delta v$. Its image in the xy-plane has two straight sides and two curved sides, and resembles a rectangle with sides of lengths $u\ \Delta v$ and Δu, which has area $u\ \Delta u\ \Delta v$.

In the next example a mapping goes from a plane to a surface that might not be a plane.

EXAMPLE 5 Let $f(x,y)$ be a function defined in a region $\mathcal{R}$ in the xy-plane. Let $\mathcal{S}$ be the surface $z = f(x,y)$ above or below $\mathcal{R}$. Let $F(x,y) = (x, y, f(x,y))$, so F is a mapping from $\mathcal{R}$ to $\mathcal{S}$. Find its magnification.

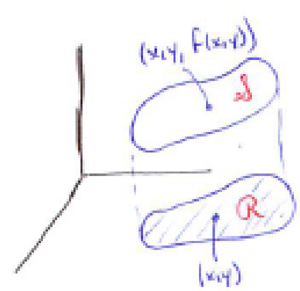


Figure 16.8.8

SOLUTION The mapping F is shown in Figure 16.8.8. If F is the inverse of the projection that will be used in Section 17.7: We have

$$\mathbf{r}(x,y) = x\mathbf{i} + y\mathbf{j} + f(x,y)\mathbf{k}.$$

The magnification is the length of $\partial\mathbf{r}/\partial x \times \partial\mathbf{r}/\partial y$. The cross product is

$$\det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial x}{\partial x} & \frac{\partial y}{\partial x} & \frac{\partial f}{\partial x} \\ \frac{\partial x}{\partial y} & \frac{\partial y}{\partial y} & \frac{\partial f}{\partial y} \end{pmatrix} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & \frac{\partial f}{\partial x} \\ 0 & 1 & \frac{\partial f}{\partial y} \end{vmatrix} = -\frac{\partial f}{\partial x}\mathbf{i} - \frac{\partial f}{\partial y}\mathbf{j} + \mathbf{k},$$

so the magnification is

$$\left|-\frac{\partial f}{\partial x}\mathbf{i} - \frac{\partial f}{\partial y}\mathbf{j} + \mathbf{k}\right| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + 1}.$$

This formula will be used in the next chapter when computing integrals over surfaces. ◇

Summary

We introduced the notion of a mapping, which is a one-to-one function from one set to another or even to the same set. Every mapping has a local magnification. For a mapping from the uv-plane to the xy-plane, $F(u,v) = (f(u,v), g(u,v))$, the magnification is the absolute value of the Jacobian determinant

$$\begin{vmatrix} \frac{\partial f}{\partial u} & \frac{\partial g}{\partial u} \\ \frac{\partial f}{\partial v} & \frac{\partial g}{\partial v} \end{vmatrix}.$$

In general, the magnification is $\left|\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}\right|$ where $\mathbf{r}(u,v) = \overrightarrow{OF}(u,v)$. We used a parallelogram to approximate the image of a small rectangle. A mapping provides a way to introduce a coordinate system; for instance, if $F(u,v) = (x,y)$, (u,v) can be used as coordinates of the point when rectangular coordinates are (x,y).

EXERCISES for Section 16.8

1. Use the Jacobian to find the magnification of the mapping $F(u, v) = (2u, 3v)$. As we noticed in Example 1, it should be 6.

2. Let $F(u, v) = (2u + 3v, 3u - 4v) = (x, y)$.

(a) Solve for u and v in terms of x and y.

(b) Why is F a mapping?

(c) What is its magnification?

3. Let F be the mapping defined in Exercise 2.

(a) Find the formula for $\mathsf{inv}F$, which goes from the xy-plane to the uv-plane.

(b) Compute $F(F^{-1}(x, y))$.

4. Let $F(u, v) = (u - v, u + v)$.

(a) Draw $F(\mathcal{R})$ if $\mathcal{R}$ is the square whose vertices are $(0, 0)$, $(1, 0)$, $(1, 1)$, and $(0, 1)$.

(b) Find the Jacobian of F.

(c) Using (b), determine the area of $F(R)$. Assume that if the magnification of a mapping is constant with value k, it magnifies all areas by k. This is justified in Exercise 9 in Section 17.9.

5. Let $F(u, v) = (e^u \cos(v), e^u \sin(v))$.

(a) Draw $F(\mathcal{R})$ for the rectangle $\mathcal{R}$ with $0 \le u \le 1$ and $0 \le \theta \le \pi/6$.

(b) Find the Jacobian of F.

6. Let $F(u, v) = (2u + 3v, u - v) = (x, y)$.

(a) Find the Jacobian of F.

(b) Find $\mathsf{inv}F$ by solving for u and v in terms of x and y.

(c) Find the Jacobian of $\mathsf{inv}F$.

(d) The product of the Jacobians of F and $\mathsf{inv}F$ is 1. Is that to be expected or is it a coincidence? (See Exercise 24.)

7. The set $\mathcal{S}$ is bounded by the curves $y = x + 3$, $y = x + 4$, $y = -x + 5$, and $y = -x + 6$.

(a) Find a rectangle $\mathcal{R}$ in the uv-plane and a mapping F from $\mathcal{R}$ to $\mathcal{S}$.

(b) Find the magnification of F.

(c) Use (c) to find the area of the parallelogram $\mathcal{S}$.

8.

(a) Find the area of the parallelogram in Example 2 by using the mapping F constructed there.

(b) Find the same area by expressing it in terms of the cross product of two vectors.

9. Find a mapping from uv-space to xy-space such that the image of the square with vertices $(0, 0)$, $(1, 0)$, $(1, 1)$, and $(0, 1)$ is the parallelogram with vertices $(0, 0)$, $(2, 1)$, $(2, 5)$, and $(0, 4)$.

Exercises 10 to 12 help us to understand why mappings are required to be one-to-one.

10. Let a, b, c, and d be constants such that $ad - bc$ is not zero. Define $F(u, v)$ to be $(au + bv, cu + dv) = (x, y)$.

(a) Show that $F(0, 0) = (0, 0)$.

(b) Show that F takes a line to a line.

(c) Show that F is a mapping, that is, show that it is one-to-one. (See Exercise 12.)

(d) Find the magnification of F.

11. Let $F(u, v) = (2u + 6v, u + 3v)$.

(a) Sketch $F(u, v)$ for $(u, v) = (0, 1)$, $(1, 0)$, $(2, 3)$, and $(4, 1)$.

(b) Show that for all (u, v), $F(u, v)$ lies on a line L in the xy-plane.

12. Let a, b, c, and d be constants. Define $F(u, v)$ to be $(au+bv, cu+dv) = (x, y)$. In Exercise 10 it was shown that F is a mapping when $ad - bc \neq 0$. Show that if $ad - bc = 0$, then F is not a mapping.

13. Let $\mathbf{r}$ be a function of u and v, $\mathbf{r}(u, v)$.

(a) Define $\frac{\partial \mathbf{r}}{\partial u}$ at (u_0, v_0) as a limit.

(b) Explain why $\frac{\partial \mathbf{r}}{\partial u}$ is approximately $\frac{\Delta \mathbf{r}}{\Delta u}$ for small Δu.

14. Let $x = f(u)$ be a scalar function of a single variable. Assume its domain is $[a, b]$ on the u-axis, its range is $[c, d]$ of the x-axis, and f is one-to-one. (See Figure 16.8.9.)

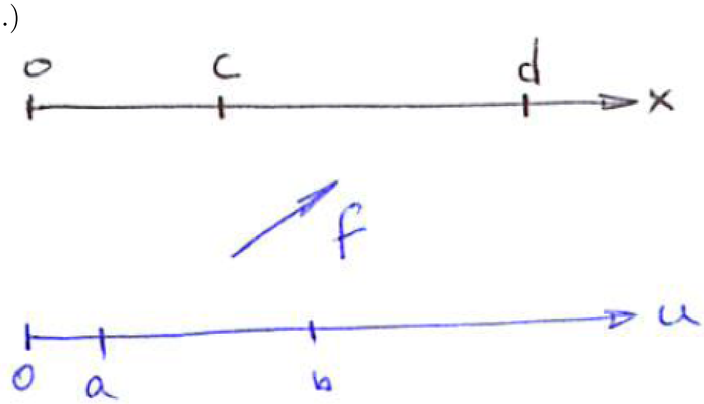


Figure 16.8.9

(a) How would you define its magnification at u_0 in $[a, b]$?

(b) How would you calculate it?

(c) Let $f(k) = u^3$. What is its magnification at $x = 2$?

15.

(a) Sketch the parabolas $\mathcal{R}$, whose equation is $v = u^2/4$, and $\mathcal{S}$, whose equation is $y = 4x^2$.

(b) Do they seem to have the same shape, with one being a larger version of the other?

(c) Find a mapping F from uv-space to xy-space such that $F(\mathcal{R}) = \mathcal{S}$ with $F(u, v) = (ku, kv)$ for some constant k.

(d) What does (c) imply about (b)?

16. Let $F(x, y) = (x, y, xy)$ for x, $y > 0$.

(a) Describe F geometrically, with a diagram.

(b) Explain why F is one-to-one.

(c) Find the Jacobian of F.

17. Let $F(uv) = (u+v, 3u-v)$ describe the mapping from uv-space to xy-space.

(a) Find the magnification of F.

(b) Sketch the image of at least five points from the circle $u^2 + v^2 = 4$.

(c) Use the data in (b) to sketch the image of the disk $u^2 + v^2 \leq 4$.

(d) What is the area of the image of the disk $u^2 + v^2 \leq 4$?

18. Let $\mathcal{S}$ be the region with in xy-space bounded by the lines $y = x$, $y = 2x$, $x + y = 3$, and $x + y = 5$.

(a) Draw $\mathcal{S}$.

(b) Find a rectangle $\mathcal{R}$ in uv-space and a mapping F such that $F(\mathcal{R}) = \mathcal{S}$. (Start by expressing the lines as level curves of two suitable functions.)

19. Let $\mathcal{R}$ be the rectangle in uv-space given by $0 \leq u \leq 2\pi$, $0 < v < \pi$. Let $F(u, v) = (a\sin(v)\cos(u), a\sin(v)\sin(u), a\cos(v))$, a point in xyz-space.

(a) Show that $F(\mathcal{R})$ is the sphere of radius a and center at the origin, but without its north pole at $(\rho, \theta, \phi) = (a, 0, 0)$, and its south pole at $(\rho, \theta, \phi) = (a, 0, \pi)$.

(b) Show that F is one-to-one.

(c) Show that on the closed rectangle, $0 \leq u \leq 2\pi$, $0 \leq v \leq \pi$, F is not one-to-one.

(d) Find the Jacobian of F.

(e) What coordinate system is F putting on the sphere? (Look at $x^2 + y^2 + z^2$.)

20.

(a) Make a large drawing of the octant of the sphere of radius a with center $(0, 0, 0)$ where x, y, and z are non-negative. On this section the latitude varies from 0, at the equator, to $\pi/2$ at the north pole, $(0, 0, a)$. The longitude varies from 0, at the xz-plane, to $\pi/2$, at the yz-plane. Let p be the latitude and q the longitude of a point on this surface.

(b) Show that the rectangular coordinates of the point with latitude p and longitude q are $x = a\cos(p)\cos(q)$, $y = a\cos(p)\sin(q)$, and $z = a\sin(p)$.

(c) Show that the magnification of the mapping from the pq-plane to the sphere is $a^2\cos(p)$. (Review Example 5, which involves a mapping from a plane to a surface.)

21. In Exercise 20 the magnification was computed using a determinant. Find the magnification directly by estimating the area of the image on the sphere of a small rectangular patch in the pq-plane whose sides have lengths Δp and Δq. (See Figure 16.8.7 and the comments after Example 4.)

22. (The formula developed in this exercise is used in Exercise 23.) Let **A**, **B**, and **C** be vectors in space such that they span a parallelepiped, as shown in Figure 16.8.10.

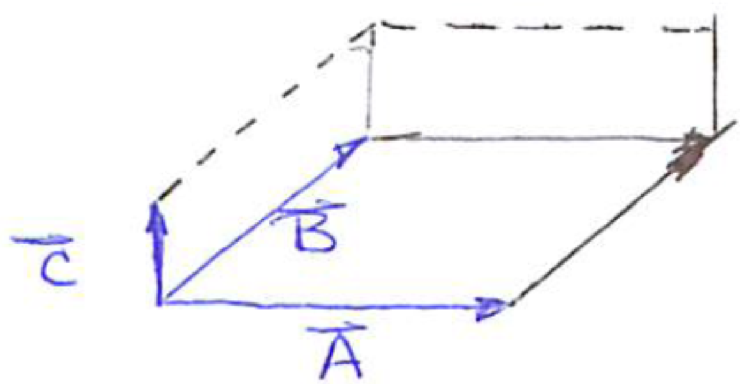


Figure 16.8.10

(a) Explain why the volume of the parallelepiped is $|\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})|$.

(b) Let $\mathbf{A} = \langle a_1, a_2, a_3\rangle$, $\mathbf{B} = \langle b_1, b_2, b_3\rangle$, and $\mathbf{C} = \langle c_1, c_2, c_3\rangle$. Show that the volume in (a) is the absolute value of the determinant

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}.$$

23. Let $\mathbf{F}$ be a mapping from uvw-space to xyz-space: $\mathbf{F}(u,v,w) = (f(u,v,w), g(u,v,w), h$

(a) Define the magnification of F at the point (u_0, v_0, w_0) in terms of a limit.

(b) Show that that limit is the absolute value of the determinant

$$\begin{vmatrix} \frac{\partial f}{\partial u} & \frac{\partial g}{\partial u} & \frac{\partial h}{\partial u} \\ \frac{\partial f}{\partial v} & \frac{\partial g}{\partial v} & \frac{\partial h}{\partial v} \\ \frac{\partial f}{\partial w} & \frac{\partial g}{\partial w} & \frac{\partial h}{\partial w} \end{vmatrix}.$$

This determinant is called the **Jacobian** of the mapping F. (Before starting, review how magnification of a mapping from uv-space to xy-space was found and expressed as a determinant. Instead of a cross product, you may need Exercise 22.)

24. Show that the product of the Jacobians of F and $\mathsf{inv}F$ is 1. (x and y are functions of u and v, but u and v are, in turn, functions of x and y. Use the chain rule, keeping in mind such things as $\partial x/\partial x = 1$ and $\partial x/\partial y = 0$.)

25. Let F be a mapping from uv-space to xy-space and G a mapping from xy-space to st-space. Then the composition $H = G \circ F$ is a mapping from uv-space to st-space.

(a) How do you think the magnification of H is related to the magnifications of F and of G?

(b) Show that your opinion in (a) is correct.

($F(u,v) = (x,y)$, where x and y are functions of u and v, and $G(x,y) = (s,t)$, where s and t are functions of x and y.)

16.9 The Chain Rule and Thermodynamics

Some basic equations of thermodynamics follow from the chain rule and the equality of mixed partial derivatives. We will describe the mathematics within the thermodynamics context. This section may serve as a review of the chain rule, or as a reference.

Implications of the Chain Rule

We start with a function of three variables, $f(x,y,z)$, which we assume has first partial derivatives

$$\left.\frac{\partial f}{\partial x}\right|_{y,z} \qquad \left.\frac{\partial f}{\partial y}\right|_{x,z} \qquad \left.\frac{\partial f}{\partial z}\right|_{x,y}.$$

The subscripts denote the variables held fixed. Without the subscripts which may become more difficult when there are many variables.

Assume that z is a function of x and y, $z = g(x,y)$. Then $f(x,y,z) = f(x,y,g(x,y))$ is a function of two variables. We call it $h(x,y)$: $h(x,y) = f(x,y,g(x,y))$. There are only two first partial derivatives of h:

$$\left.\frac{\partial h}{\partial x}\right|_{y} \quad \text{and} \quad \left.\frac{\partial h}{\partial y}\right|_{x}.$$

Let the value of $f(x,y,z)$ be called u, $u = f(x,y,z)$. But x, y, and z are functions of x and y: $x = x$, $y = y$, and $z = g(x,y)$. So $u = h(x,y)$.

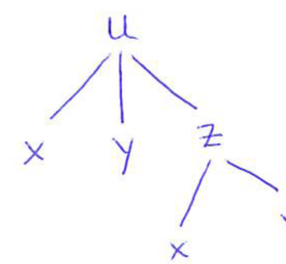


Figure 16.9.1 A change in x affects f directly and also indirectly because it causes a change in z, which also affects f.

Figure 16.9.1 provides a pictorial view of the relationship between the variables. Both x and y appear as middle and independent variables. We have $u = f(x,y,z)$ and also $u = h(x,y)$. By the chain rule

$$\left.\frac{\partial h}{\partial x}\right|_{y} = \left.\frac{\partial f}{\partial x}\right|_{y,z} \left.\frac{\partial x}{\partial x}\right|_{y} + \left.\frac{\partial f}{\partial y}\right|_{x,z} \left.\frac{\partial y}{\partial x}\right|_{y} + \left.\frac{\partial f}{\partial z}\right|_{x,y} \left.\frac{\partial g}{\partial x}\right|_{y}.$$

We know $\partial x/\partial x = 1$. Because x and y are independent variables, $\partial y/\partial x = 0$ and we have

$$\left.\frac{\partial h}{\partial x}\right|_{y} = \left.\frac{\partial f}{\partial x}\right|_{y,z} + \left.\frac{\partial f}{\partial z}\right|_{x,y} \left.\frac{\partial g}{\partial x}\right|_{y}, \tag{16.9.1}$$

or

$$\frac{\partial h}{\partial x} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial g}{\partial x}. \tag{16.9.2}$$

When the subscripts are omitted we have to look back at the definitions of f, g, and h to see which variables are held fixed.

EXAMPLE 1 We check (16.9.2) when

$$f(x,y,z) = x^2y^3z^5 \quad \text{and} \quad g(x,y) = 2x + 3y.$$

SOLUTION Computing $\partial h/\partial x$ directly gives

$$\begin{aligned}
\frac{\partial h}{\partial x} &= \frac{\partial}{\partial x}\left(x^2y^3(2x+3y)^5\right) \\
&= y^3\frac{\partial}{\partial x}\left(x^2(2x+3y)^5\right) \\
&= y^3\left(2x(2x+3y)^5 + x^2\left(5(2x+3y)^4(2)\right)\right) \\
&= 2xy^3(2x+3y)^5 + 10x^2y^3(2x+3y)^4. \qquad (16.9.3)
\end{aligned}$$

Let us find $\frac{\partial h}{\partial x}$ with the aid of (16.9.2). We have $h(x,y) = f(x,y,g(x,y)) = x^2y^3(2x+3y)^5$. Then $\dfrac{\partial f}{\partial x} = 2xy^3z^5$ and $\dfrac{\partial f}{\partial z} = 5x^2y^3z^4$. Also $\dfrac{\partial g}{\partial x} = 2$. Thus

$$\begin{aligned}
\frac{\partial h}{\partial x} &= \frac{\partial f}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial g}{\partial x} \\
&= 2xy^3z^5 + (5x^2y^3z^4)(2) \\
&= 2xy^3(2x+3y)^5 + 10x^2y^3(2x+3y)^4,
\end{aligned}$$

which agrees with (16.9.3). ◇

What If $z = g(x,y)$ Makes $f(x,y,z)$ Constant?

Assume that when z is replaced by $g(x,y)$, $h(x,y) = f(x,y,g(x,y))$ is constant: $h(x,y) = f(x,y,g(x,y)) = C$. This happens when we use the equation $f(x,y,z) = C$ to determine z implicitly as a function of x and y.

Then
$$\left.\frac{\partial h}{\partial x}\right|_y = 0 \quad \text{and} \quad \left.\frac{\partial h}{\partial y}\right|_x = 0.$$

(16.9.4) will be the foundation for deriving (16.9.8) and (16.9.9), key mathematical relationships used in thermodynamics.

In this case, which occurs frequently in thermodynamics, (16.9.1) becomes

$$\boxed{0 = \left.\frac{\partial f}{\partial x}\right|_{y,z} + \left.\frac{\partial f}{\partial z}\right|_{x,y}\left.\frac{\partial g}{\partial x}\right|_y.} \qquad (16.9.4)$$

Solving (16.9.4) for $\left.\dfrac{\partial g}{\partial x}\right|_y$ we obtain

$$\boxed{\left.\frac{\partial g}{\partial x}\right|_y = \frac{-\left.\frac{\partial f}{\partial x}\right|_{y,z}}{\left.\frac{\partial f}{\partial z}\right|_{x,y}}.} \qquad (16.9.5)$$

Equation (16.9.5) expresses the partial derivative of $g(x,y)$ with respect to x in terms of the partial derivatives of the original function $f(x,y,z)$.

EXAMPLE 2 Let $f(x,y,z) = x^3y^5z^7$. Define $z = g(x,y)$ implicitly by $x^3y^5(g(x,y))^7 = 1$. That is, $z = g(x,y) = x^{-3/7}y^{-5/7}$. Verify (16.9.5).

SOLUTION We have $\left.\frac{\partial g}{\partial x}\right|_y = \frac{-3}{7}x^{-10/7}y^{-5/7}$, $\left.\frac{\partial f}{\partial x}\right|_{y,z} = 3x^2y^5z^7$, and $\left.\frac{\partial f}{\partial z}\right|_{x,x} = 7x^3y^5z^6$. Substituting in (16.9.5), we have

$$\begin{aligned}
\frac{-\left.\frac{\partial f}{\partial x}\right|_{y,z}}{\left.\frac{\partial f}{\partial z}\right|_{x,y}} &= \frac{-\left(3x^2y^5z^7\right)}{7x^3y^5z^6} \\
&= -\frac{3}{7}x^{-1}z \\
&= -\frac{3}{7}x^{-1}x^{-3/7}y^{-5/7} \qquad \text{(because } x^3y^5z^7 = 1\text{)} \\
&= -\frac{3}{7}x^{-10/7}y^{-5/7} \\
&= \left.\frac{\partial g}{\partial x}\right|_y
\end{aligned}$$

so 16.9.5) is satisfied. ◇

The Reciprocity Relations

In a thermodynamics text you will see equations of the form

$$\left.\frac{\partial x}{\partial z}\right|_y = \frac{1}{\left.\frac{\partial z}{\partial x}\right|_y}.$$

We will explain where this comes from, presenting the mathematical details often glossed over in the applied setting. There is a function $f(x,y,z)$ with constant value C, $f(x,y,z) = C$. It is assumed that this equation determines z as a function of x and y, or, similarly, determines x as a function of y and z, or y as a function of x and z. There are six first partial derivatives:

See Exercise 5.

$$\left.\frac{\partial z}{\partial x}\right|_y, \quad \left.\frac{\partial z}{\partial y}\right|_x, \quad \left.\frac{\partial x}{\partial y}\right|_z, \quad \left.\frac{\partial x}{\partial z}\right|_y, \quad \left.\frac{\partial y}{\partial x}\right|_z, \quad \left.\frac{\partial y}{\partial z}\right|_x. \tag{16.9.6}$$

An equation analogous to (16.9.5) holds for each. For instance,

$$\left.\frac{\partial x}{\partial z}\right|_y = \frac{-\left.\frac{\partial f}{\partial z}\right|_{x,y}}{\left.\frac{\partial f}{\partial x}\right|_{y,z}}. \tag{16.9.7}$$

This is to be expected, for $\frac{\Delta z}{\Delta x}$ is the reciprocal of $\frac{\Delta x}{\Delta z}$.

Combining (16.9.5) and (16.9.7) shows that

$$\boxed{\left.\frac{\partial x}{\partial z}\right|_y = \frac{1}{\left.\frac{\partial z}{\partial x}\right|_y}.} \tag{16.9.8}$$

Equation (16.9.8) is an example of a **reciprocity relation**: The partial derivative of one variable with respect to a second variable is the reciprocal of the partial derivative of the second variable with respect to the first variable.

EXAMPLE 3 Let $f(x, y, z) = 2x + 3y + 5z = 12$. Verify that $\partial z/\partial x$ is the reciprocal of $\partial x/\partial z$.

SOLUTION Since $2x + 3y + 5z = 12$, $z = (12 - 2x - 3y)/5$. Then $\partial z/\partial x = -2/5$.

Also, $x = (12 - 3y - 5z)/2$, so $\partial x/\partial z = -5/2$, which is, as it should be, the reciprocal of $\partial z/\partial x$. ◇

The **Cyclic Relation**, also known as the Triple Product Rule, the Cyclic Chain Rule, or Euler's Chain Rule.

The Cyclic Relations

With the aid of equations like (16.9.7) we can establish the surprising relation

$$\boxed{\left.\frac{\partial x}{\partial y}\right|_z \left.\frac{\partial y}{\partial z}\right|_x \left.\frac{\partial z}{\partial x}\right|_y = -1.} \tag{16.9.9}$$

Equation (16.9.9) results from the use of three versions of (16.9.7). The left-hand side of (16.9.9) can be expressed as

$$\left(\frac{-\left.\frac{\partial f}{\partial y}\right|_{x,z}}{\left.\frac{\partial f}{\partial x}\right|_{y,z}}\right)\left(\frac{-\left.\frac{\partial f}{\partial z}\right|_{x,y}}{\left.\frac{\partial f}{\partial y}\right|_{x,z}}\right)\left(\frac{-\left.\frac{\partial f}{\partial x}\right|_{y,z}}{\left.\frac{\partial f}{\partial z}\right|_{x,y}}\right) \tag{16.9.10}$$

Cancellation reduces (16.9.10) to -1.

EXAMPLE 4 Let $f(x,y,z) = 2x+3y+5z = 12$. The equation determines implicitly each variable in terms of the others. Verify (16.9.9) in this case.
SOLUTION From $2x+3y+5z=12$,

$$x = \frac{12-3y-5z}{2} \qquad y = \frac{12-2x-5z}{3} \qquad z = \frac{12-2x-3y}{5}$$

Then $\partial x/\partial y = -3/2$, $\partial y/\partial z = -5/3$, and $\partial z/\partial x = -2/5$, and we have

$$\frac{\partial x}{\partial y}\frac{\partial y}{\partial z}\frac{\partial z}{\partial x} = \left(\frac{-3}{2}\right)\left(\frac{-5}{3}\right)\left(\frac{-2}{5}\right) = -1$$

◇

If two of the three partial derivatives in (16.9.9) are easy to calculate, then we can use (16.9.9) to find the third, which may otherwise be hard to calculate. We illustrate this use with an example from thermodynamics, where T denotes temperature, p pressure, and v the volume per unit mass.

v is the reciprocal of density

Equations (16.9.4), (16.9.8), and (16.9.9) are the essential mathematical relationships used in thermodynamics. We now show their use.

EXAMPLE 5 In van der Waal's equation p, T, and v are related by

van der Waal's equation is only one example of an **equation of state**. See also Exercises 11 and 12.

$$p = \frac{RT}{v-b} - \frac{a}{v^2}; \tag{16.9.11}$$

R, a and b are constants. Use a cyclic relation to find $(\partial v/\partial T)_p$.
SOLUTION We use the cyclic relation

Exercises 13 and 14 describe other ways to solve Example 5.

$$\left.\frac{\partial v}{\partial T}\right|_p \left.\frac{\partial T}{\partial p}\right|_v \left.\frac{\partial p}{\partial v}\right|_T = -1. \tag{16.9.12}$$

Looking at (16.9.11), we see that $(\partial p/\partial T)_v$ is easier to calculate than $(\partial T/\partial p)_v$. So (16.9.12) becomes

$$\frac{\left.\frac{\partial v}{\partial T}\right|_p \left.\frac{\partial p}{\partial v}\right|_T}{\left.\frac{\partial p}{\partial T}\right|_v} = -1$$

and therefore

$$\left.\frac{\partial v}{\partial T}\right|_p = -\frac{\left.\frac{\partial p}{\partial T}\right|_v}{\left.\frac{\partial p}{\partial v}\right|_T}. \tag{16.9.13}$$

Since p is given as a function of v and T, the numerator and denominator in (16.9.13) are

$$\left(\frac{\partial p}{\partial T}\right)_v = \frac{R}{v-b} \quad \text{and} \quad \left(\frac{\partial p}{\partial v}\right)_T = \frac{-RT}{(v-b)^2} + \frac{2a}{v^3}.$$

Thus, by (16.9.13),

$$\left(\frac{\partial v}{\partial T}\right)_p = \frac{-R/(v-b)}{-RT/(v-b)^2 + 2a/v^3}.$$

◇

Using the Equality of the Mixed Partial Derivatives

Having shown how the chain rule provides some of the basic equations in thermodynamics, let us show how the equality of the mixed partials leads to other basic equations.

We have a thermodynamic process in which the pressure is denoted by p, the temperature by T, and the volume per unit mass by v. Other common variables are

u	thermal energy per unit mass
s	entropy per unit mass
a	Helmholtz free energy per unit mass
g	Gibbs free energy per unit mass
h	enthalpy per unit mass

That is a total of eight variables. If they were independent, the states would be part of an eight-dimensional space, they are not. In fact any two determine all the others.

For instance, u may be viewed as a function of s and v, and we have $\left.\frac{\partial u}{\partial s}\right|_v$, which is the definition of temperature, T. Thermodynamic texts either state or derive the **Gibbs relation**:

When you look at a thermometer, you are gazing at the value of a partial derivative.

$$du = T\ ds - p\ dv.$$

It tells us that u can be viewed as a function of s and v, and that

$$\left.\frac{\partial u}{\partial s}\right|_v = T \quad \text{and} \quad \left.\frac{\partial u}{\partial v}\right|_s = -p.$$

Equating the mixed second partial derivatives then gives us

$$\begin{aligned}
\frac{\partial^2 u}{\partial v \partial s} &= \frac{\partial^2 u}{\partial s \partial v} && \text{(equality of mixed partials of } u(s,v))\\
\frac{\partial}{\partial v}\left(\frac{\partial u}{\partial s}\right) &= \frac{\partial}{\partial s}\left(\frac{\partial u}{\partial v}\right) \\
\left.\frac{\partial T}{\partial v}\right|_s &= \left.\frac{\partial(-p)}{\partial s}\right|_v && \left(\text{because } \left.\frac{\partial u}{\partial s}\right|_v = T \text{ and } \left.\frac{\partial u}{\partial v}\right|_s = -p\right)\\
\left.\frac{\partial T}{\partial v}\right|_s &= -\left.\frac{\partial p}{\partial s}\right|_v .
\end{aligned}$$

Several thermodynamic statements that equate two partial derivatives are obtained this way. The starting point is

$dz = M\,dx + N\,dy$ is an **exact differential**.

$$dz = M\ dx + N\ dy$$

where M is $\left.\frac{\partial z}{\partial x}\right|_y$ and N is $\left.\frac{\partial z}{\partial y}\right|_x$. Then, because

$$\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial^2 z}{\partial y \partial x},$$

it follows that

$$\left.\frac{\partial M}{\partial y}\right|_x = \left.\frac{\partial N}{\partial x}\right|_y .$$

Summary

We showed how the chain rule in the special case where an intermediate variable is also a final variable justifies certain identities, namely, the reciprocal and cyclic relations used in thermodynamics. Then we showed how the equality of the mixed partial derivatives is used to derive other equations.

EXERCISES for Section 16.9

1. Let $u = x^2 + y^2 + z^2$ and let $z = x + y$.

(a) The symbol $\dfrac{\partial u}{\partial x}$ has two interpretations. What are they?

(b) Evaluate $\dfrac{\partial u}{\partial x}$ in both cases.

(c) Using subscripts, distinguish the two partial derivatives.

2. Let $z = rst$ and let $r = st$.

(a) The symbol $\dfrac{\partial z}{\partial t}$ has two interpretations. What are they?

(b) Evaluate $\dfrac{\partial z}{\partial t}$ in both cases.

(c) What notation would distinguish these two partial derivatives?

3. Let $u = f(x, y, z)$ and $z = g(x, y)$. Then u is indirectly a function of x and of y. Express $\left.\dfrac{\partial u}{\partial x}\right|_y$ in terms of partial derivatives of f. Supply all the steps.

4. Assume that the equation $f(x, y, z) = C$, C a constant, determines x as a function of y and z: $x = h(y, z)$. Express $\left.\dfrac{\partial x}{\partial y}\right|_z$ in terms of partial derivatives of f. Supply all the steps.

5. What is the product of the six partial derivatives in (16.9.6)?

6. Using f from Example 2, verify the analog of (16.9.7) for $\left.\dfrac{\partial z}{\partial y}\right|_x$.

7. Let $f(x, y, z) = 2x + 4y + 3z$. The equation $f(x, y, z) = 7$ determines one variable as a function of the other two. Verify (16.9.7), where z is viewed as a function of x and y.

8. Obtain the cyclic relation

$$\left.\frac{\partial x}{\partial z}\right|_y \left.\frac{\partial z}{\partial y}\right|_x \left.\frac{\partial y}{\partial x}\right|_z = -1.$$

(Duplicate the steps leading to (16.9.9).)

9. Verify (16.9.9) for $f(x, y, z) = x^3y^5z^7 = 1$.

10. Verify (16.9.9) for $f(x, y, z) = 2x + 4y + 3z = 7$.

11. The equation of state for an ideal gas is $pv = RT$, where R is a constant. Find $(\partial v/\partial T)_p$.

12. The Redlich-Kwang equation

$$p = \frac{RT}{v-b} - \frac{a}{v(v+b)T^{1/2}}.$$

is an improvement upon the van der Waal's equation of state (16.9.11) for gases and liquids. Find $(\partial v/\partial T)_p$.

13. Find $(\partial v/\partial T)_p$ in Example 5 by differentiating both sides of (16.9.11) with respect to T, holding p constant.

14. One way to find $(\partial v/\partial T)_p$ in Example 5 is by first finding an equation that expresses v in terms of T and p. What difficulty occurs in this approach?

15. In Example 5, find $(\partial v/\partial p)_T$, $(\partial T/\partial v)_p$, and $(\partial T/\partial p)_v$.

16. In thermodynamics there is the Gibbs relation

$$dh = T\ ds + v\ dp.$$

It is understood that $\left.\frac{\partial h}{\partial s}\right|_p = T$ and $\left.\frac{\partial h}{\partial p}\right|_s = v$. Deduce that $\left.\frac{\partial T}{\partial p}\right|_s = \left.\frac{\partial v}{\partial s}\right|_p$.

17. Consider the thermodynamic equation

$$\left.\frac{\partial E}{\partial T}\right|_v = \left.\frac{\partial E}{\partial T}\right|_p + \left.\frac{\partial E}{\partial p}\right|_T \left.\frac{\partial p}{\partial T}\right|_v . \tag{16.9.14}$$

(a) What are the dependent variables?

(b) What are the independent variables?

(c) What are the intermediate variables?

(d) Draw a diagram showing the paths from the dependent variables to the independent variables.

(e) Use the chain rule to complete the derivation of (16.9.14).

18. Show that $\left.\dfrac{\partial p}{\partial T}\right|_v = \dfrac{-\left.\dfrac{\partial v}{\partial T}\right|_p}{\left.\dfrac{\partial v}{\partial p}\right|_T}$.

19. Show that

(a) $\left.\dfrac{\partial E}{\partial v}\right|_p = \left.\dfrac{\partial E}{\partial T}\right|_p \left.\dfrac{\partial T}{\partial v}\right|_p$

(b) $\left.\dfrac{\partial E}{\partial p}\right|_v = \left.\dfrac{\partial E}{\partial T}\right|_p \left.\dfrac{\partial T}{\partial p}\right|_v + \left.\dfrac{\partial E}{\partial p}\right|_T$.

20. Show that $\left.\dfrac{\partial p}{\partial T}\right|_v \left.\dfrac{\partial T}{\partial p}\right|_v = 1$. (Express each of the partial derivatives as a quotient of partial derivatives, as in Exercise 18.)

21. Show that van der Waal's equation, (16.9.11) in Example 5, leads to

$$\left.\frac{\partial p}{\partial T}\right|_v \left.\frac{\partial T}{\partial v}\right|_p \left.\frac{\partial v}{\partial p}\right|_T = -1.$$

22. Let $u = F(x, y, z)$ and $z = f(x, y)$. Thus u is a composite function of x and y: $u = G(x, y) = F(x, y, f(x, y))$. Assume that $G(x, y) = x^2 y$. Obtain a formula for $\dfrac{\partial f}{\partial x}$ in terms of $\dfrac{\partial F}{\partial x}$, $\dfrac{\partial F}{\partial y}$, and $\dfrac{\partial F}{\partial z}$. All three need not appear in your answer.

23. Let $u = F(x, y, z)$ and $x = f(y, z)$. Thus u is a composite function of y and z: $u = G(y, z) = F(f(y, z), y, z)$. Assume that $G(y, z) = 2y + z^2$. Obtain a formula for $\dfrac{\partial f}{\partial z}$ in terms of $\dfrac{\partial F}{\partial x}$, $\dfrac{\partial F}{\partial y}$, and $\dfrac{\partial F}{\partial z}$. All three need not appear in your answer.

24. Two functions u and v of x and y are defined implicitly by

$$F(u, v, x, y) = 0 \quad \text{and} \quad G(u, v, x, y) = 0.$$

Assuming all necessary differentiability, find a formula for $\dfrac{\partial u}{\partial x}$ in terms of partial derivatives of F and G.

16.S Chapter Summary

This chapter extended to functions of two or more variables the notion of the derivative. For a function of several variables a partial derivative is the derivative with respect to one of the variables, when the other variables are held constant.

The definition rests on a limit. The partial derivative with respect to x of $f(x, y)$ at (a, b) is

$$\frac{\partial f}{\partial x}(a,b) = \lim_{\Delta x \to 0} \frac{f(a+\Delta x, b) - f(a,b)}{\Delta x}.$$

As there are higher-order derivatives, there are higher-order partial derivatives:

$$\frac{\partial^2 f}{\partial x \partial x} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right), \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right), \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right), \text{ and } \frac{\partial^2 f}{\partial y \partial y} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right).$$

For functions usually encountered in applications, the two mixed partials, $\partial^2 f/\partial x \partial y$ and $\partial^2 f/\partial y \partial x$, are equal so the order of differentiation does not matter.

Also, for common functions we can differentiate under the integral sign:

$$\text{if } g(y) = \int_a^b f(x,y)\,dx, \text{ then } \frac{dg}{dy} = \int_a^b \frac{\partial f}{\partial y}(x,y)\,dx.$$

For a function of one variable, $f(x)$, with a continuous derivative,

$$\Delta f = f(a+\Delta x) - f(a) = f'(c)\Delta x = (f'(a)+\epsilon)\Delta x = f'(a)\Delta x + \epsilon\Delta x. \quad (16.S.1)$$

Here c is in $[a, a+\Delta x]$ and $\epsilon \to 0$ as $\Delta x \to 0$. The analog of (16.S.1) for a function of two or more variables is the basis for the chain rule for functions of several variables:

$$\begin{aligned}
\Delta f &= f(a+\Delta x, b+\Delta y) - f(a,b) \\
&= f(a+\Delta x, b+\Delta y) - f(a, b+\Delta y) + f(a, b+\Delta y) - f(a,b) \\
&= \frac{\partial f}{\partial x}(a,b)\Delta x + \frac{\partial f}{\partial y}(a,b)\Delta y + \epsilon_1 \Delta x + \epsilon_2 \Delta y,
\end{aligned}$$

where ϵ_1 and $\epsilon_2 \to 0$ as $\Delta x \to 0$ and $\Delta y \to 0$.

The chain rule showed that if $g(u)$ and $h(u)$ are differentiable functions, then $y = g(x+kt) + h(x-kt)$, k constant, satisfies the partial differential equation (PDE) $\partial^2 y/\partial t^2 = k^2 \partial^2 y/\partial x^2$. It was used by Maxwell to support his conjecture that light is an electromagnetic phenomenon.

See the CIE section on Maxwell's equations at the end of Chapter 18.

The gradient, a vector function, was defined as $\nabla f = \langle f_x, f_y \rangle$ or, for a function of three variables, $\nabla f = \langle f_x, f_y, f_z \rangle$. The gradient points in the direction a function increases most rapidly. The rate at which $f(x, y)$ changes in the direction of a unit vector $\mathbf{u}$ is $\nabla f \cdot \mathbf{u}$. The gradient evaluated at a point on the level curve $f(x, y) = k$ or level surface $f(x, y, z) = k$ is perpendicular to the level curve that passes through that point. At a critical point the gradient is zero.

For a function of one variable the sign of the second derivative helps tell whether a critical point is a maximum or a minimum. For a function of two variables, the test involves three second-order partial derivatives. The signs of f_{xx} and $f_{xx}f_{yy} - (f_{xy})^2$ are important.

The Lagrange method of finding an extremum of f subject to constraints $g_1 = 0$, $g_2 = 0$, …, $g_n = 0$ depends on the observation that at an extremum ∇f can be written as $\lambda_1 \nabla g_1 + \lambda_2 \nabla g_2 + \cdots + \lambda_n \nabla g_n$.

In Section 16.8 partial derivatives were shown to be involved in finding the magnification of a mapping.

The final section, using the chain rule and the equality of mixed partial derivatives, developed some fundamental equations in thermodynamics.

EXERCISES for 16.S

1. Let $f(x, y) = x^2 - y^2$ and $g(x, y) = 2xy$. Show that

(a) $\frac{\partial f}{\partial x} = \frac{\partial g}{\partial y}$

(b) $\frac{\partial f}{\partial y} = -\frac{\partial g}{\partial x}$

(c) $\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0$

(d) $\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} = 0$

2. Repeat Exercise 1 for $f(x, y) = \ln\left(\sqrt{x^2 + y^2}\right)$ and $g(x, y) = \arctan(y/x)$.

3. In estimating the value of a right circular cylinder tree trunk, there may be a 5 percent error in estimating the diameter and a 3 percent error in measuring the height. How large an error may occur in estimating the volume?

4. Let T denote the time it takes for a pendulum to complete a back-and-forth swing. If its length is L and g is the acceleration due to gravity, then

$$T = 2\pi\sqrt{\frac{L}{g}}.$$

A 3 percent error may be made in measuring L and a 2 percent error in measuring g. How large an error may we make in estimating T?

5. Let $u = f(x, y, z)$ and $\mathbf{r} = \mathbf{G}(t)$. Then u is a composite function of t. Show that

$$\frac{du}{dt} = \nabla f \cdot \mathbf{G}'(t),$$

where ∇f is evaluated at $\mathbf{G}(t)$. For instance, let $u = f(x, y, z)$ and let $\mathbf{G}$ describe the path of a particle. Then the rate of change in the temperature on the path is the dot product of the temperature gradient ∇f and the velocity vector $\mathbf{v} = \mathbf{G}'$.

In Exercises 6 to 12 assume the functions are defined throughout the xy-plane and have continuous partial derivatives.

6. The function $3x + g(y)$, for any differentiable function $g(y)$, satisfies the partial differential equation $\partial f/\partial x = 3$. Are there any other solutions to that equation? Explain your answer.

7. Find all functions f such that $\partial f/\partial x = 3$ and also $\partial f/\partial x = 3$.

8. Show that there is no function f such that $\partial f/\partial x = 3y$ and $\partial f/\partial y = 4x$.

9. Find all functions such that $f_{xx}(x, y) = 0$.

10. Find all functions such that $f_{xx}(x, y) = 0$ and $f_{yy}(x, y) = 0$.

11. Find all functions such that $f_{xy}(x, y) = 0$.

12. Find all functions such that $f_{xy}(x, y) = 1$.

13. Show that for a polynomial $P(x, y)$, P_{yx} equals P_{xy}. (It is enough to show it for an arbitrary monomial $ax^m y^n$, where a is constant and m and n are non-negative integers. The cases where m or n is 0 should be treated separately.)

14. A hiker is at the origin on a hill whose surface has the equation $z = x$. If he walks south, along the positive x-axis the slope of his path would be steep, 1, with angle $\pi/4$. If he walked along the y-axis, the slope would be 0.

(a) If he walked NE what would the slope of his path be?

(b) In what direction should he walk so his path would have a slope of 0.2?

15. Let f and g be functions of x and y that have continuous second derivatives. Assume the first partial derivatives of f and g satisfy

$$\frac{\partial f}{\partial x} = \frac{\partial g}{\partial y} \quad \text{and} \quad \frac{\partial f}{\partial y} = -\frac{\partial g}{\partial x}. \tag{16.S.2}$$

Show that
$$\frac{\partial^2 f}{\partial x^2}+\frac{\partial^2 f}{\partial y^2}=0 \quad\text{and}\quad \frac{\partial^2 g}{\partial x^2}+\frac{\partial^2 g}{\partial y^2}=0. \tag{16.S.3}$$

16. Let $V(x, y, z) = xyz$ be the volume of a box of sides x, y, and z. Compute ΔV and dV and show them in Figure 16.S.1.

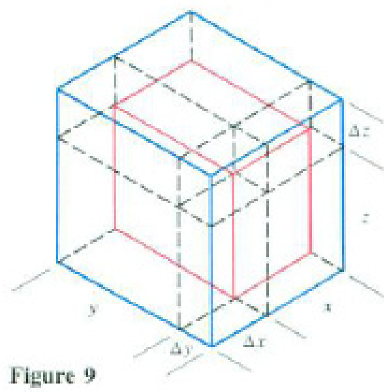


Figure 16.S.1

In Exercises 17 to 20 concern the definition of $\lim_{(x,y)\to P_0} f(x, y)$.

17. Let $f(x, y) = x + y$.

(a) Show that if $P = (x, y)$ lies within a distance 0.01 of $(1, 2)$, then $|x-1| < 0.01$ and $|y - 2| < 0.01$. (See Figure 16.S.2).

(b) Show that if $|x - 1| < 0.01$ and $|y - 2| < 0.01$, then $|f(x, y) - 3| < 0.02$.

(c) Find a number $\delta > 0$ such that if $P = (x, y)$ is in the disk with center $(1, 2)$ and radius δ, then $|f(x, y) - 3| < 0.001$.

(d) Show that for any positive number ϵ, no matter how small, there is a positive number δ such that when $P = (x, y)$ is in the disk with radius δ and center $(1, 2)$, then $|f(x, y) - 3| < \epsilon$. Give δ as a function of ϵ.

(e) What may we conclude from (d)?

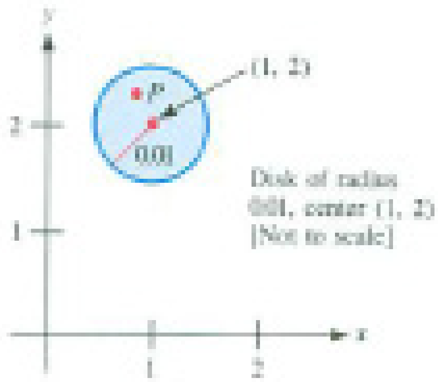


Figure 16.S.2

18. Let $f(x, y) = 2x + 3y$.

(a) Find a disk with center $(1, 1)$ such that whenever P is in it, $|f(P) - 5| < 0.01$

(b) Let ϵ be a positive number. Show that there is a disk with center $(1, 1)$ such that whenever P is in it, $|f(P) - 5| < \epsilon$. Give δ as a function of ϵ.

(c) What may we conclude from (b)?

19. Let $f(x, y) = x^2y/(x^4 + 2y^2)$.

(a) What is the domain of f?

(b) Fill in the three missing values in the table:

(x, y)	$(0.01, 0.01)$	$(0.01, 0.02)$	$(0.001, 0.003)$
$f(x, y)$			

(c) From (b), do you think $\lim_{P\to(0,0)} f(P)$ exists? If so, what is it?

(d) Fill in the three missing values in the table:

(x, y)	$(0.5, 0.25)$	$(0.1, 0.01)$	$(0.001, 0.000001)$
$f(x, y)$			

(e) From (d), do you think $\lim_{P\to(0,0)} f(P)$ exists? If so, what is it?

(f) Does $\lim_{P\to(0,0)} f(P)$ exist? If so, what is it? Explain.

20. Let $f(x, y) = 5x^2y/(2x^4 + 3y^2)$.

(a) What is the domain of f?

(b) As P approaches $(0,0)$ on the line $y = 2x$, what happens to $f(P)$?

(c) As P approaches $(0,0)$ on the line $y = 3x$, what happens to $f(P)$?

(d) As P approaches $(0,0)$ on the parabola $y = x^2$, what happens to $f(P)$?

(e) Does $\lim_{P\to(0,0)} f(P)$ exist? If so, what is it? Explain.

21. This exercise outlines a proof that the mixed partials of $f(x,y)$ are generally equal. It suffices to show that $f_{xy}(0,0) = f_{yx}(0,0)$. We assume that the first and second partial derivatives are continuous in some disk with center $(0,0)$.

(a) Why is $f_{xy}(0,0)$ equal to

$$\lim_{k\to 0} \frac{f_x(0,k) - f_x(0,0)}{k}? \tag{16.S.4}$$

(b) Why is the limit in (16.S.4) equal to

$$\lim_{k\to 0} \left(\lim_{h\to 0} \frac{(f(h,k) - f(0,k)) - (f(h,0) - f(0,0))}{hk} \right)? \tag{16.S.5}$$

(c) Let $u(y) = f(h,y) - f(0,y)$. Show that the fraction in (16.S.5) equals

$$\frac{u(k) - u(0)}{hk},$$

and it equals $u'(K)/h$ for some K between 0 and k.

(d) Why is $u'(K) = f_y(h,K) - f_y(0,K)$?

(e) Why is $u'(K)/h$ equal to $(f_y)_x(H,K)$ for some H between 0 and h?

(f) Deduce that $f_{xy}(0,0) = f_{yx}(0,0)$.

(g) Did this derivation use the continuity of f_{yx}? If so, how?

(h) Did this derivation use the continuity of f_{xy}? If so, how?

(i) Did we need to assume f_{xy} exists? If so, where was the assumption used?

(j) Did we need to assume f_{yx} exists? If so, where was the assumption used?

22. The assertion that we can differentiate across the integral sign, says that

$$\frac{d}{dt} \int_a^b f(x,t)\, dx = \int_a^b \frac{\partial}{\partial t} f(x,t)\, dx. \tag{16.S.6}$$

(a) Why is the derivative on the left an ordinary derivative but the derivative on the right is a partial derivative?

(b) Using the definitions of ordinary derivatives and partial derivatives as limits, show what (16.S.6) says about limits.

(c) Verify (16.S.6) for $f(x,t) = x^7t^4$.

(d) Verfiy (16.S.6) for $f(x,t) = \cos(xt)$.

23. A consumer has a budget of B dollars and may purchase n different items. The price of the ith item is p_i dollars. When the consumer buys x_i units of the ith item, the total cost is $\sum_{i=1}^n p_i x_i$. Assume that $\sum_{i=1}^n p_i x_i = B$ and that the consumer wishes to maximize her utility $u(x_1, x_2 \dots x_n)$.

(a) Show that when $x_1, \dots, x_n$, are chosen to maximize utility, then

$$\frac{\partial u/\partial x_i}{p_i} = \frac{\partial u/\partial x_j}{p_j}.$$

(b) Explain the result in (a) using economic intuition. (Consider a slight change in x_i and x_j, with the other x_k's held fixed.)

24. The following is quoted from a bioeconomics text (Colin W. Clark in *Mathematical Bioeconomics*, Wiley, New York, 1976):

> [S]uppose there are N fishing grounds. Let $H^i = H^i(R^i, E^i)$ denote the production function for the total harvest H^i on the ith ground as a function of the recruited stock level R^i and effort E^i on the ith ground. The problem is to determine the least total cost $\sum_{i=1}^N c_i E^i$ at which a given total harvest $H = \sum_{i=1}^n H^i$ can be achieved. This problem can be easily solved by Lagrange multipliers. The result is simply
>
> $$\frac{1}{c_i}\frac{\partial H^i}{\partial E^i} = \text{constant}$$
>
> [independent of i].

Verify the assertion. The c_i's are constants. The superscripts are not exponents but are used to name the functions.

25. (*Computer science*) This exercise is based on J. D. Ullman, *Principles of Database Systems*, pp. 82–83, Computer Science Press, Potomac, Md., 1980. It arises in the design of efficient bucket sorts. (A **bucket sort** is a way of rearranging information in a database.) Let p_1, p_2, $\ldots$, p_k and B be positive constants. Let b_1, b_2, $\ldots$, b_k be k nonnegative variables satisfying $\sum_{j=1}^{k} b_j = B$. The quantity $\sum_{j=1}^{k} p_j \cdot 2^{B-b_j}$ represents the expected search time. What values of b_1, b_2, $\ldots$, b_k does the method of Lagrange multipliers give for the minimum expected search time?

26. Assume that $f(x, y, z)$ has an extreme value at P_0 on the level surface $g(x, y, z) = k$. Why are ∇g adn ∇f both evaluated at P_0 both perpendicular to the surface at P_0?

Calculus is Everywhere # 23
The Wave in a Rope

We will develop what may be the most famous partial differential equation. In the CIE of the next chapter we will solve it and then apply it in the final chapter.

As Morris Kline writes in *Mathematical Thought from Ancient to Modern Times*, "The first real success with partial differential equations came in renewed attacks on the vibrating string problem, typified by the violin string. The approximation that the vibrations are small was imposed by d'Alembert (1717-1783) in his papers of 1746."

Imagine shaking the end of a rope up and down gently, as in Figure C.23.1. That starts a wave moving along the rope. The molecules in the rope move up and down while the wave travels to the right. For a sound wave, the wave travels at 700 miles per hour, but the air vibrates back and forth. (When someone says "good morning" to us, we feel no wind.)

Figure C.23.1

To develop the mathematics of the wave in a weightless rope, we make some simplifying assumptions. We suppose that each molecule moves only up and down, the distance it moves is very small, and the slope of the curve assumed by the rope remains close to zero.

At time t the vertical position of the molecule whose x-coordinate is x is $y = y(x,t)$, for it depends on both x and t. For a short section of the rope at time t, shown as PQ in Figure C.23.2, assuming that the tension T is the same throughout the rope, we apply Newton's Second Law, which implies force equals mass times acceleration, to the mass in PQ.

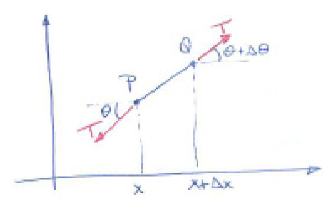


Figure C.23.2

If the linear density of the rope is λ, the mass of the segment is λ times its length. Because displacements are small we will approximate the length by Δx. The upward force exerted by the rope on the segment is $T\sin(\theta+\Delta\theta)$ and the downward force is $T\sin(\theta)$. The net vertical force is $T\sin(\theta+\Delta\theta) - T\sin(\theta)$. Thus

$$\underbrace{T\sin(\theta+\Delta\theta) - T\sin(\theta)}_{\text{net vertical force}} = \underbrace{\lambda\Delta x}_{\text{mass}} \quad \underbrace{\frac{\partial^2 y}{\partial t^2}}_{\text{acceleration}} \quad . \tag{C.23.1}$$

Because y is a function of x and t, we have a partial derivative, not an ordinary derivative.

We express $\sin(\theta)$ and $\sin(\theta+\Delta\theta)$ in terms of the partial derivative $\partial y/\partial x$. Because θ is near 0, $\cos(\theta)$ is near 1. Thus $\sin(\theta)$ is approximately $\sin(\theta)/\cos(\theta) = \tan(\theta)$, the slope of the rope at time t above (or below) x, which is $\partial y/\partial x$ at x and t. Similarly, $\sin(\theta+\Delta\theta)$ is approximately $\partial y/\partial x$ at $x+\Delta x$ and t. So

(C.23.1) is approximated by

$$T\frac{\partial y}{\partial x}(x+\Delta x,t) - T\frac{\partial y}{\partial x}(x,t) = \lambda\Delta x\frac{\partial^2 y}{\partial t^2}(x,t).$$

Dividing by Δx gives

$$\frac{T\left(\frac{\partial y}{\partial x}(x+\Delta x,t) - \frac{\partial y}{\partial x}(x,t)\right)}{\Delta x} = \lambda\frac{\partial^2 y}{\partial t^2}(x,t).$$

Letting Δx approach 0, we obtain

$$T\frac{\partial^2 y}{\partial x^2}(x,t) = \lambda\frac{\partial^2 y}{\partial t^2}(x,t).$$

Since both T and λ are positive, $T/\lambda = c^2$ for some constant c, and we can write (23) in the traditional form

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 y}{\partial t^2}.$$

This is the famous **wave equation**. It relates the acceleration of the molecule, $\partial^2 y/\partial t^2$, to the geometry of the curve, expressed by $\partial^2 y/\partial x^2$. Since we assume that the slope of the rope remains near 0, $\frac{\partial^2 y}{\partial x^2}$ is approximately

$$\frac{\frac{\partial^2 y}{\partial x^2}}{\left(\sqrt{1+\left(\frac{\partial y}{\partial x}\right)^2}\right)^3},$$

which is curvature. At the curvier part of the rope, the acceleration is greater. The wave moves along the rope, but the molecules of the rope move up and down perpendicular to the directio of the waves.

As the CIE in the next chapter shows, the constant c turns out to be the velocity of the wave.

EXERCISES

1. Figure C.23.3 shows a vibrating string whose ends are fixed at A and B. Assume that each part of the string moves parallel to the y-axis. Let $y = f(x,t)$ be the height of the string at the point with abscissa x at time t, as shown in the figure. The partial derivatives are denoted $\partial y/\partial x$ and $\partial y/\partial t$.

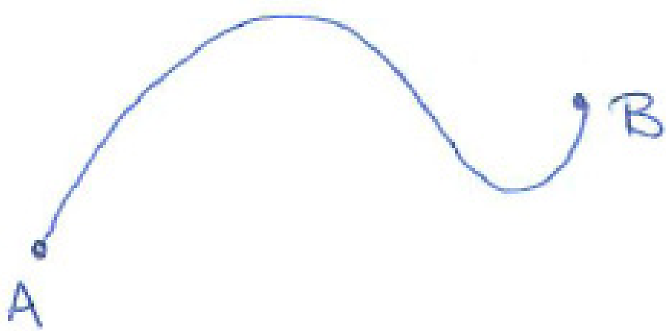


Figure C.23.3

(a) What is the meaning of y_x?

(b) What is the meaning of y_t?

Chapter 17

Plane and Solid Integrals

In Chapter 2 we introduced the derivative, one of the two main concepts in calculus. In Chapter 16 we extended the derivative to higher dimensions. In the present chapter, we generalize the concept of the definite integral, introduced in Chapter 6, to higher dimensions.

Instead of using the notation of Chapter 6 to define the definite integral, we will restate the definition so that it generalizes to higher dimensions.

We started with an interval $[a, b]$, which we will call I, and a continuous function f defined at each point P of I. Then we cut I into n short intervals I_1, I_2, ..., I_n, and chose a point P_1 in I_1, P_2 in I_2, ..., P_n in I_n. See Figure 17.0.1. Denoting the length of I_i by L_i, we formed

$$\sum_{i=1}^{n} f(P_i)L_i.$$

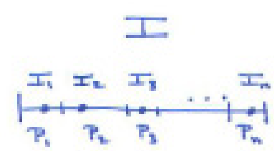


Figure 17.0.1

The limit of the sums as all the subintervals are chosen shorter and shorter is the definite integral of f over interval I. We denoted it $\int_a^b f(x)\ dx$. We now denote it $\int_I f(P)dL$. This notation tells us that we are integrating a function, f, over an interval I. The dL reminds us that the integral is the limit of approximations formed as the sum of products of the function value and the length of a short interval.

We will define integrals of functions over plane regions, such as squares and disks, over solid regions, such as cubes and balls, and over surfaces such as the surface of a ball, in the same way. Integrals can be used to compute total mass of an object if we know its density at each point, or total gravitational attraction, or center of gravity, and so on.

It is one thing to define these higher-dimensional integrals. It is another to calculate them. Most of our attention will be devoted to seeing how to compute them with the aid of so-called iterated integrals, which involve integrals over intervals, the type defined in Chapter 6.

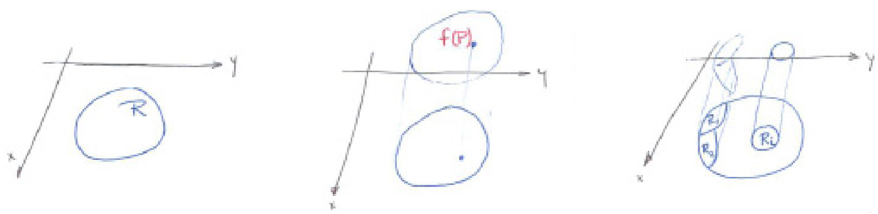
17.1 The Double Integral: Integrals Over Plane Areas

We suggest you re-read the introduction to this chapter and the definition of the definite integral $\int_a^b f(x)\,dx$ before going on.

The goal of this section is to introduce the integral of a function defined in a region of a plane.

Volume Approximated by Sums

Let R be a region in the xy-plane, bounded by curves. For convenience, assume R is convex, for example, an ellipse, a disk, a parallelogram, a rectangle, or a square. We draw R in perspective in Figure 17.1.1(a). Suppose that there is



(a) (b) (c)

Figure 17.1.1

a surface above R and that $f(P)$ is the height of the surface above any point P on R, as shown in Figure 17.1.1(b)

If we know $f(P)$ for every point P could we estimate the volume, V, of the solid under the surface and above R? As we used rectangles to estimate the area of a region in Section 6.1, we will use cylinders to estimate the volume of a solid. The volume of a cylinder is the product of its height and the area of its base.

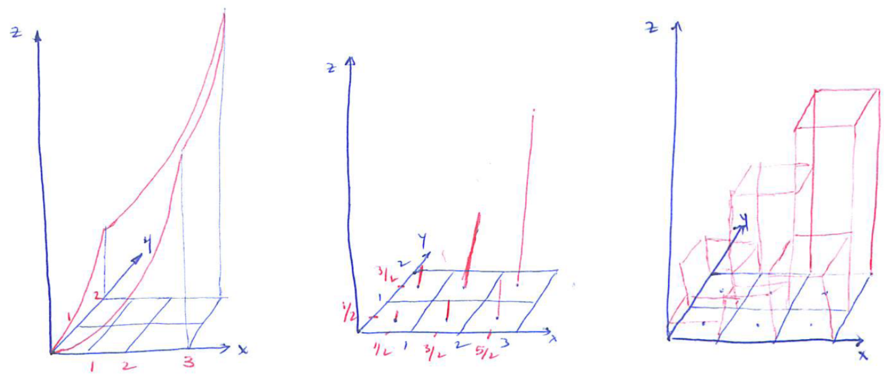
We cut R into n small regions R_1, R_2, ..., R_n. Each R_i has area A_i. Choose points P_1 in R_1, P_2 in R_2, ..., P_n in R_n. Then we put a cylinder over each region R_i. Its height will be $f(P_i)$. There will then be n cylinders. Their total volume is

$$\sum_{i=1}^{n} f(P_i)A_i. \tag{17.1.1}$$

If f is a continuous function, as we choose the regions R_1, R_2, ..., R_n, smaller and smaller, the sum (17.1.1) approaches the volume V.

EXAMPLE 1 Estimate the volume of the solid under the paraboloid $z = x^2 + y^2$ and above the rectangle R whose vertices are $(0,0)$, $(3,0)$, $(3,2)$, and $(0,2)$.

SOLUTION Figure 17.1.2(a) shows the solid.



(a) (b) (c)

Figure 17.1.2

The highest point is above $(3, 2)$, where $z = 13$. So the solid fits in a box whose height is 13 and whose base has area 6. So we know its volume is at most $13 \cdot 6 = 78$.

To estimate the volume we cut the rectangular box into six 1-by-1 squares and evaluate $z = x^2 + y^2$ at, say, the center of each square, as shown in Figure 17.1.2(b).

Then we form a cylinder for each square. The base is the square and the height is the value of $x^2 + y^2$ at the center of the square, as shown in Figure 17.1.2(c).

Then the total volume is

$$\begin{aligned}
&\underbrace{\frac{1}{2}}_{\text{height}} \cdot \underbrace{1}_{\text{area of base}} + \underbrace{\frac{5}{2}}_{\text{height}} \cdot \underbrace{1}_{\text{area of base}} + \underbrace{\frac{13}{2}}_{\text{height}} \cdot \underbrace{1}_{\text{area of base}} \\
+ &\underbrace{\frac{5}{2}}_{\text{height}} \cdot \underbrace{1}_{\text{area of base}} + \underbrace{\frac{9}{2}}_{\text{height}} \cdot \underbrace{1}_{\text{area of base}} + \underbrace{\frac{17}{2}}_{\text{height}} \cdot \underbrace{1}_{\text{area of base}} \quad = \quad 25.
\end{aligned}$$

This estimate is 25 cubic units. By cutting the base into smaller pieces and using more cylinders we could make a more accurate estimate. ◇

Density

Before we consider a total mass problem we must define the concept of density. We have a piece of sheet metal, which we view as part of a plane, that is homogeneous, the same everywhere. Let R be any region in it, of area A and mass m. The quotient m/A is the same for all regions R, and is called the density.

The material, unlike sheet metal, may not be uniform. As R varies, the quotient m/A, or average density in R, also varies. Physicists define **density at a point** as follows.

Figure 17.1.3

They consider a small disk R of radius r and center at P, as in Figure 17.1.3. Let $m(r)$ be the mass in the disk and $A(r)$ be its area (πr^2). Then

$$\text{Density at } P = \lim_{r\to 0} \frac{m(r)}{A(r)}.$$

Thus density is denoted $\sigma(P)$, which is read as "sigma of P."

With the physicists, we will assume the density $\sigma(P)$ exists at each point and that it is a continuous function. This implies that if R is a small region of area A and P is a point in it then the product $\sigma(P)A$ is an approximation of the mass in R.

σ is Greek for our letter "s", the initial letter of "surface." $\sigma(P)$ denotes the density of a surface at P.

Total Mass Approximated by Sums

Assume that a flat region R is occupied by a material of varying density. The density at point P in R is $\sigma(P)$. Estimate M, the total mass in R.

We cut R into n small regions R_1, R_2, …, R_n. Each R_i has area A_i. We choose points P_1 in R_1, P_2 in $R_2, \ldots, P_n$ in R_n. Then we estimate the mass in each region R_i, as shown in Figure 17.1.4. The mass in R_i is approximately

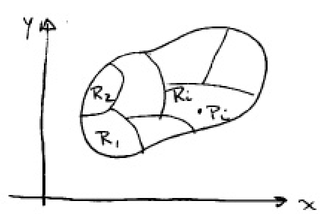


Figure 17.1.4 This example has $n = 7$ subregions.

$$\underbrace{\sigma(P_i)}_{\text{density}} \cdot \underbrace{A_i}_{\text{area}}.$$

Thus

$$\sum_{i=1}^{n} \sigma(P_i)A_i \tag{17.1.2}$$

is the total estimate. If σ is a continuous function, then as we divide R into smaller and smaller regions, the sums (17.1.2) approach the total mass M.

A **lamina** is a thin plate, sheet, or layer.

EXAMPLE 2 A rectangular **lamina**, of varying density occupies the rectangle with corners at $(0,0)$, $(3,0)$, $(3,2)$, and $(0,2)$ in the xy-plane. Its density at (x,y) is x^2+y^2 grams per square centimeter. Estimate its mass by cutting it into six 1-by-1 squares and evaluating the density at the center of each square.

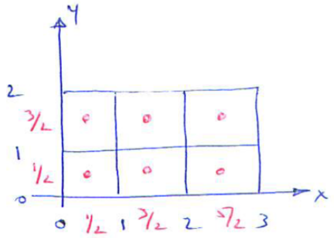


Figure 17.1.5

SOLUTION The six squares are shown in Figure 17.1.5. The density at the center of the square bounded by $x = 1$, $x = 2$, $y = 0$, and $y = 1$ is $\left(\frac{3}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{5}{2}$. Since its area is $1 \times 1 = 1$, an estimate of its mass is

$$\underbrace{\frac{5}{2}}_{\text{density}} \cdot \underbrace{1}_{\text{area}} = \frac{5}{2} \text{ grams.}$$

Similar estimates for the remaining small squares give a total estimate of

$$\frac{1}{2} \cdot 1 + \frac{5}{2} \cdot 1 + \frac{13}{2} \cdot 1 + \frac{5}{2} \cdot 1 + \frac{9}{2} \cdot 1 + \frac{17}{2} \cdot 1 = 25 \text{ grams}$$

This sum is identical to the sum (17.1.2), which estimates a volume. ◇

The arithmetic in Examples 1 and 2 shows that unrelated problems, one about volume, the other about mass, lead to the same estimates. Moreover, as the rectangle is cut into smaller pieces, the estimates would become closer and closer to the volume or the mass. These estimates, similar to the estimates $\sum_{i=1}^{n} f(P_i)L_i$ that appear in the definition of the definite integral $\int_a^b f(x)\ dx$, bring us to the definition of a double integral. It is called a double integral because the domain of the function is in the two-dimensional plane.

The Double Integral

The definition of the double integral is almost the same as that of $\int_a^b f(x)\ dx$, the integral over an interval. The differences are:

1. instead of dividing an interval into smaller intervals, we divide a planar region into smaller planar regions,

2. instead of a function defined on an interval, we have a function defined on a planar region, and

3. we need a quantitative way to say that a little region is small.

To meet the need described in (3) we define the diameter of a planar region. The **diameter** of a region bounded by a curve is the maximum distance between two points in the region. For instance, the diameter of a square of side s is $s\sqrt{2}$ and the diameter of a disk is the same as its diameter that we know from geometry.

DEFINITION *(Double Integral)* Let R be a planar region bounded by curves and f a continuous function defined on R. Partition R into smaller regions R_1, R_2, ..., R_n of respective areas A_1, A_2, ...,

A_n. Choose points P_1 in R_1, P_2 in R_2, ..., P_n in R_n and form the approximating Riemann sum

$$\sum_{i=1}^{n} f(P_i)A_i. \tag{17.1.3}$$

Form a sequence of partitions such that the diameters of the regios R_i approach 0. Then the sums (17.1.3) approach a limit, which is called the integral of f over R or the double integral of f over R. It is denoted

$$\int_R f(P)\, dA.$$

It is proved in advanced courses that when f is continuous, the sums approach a limit.

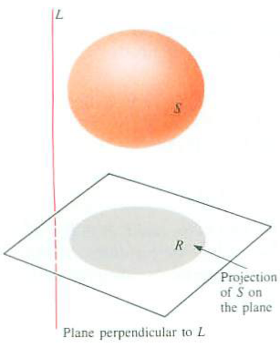


Figure 17.1.6 The highlighted segment of the vertical line through P has length $c(P)$.

Volume Expressed as a Double Integral

Consider a solid S and its projection R on a plane, as in Figure 17.1.6. Assume that for each point P in R the line through P perpendicular to R intersects S in a line segment of length $c(P)$. Then

The double integral of cross section is the volume

$$\text{Volume of } S = \int_R c(P)\, dA.$$

Mass Expressed as a Double Integral

For a plane distribution of mass through a region R, as shown in Figure 17.1.7, the density may vary. Denote the density at P by $\sigma(P)$ in grams per square centimeter. Then

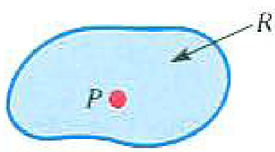


Figure 17.1.7

The double integral of density is the total mass

$$\text{Mass in } R = \int_R \sigma(P)\, dA$$

Average Value as a Double Integral

The average value of $f(x)$ for x in the interval $[a, b]$ was defined in Section 6.3 as

$$\frac{\int_a^b f(x)\ dx}{\text{length of interval}}.$$

We make a similar definition for a function defined on a two-dimensional region.

DEFINITION *(Average value)* The **average value** of f over the region R is

$$\frac{\int_R f(P)\ dA}{\text{Area of } R}.$$

If $f(P)$ is positive for all P in R, there is a geometric interpretation of the average of f over R. Let S be the solid below the graph of f (a surface) and above R. The average value of f over R is the height of the cylinder whose base is R and whose volume is the same as the volume of S. (See Figure 17.1.8.)

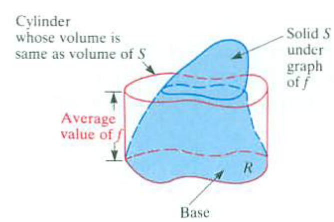


Figure 17.1.8

The integral $\int_R f(P)\ dA$ is called an integral over a plane region to distinguish it from $\int_a^b f(x)\ dx$, which is called, an integral over an interval. Another notation is $\int\int_R f(x, y)\ dA$. We prefer the notation $\int_R f(P)\ dA$ because it does not mention a coordinate system.

The integral of $f(P) = 1$ over a region is its area. The approximating sum $\sum_{i=1}^n f(P_i)A_i$ equals $\sum_{i=1}^n 1 \cdot A_i = A_1 + A_2 + \cdots + A_n$, which is the area of the region R that is being partitioned. Since every approximating sum has this same value, it follows that

$$\lim_{n\to\infty} \sum_{i=1}^{n} f(P_i)A_i = \text{Area of } R.$$

Consequently

The double integral of the function 1 is the area

$$\int_R 1\ dA = \text{Area of } R.$$

This formula will come in handy on several occasions. The 1 is often omitted, in which case we write $\int_R\ dA = $ Area of R. A table summarizing some of the main applications of the double integral $\int_R f(P)\ dA$ can be found in the Summary section for this chapter (see Table 17.S.1 on page 1568).

Properties of Double Integrals

Integrals over plane regions have properties similar to those of integrals over intervals:

1. $\int_R cf(P)\ dA = c\int_R f(P)\ dA$ for any constant c.
2. $\int_R [f(P) + g(P)]\ dA = \int_R f(P)\ dA + \int_R g(P)\ dA$.
3. If $f(P) \leq g(P)$ for all points P in R, then $\int_R f(P)\ dA \leq \int_R g(P)\ dA$.
4. If R is broken into two regions, R_1 and R_2, overlapping at most on their boundaries, then

$$\int\limits_R f(P)\ dA = \int\limits_{R_1} f(P)\ dA + \int\limits_{R_2} f(P)\ dA.$$

For instance, consider **(3)** when $f(P)$ and $g(P)$ are both positive. Then $\int_R f(P)\ dA$ is the volume under the surface $z = f(P)$ and above R in the xy-plane. Similarly $\int_R g(P)\ dA$ is the volume under $z = f(P)$ and above R. Then **(3)** asserts that the volume of a solid is not larger than the volume of a solid that contains it. (See Figure 17.1.9.)

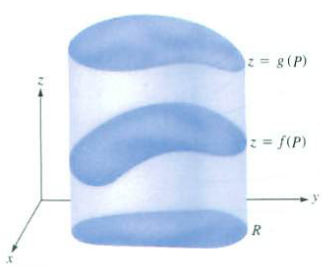


Figure 17.1.9

Summary

Just as $\int_a^b f(x)\ dx$, an integral over an interval is defined as the limit of sums of the form $\sum_{i=1}^n f(P_i)L_i$, $\int_R f(P)\ dA$, a double integral, is defined as the limit of sums of the form $\sum_{i=1}^n f(P_i)A_i$. They arise in computing volumes, total mass, or average value. We used a sum $\sum_{i=1}^n f(P_i)A_i$ to estimate a double integral.

EXERCISES for Section 17.1

1. In the estimates for the volume in Example 1, the centers of the squares were used as the P_i's. Make an estimate for the volume in Example 1 by using the same partition but taking as P_i

(a) the lower left corner of R_i,

(b) the upper right corner of R_i.

(c) What do (a) and (b) tell about the volume of the solid?

2. Estimate the mass in Example 2 using the partition of R into six squares and taking as the P_i's

(a) upper left corners,

(b) lower right corners.

3. Let R be a set in the plane whose area is A. Let f be the function such that $f(P) = 5$ for every point P in R.

(a) What can be said about any approximating sum $\sum_{i=1}^{n} f(P_i)A_i$ formed for this R and this f?

(b) What is the value of $\int_R f(P)\ dA$?

4. Let R be the square with vertices $(1, 1)$, $(5, 1)$, $(5, 5)$, and $(1, 5)$. Let $f(P)$ be the distance from P to the y-axis.

(a) Estimate $\int_R f(P)\ dA$ by partitioning R into four squares and using their centers as sampling points.

(b) Show that $16 \leq \int_R f(P)\ dA \leq 80$.

5. Let f and R be as in Example 1. Use the estimate of $\int_R f(P)\ dA$ obtained in the text to estimate the average of f over R.

6. Assume that for P in R, $m \leq f(P) \leq M$, where m and M are constants. Let A be the area of R. By examining approximating sums, show that

$$mA \leq \int_R f(P)\, dA \leq MA.$$

7.

(a) Let R be the rectangle with vertices $(0,0)$, $(2,0)$, $(2,3)$, and $(0,3)$. Let $f(x,y) = \sqrt{x+y}$. Estimate $\int_R \sqrt{x+y}\, dA$ by partitioning R into six 1-by-1 squares and choosing the sampling points to be their centers.

(b) Use (a) to estimate the average value of f over R.

8.

(a) Let R be the square with vertices $(0,0)$, $(0.8,0)$, $(0.8,0.8)$, and $(0,0.8)$. Let $f(P) = f(x,y) = e^{xy}$. Estimate $\int_R e^{xy}\, dA$ by partitioning R into sixteen squares and choosing the sampling points to be their centers.

(b) Use (a) to estimate the average value of $f(P)$ over R.

(c) Show that $0.64 \leq \int_R f(P)\, dA \leq 0.64e^{0.64}$.

9.

(a) Let R be the triangle with vertices $(0,0)$, $(4,0)$, and $(0,4)$ shown in Figure 17.1.10. Let $f(x,y) = x^2 y$. Use the partition into four triangles and sampling points shown in the diagram to estimate $\int_R f(P)\, dA$.

(b) What is the maximum value of $f(x,y)$ in R?

(c) From (b) obtain an upper bound on $\int_R f(P)\, dA$.

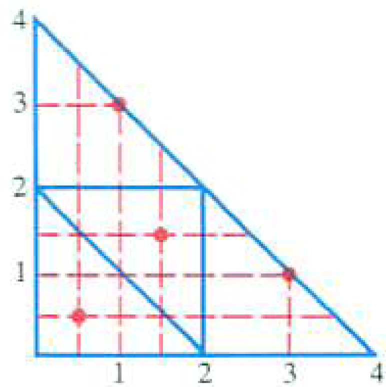


Figure 17.1.10

10.

(a) Sketch the surface $z = \sqrt{x^2 + y^2}$.

(b) Let V be the region in space below the surface in (a) and above the square R with vertices $(0,0)$, $(1,0)$, $(1,1)$, and $(0,1)$. Let v be the volume of V. Show that $v \leq \sqrt{2}$.

(c) Using a partition of R with sixteen squares, find an estimate for v that is too large.

(d) Using the partition in (c), find an estimate for v that is too small.

11. The amount of rain that falls at point P during one year is $f(P)$ inches. Let R be a region, and assume areas are measured in square inches.

(a) What is the meaning of $\displaystyle\int_R f(P)\,dA$?

(b) What is the meaning of $\dfrac{\int_R f(P)\,dA}{\text{Area of } R}$?

12. A region R in the plane is divided into two regions R_1 and R_2. The function $f(P)$ is defined throughout R. Assume that you know the areas of R_1 and R_2, A_1 and A_2, and the average of f over R_1 and the average of f over R_2, they are f_1 and f_2. Find the average of f over R. (See Figure 17.1.11(a).)

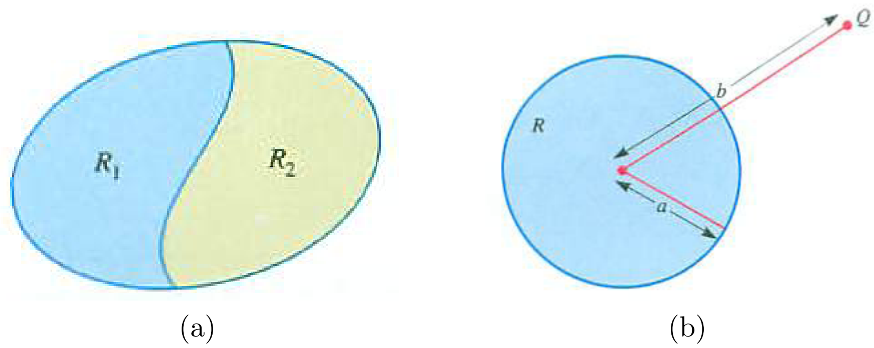


Figure 17.1.11

13. Figure 17.1.12(a) shows parts of some level curves of a function $z = f(x, y)$ and a square R. Estimate $\int_R f(P)\, dA$, and describe your reasoning.

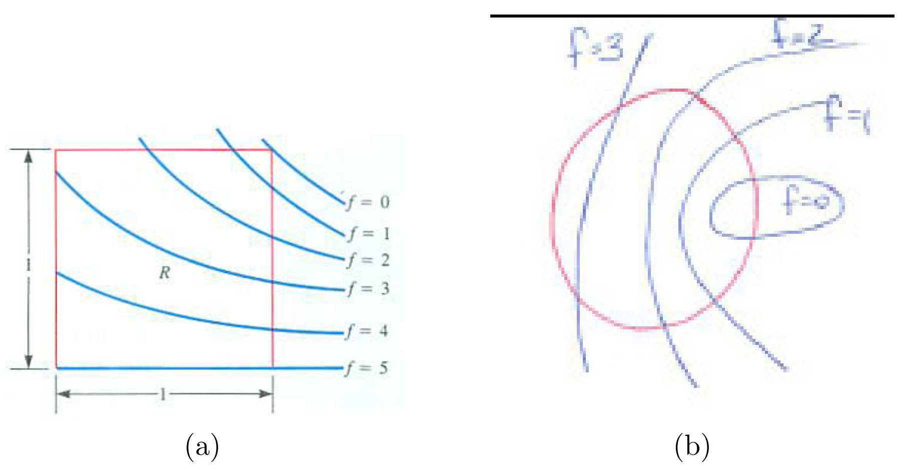


Figure 17.1.12

14. Figure 17.1.12(b) shows parts of some level curves of a function $z = f(x, y)$ and a unit disk R. Estimate $\int_R f(P)\, dA$, and describe your reasoning.

15. A point Q on the xy-plane is at a distance b from the center of a disk R of radius $a(a < b)$ in the xy-plane. For P in R let $f(P) = 1/\overrightarrow{PQ}$. Find positive numbers c and d such that:

$$c < \int_R f(P)\, dA < d.$$

The numbers c and d depend on a and b. See Figure 17.1.11(b).

16.

(a) Let R be a disk of radius 1. Let $f(P)$, for P in R, be the distance from P to the center of the disk. By cutting R into narrow circular rings with centers at the center of the disk, evaluate $\int_R f(P)\, dA$.

(b) Find the average of $f(P)$ over R.

Exercises 17 and 18 introduce **Monte Carlo methods** for estimating a double integral using randomly chosen points. They tend to be inefficient because the error decreases on the order of $1/\sqrt{n}$, where n is the number of random points. That is a slow rate. They are used only when it's possible to choose n very large.

17. This exercise involves estimating an integral by choosing points randomly. A computing machine can be used to generate random numbers and thus random points in the plane that can be used to estimate definite integrals, as we now show. Say that a region R lies in the square whose vertices are $(0, 0)$, $(2, 0)$, $(2, 2)$, and $(0, 2)$, and a complicated function f is defined in R. The machine generates 100 random points (x, y) in the square. Of these, 73 lie in R. The average value of f for these 73 points is 2.31.

(a) What is a reasonable estimate of the area of R?

(b) What is a reasonable estimate of $\int_R f(P)\, dA$?

18. Let R be the disk bounded by the unit circle $x^2 + y^2 = 1$ in the xy plane. Let $f(x, y) = e^{x^2 y}$ be the temperature at (x, y).

(a) Estimate the average value of f over R by evaluating $f(x, y)$ at twenty random points in R. (Adjust your program to select each of x and y randomly in the interval $[-1, 1]$. In this way you construct a random point (x, y) in the square whose vertices are $(1, 1)$, $(-1, 1)$, $(-1, -1)$, $(1, -1)$. Consider only those points that lie in R.)

(b) Use (a) to estimate $\int_R f(P)\, dA$.

(c) Show why $\pi/e \leq f_R f(P)\, dA \leq \pi e$.

19. Sam is heckling again.

Sam: As usual, the authors made this harder than necessary. They didn't need to introduce diameters. Instead they could have used good old area. They could have taken the limit as all the areas of the little pieces approached 0. I'll send them a note.

Is Sam right?

20. The unit square can be partitioned into nine congruent squares.

(a) What is the diameter of the small squares?

(b) It is possible to partition the square into nine regions whose largest diameter is 5/11. Show that 5/11 is smaller than the diameter in (a).

(c) Does a region of diameter d always fit in a disk of diameter d?

Skill Drill: Derivatives

In Exercises 21 and 24 a and b are constants. In each case verify that the derivative of the first function is the second function.

21. $\frac{ax^2+b}{x^3}$, $\frac{-(ax^2+3b)}{x^4}$.

22. $\frac{x}{2}+\frac{\sin(2ax)}{4a}$, $\cos^2(ax)$.

23. $\frac{1}{a}\sin(ax)-\frac{1}{3a}\sin^3(ax)$, $\cos^3(ax)$.

24. $\frac{b}{a^2(ax+b)}+\frac{1}{a^2}\ln|ax+b|$, $\frac{x}{(ax+b)^2}$.

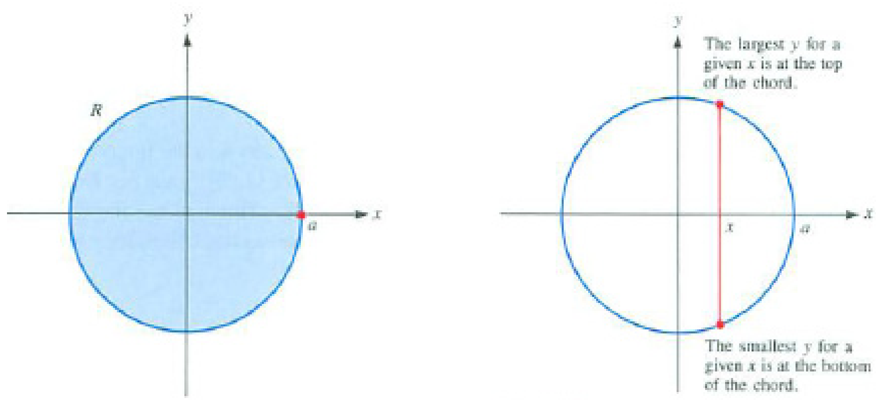
17.2 Computing $\int_R f(P)\,dA$ Using Rectangular Coordinates

In this section, we will show how to use rectangular coordinates to evaluate $\int_R f(P)\,dA$, the integral of a function f over a plane region R. The method requires that both R and f be described in rectangular coordinates. We first show how to describe plane regions R in rectangular coordinates.

Describing R in Rectangular Coordinates

Some examples illustrate how to describe planar regions by their cross sections in rectangular coordinates.

EXAMPLE 1 Describe a disk R of radius a in rectangular coordinates.



(a) (b)

Figure 17.2.1

SOLUTION Introduce an xy coordinate system with its origin at the center of the disk, as in Figure 17.2.1(a). The figure shows that x ranges from $-a$ to a. We now tell how y varies for each x in $[-a, a]$.

Figure 17.2.1(b) shows a cross section at x. The circle has the equation $x^2+y^2 = a^2$. The top half has the equation $y = \sqrt{a^2-x^2}$ and the bottom half, $y = -\sqrt{a^2-y^2}$. So, for x in $[-a, a]$, y varies from $-\sqrt{a^2-x^2}$ to $\sqrt{a^2-x^2}$. (As a check, test $x = 0$. Does y vary from $-\sqrt{a^2-0^2} = -a$ to $\sqrt{a^2-0^2} = a$? It does, asFigure 17.2.1(b) shows.)

Region R is described by vertical cross sections as

$$-a \le x \le a, \qquad -\sqrt{a^2-x^2} \le y \le \sqrt{a^2-x^2}.$$

◇

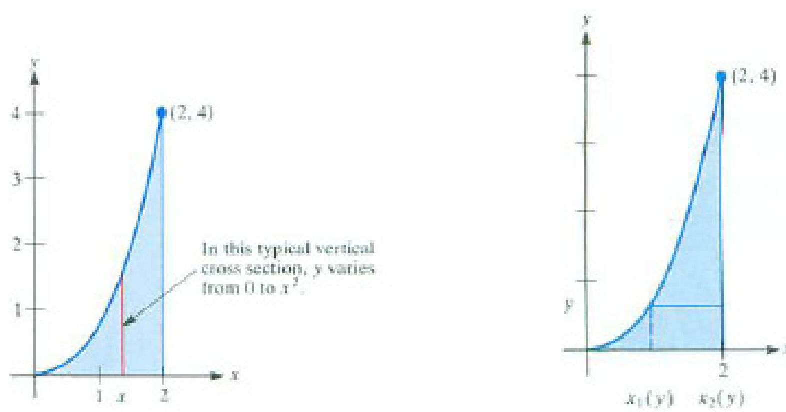
EXAMPLE 2 Let R be the region bounded by $y = x^2$, the x-axis, and the line $x = 2$. Describe R by cross sections parallel to the y-axis.

SOLUTION Figure 17.2.2(a) shows that for points (x, y) in R, x ranges from 0 to 2. To describe R completely, we describe the behavior of y for x in the interval $[0, 2]$.

The cross section above $(x, 0)$ extends from the x-axis to the curve $y = x^2$, so the y coordinate varies from 0 to x^2. The description of R by vertical cross sections is

$$0 \leq x \leq 2, \qquad 0 \leq y \leq x^2.$$

◇



(a) (b)

Figure 17.2.2

EXAMPLE 3 Describe the region R of Example 2 by cross sections parallel to the x-axis, that is, horizontal cross sections.

SOLUTION Figure 17.2.2(b) shows that y varies from 0 to 4. For y in $[0, 4]$, x varies from a smallest value $x_1(y)$ to a largest value $x_2(y)$. For each value of y in $[0, 4]$, $x_2(y) = 2$. To find $x_1(y)$, use the fact that the point $(x_1(y), y)$ is on the curve $y = x^2$, that is,

$$x_1(y) = \sqrt{y}.$$

The description of R in terms of horizontal cross sections is

$$0 \leq y \leq 4, \qquad \sqrt{y} \leq x \leq 2.$$

◇

EXAMPLE 4 Describe the region R whose vertices are $(0,0)$, $(6,0)$, $(4,2)$, and $(0,2)$ by vertical cross sections and then by horizontal cross sections. See Figure 17.2.3.

SOLUTION Figure 17.2.3 shows x varies between 0 and 6. For x in $[0,4]$, y ranges from 0 to 2. For x in $[4,6]$, y ranges from 0 to the value of y on the line through $(4,2)$ and $(6,0)$, which has equation $y = 6 - x$. The description of R by vertical cross sections requires two descriptions:

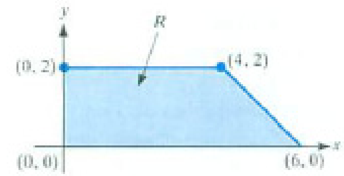


Figure 17.2.3

$$0 \le x \le 4, \quad 0 \le y \le 2$$

and

$$4 \le x \le 6, \quad 0 \le y \le 6 - x.$$

Using horizontal cross sections provides a simpler description. First, y goes from 0 to 2. For y in $[0,2]$, x goes from 0 to the value of x on the line $y = 6-x$. Solving this equation for x yields $x = 6 - y$.

The description in terms of horizontal cross sections is shorter:

$$0 \le y \le 2, \qquad 0 \le x \le 6 - y.$$

◇

These examples are typical. First, determine the range of one coordinate, and then see how the other coordinate varies for any fixed value of the first coordinate.

SHERMAN: Woody suggests deleting this paragraph. I like it because it summarizes all 4 examples. Your thoughts?

Evaluating $\int_R f(P)\,dA$ by Iterated Integrals

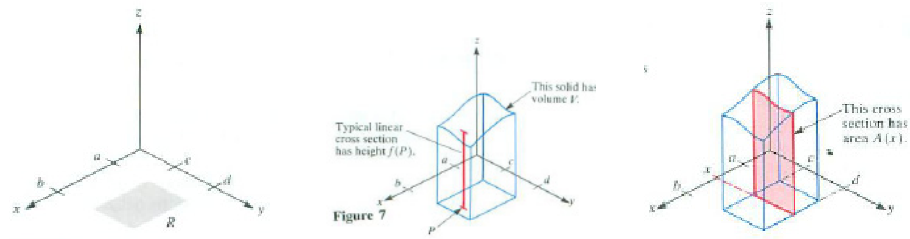
We will offer an intuitive development of a formula for computing double integrals over plane regions.

We develop a way for computing a double integral over a rectangle. After applying this in Example 5, we make the modification needed to evaluate double integrals over more general regions.

Consider a rectangular region R whose description by cross sections is

$$a \le x \le b, \qquad c \le y \le d,$$

as shown in Figure 17.2.4(a). If $f(P) \ge 0$ for all P in R, then $\int_R f(P)\,dA$ is the volume V of the solid whose base is R and which has, above P, height $f(P)$. (See Figure 17.2.4(b).) Let $A(x)$ be the area of the cross section made by a



(a) (b) (c)

Figure 17.2.4

plane perpendicular to the x-axis and having abscissa x, as in Figure 17.2.4(c). As was shown in Section 5.1,

$$V = \int_b^a A(x)\ dx.$$

The area $A(x)$ is expressible as a definite integral

$$A(x) = \int_c^d f(x, y)\ dy.$$

Here x is held fixed throughout this integration to find $A(x)$. This provides a way to evaluate $V = \int_R f(P)\ dA$, namely,

$$\int_R f(P)\ dA = V = \int_a^b A(x)\ dx = \int_a^b \left(\int_c^d f(x, y)\ dy \right)\ dx.$$

$$\boxed{\int_R f(P)\ dA = \int_a^b \left(\int_c^d f(x, y)\ dy \right)\ dx.}$$

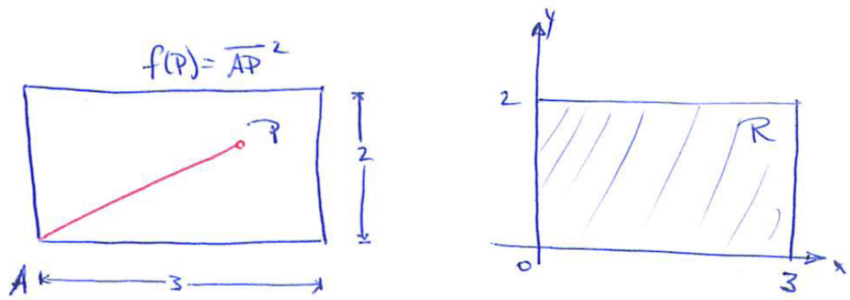
Cross sections by planes perpendicular to the y-axis could be used. Then similar reasoning shows that

$$\int_R f(P)\ dA = \int_c^d \left(\int_a^b f(x, y)\ dx \right)\ dy.$$

The quantities $\int_a^b \left(\int_c^d f(x,y)\ dy \right)\ dx$ and $\int_c^d \left(\int_a^b f(x,y)\ dx \right)\ dy$ are called **iterated integrals**. Usually the parentheses are omitted and they the iterated integrals are written $\int_a^b \int_c^d f(x,y)\ dy\ dx$ and $\int_c^d \int_a^b f(x,y)\ dx\ dy$. The order of dx and dy matters. The differential on the left tells which integration is performed first.

EXAMPLE 5 Compute the double integral $\int_R f(P)\ dA$, where R is the rectangle shown in Figure 17.2.5(a) and f is defined by $f(P) = \overline{AP}^2$.

SHERMAN: Is $\overline{AP}$ the length or the segment? I believe it's normal for this to be the segment. The length is AP. Stewart uses AP as the segment and $|AP|$ as the length. I see other notations in other books. I believe we have been consistent; does it really matter which we use? I am concerned that AP^2 would be hard to interpret.



(a) (b)

Figure 17.2.5

SOLUTION Introduce xy coordinates as in Figure 17.2.5(b). In rectangular coordinates

$$f(x,y) = \overline{AP}^2 = x^2 + y^2.$$

Because x takes all values from 0 to 3 and for each x the number y takes all values between 0 and 2,

$$\int_R f(P)\ dA = \int_0^4 \left(\int_0^2 (x^2 + y^2)\ dy \right)\ dx.$$

The double integral appeared in Example 1 for a volume and in Example 2 for the mass of a solid.

We must first compute the inner integral

The cross-sectional area $A(x)$.

$$\int_0^2 (x^2 + y^2)\ dy, \qquad \text{where } x \text{ is fixed in } [0,3].$$

To apply the Fundamental Theorem of Calculus, find a function $F(x,y)$ such that

$$\frac{\partial F}{\partial y} = x^2 + y^2.$$

Because x is constant during this first integration,

$$F(x,y) = x^2y + \frac{y^3}{3}$$

is such a function. By the Fundamental Theorem of Calculus,

$$\int_0^2 (x^2+y^2)\ dy \;=\; \left(x^2y + \frac{y^3}{3}\right)\Bigg|_{y=0}^{y=2} = \left(x^2\cdot 2 + \frac{2^3}{3}\right) - \left(x^2\cdot 0 + \frac{0^3}{3}\right) = 2x^2 + \frac{8}{3}.$$

The notation $|_{y=0}^{y=2}$ reminds us that y is replaced by 0 and 2.

The formula $2x^2 + \frac{8}{3}$ is the area $A(x)$ discussed earlier in this section. Compute

$$\int_0^3 A(x)\ dx = \int_0^3 (2x^2 + \frac{8}{3})\ dx.$$

By the Fundamental Theorem of Calculus,

$$\int_0^3 \left(2x^2 + \frac{8}{3}\right)\ dx = \left(\frac{2x^3}{3} + \frac{8x}{3}\right)\Bigg|_0^3 = (18+8) - (0+0) = 26.$$

How does this compare with the estimates in Section 17.1?

Hence the two-dimensional double integral has the value 26. The volume of the region in Example 1 of Section 17.1 is 26 cubic centimeters. The mass in Example 2 is 26 grams. ◇

If R is not a rectangle, the iterated integral that equals $\int_R f(P)\ dA$ differs from that for the case where R is a rectangle only in the intervals of integration. If R has the description

$$a \le x \le b \qquad y_1(x) \le y \le y_2(x),$$

by cross sections parallel to the y-axis, as in Figure 17.2.6, then

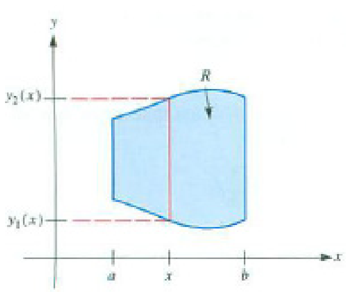


Figure 17.2.6

$$\boxed{\int_R f(P)\ dA = \int_a^b \left(\int_{y_1(x)}^{y_2(x)} f(x,y)\ dy\right) dx.}$$

Similarly, if R has the description

$$c \le y \le d \qquad x_1(y) \le x \le x_2(y),$$

by cross sections parallel to the x-axis, then

$$\int\limits_R f(P)\,dA = \int\limits_c^d \left(\int\limits_{x_1(y)}^{x_2(y)} f(x,y)\,dx \right) dy.$$

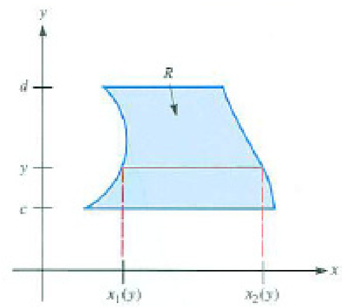
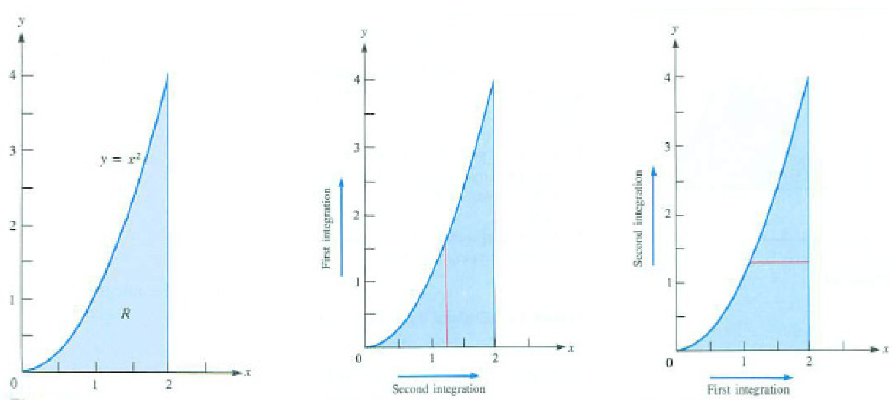


Figure 17.2.7

The intervals of integration are determined by R; the function f influences only the integrand. See Figure 17.2.7.

EXAMPLE 6 Evaluate $\int_R 3xy\,dA$ over the region R shown in Figure 17.2.8(a).
In the next example R is the region described in Examples 2 and 3.
SOLUTION We can use cross sections parallel either to the y-axis or to the x-axis (see Figure 17.2.8(b) and (c)).



(a) (b) (c)

Figure 17.2.8 The red slices show which integration is performed first.

SHERMAN: I've rewritten the text before and after Example 6, including moving the interpretation to the end. OK?

If, as shown in Figure 17.2.8(b), cross sections parallel to the y-axis are used, then R is described by

$$0 \le x \le 2 \qquad 0 \le y \le x^2.$$

Thus

$$\int\limits_R 3xy\,dA = \int\limits_0^2 \left(\int\limits_0^{x^2} 3xy\,dy \right) dx.$$

To compute the iterated integral, we start with the integral in which x is fixed and y goes from 0 to x^2. With x fixed,

$$\int\limits_0^{x^2} 3xy\,dy = \left(3x\frac{y^2}{2} \right)\Bigg|_{y=0}^{y=x^2} = 3x\frac{(x^2)^2}{2} - 3x\frac{0^2}{2} = \frac{3x^5}{2}.$$

Then

$$\int_0^2 \frac{3x^5}{2}\,dx = \left.\frac{3x^6}{12}\right|_0^2 = 16.$$

The region R can also be described in terms of cross sections parallel to the x-axis:

$$0 \le y \le 4 \qquad \sqrt{y} \le x \le 2.$$

See Figure 17.2.8(c). Then the double integral is evaluated by a different iterated integral,

$$\int_R 3xy\,dA = \int_0^4 \left(\int_{\sqrt{y}}^2 3xy\,dx \right) dy,$$

which, as the reader may verify, equals 16. See Figure 17.2.8(b). ◇

The fact that $\int_R 3xy\,dA = 16$ has three interpretations:

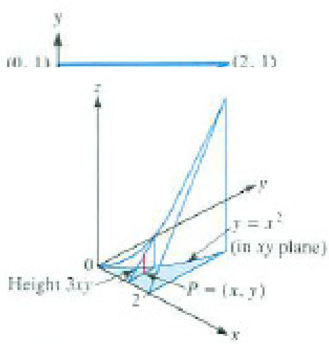


Figure 17.2.10

1. If at each point $P = (x, y)$ in R we erect a line segment above P of length $3xy$ cm, then the volume of this solid is 16 cm^3 (See Figure 17.2.10.)

2. If the density of matter at (x, y) in R is $3xy$ kg/cm^2, then the total mass in R is 16 kg.

3. To find the average temperature in R we need the area of R:

$$\int_R dA = \int_0^2 \int_0^{x^2} dy\,dx = \int_0^2 x^2\,dx = \left.\frac{1}{3}x^3\right|_0^2 = \frac{8}{3}.$$

   So, if the temperature at (x, y) in R is $3xy$ then the average temperature in R is $\frac{16}{8/3} = \frac{3}{2}$.

In Example 6 we could evaluate $\int_R f(P)\,dA$ by cross sections in either direction. In the next example we don't have that choice.

EXAMPLE 7 A triangular lamina is located as in Figure 17.2.9. Its density at (x, y) is e^{y^2}. Find its mass, that is, $\int_R f(P)\,dA$, where $f(x, y) = e^{y^2}$.

SOLUTION The description of R by vertical cross sections is

$$0 \le x \le 2, \qquad \frac{x}{2} \le y \le 1.$$

Hence

$$\int_R f(P)\,dA = \int_0^2 \left(\int_{x/2}^1 e^{y^2}\,dy \right) dx.$$

Since e^{y^2} does not have an elementary antiderivative, the Fundamental Theorem of Calculus is useless in computing

$$\int_{x/2}^1 e^{y^2}\,dy.$$

So we try horizontal cross sections instead. The description of R is now

$$0 \le y \le 1, \qquad 0 \le x \le 2y.$$

This leads to a different iterated integral, namely

$$\int_R f(P)\,dA = \int_0^1 \left(\int_0^{2y} e^{y^2}\,dx \right) dy.$$

For the first integration, $\int_0^2 e^{y^2}\,dx$, y is fixed, so the integrand is constant. Thus

$$\int_0^{2y} e^{y^2}\,dx \quad = \quad e^{y^2} \int_0^{2y} 1\,dx = e^{y^2} x \Big|_{x=0}^{x=2y} = e^{y^2} 2y.$$

The second definite integral in the iterated integral is thus $\int_0^1 e^{y^2} 2y\,dy$, which can be evaluated by the Fundamental Theorem of Calculus, since $d(e^{y^2})/dy = e^{y^2} 2y$:

$$\int_0^1 e^{y^2} 2y\,dy = e^{y^2}\Big|_0^1 = e^{1^2} - e^{0^2} = e - 1.$$

The total mass is $e - 1$. $\diamond$

Computing a definite integral over a plane region R involves a wise choice of an xy-coordinate system, a description of R and f relative to this coordinate system, and the computation of two definite integrals over intervals. The order of the integrations may affect the difficulty of the computation. The order is determined by the description of R by cross sections.

SHERMAN: Should xy coordinate system be hyphenated? At present, we don't, at present. More importantly, have we ever really defined xy-coordinate system? Could (most of) these be replaced by xy-plane?

Summary

We showed that the integral of $f(P)$ over a plane region R can be evaluated by an iterated integral, where the limits of integration are determined by R. If a line parallel to the y-axis meets R in at most two points then R has the description

$$a \leq x \leq b, \qquad y_1(x) \leq y \leq y_2(x)$$

and

$$\int\limits_R f(P)\ dA = \int\limits_a^b \left(\int\limits_{y_1(x)}^{y_2(x)} f(x,y)\ dy \right) dx.$$

If a line parallel to the x-axis meets R in at most two points, then, similarly, R can be described in the form

$$c \leq y \leq d \qquad x_1(y) \leq x \leq x_2(y)$$

and

$$\int\limits_R f(P)\ dA = \int\limits_c^d \left(\int\limits_{x_1(y)}^{x_2(y)} f(x,y)\ dx \right) dy.$$

A Few Words on Notation

We use the notation $\int f(P)\ dA$ or $\int_R f(P)\ dA$ for a (double) integral over a plane region, and later in this chapter, $\int f(P)\ dS$ or $\int_{\mathcal{S}} f(P)\ dS$ for an integral over a surface, and $\int f(P)\ dV$ or $\int_R f(P)\ dV$ for an integral over a region in space. The symbols dA, dS, and dV indicate the type of set over which the integral is defined.

Many people use repeated integral signs to distinguish dimensions. For instance they write $\int f(P)\ dA$ as $\iint f(P)\ dA$ or $\iint f(x,y)\ dx\ dy$. Similarly, they denote an integral over a region in space by $\iiint f(P)\ dx\ dy\ dz$.
We use the single integral sign and P for point for all integrals for three reasons:

1. it is free of any coordinate system
2. it is compact (uses the fewest symbols): $\int$ for "integral", $f(P)$ or f for the integrand, and dA, dS, or dV for the set
3. it allows the symbols $\iint$ and $\iiint$ to be reserved for iterated integrals.

Iterated integrals are a different mathematical object. Each integral in an iterated integral over an interval. We write dx (or dy or dz) when dealing with an integral over an interval.

EXERCISES for Section 17.2

Exercises 1 to 12 provide practice in describing plane regions by cross sections in rectangular coordinates. Describe the region by (a) vertical cross sections and (b) horizontal cross sections.

1. The triangle whose vertices are $(0,0)$, $(2,1)$, $(0,1)$.

2. The triangle whose vertices are $(0,0)$, $(2,0)$, $(1,1)$.

3. The parallelogram with vertices $(0,0)$, $(1,0)$, $(2,1)$, $(1,1)$.

4. The parallelogram with vertices $(2,1)$, $(5,1)$, $(3,2)$, $(6,2)$.

5. The disk of radius 5 and center $(0,0)$.

6. The trapezoid with vertices $(1,0)$, $(3,2)$, $(3,3)$,$(1,6)$.

7. The triangle bounded by the lines $y = x$, $x + y = 2$, and $x + 3y = 8$.

8. The region bounded by the ellipse $4x^2 + y^2 = 4$.

9. The triangle bounded by the lines $x = 0$, $y = 0$, and $2x + 3y = 6$.

10. The region bounded by the curves $y = e^x$, $y = 1 - x$, and $x = 1$.

11. The quadrilateral bounded by the lines $y = 1$, $y = 2$, $y = x$, and $y = x/3$.

12. The quadrilateral bounded by the lines $x = 1$, $x = 2$, $y = x$, and $y = 5 - x$.

In Exercises 13 to 16 draw the regions and describe them by horizontal cross sections.

13. $0 \le x \le 2$, $2x \le y \le 3x$

14. $1 \le x \le 2$, $x^3 \le y \le 2x^2$

15. $0 \le x \le \pi/4$, $0 \le y \le \sin x$ and $\pi/4 \le x \le \pi/2$, $0 \le y \le \cos x$

16. $1 \le x \le e$, $(x-1)/(e-1) \le y \le \ln x$

In Exercises 17 to 22 evaluate

17. $\int_0^1 \left(\int_0^x (x + 2y)\, dy\right)\, dx$

18. $\int_1^2 \left(\int_x^{2x}\, dy\right)\, dx$

19. $\int_0^2 \left(\int_0^{x^2} xy^2\, dy\right)\, dx$

20. $\int_1^2 \left(\int_0^y e^{x+y}\, dx\right)\, dy$

21. $\int_1^2 \left(\int_0^{\sqrt{y}} yx^2\, dx\right)\, dy$

22. $\int_0^1 \left(\int_0^x y \sin(\pi x)\, dy\right)\, dx$

23. Complete the calculation of the second iterated integral in Example 6.

24.

(a) Sketch the solid region S below the plane $z = 1 + x + y$ and above the triangle R in the xy-plane with vertices $(0,0)$, $(1,0)$, $(0,2)$.

(b) Describe R in terms of coordinates.

(c) Set up an iterated integral for the volume of S.

(d) Evaluate the expression in (c), and show in the manner of Figure 17.2.8(a) and 17.2.8(b) which integration you performed first.

(e) Carry out (c) and (d) in the other order of integration.

25. Let S be the solid region below the paraboloid $z = x^2 + 2y^2$ and above the rectangle in the xy plane with vertices $(0,0)$, $(1,0)$, $(1,2)$, $(0,2)$. Carry out the steps of Exercise 24.

26. Let S be the solid region below the saddle $z = xy$ and above the triangle in the xy plane with vertices $(1,1)$, $(3,1)$, and $(1,4)$. Carry out the steps of Exercise 24 .

27. Let S be the solid region below the saddle $z = xy$ and above the region in the first quadrant of the xy plane bounded by the parabolas $y = x^2$ and $y = 2x^2$ and the line $y = 2$. Carry out the steps of Exercise 24.

28. Find the mass of a lamina occupying the bounded region bounded by $y = 2x^2$ and $y = 5x - 3$ and whose density at (x, y) is xy.

29. Find the mass of a thin lamina occupying the triangle whose vertices are $(0,0)$, $(1,0)$, $(1,1)$ and whose density at (x, y) is $1/(1 + x^2)$.

30. The temperature at (x, y) is $T(x, y) = \cos(x + 2y)$. Find the average temperature in the triangle with vertices $(0,0)$, $(1,0)$, $(0,2)$.

31. The temperature at (x, y) is $T(x, y) = e^{x-y}$. Find the average temperature in the region in the first quadrant bounded by the triangle with vertices $(0,0)$, $(1,1)$, and $(3,1)$.

In Exercises 32 to 35 replace the iterated integral by an equivalent one with the order of integration reversed. First sketch the region R of integration.

32. $\int_0^2 \left(\int_0^{x^2} x^3 y \, dy \right) \, dx$

33. $\int_0^{\pi/2} \left(\int_0^{\cos x} x^2 \, dy \right) \, dx$

34. $\int_0^1 \left(\int_{x/2}^{x} xy \, dy \right) \, dx + \int_1^2 \left(\int_{x/2}^{1} xy \, dy \right) \, dx$

35. $\int_{-1/\sqrt{2}}^{0} \left(\int_{-x}^{\sqrt{1-x^2}} x^3 y \, dy \right) \, dx + \int_0^1 \left(\int_0^{\sqrt{1-x^2}} x^3 y \, dy \right) \, dx$

In Exercises 36 to 39 evaluate the iterated integrals. First sketch the region of integration.

36. $\int_0^1 \left(\int_x^1 \sin(y^2)\ dy \right)\ dx$

37. $\int_0^1 \left(\int_{\sqrt{x}}^1 \frac{dy}{\sqrt{1+y^3}} \right)\ dx$

38. $\int_0^1 \left(\int_{\sqrt[3]{y}}^1 \sqrt{1+x^4}/dx \right)\ dy$

39. $\int_1^2 \left(\int_1^y \frac{\ln x}{x}\ dx \right)\ dy + \int_2^4 \left(\int_{y/2}^2 \frac{\ln x}{x}\ dx \right)\ dy$

40. Let $f(x,y) = y^2 e^{y^2}$ and let R be the triangle bounded by $y = a$, $y = x/2$, and $y = x$. Assume that a is positive.

(a) Set up two iterated integrals for $\int_R f(P)\ dA$.

(b) Evaluate the easier one.

41. Let R be the finite region bounded by the curve $y = \sqrt{x}$ and the line $y = x$. Let $f(x,y) = (\sin(y))/y$ if $y \neq 0$ and $f(x,0) = 1$. Compute $\int_R f(P)\ dA$.

17.3 Computing $\int_R f(P)\, dA$ Using Polar Coordinates

This section shows how to evaluate $\int_R f(P)\, dA$ using polar coordinates. The method is appropriate when the region R has a simple description in polar coordinates, for instance if it is a disk or cardioid. As in Section 17.2, we first examine how to describe cross sections in polar coordinates. Then we describe the iterated integral in polar coordinates that equals $\int_R f(P)\, dA$.

Describing R in Polar Coordinates

In describing a region R in polar coordinates, we first determine the range of θ and then see how r varies for a fixed value of θ. (The reverse order is seldom useful.) Some examples show how to find how r varies.

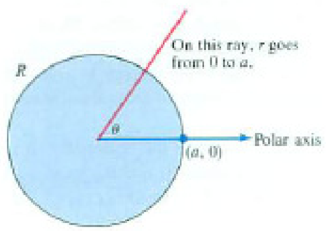


Figure 17.3.1

EXAMPLE 1 Let R be the disk of radius a and center at the pole of a polar coordinate system. (See Figure 17.3.1.) Describe R with cross sections by rays emanating from the pole.

SOLUTION To sweep out R, θ goes from 0 to 2π. On the ray for a fixed angle θ, r goes from 0 to a. (See Figure 17.3.1.) The description is

$$0 \leq \theta \leq 2\pi, \qquad 0 \leq r \leq a.$$

◇

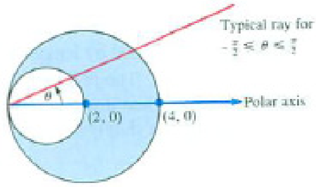


Figure 17.3.2

EXAMPLE 2 Let R be the region between the circles $r = 2\cos\theta$ and $r = 4\cos\theta$. Describe R in terms of cross sections by rays from the pole. (See Figure 17.3.2.)

SOLUTION To sweep out this region, use the rays from $\theta = -\pi/2$ to $\theta = \pi/2$. For each θ, r varies from $2\cos\theta$ to $4\cos\theta$. The description is

$$-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}, \qquad 2\cos\theta \leq r \leq 4\cos\theta.$$

◇

As Examples 1 and 2 suggest, polar coordinates provide simple descriptions for regions bounded by circles. The next example shows that polar coordinates may also provide simple descriptions of regions bounded by straight lines, especially if some of them pass through the origin.

EXAMPLE 3 Let R be the triangular region whose vertices, in rectangular coordinates, are $(0, 0)$, $(1, 1)$, and $(0, 1)$. Describe R in polar coordinates.

SOLUTION Inspection of R in Figure 17.3.3 shows that θ varies from $\pi/4$ to $\pi/2$. For each θ, r goes from 0 until the point (r,θ) is on the line $y=1$, that is, on the line $r\sin(\theta)=1$. Thus the upper limit of r for each θ is $1/\sin(\theta)$. The description of R is

$$\frac{\pi}{4} \le \theta \le \frac{\pi}{2}, \qquad 0 \le r \le \frac{1}{\sin(\theta)}.$$

◇

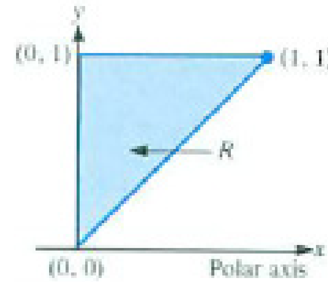


Figure 17.3.3

In general, cross sections by rays lead to descriptions of plane regions of the form:

$$\alpha \le \theta \le \beta, \qquad r_1(\theta) \le r \le r_2(\theta).$$

A Basic Difference Between Rectangular and Polar Coordinates

Before we can set up an iterated integral in polar coordinates for $\int_R f(P)\,dA$ we contrast certain properties of rectangular and polar coordinates.

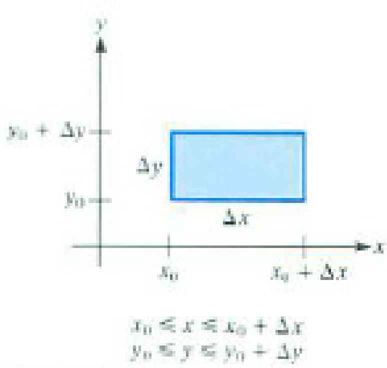
Consider points (x,y) in the plane that satisfy

$$x_0 \le x \le x_0 + \Delta x \qquad \text{and} \qquad y_0 \le y \le y_0 + \Delta y,$$

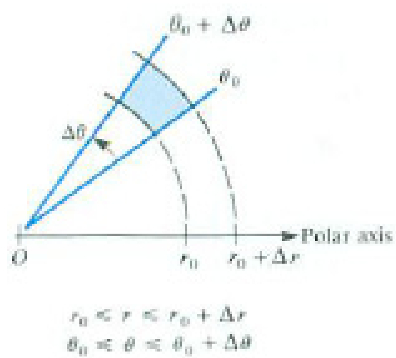
where x_0, Δx, y_0 and Δy are numbers with Δx and Δy positive. The set is a rectangle of sides Δx and Δy shown in Figure 17.3.4(a). The area of the rectangle is the product of Δx and Δy:

$$\text{Area} = \Delta x\ \Delta y.$$

We get a different area in polar coordinates.



(a)



(b)

Figure 17.3.4

The set in the plane consisting of the points (r, θ) such that

$$r_0 \leq r \leq r_0 + \Delta r \qquad \text{and} \qquad \theta_0 \leq \theta \leq \theta_0 + \Delta\theta,$$

where r_0, Δr, θ_0 and $\Delta\theta$ are numbers, with r_0, Δr, θ_0 and $\Delta\theta$ all positive, is shown in Figure 17.3.4(b).

The exact area is found in Exercise 30.

When Δr and $\Delta\theta$ are small, the set is approximately a rectangle, one side of which has length Δr and the other, $r_0\ \Delta\theta$. So its area is approximately $r_0\ \Delta r\ \Delta\theta$. In this case,

$$\text{Area} \approx r_0\ \Delta r\ \Delta\theta.$$

The area is not the product of Δr and $\Delta\theta$. (It could not be since $\Delta\theta$ is in radians, a dimensionless quantity so $\Delta r\ \Delta\theta$ has the dimension of length, not of area.)

It is necessary to replace dA by $r\ dr\ d\theta$, not by $dr\ d\theta$.

How to Evaluate $\int_R f(P)\ dA$ by an Iterated Integral in Polar Coordinates

The method for computing $\int_R f(P)\ dA$ with polar coordinates involves an iterated integral where the dA is replaced by $r\ dr\ d\theta$. A more detailed explanation of why the r must be added is given at the end of this section.

Notice the factor r in the integrand.

Evaluating $\int_R f(P)\ dA$ in Polar Coordinates

1. Express $f(P)$ as $f(r, \theta)$.
2. Describe the region R in polar coordinates:
$$\alpha \leq \theta \leq \beta, \qquad r_1(\theta) \leq r \leq r_2(\theta).$$
3. Evaluate the iterated integral
$$\int_\alpha^\beta \left(\int_{r_1(\theta)}^{r_2(\theta)} f(r, \theta) r\ dr \right) d\theta.$$

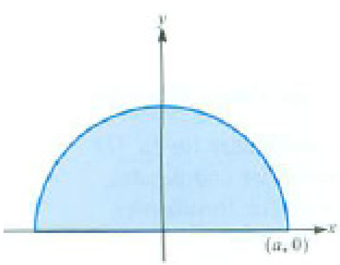


Figure 17.3.5

EXAMPLE 4 Let R be the semicircle of radius a shown in Figure 17.3.5. Let $f(P)$ be the distance from a point P to the x-axis. Evaluate $\int_R f(P)\ dA$

by an iterated integral in polar coordinates.
SOLUTION In polar coordinates, R has the description

$$0 \le \theta \le \pi, \qquad 0 \le r \le a.$$

SHERMAN: Is the extra annotation in the displayed equation overkill? (Don't worry about the extra space, production will improve this.)

The distance from P to the x-axis is y in rectangular coordinates. Since $y = r\sin(\theta)$, $f(P) = r\sin(\theta)$. Thus,

Notice the extra r in the integrand.

$$\int_R f(P)\,dA = \int_0^\pi \left(\int_0^a (r\sin(\theta)) \underbrace{r}_{\text{Remembe this } r.} \,dr \right) d\theta.$$

The calculation of the iterated integral is like that for an iterated integral in rectangular coordinates. First, evaluate the inside integral:

From here on the calculation is like those in the preceding section.

$$\int_0^a r^2 \sin(\theta)\,dr = \sin(\theta) \int_0^a r^2\,dr = \sin(\theta) \left(\frac{r^3}{3} \right) \Big|_0^a = \frac{a^3 \sin(\theta)}{3}.$$

The outer integral is therefore

$$\begin{aligned} \int_0^\pi \frac{a^3 \sin(\theta)}{3}\,d\theta &= \frac{a^3}{3} \int_0^\pi \sin(\theta)\,d\theta = \frac{a^3}{3}(-\cos(\theta))\Big|_0^\pi \\ &= \frac{a^3}{3}[(-\cos(\pi)) - (-\cos(0))] = \frac{a^3}{3}(1+1) = \frac{2a^3}{3}. \end{aligned}$$

Thus

$$\int_R f(P)\,dA = \frac{2a^3}{3}.$$

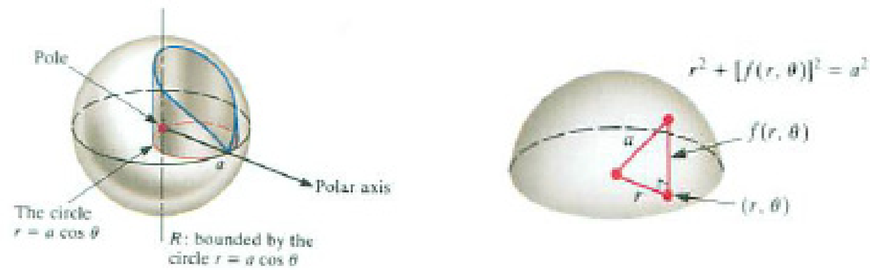
◇

EXAMPLE 5 A ball of radius a has its center at the pole of a polar coordinate system. Find the volume of the part of the ball that lies above the plane region R bounded by the curve $r = a\cos(\theta)$. (See Figure 17.3.6(a).)
SOLUTION It is necessary to describe R and f in polar coordinates, where $f(P)$ is the length of a cross section of the solid made by a vertical line through P. R is described as follows: r goes from 0 to $a\cos(\theta)$ for each θ in $[-\pi/2, \pi/2]$, that is,

$$-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}, \qquad 0 \le r \le a\cos\theta.$$

We have to express $f(P)$ in polar coordinates. Figure 17.3.6(b) shows the top half of a ball of radius a. By the Pythagorean Theorem,

$$r^2 + (f(r,\theta))^2 = a^2.$$



(a) (b)

Figure 17.3.6

Thus

$$f(r,\theta) = \sqrt{a^2 - r^2}.$$

Consequently,

$$\text{Volume} = \int_R f(P)\ dA = \int_{-\pi/2}^{\pi/2} \left(\int_0^{a\cos(\theta)} \sqrt{a^2 - r^2}\ r\ dr \right)\ d\theta.$$

Exploiting symmetry, compute half the volume, keeping θ in $[0, \pi/2]$, and then double the result:

$$\begin{aligned}
\int_0^{a\cos(\theta)} \sqrt{a^2 - r^2}\ r\ dr &= \left. \frac{-(a^2 - r^2)^{3/2}}{3} \right|_0^{a\cos(\theta)} = -\left(\frac{(a^2 - a^2\cos^2(\theta))^{3/2}}{3} - \frac{(a^2)^{3/2}}{3} \right) \\
&= \frac{a^3}{3} - \frac{(a^2 - a^2\cos^2(\theta))^{3/2}}{3} = \frac{a^3}{3} - \frac{a^3(1 - \cos^2(\theta))^{3/2}}{3} \\
&= \frac{a^3}{3}(1 - \sin^3(\theta)).
\end{aligned}$$

(The trigonometric formula $\sin(\theta) = \sqrt{1 - \cos^2(\theta)}$ is true when $0 \leq \theta \leq \pi/2$, but not when $-\pi/2 \leq \theta \leq 0$.)

Then comes the second integration:

$$\begin{aligned}
\int_0^{\pi/2} \frac{a^3}{3}(1-\sin^3(\theta))\, d\theta &= \frac{a^3}{3}\int_0^{\pi/2} (1-(1-\cos^2(\theta))\sin(\theta))\, d\theta \\
&= \frac{a^3}{3}\int_0^{\pi/2} 1-\sin(\theta)+\cos^2(\theta)\sin(\theta)\, d\theta \\
&= \frac{a^3}{3}\left(\theta+\cos(\theta)-\frac{\cos^3(\theta)}{3}\right)\Bigg|_0^{\pi/2} \\
&= \frac{a^3}{3}\left[\frac{\pi}{2}-\left(1-\frac{1}{3}\right)\right] = a^3\left(\frac{3\pi-4}{18}\right).
\end{aligned}$$

The total volume is twice as large so

$$V = a^3\left(\frac{3\pi-4}{9}\right).$$

◇

EXAMPLE 6 A circular disk of radius a is formed of a material that has density $\sigma(P)$ at each point P equal to the distance from P to the center.

(a) Set up an iterated integral in rectangular coordinates for the total mass of the disk.

(b) Set up an iterated integral in polar coordinates for the total mass of the disk.

(c) Compute the easier one.

SOLUTION The disk is shown in Figure 17.3.7.

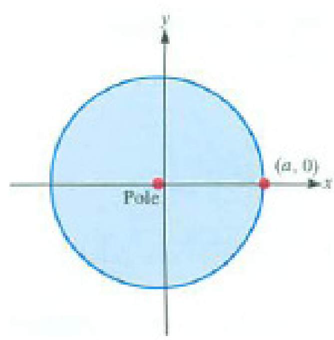


Figure 17.3.7

(a) (Rectangular coordinates) The density $\sigma(P)$ at the point $P=(x,y)$ is $\sqrt{x^2+y^2}$. The disk has the description

$$-a \le x \le a, \qquad -\sqrt{a^2-x^2} \le y \le \sqrt{a^2-x^2}.$$

Thus

$$\text{Mass} = \int_R \sigma(P)\, dA = \int_{-a}^{a}\left(\int_{-\sqrt{a^2-x^2}}^{\sqrt{a^2-x^2}} \sqrt{x^2+y^2}\, dy\right) dx.$$

(b) (Polar coordinates) The density $\sigma(P)$ at $P = (r, \theta)$ is r. The disk has the description

$$0 \leq \theta \leq 2\pi, \qquad 0 \leq r \leq a.$$

Thus

$$\text{Mass} = \int_R \sigma(P)\, dA \quad = \quad \int_0^{2\pi} \left(\int_0^a r \cdot r\, dr \right) d\theta = \int_0^{2\pi} \left(\int_0^a r^2\, dr \right) d\theta.$$

(c) Even the first integration in the iterated integral in (a) would be difficult. However, the iterated integral in (b) is straightforward: The first integration gives

$$\int_0^a r^2\, dr = \left. \frac{r^3}{3} \right|_0^a = \frac{a^3}{3}.$$

The second integration gives

$$\int_0^{2\pi} \frac{a^3}{3}\, d\theta = \left. \frac{a^3\theta}{3} \right|_0^{2\pi} = \frac{2\pi a^3}{3}.$$

The total mass is $2\pi a^3/3$.

◇

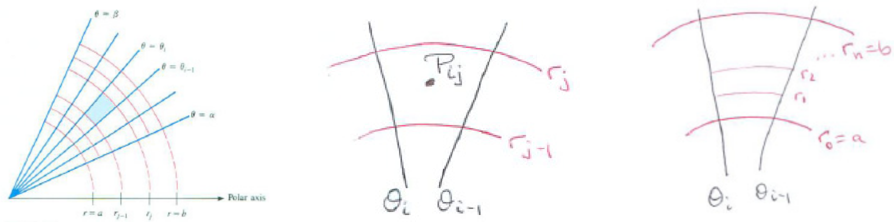
A Fuller Explanation of the Extra r in the Integrand

To estimate $\int_R f(P)\, dA$ for the region in the plane bounded by the circles $r = a$ and $r = b$ and the rays $\theta = \alpha$ and $\theta = \beta$, break the region into n^2 pieces with the aid of the partition $r_0 = a$, r_1, …, r_i, … $r_n = b$ and $\theta_0 = \alpha$, θ_1, …, θ_j, …, $\theta_n = \beta$ For convenience, assume that all $r_i - r_{i-1}$ are equal to Δr and all $\theta_j - \theta_{j-1}$ are equal to $\Delta\theta$. (See Figure 17.3.8(a).) A typical piece, shown in Figure 17.3.8(b), has area, exactly

$$A_{ij} = \frac{(r_j + r_{j-1})}{2}(r_j - r_{j-1})(\theta_i - \theta_{i-1}),$$

as shown in Exercise 6.

Let $P_{ij} = \left(\frac{1}{2}(r_j + r_{j-1}), \frac{1}{2}(\theta_i + \theta_{i-1})\right)$. Then the sum of the n^2 terms of the form $f(P_{ij})A_{ij}$ is an estimate of $\int_R f(P)\, dA$. Let us look closely at the



(a) (b) (c)

Figure 17.3.8

summand for the n pieces between the rays $\theta = \theta_{i-1}$ and $\theta = \theta_i$, as shown in Figure 17.3.8(c). The sum is

$$\sum_{i=1}^{n} f\left(\frac{r_j + r_{j-1}}{2}, \frac{\theta_i + \theta_{i-1}}{2}\right) \frac{r_j + r_{j+1}}{2} \Delta r \Delta\theta. \tag{17.3.1}$$

In (17.3.1), θ_i, θ_{i-1}, Δr, and $\Delta\theta$ are constants. If we define $g(r,\theta)$ to be $f(r,\theta)r$, then the sum is

$$\left(\sum_{j=1}^{n} g\left(\frac{r_j + r_{j+1}}{2}, \frac{\theta_i + \theta_{i-1}}{2}\right) \Delta r\right) \ \Delta\theta. \tag{17.3.2}$$

The sum in parentheses in (17.3.2) is an estimate of

$$\int_a^b g\left(r, \frac{\theta_i + \theta_{i-1}}{2}\right) dr.$$

Thus the sum (17.3.1), corresponding to the region between the rays $\theta = \theta_i$ and $\theta = \theta_{i-1}$, is

$$\sum_{i=1}^{n} \int_a^b g\left(r, \frac{\theta_i + \theta_{i-1}}{2}\right) dr \ \Delta\theta. \tag{17.3.3}$$

Let $h(\theta) = \int_a^b g(r,\theta)dr$. Then (17.3.3) equals

$$\sum_{i=1}^{n} h\left(\frac{\theta_i + \theta_{i-1}}{2}\right) \ \Delta\theta.$$

This is an estimate of $\int_a^b h(\theta)d\theta$. Hence the sum of n^2 terms of the form $f(P_{ij})A_{ij}$ is an approximation of

$$\int_\alpha^\beta h(\theta)\ d\theta \ = \ \int_\alpha^\beta \left(\int_a^b g(r,\theta)\ dr\right) d\theta = \int_\alpha^\beta \left(\int_a^b f(r,\theta) r\ dr\right) d\theta \tag{17.3.4}$$

The extra factor r appears when we obtained the first integral, $\int_a^b f(r,\theta)r\ dr$. The sum of the n^2 terms A_{ij}, which we knew approximated the double integral $\int_R f(P)\ dA$, we now see approximates also the iterated integral (17.3.4). Taking limits as $n \to \infty$ shows that the iterated integral equals the double integral.

Summary

We saw how to calculate an integral $\int_R f(P)\ dA$ by introducing polar coordinates. To do this, the plane region R is described in polar coordinates as

$$\alpha \leq \theta \leq \beta, \qquad r_1(\theta) \leq r \leq r_2(\theta).$$

Then

$$\int\limits_R f(P)\ dA = \int\limits_{\alpha}^{\beta} \int\limits_{r_1(\theta)}^{r_2(\theta)} f(r,\theta) r\ dr\ d\theta.$$

The extra r in the integrand is due to the fact that a small region corresponding to changes dr and $d\theta$ has area area approximately $r\ dr\ d\theta$ (not $dr\ d\theta$). Polar coordinates are convenient when either the function f or the region R has a simple description in terms of r and θ.

EXERCISES for Section 17.3

In Exercises 1 to 6 draw and describe the regions in the form $\alpha \leq \theta \leq \beta$, $r_1(\theta) \leq r \leq r_2(\theta)$.

1. The region inside the curve $r = 3 + \cos(\theta)$.

2. The region between the curve $r = 3 + \cos(\theta)$ and the curve $r = 1 + \sin(\theta)$.

3. The triangle whose vertices have the rectangular coordinates $(0, 0)$,$(1, 1)$, and $(1, \sqrt{3})$.

4. The circle bounded by the curve $r = 3\sin(\theta)$.

5. The region shown in Figure 17.3.9.

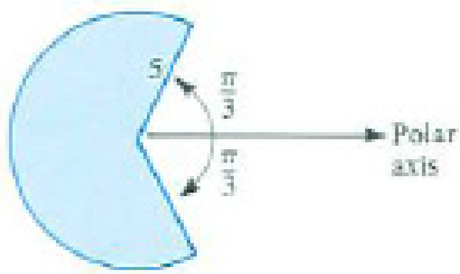


Figure 17.3.9

6. The region in the loop of the three-leaved rose, $r = \sin(3\theta)$, that lies in the first quadrant.

7.

(a) Draw the region R bounded by the lines $y = 1$, $y = 2$, $y = x$, $y = x/\sqrt{3}$.

(b) Describe R in terms of horizontal cross sections,

(c) Describe R in terms of vertical cross sections,

(d) Describe R in terms of cross sections by polar rays.

8.

(a) Draw the region R whose description is given by

$$-2 \leq y \leq 2, \qquad -\sqrt{4 - y^2} \leq x \leq \sqrt{4 - y^2}.$$

(b) Describe R by vertical cross sections.

(c) Describe R by cross sections formed by using polar rays.

9. Describe in polar coordinates the square whose vertices have rectangular coordinates $(0,0)$, $(1,0)$, $(1,1)$, $(0,1)$.

10. Describe the trapezoid whose vertices have rectangular coordinates $(0,1)$, $(1,1)$, $(2,2)$, $(0,2)$.

(a) by horizontal cross sections

(b) by vertical cross sections

(c) in polar coordinates

In Exercises 5 to 14 draw the region R and evaluate $\int_R r^2 \, dA$.

11. $-\pi/2 \le \theta \le \pi/2$, $0 \le r \le \cos(\theta)$

12. $0 \le \theta \le \pi/2$, $0 \le r \le \sin^2(\theta)$

13. $0 \le \theta \le 2\pi$, $0 \le r \le 1 + \cos(\theta)$

14. $0 \le \theta \le \pi/6$, $0 \le r \le \sin 2(\theta)$

In Exercises 15 to 18 draw R and evaluate $\int_R y^2 \, dA$.

15. The circle of radius a, center at the pole.

16. The circle of radius a with center at $(a,0)$ in polar coordinates.

17. The region within the cardioid $r = 1 + \sin(\theta)$.

18. The region within one leaf of the four-leaved rose $r = \sin(2\theta)$.

The average of a function $f(P)$ over a region R in the plane is defined as $\int_R f(P) \, dA$ divided by the area of R. In Exercises 19 to 22, find this average.

19. $f(P)$ is the distance from P to the pole; R is one leaf of the three-leaved rose, $r = \sin(3\theta)$.

20. $f(P)$ is the distance from P to the x-axis; R is the region between the rays $\theta = \pi/6$, $\theta = \pi/4$, and the circles $r = 2$, $r = 3$.

21. $f(P)$ is the distance from P to a fixed point on the border of a disk R of radius a. (Choose the pole wisely.)

22. $f(P)$ is the distance from P to the x-axis; R is the region within the cardioid $r = 1 + \cos(\theta)$.

In Exercises 23 to 26 evaluate the iterated integral using polar coordinates.

23. $\int_0^1 \left(\int_0^x \sqrt{x^2 + y^2} \, dy \right) \, dx$

24. $\int_0^1 \left(\int_0^{\sqrt{1-x^2}} x^3\,dy\right)\,dx$

25. $\int_0^1 \left(\int_x^{\sqrt{1-x^2}} xy\,dy\right)\,dx$

26. $\int_1^2 \left(\int_{x/\sqrt{3}}^{\sqrt{3}x} (x^2+y^2)^{3/2}\,dy\right)\,dx$

27. Evaluate the integrals over the regions.

(a) $\int_R \cos(x^2+y^2)\,dA$ when R is the portion in the first quadrant of the disk of radius a centered at the origin.

(b) $\int_R \sqrt{x^2+y^2}\,dA$ when R is the triangle bounded by the line $y=x$, the line $x=2$, and the x-axis.

28. Find the volume of the region above the paraboloid $z=x^2+y^2$ and below the plane $z=x+y$.

29. The area of a region R is equal to $\int_R 1\,dA$. Use this to find the area of a disk of radius a. (Use an iterated integral in polar coordinates.)

30. Find the area of the shaded region in Figure 17.3.4(b) as follows:

(a) Find the area of the ring between two circles, one of radius r_0, the other of radius $r_0+\Delta r$.

(b) What fraction of the area in (a) is included between two rays whose angles differ by $\Delta\theta$?

(c) Show that the area of the shaded region in Figure 17.3.4(b) is

$$\left(r_0+\frac{\Delta r}{2}\right)\Delta r\Delta\theta.$$

31. In Example 5 we computed half of

$$\int_{-\pi/2}^{\pi/2}\left(\int_0^{a\cos(\theta)} \sqrt{a^2-r^2}\;r\,dr\right)\,d\theta$$

and doubled the result. Evaluate it directly. The result should still be $a^3(3\pi-4)/9$. Use trigonometric formulas with care.

Exercise 32 shows that $\int_0^\infty e^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}$. Before beginning this exercise, read these two quotes that concern this improper integral.

> Once when lecturing to a class he [the physicist Lord Kelvin] used the word "mathematician" and then interrupting himself asked the class: "Do you know what a mathematician is?" Stepping to his blackboard he wrote upon it: $\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}$. Then putting his finger on what he had written, he turned to his class and said, "A mathematician is one to whom this is as obvious as that twice two makes four is to you."

S. P. Thompson, in *Life of Lord Kelvin* (Macmillan, London, 1910).

SHERMAN: Lots of changes here. Some of your suggested edits appeared to me to be wrong, e.g., moving citation. Please check, and restate any more changes that are needed.

> Many things are not accessible to intuition at all, the value of $\int_0^{\infty} e^{-x^2}\,dx$ for instance.

J. E. Littlewood, in "Newton and the attraction of the sphere", *Mathematical Gazette*, vol. 63, 1948.

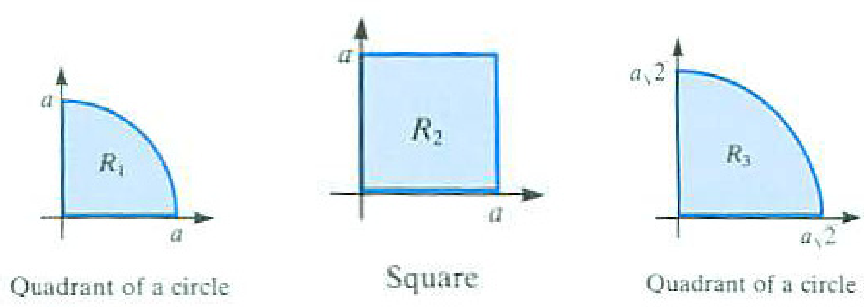
32. Let $f(P) = e^{-r^2}$ where r is the distance from P to the origin. Hence, $f(r,\theta) = e^{-r^2}$ in polar coordinates and, in rectangular coordinates, $f(x,y) = e^{-x^2-y^2}$. In Figure 17.3.10, R_1 is inside R_2 and R_2 is inside R_3.

(a) Show that $\int_{R_1} f(P)\,dA = \frac{\pi}{4}\left(1 - e^{-a^2}\right)$ and that $\int_{R_3} f(P)\,dA = \frac{\pi}{4}\left(1 - e^{-2a^2}\right)$.

(b) By considering $\int_{R_2} f(P)\,dA$ and the results in (a), show that

$$\frac{\pi}{4}\left(1 - e^{-a^2}\right) < \left(\int_0^{\infty} e^{-x^2}\,dx\right)^2 < \frac{\pi}{4}\left(1 - e^{-2a^2}\right).$$

(c) Show that $\int_0^{\infty} e^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}$.



(a) (b) (c)

Figure 17.3.10

33. Figure 17.3.11 shows the bell curve or normal curve. From Exercise 32, $\int_0^{\infty} e^{-x^2}\,dx = \sqrt{\pi}/2$. Show that the area under the curve in Figure 17.3.11 is $\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-x^2}\,dx = 1$.

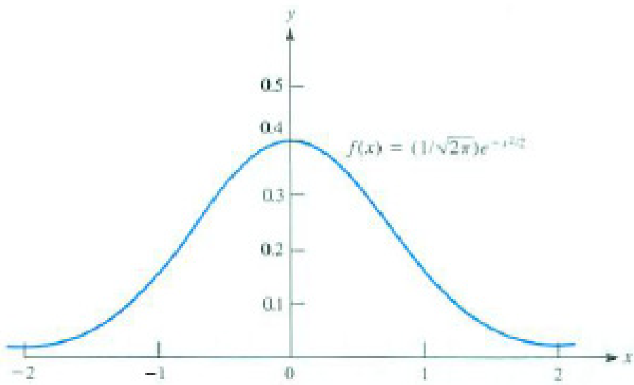


Figure 17.3.11

34. (The spread of epidemics.) In the theory of a spreading epidemic it is assumed that the probability that a contagious person infects someone D miles away depends only on D. Suppose that a population is uniformly distributed in a circular city whose radius is 1 mile. Assume that the probability of infection is proportional to $2-D$. For a point Q let $f(P) = 2 - \overline{PQ}$. Let R be the region occupied by the city.

(a) Why is the exposure of a person residing at Q proportional to $\int_R f(P)\,dA$, assuming that contagious people are uniformly distributed throughout the city?

(b) Compute the definite integral (i) when Q is the center of town and (ii) when Q is on the edge of town.

(c) In view of (b), which is the safer place?

Transportation problems lead to integrals over plane sets, as Exercises 35 to 38 illustrate.

35. Show that the average travel distance from the center of a disk of area A to points in the disk is $2\sqrt{A}/(3\sqrt{\pi}) \approx 0.376\sqrt{A}$.

36. Show that the average travel distance from the center of a regular hexagon of area A to points in the hexagon is

$$\frac{\sqrt{2A}}{3^{3/4}}\left(\frac{1}{3} + \frac{\ln 3}{4}\right) \approx 0.377\sqrt{A}.$$

37. Show that the average travel distance from the center of a square of area A to points in the square is $(\sqrt{2} + \ln(\tan(3\pi/8)))\sqrt{A/6} \approx 0.383\sqrt{A}$.

38. Show that the average travel distance from the centroid of an equilateral triangle of area A to points in the triangle is

$$\frac{\sqrt{A}}{3^{9/4}}\left(2\sqrt{3}+\ln(\tan(\frac{5\pi}{12}))\right)\approx 0.404\sqrt{A}$$

The centroid of a triangle is the intersection point of its medians.

In Exercises 35 to 38 distance is the ordinary straight-line distance. In cities the usual street pattern suggests that the metropolitan distance between the points (x_1, y_1) and (x_2, y_2) should be measured by $|x_1 - x_2| + |y_1 - y_2|$.

39. Show that if in Exercise 35 metropolitan distance is used, then the average is $8\sqrt{A}/(3\pi^{3/2}) \approx 0.470\sqrt{A}$.

40. Show that if in Exercise 38 metropolitan distance is used, then the average is $\sqrt{A/2}$. In most cities the metropolitan average tends to be about 25 percent larger than the direct-distance average.

41.

Sam: The formula in this section for integrating in polar coordinates is wrong. I'll get the right formula. We don't need the factor r.

Jane: But the book's formula gives correct answers.

Sam: I don't care. Let $f(r, \theta)$ be positive and I'll show how to integrate over the set R bounded by $r = b$ and $r = a$, $b > a$, and $\theta = \beta$ and $\theta = \alpha$, $\beta > \alpha$. We have $\int_R f(P)\ dA$ is the volume under the graph of f and above R. Right?

Jane: Right.

Sam: The area of the cross section corresponding to a fixed angle θ is $\int_a^b f(r, \theta)\ dr$. Right?

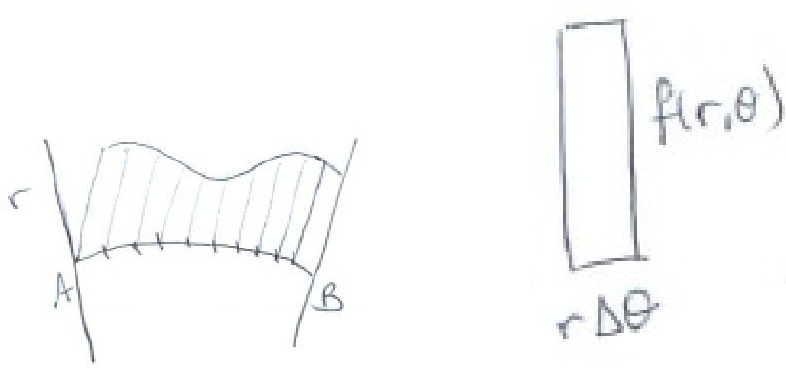
Jane: Right.

Sam: So I just integrate cross-sectional areas as θ goes from α to β, and the volume is therefore $\int_\alpha^\beta (\int_a^b f(r, \theta)\ dr)\ d\theta$. Perfectly straightforward. I hate to overthrow a formula that's been around for three centuries.

What does Jane say next?

42.

Jane: You looked at fixed θ. I'll use a fixed r. Look at the area under the graph of f and above the circle of radius r. I'll draw this fence for you (see Figure 17.3.12(a).



(a) (b)

Figure 17.3.12

To estimate its area I'll cut the arc AB into n sections of equal length by angle $\theta_0 = a, \ldots, \theta_n = \beta$.

Then I break AB into n short arcs, each of length $r\Delta\theta$. (Remember, Sam, how radians are defined.) The approximation to the shaded area looks like Figure 17.3.12(b) and resembles a rectangle of height $f(r,\theta)$ and width $r\ \Delta\theta$. So the local approximation to the area is $f(r,\theta)\ r\ \Delta\theta$ and the area of the fence is $\int_\alpha^\beta f(r,\theta)\ r\ \ d\theta$. Here r is fixed. Then I integrate this cross-sectional area as r goes from a to b. The total volume is then $\int_a^b \left(\int_\alpha^\beta f(r,\theta)\ r\ d\theta\right)\ dr$. That gives the volume, which equals $\int_R f(r,\theta)\ dA$.

Sam: All right.

Jane: At least it gives the factor r.

Sam: Maybe we're both right.

What does Jane say?

17.4 The Triple Integral: Integrals Over Solid Regions

In this section we define integrals over solid regions in space and show how to compute them by iterated integrals using rectangular coordinates. Throughout we assume the regions are bounded by smooth surfaces and the functions are continuous.

The Triple Integral

Let R be a region in space bounded by a surface. For instance, R could be a ball, a cube, or a tetrahedron. Let f be a function defined on R.

For a positive integer n break R into n regions $R_1, R_2, \dots R_n$. Choose a point P_1 in R_1, P_2 in R_2, ..., P_n in R_n. Let the volume of R_i be V_i. Then

$$\lim_{\text{diameters of the } R_i \text{ approach } 0} \sum_{i=1}^{n} f(P_i)V_i \quad \text{exists.}$$

It is denoted

$$\int_R f(P)\, dV \tag{17.4.1}$$

and is called the integral of f over R or the **triple integral** of f over R.

EXAMPLE 1 If $f(P) = 1$ for each point P in a solid region R, compute $\int_R f(P)\, dV$.

SOLUTION Each approximating sum $\sum_{i=1}^{n} f(P_i)V_i$ has the value

$$\sum_{i=1}^{n} 1 \cdot V_i = V_1 + V_2 + \cdots + V_n = \text{Volume of } R.$$

Hence

$$\int_R f(P)\, dV = \text{Volume of } R,$$

which is useful for computing volumes. ◇

Average of a function

The **average value** of a function f defined on a region R in space is defined as

$$\frac{\int_R f(P)\, dV}{\text{Volume of } R}.$$

This is the analog of the definition of the average value of a function over an interval (Section 6.3) or the average value of a function over a plane region (Section 17.1).

If a mass is distributed in a region R its density at a point P is defined as a limit. For positive r let $V(r)$ be the volume of a ball of radius r centered at P, and $m(r)$ the mass in it. Then the density at P is

$$\lim_{r\to 0}\frac{m(r)}{V(r)}.$$

If f describes the density of matter in R, then the average value of f is the density of a homogeneous solid occupying R and having the same total mass as the given solid. That is, if the average density

$$\frac{\int_R f(P)\ dV}{\text{Volume of } R}.$$

is multiplied by the volume of R, the product is

$$\int\limits_R f(P)\ dV,$$

which is the total mass.

An Interpretation of $\int_R f(P)\ dV$.

Triple integrals appear in the study of gravitation, rotating bodies, centers of gravity, and in electro-magnetic theory. The simplest way to think of them is to interpret $f(P)$ as the density at P of some distribution of matter. Then $\int_R f(P)\ dV$ is the total mass in a region R.

We can't picture $\int_R f(P)\ dV$ as measuring the volume of something. We could do this for $\int_R f(P)\ dA$, because we could use two dimensions for describing the region of integration and then the third dimension for the values of the function, obtaining a surface in three-dimensional space. However, with $\int_R f(P)\ dV$, we use up three dimensions describing the region of integration. We need four-dimensional space to show the values of the function.

Describing a Solid Region

In order to evaluate triple integrals, it is necessary to describe solid regions in terms of coordinates.

A description of a solid region in rectangular coordinates has the form

$$a \le x \le b, \qquad y_1(x) \le y \le y_2(x), \qquad z_1(x,y) \le z \le z_2(x,y).$$

The inequalities on x and y describe the projection of the region on the xy-plane. The inequalities for z then tell how z varies on a line parallel to the z-axis and passing through the point (x, y) in the projection. (See Figure 17.4.1.)

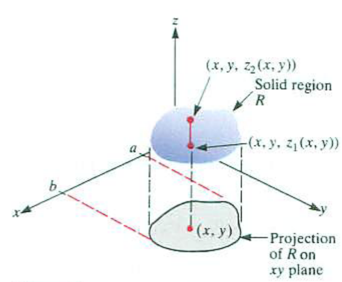
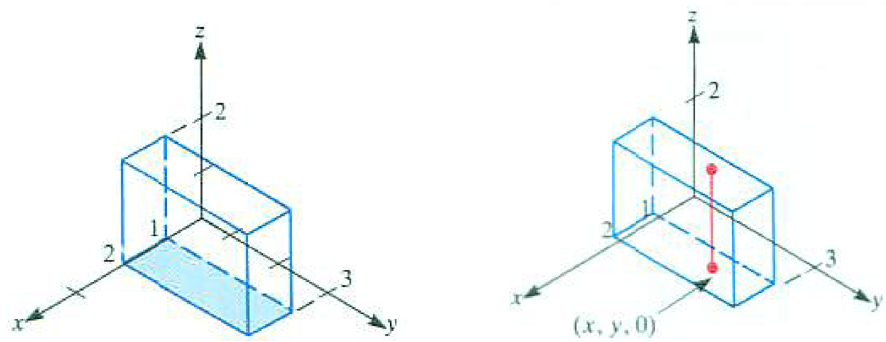


Figure 17.4.1

EXAMPLE 2 Describe in terms of x, y, and z the rectangular box shown in Figure 17.4.2(a).



(a) (b)

Figure 17.4.2

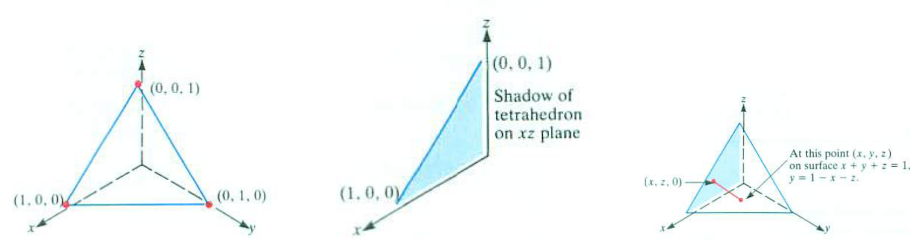
SOLUTION The projection of the box on the xy-plane has a description $1 \le x \le 2$, $0 \le y \le 3$. For each point in it, z varies from 0 to 2, as shown in Figure 17.4.2(b). So the description of the box is

$$1 \le x \le 2, \qquad 0 \le y \le 3, \qquad 0 \le z \le 2,$$

which is read from left to right as "x goes from 1 to 2; for each x, y goes from 0 to 3; for each x and y, z goes from 0 to 2."

We could have changed the order of x and y in the description or projected the box on one of the other two coordinate planes. There are six possible descriptions. ◇

EXAMPLE 3 Describe by cross sections the tetrahedron bounded by the planes $x = 0$, $y = 0$, $z = 0$, and $x + y + z = 1$, as shown in Figure 17.4.3(a) and it is described by



(a) (b) (c)

Figure 17.4.3

SOLUTION Project the tetrahedron onto the xz plane. The projection is shown in Figure 17.4.3(b). A description of the shadow is

$$0 \leq x \leq 1, \qquad 0 \leq z \leq 1 - x,$$

since the slanted edge has the equation $x + z = 1$. For each point (x, z) in the projection, y ranges from 0 up to the value of y that satisfies the equation $x+y+z = 1$, that is, up to $y = 1-x-z$. (See Figure 17.4.3(c).) A description of the tetrahedron is

$$0 \leq x \leq 1, \qquad 0 \leq z \leq 1 - x, \qquad 0 \leq y \leq 1 - x - z.$$

That is, x goes from 0 to 1; for each x, z goes from 0 to $1 - x$; for each x and z, y goes from 0 to $1 - x - z$. ◇

EXAMPLE 4 Describe in rectangular coordinates the ball of radius 4 whose center is at the origin.

SOLUTION The projection of the ball on the xy-plane is the disk of radius 4 and center $(0, 0)$. Its description is

$$-4 \leq x \leq 4, \qquad -\sqrt{16 - x^2} \leq y \leq \sqrt{16 - x^2}.$$

Hold (x, y) fixed in the xy plane and consider the way z varies on the line parallel to the z-axis that passes through the point $(x, y, 0)$. Since the sphere that bounds the ball has equation

$$x^2 + y^2 + z^2 = 16,$$

for each (x, y), z varies from

$$-\sqrt{16 - x^2 - y^2} \qquad \text{to} \qquad \sqrt{16 - x^2 - y^2}.$$

This describes the line segment shown in Figure 17.4.4.

The ball, therefore, has a description

$$-4 \leq x \leq 4, \quad -\sqrt{16 - x^2} \leq y \leq \sqrt{16 - x^2}, \quad \sqrt{16 - x^2 - y^2} \leq z \leq \sqrt{16 - x^2 - y^2}.$$

◇

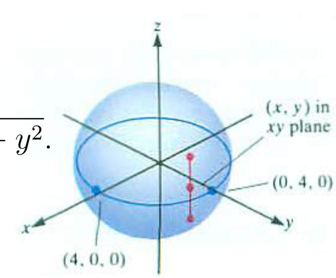


Figure 17.4.4

Iterated Integrals for $\int_R f(P)\, dV$

The iterated integral in rectangular coordinates for $\int_R f(P)\, dV$ is similar to that for evaluating integrals over plane sets. It involves three integrations instead of two. The limits of integration are determined by the description of R in rectangular coordinates. If R has the description

$$a \leq x \leq b, \qquad y_1(x) \leq y \leq y_2(x), \qquad z_1(x,y) \leq z \leq z_2(x,y),$$

then

$$\int\limits_R f(P)\ dV = \int\limits_a^b \left(\int\limits_{y_1(x)}^{y_2(x)} \left(\int\limits_{z_1(x,y)}^{z_2(x,y)} f(x,y,z)\ dz \right)\ dy \right)\ dx.$$

An example illustrates how this formula is applied. In Exercise 31 an argument for its plausibility is presented.

EXAMPLE 5 Compute $\int_R z\ dV$, where R is the tetrahedron in Example 3.

SOLUTION One description of the tetrahedron is

$$0 \leq y \leq 1, \qquad 0 \leq x \leq 1-y, \qquad 0 \leq z \leq 1-x-y.$$

Hence

$$\int\limits_R z\ dV = \int\limits_0^1 \left(\int\limits_0^{1-y} \left(\int\limits_0^{1-x-y} z\ dz \right)\ dx \right)\ dy.$$

Compute the inner integral first, treating x and y as constants. By the Fundamental Theorem of Calculus,

$$\int\limits_0^{1-x-y} z\ dz = \left. \frac{z^2}{2} \right|_{z=0}^{z=1-x-y} = \frac{(1-x-y)^2}{2}.$$

The next integration, where y is fixed, is

$$\int\limits_0^{1-y} \frac{(1-x-y)^2}{2}\ dx \quad = \quad -\left. \frac{(1-x-y)^3}{6} \right|_{x=0}^{x=1-y} = -\frac{0^3}{6} + \frac{(1-y)^3}{6} = \frac{(1-y)^3}{6}.$$

The third integration is

$$\int\limits_0^1 \frac{(1-y)^3}{6}\ dy = -\left. \frac{(1-y)^4}{24} \right|_0^1 = -\frac{0^4}{24} + \frac{1^4}{24} = \frac{1}{24}.$$

This completes the calculation that

$$\int\limits_R z\ dV = \frac{1}{24}.$$

◇

A Word about Four-Dimensional Space

We can think of two-dimensional space as the set of ordered pairs (x, y) of real numbers. The set of ordered triplets of real numbers (x, y, z) represents three-dimensional space. The set of ordered quadruplets of real numbers (x, y, z, t) represents four-dimensional space.

In two-dimensional space the set of points of the form $(x, 0)$, the y-axis, meets the set of points of the form $(0, y)$, the y-axis, in a point, namely the origin $(0, 0)$. In 4-space the set of points of the form $(x, y, 0, 0)$ forms a plane congruent to the xy-plane. The set of points of the form $(0, 0, z, t)$ forms another. Their intersection is a single point $(0, 0, 0, 0)$. Can you picture two endless planes meeting in a single point? If so, please tell us how.

Summary

We defined $\int_{\mathcal{R}} f(P)\ dV$, where $\mathcal{R}$ is a region in space. The volume of a solid region $\mathcal{R}$ is $\int_{\mathcal{R}} dV$ and, if $f(P)$ is the density of matter at P, then $\int_{\mathcal{R}} f(P)\ dV$ is the total mass. We also showed how to evaluate integrals by introducing rectangular coordinates and computing an iterated integral.

There are six possible orders for the three variables x, y, and z. If $\mathcal{R}$ is described by

SHERMAN: I have made some changes here.

$$a \le x \le b, \qquad y_1(x) \le y \le y_2(x), \qquad z_1(x, y) \le z \le z_2(x, y).$$

Then

$$\int_R f(P)\ dV = \int_a^b \left(\int_{y_1(x)}^{y_2(x)} \left(\int_{z_1(x,y)}^{z_2(x,y)} f(x, y, z)\ dz \right)\ dy \right)\ dx.$$

To use any of the other five orders requires a corresponding description of $\mathcal{R}$.

EXERCISES for Section 17.4

Exercises 1 to 4 concern the definition of $\int_R f(P)\,dV$.

1. A cube of side 4 centimeters is made of a material of varying density. Near one corner, A, it is very light; at the opposite corner it is very dense. The density $f(P)$ (in grams per cubic centimeter) at a point P in the cube is the square of the distance from A to P (in centimeters). See Figure 17.4.5.

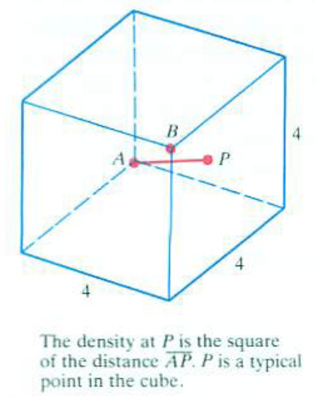


Figure 17.4.5

(a) Find upper and lower estimates for the mass of the cube by partitioning it into eight cubes.

(b) Using the same partition, estimate the mass of the cube, but select as the P_i's the centers of the cubes.

(c) What does (b) say about the average density in the cube?

2. How would you define the average distance from points of a set in space to a point P_0?

3. If R is a ball of radius r and $f(P) = 5$ for each point in R, compute $\int_R f(P)\,dV$ by examining approximating sums. The ball has volume $(4/3)\pi r^3$.

4. If R is a three-dimensional set and $f(P)$ is never more than 8 for all P in R,

(a) what can we say about the maximum value of $\int_R f(P)\,dV$?

(b) what can we say about the average of f over R?

In Exercises 5 to 10 draw the solids.

5. $1 \le x \le 3$, $0 \le y \le 2$, $0 \le z \le x$

6. $0 \le x \le 1$, $0 \le y \le 1$, $1 \le z \le 1 + x + y$

7. $0 \le y \le 1$, $0 \le x \le y^2$, $y \le z \le 2y$

8. $0 \leq y \leq 1,\ y^2 \leq x \leq y,\ 0 \leq z \leq x + y$

9. $-1 \leq z \leq 1,\ -\sqrt{1-z^2} \leq x \leq \sqrt{1-z^2},\ -\frac{1}{2} \leq y \leq \sqrt{1-x^2-z^2}$

10. $0 \leq z \leq 3,\ 0 \leq y \leq \sqrt{9-z^2},\ 0 \leq x \leq \sqrt{9-y^2-z^2}$

In Exercises 11 to 14 evaluate.

11. $\int_0^1 \left(\int_0^2 \left(\int_0^x z\ dz \right)\ dy \right)\ dx$

12. $\int_0^1 \left(\int_{x^3}^{x^2} \left(\int_0^{x+y} z\ dz \right)\ dy \right)\ dx$

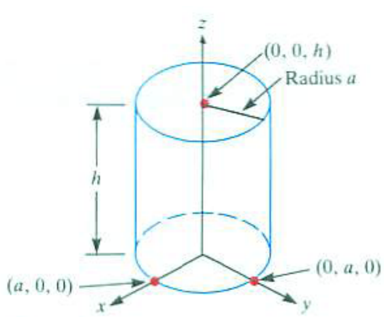
13. $\int_2^3 \left(\int_x^{2x} \left(\int_0^1 (x+z)\ dz \right)\ dy \right)\ dx$

14. $\int_0^1 \left(\int_0^x \left(\int_0^3 (x^2+y^2)\ dz \right)\ dy \right)\ dx$

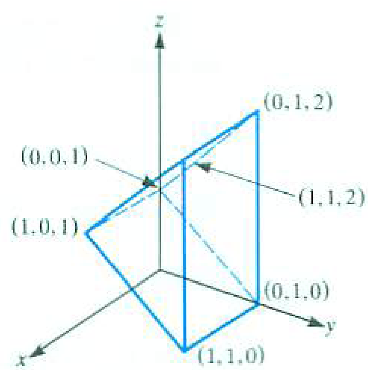
15. Describe the solid cylinder of radius a and height h shown in Figure 17.4.6(a) in rectangular coordinates

(a) in the order x, y, z,

(b) in the order x, z, y.



(a)



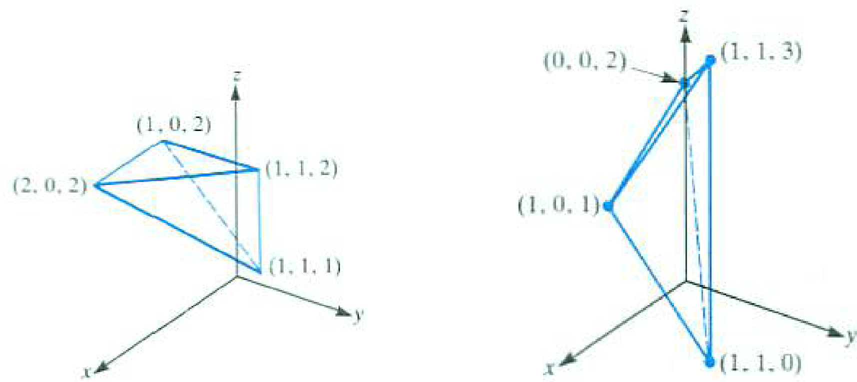
(b)

Figure 17.4.6

16. Describe the prism shown in Figure 17.4.6(b) in rectangular coordinates in two ways.

(a) First project it onto the xy-plane.

(b) First project it onto the xz-plane.



(a) (b)

Figure 17.4.7

17. Describe the tetrahedron in Figure 17.4.7(a) in rectangular coordinates in two ways:

(a) First project it onto the xy-plane.

(b) First project it onto the xz-plane.

18. Describe the tetrahedron in Figure 17.4.7(b) with vertices at $(1,1,0)$, $(1,0,1)$, $(0,0,2)$, and $(1,1,3)$.

(a) Draw its projection on the xy-plane.

(b) Obtain equations of its top and bottom planes.

19. Let R be the tetrahedron whose vertices are $(0,0,0)$, $(a,0,0)$, $(0,b,0)$, and $(0,0,c)$, where a, b, and c are positive.

(a) Sketch R.

(b) Find the equation of its top surface.

(c) Compute $\int_R z\ dV$.

20. Compute $\int_R z\ dV$, where R is the region above the rectangle whose vertices are $(0,0,0)$, $(2,0,0)$, $(2,3,0)$, and $(0,3,0)$ and below the plane $z = x + 2y$.

21. Find the mass of the cube in Exercise 1. (See Figure 17.4.1.)

22. Find the average value of the square of the distance from a corner of a cube of side a to points in the cube.

23. Find the average of the square of the distance from a point in a cube of side a to the center of the cube.

24. A solid consists of points below the surface $z = xy$ that are above the triangle whose vertices are $(0,0,0)$, $(1,0,0)$, and $(0,2,0)$. If the density at (x,y,z) is $x+y$, find the total mass.

25. Compute $\int_R xy \; dV$ for the tetrahedron of Example 3.

26.

(a) Describe in rectangular coordinates the right circular cone, C, of radius r and height h if its axis is on the positive z-axis and its vertex is at the origin. Draw the cross sections for fixed x and fixed x and y.

(b) Find $\int_C z \; dV$.

(c) Find the average value of z in the cone in (a).

27. The temperature at (x,y,z) is e^{-x-y-z}. Find the average temperature in the tetrahedron whose vertices are $(0,0,0)$, $(1,1,0)$, $(0,0,2)$, and $(1,0,0)$.

28. The temperature at (x,y,z), $y > 0$, is $e^{-x}/\sqrt{y}$. Find the average temperature in the region bounded by the cylinder $y = x^2$, the plane $y = 1$, and the plane $z = 2y$.

29. Without using an iterated integral, evaluate $\int_R x \; dV$, where R is a ball whose center is $(0,0,0)$ and whose radius is a.

30. The work done in lifting a weight of w pounds a vertical distance of x feet is wx foot-pounds. Imagine that through geological activity a mountain is formed consisting of material originally at sea level. Let the density of the material near point P in the mountain be $g(P)$ pounds per cubic foot and the height of P be $h(P)$ feet. What definite integral represents the total work expended in forming the mountain? This type of problem is important in the geological theory of mountain formation.

31. In Section 17.2 an intuitive argument was presented for the equality

$$\int_R f(P)\,dA = \int_a^b \left(\int_{y_1(x)}^{y_2(x)} f(x,y)\,dy \right) dx.$$

Here is an intuitive argument for the equality

$$\int_R f(P)\,dV = \int_{x_1}^{x_2} \left(\int_{y_1(x)}^{y_2(x)} \left(\int_{x_1(x,y)}^{x_2(x,y)} f(x,y,z)\,dz \right) dy \right) dx.$$

Interpret $f(P)$ as density.

(a) Let $R(x)$ be the plane cross section consisting of all points in R with abscissa x. Show that the average density in $R(x)$ is

$$\frac{\int_{y_1(x)}^{y_2(x)} \left(\int_{z_1(x,y)}^{z_2(x,y)} f(x,y,z)\,dz \right) dy}{\text{Area of } R(x)}.$$

(b) Show that the mass of R between the plane sections $R(x)$ and $R(x+\Delta x)$ is approximately

$$\int_{y_1(x)}^{y_2(x)} \left(\int_{z_1(x,y)}^{z_2(x,y)} f(x,y,z)\,dz \right) dy\,\Delta x.$$

(c) From (b) obtain an iterated integral in rectangular coordinates for $\int_R f(P)\,dV$.

17.5 Cylindrical and Spherical Coordinates

Rectangular coordinates provide convenient descriptions of solids bounded by planes. In this section we describe two other coordinate systems, cylindrical, ideal for describing circular cylinders, and spherical, ideal for describing spheres, balls, and cones. Both will be used in the next section to evaluate multiple integrals by iterated integrals.

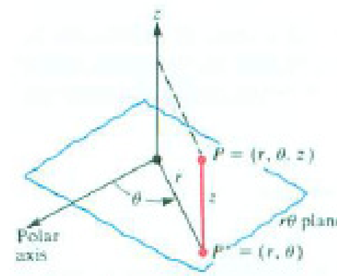
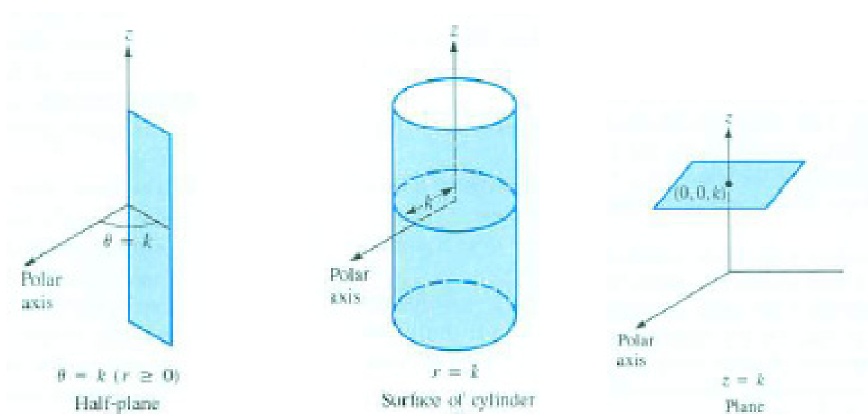


Figure 17.5.1

Cylindrical Coordinates

Cylindrical coordinates combine polar coordinates in the plane with the z of rectangular coordinates in space. A point P in space receives the name (r, θ, z) as in Figure 17.5.1. We are free to choose the direction of the polar axis; usually it will coincide with the x-axis of an (x, y, z) system. The point (r, θ, z) is directly above (or below) $P^* = (r, \theta)$ in the $r\theta$-plane. Since the set of points $P = (r, \theta, z)$ for which r is some constant is a circular cylinder, this coordinate system is convenient for describing such cylinders. As with polar coordinates, the cylindrical coordinates of a point are not unique.

Figure 17.5.2 shows the coordinate surfaces $\theta = k$, $r = k$, and $z = k$, where k is a positive number.

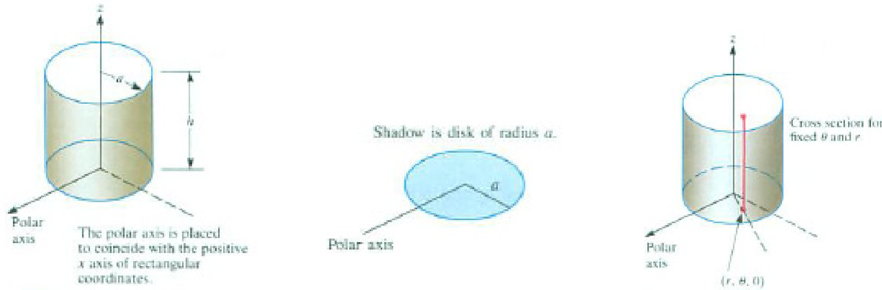


(a) (b) (c)

Figure 17.5.2 The coordinates surfaces for cylindrical coordinates: (a) $\theta = k$, (b) $r = k$, and (c) $z = k$, where k is a positive number.

EXAMPLE 1 Describe a solid cylinder of radius a and height h in cylindrical coordinates. Assume that its axis is on the positive z-axis and its lower base has its center at the pole, as in Figure 17.5.3(a).

SOLUTION The projection of the cylinder on the $r\theta$-plane is the disk of



(a) (b) (c)

Figure 17.5.3

radius a with center at the pole shown in Figure 17.5.3(b). Its description is

$$0 \le \theta \le 2\pi, \qquad 0 \le r \le a.$$

For each point (r, θ) in the projection, the line through it and parallel to the z-axis intersects the cylinder in a line segment. On the segment z varies from 0 to h. (See Figure 17.5.3(c).) Thus a description of the cylinder is

$$0 \le \theta \le 2\pi, \qquad 0 \le r \le a, \qquad 0 \le z \le h.$$

◇

EXAMPLE 2 Describe in cylindrical coordinates the region in space formed by the intersection of a solid (endless) cylinder of radius 3 with a ball of radius 5 whose center in on the axis of the cylinder. Place the cylindrical coordinate system as shown in Figure 17.5.4.

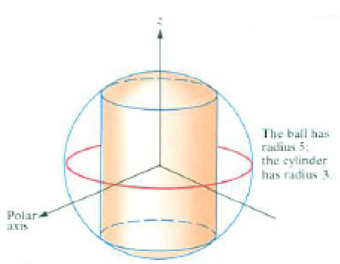


Figure 17.5.4

SOLUTION The point $P = (r, \theta, z)$ is a distance $\sqrt{r^2 + z^2}$ from the origin O, for, by the Pythagorean theorem, $r^2 + z^2 = \overline{OP}^2$. (See Figure 17.5.5.) We will use this fact in a moment.

Now we describe the solid. From the fact that cross sections of the cylinder are circles of radius 3, θ varies from 0 to 2π and z from 0 to 3. For fixed θ and r, the cross section of the solid is a line segment determined by the sphere that bounds the ball, as shown in Figure 17.5.5(b). Because the sphere has radius 5, for any point (r, θ, z) on it,

$$r^2 + z^2 = 25 \qquad \text{or} \qquad z = \pm\sqrt{25 - r^2}.$$

Thus, on the segment determined by r and θ, z varies from $-\sqrt{25 - r^2}$ to $\sqrt{25 - r^2}$.

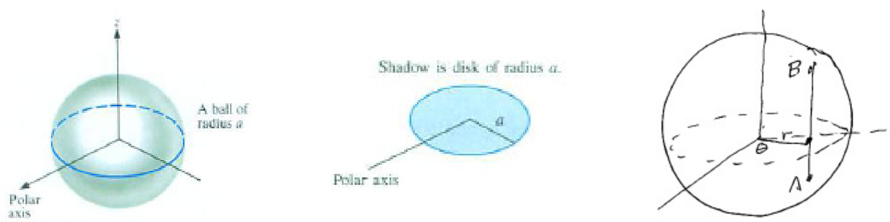
The solid has the description

$$0 \le \theta \le 2\pi, \qquad 0 \le r \le 3, \qquad -\sqrt{25 - r^2} \le z \le \sqrt{25 - r^2}.$$

◇

EXAMPLE 3 Describe a ball of radius a in cylindrical coordinates.

SOLUTION Place the origin at the center of the ball, as in Figure 17.5.5(a). The projection of the ball on the $r\theta$-plane is a disk of radius a, shown in Figure 17.5.5(b) in perspective. The projection is described by



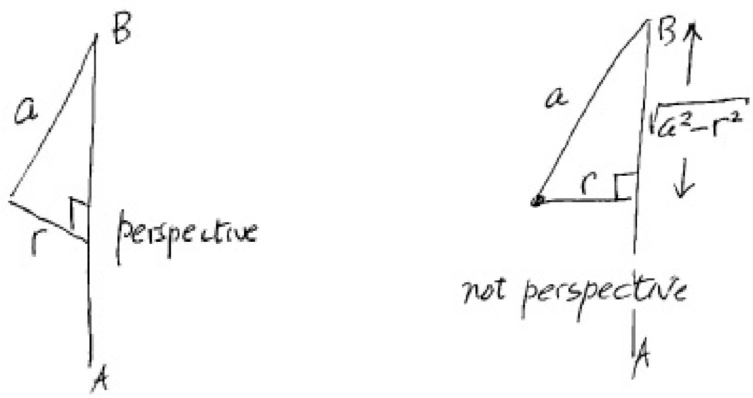
(a) (b) (c)

Figure 17.5.5

$$0 \leq \theta \leq 2\pi, \qquad 0 \leq r \leq a.$$

All that remains is to see how z varies for given r and θ. How does z vary on the iine segment AB in Figure 17.5.5(c)?

If r is a, then z varies from 0 to 0, as Figure 17.5.5(c) shows. If r is 0, then z varies from $-a$ to a. The bigger r is, the shorter AB is. Figure 17.5.6 shows the geometry, first in perspective. Using Figure 17.5.6, we see that z varies



(a) (b)

Figure 17.5.6

from $-\sqrt{a^2-r^2}$ to $\sqrt{a^2-r^2}$. You can check this by testing the easy cases, $r=0$ and $r=a$. All told,

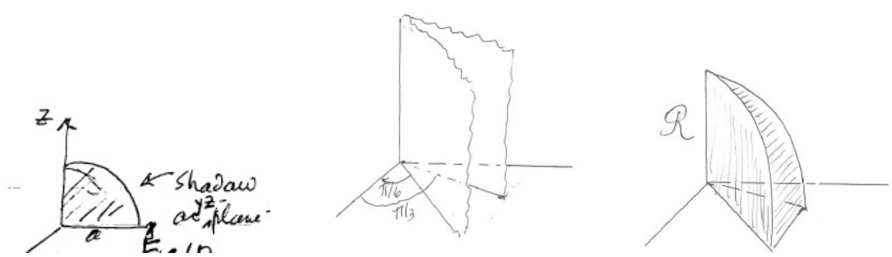
$$\underbrace{0 \le \theta \le 2\pi, \quad 0 \le r \le a,}_{\text{The projection}} \qquad \underbrace{-\sqrt{a^2-r^2} \le z \le \sqrt{a^2-r^2}}_{\text{Range of } z \text{ for each } \theta \text{ and } r}$$

◇

EXAMPLE 4 Draw the region R bounded by the surfaces $r^2+z^2=a^2$, $\theta=\pi/6$, and $\theta=\pi/3$, situated in the first octant.

SOLUTION The surface $r^2+z^2=a^2$ is a sphere with radius a centered at the origin: $x^2+y^2+z^2=a^2$. Figure 17.5.7(a) shows the part of it in the first octant.

The shading and dashed hidden line help make the diagram clearer.



(a) (b) (c)

Figure 17.5.7

Next we draw the half planes $\theta=\pi/6$ and $\theta=\pi/3$, as in Figure 17.5.7(b), again showing only the parts in the first octant. Finally we put Figure 17.5.7(a) and (b) together in (c), to see that R is a wedge from a ball.

The boundary of R has three planar surfaces ($z=0$, $\theta=\pi/6$, and $\theta=\pi/3$) and one curved surface $r^2+z^2=a^2$.

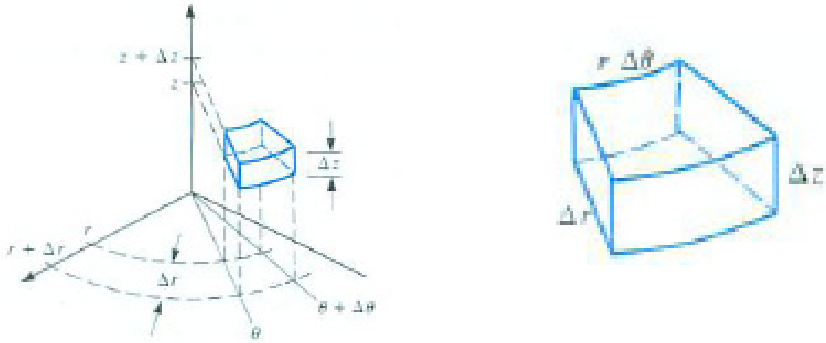
The description of R in cylindrical coordinates is

$$0 \le r \le a, \qquad 0 \le z \le \sqrt{a^2-r^2}, \qquad \pi/6 \le \theta \le \pi/3.$$

◇

The Volume Swept Out by Δr, $\Delta\theta$, and Δz

To use polar coordinates to evaluate an integral over a plane set we needed to know that the area of the region corresponding to small changes Δr and $\Delta\theta$ is roughly $r\Delta r\Delta\theta$. To evaluate integrals over solids using an iterated integral



(a) (b)

Figure 17.5.8

in cylindrical coordinates, we will need to estimate the volume of the region corresponding to small changes Δr, $\Delta\theta$, Δz in the three coordinates.

The set of points (r, θ, z) whose r coordinates are between r and $r + \Delta r$, whose θ coordinates are between θ and $\theta + \Delta\theta$, and whose z coordinates are between z and $z + \Delta z$ is shown in Figure 17.5.8(a). It is a solid with four flat surfaces and two curved surfaces.

When Δr and $\Delta\theta$ are small, the area of the flat base of the solid is approximately $r\Delta r\Delta\theta$, as shown in Section 9.2 and as we saw when working with polar coordinates in the plane. Thus, when Δr, $\Delta\theta$, and Δz are small, the volume ΔV of the solid in Figure 17.5.8(b) is approximately

$$(\text{Area of base})(\text{height}) \approx r\ \Delta r\ \Delta\theta\Delta z.$$

That is,

$$\boxed{\Delta V \approx r\Delta r\Delta\theta\Delta z.}$$

As the factor r appears in iterated integrals in polar coordinates, it also appears in iterated integrals in cylindrical coordinates.

Spherical Coordinates

The third standard coordinate system in space is **spherical coordinates**, which combines the θ of cylindrical coordinates with two others.

ρ is pronounced "row"; it is the Greek letter for r. The letter ϕ is pronounced "fee" or "fie."

In spherical coordinates a point P is described by three numbers: ρ, the distance from P to the origin O; θ, the same angle as in cylindrical coordinates; ϕ, the angle between the positive z-axis and the ray from O to P.

In physics and engineering the letter r is used instead of ρ, and ρ is put to other uses.

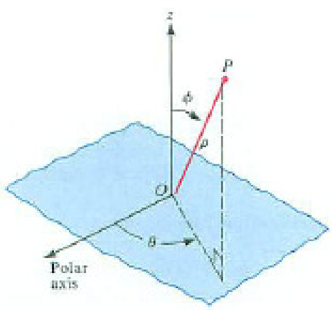
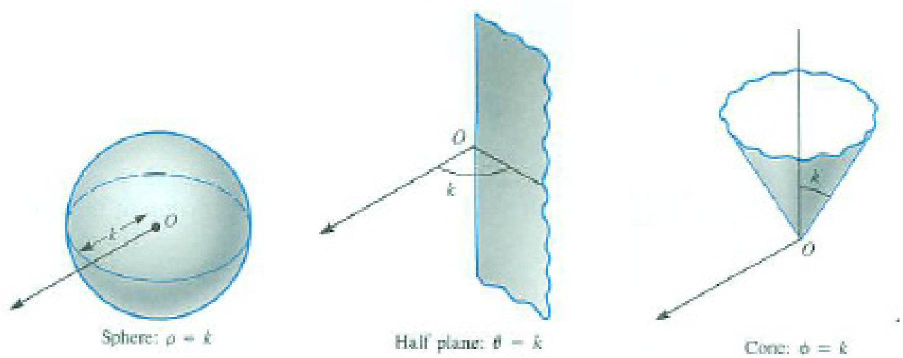


Figure 17.5.9

The point P is denoted $P = (\rho, \theta, \phi)$. The angle ϕ is the same as the direction angle of $\overline{OP}$ with $\mathbf{k}$, $0 \leq \phi \leq \pi$. (See Figure 17.5.9.) For a positive constant k the coordinate surfaces $\rho = k$ (a sphere), $\theta = k$ (a half plane), and $\phi = k$ (a cone) are shown in Figure 17.5.10.



(a) (b) (c)

Figure 17.5.10 The coordinates surfaces for spherical coordinates: (a) $\rho = k$, (b) $\theta = k$, and (c) $\phi = k$, where k is a positive number.

When ϕ and θ are fixed and ρ varies, the result is a ray, as shown in Figure 17.5.11.

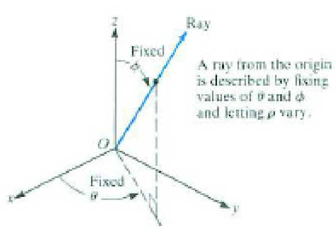


Figure 17.5.11

Relation to Rectangular Coordinates

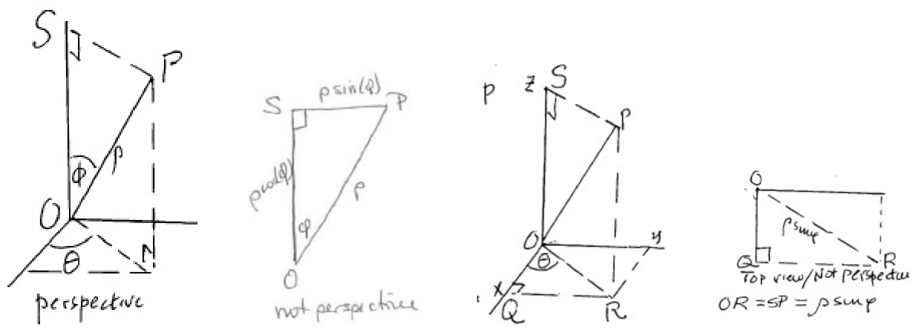
Figure 17.5.12 displays the relation between spherical and rectangular coordinates of a point $P = (\rho, \theta, \phi) = (x, y, z)$.

The right triangle OSP has hypotenuse OP and a right angle at S, and the right triangle OQR has a right angle at Q.

We have $z = \rho\cos(\phi)$, $\overline{OR} = \overline{SP} = \rho\sin(\phi)$, and $x = \overline{OR}\cos(\theta) = \rho\sin(\phi)\cos(\theta)$ and $y = \overline{OR}\sin(\theta) = \rho\sin(\phi)\sin(\theta)$.

EXAMPLE 5 Figure 17.5.13 shows a point given in spherical coordinates. Find its rectangular coordinates.

SOLUTION Because $\rho = 2$, $\theta = \pi/3$, $\phi = \pi/6$,



(a) (b) (c) (d)

Figure 17.5.12

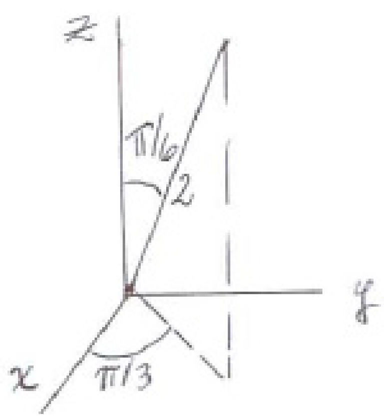


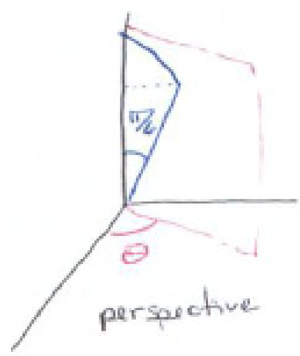
Figure 17.5.13

$$\begin{aligned} x &= 2\sin(\frac{\pi}{6})\cos(\frac{\pi}{3}) = 2\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{2} \\ y &= 2\sin(\frac{\pi}{6})\sin(\frac{\pi}{3}) = 2\cdot\frac{1}{2}\cdot\frac{\sqrt{3}}{2} = \frac{\sqrt{3}}{2} \\ z &= 2\cos(\frac{\pi}{6}) = 2\frac{\sqrt{3}}{2} = \sqrt{3}. \end{aligned}$$

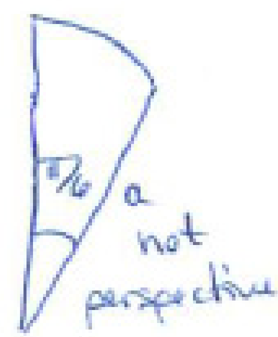
As a check, $x^2 + y^2 + z^2$ should equal ρ^2, and it does, for $(1/2)^2 + (\frac{\sqrt{3}}{2})^2 + (\sqrt{3})^2 = \frac{1}{4} + \frac{3}{4} + 3 = 4 = 2^2$. $\diamond$

The next example uses spherical coordinates to describe a cone topped by part of a ball.

EXAMPLE 6 The region R consists of the portion of a ball of radius a that lies within a cone of half angle $\pi/6$. The vertex of the cone is at the center of the ball.



(a)



(b)

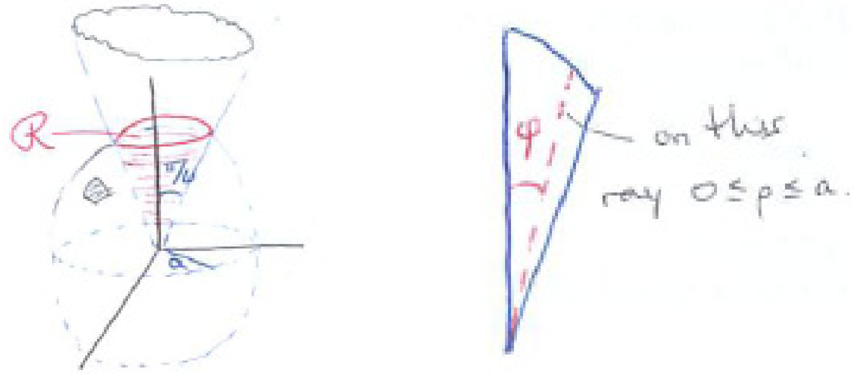
Figure 17.5.14

SOLUTION R is shown in Figure 17.5.15(a). It resembles an ice cream cone, the dry cone topped with spherical ice cream.

Because R is a solid of revolution (around the z-axis), $0 \leq \theta \leq 2\pi$. The section of R corresponding to an angle θ is the intersection of R with a half plane, shown in Figure 17.5.15(b).

In this sector of a disk, ϕ goes from 0 to $\pi/6$, independent of θ. Finally, θ and ϕ determine a ray on which ρ goes from 0 to a, as in Figure 17.5.15(b). So the description in spherical coordinates is

$$0 \leq \theta \leq 2\pi, \qquad 0 \leq \phi \leq \frac{\pi}{6}, \qquad 0 \leq \rho \leq a.$$



(a) (b)

Figure 17.5.15

◇

The next example describes a ball in rectangular and spherical coordinates.

EXAMPLE 7 Describe a ball of radius a in (a) rectangular and (b) spherical coordinates.

SOLUTION In each case we put the origin of the coordinate system at the center of the ball.

(a) Rectangular coordinates: The projection of the ball on the xy-plane is a disk of radius a, described by

$$-a \leq x \leq a, \qquad -\sqrt{a^2 - x^2} \leq y \leq \sqrt{a^2 - x^2}.$$

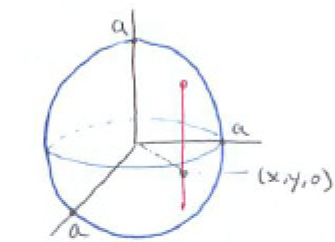


Figure 17.5.16

For (x, y) in the projection, z varies along the segment AB in Figure 17.5.16.

The equation of the sphere is $x^2 + y^2 + z^2 = a^2$. Therefore, at A, z is $-\sqrt{z^2 - x^2 - y^2}$, and, at B, z is $\sqrt{a^2 - x^2 - y^2}$. The entire description is

$$-a \leq x \leq a, \quad -\sqrt{a^2 - x^2} \leq y \leq \sqrt{a^2 - x^2}, \quad -\sqrt{a^2 - x^2 - y^2} \leq z \leq \sqrt{z^2 - x^2 - y^2}.$$

(b) Spherical coordinates: This time the projection on the xy-plane plays no role. Instead, we begin with

$$0 \leq \theta \leq 2\pi, \qquad 0 \leq \phi \leq \pi,$$

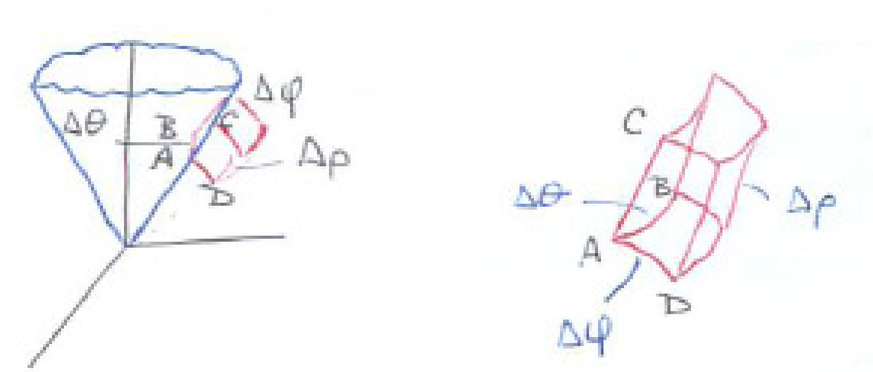
which sweeps out all the rays from the origin. On each ray ρ goes from 0 to a. The complete description involves only constants as bounds:

$$0 \leq \theta \leq 2\pi, \qquad 0 \leq \phi \leq \pi, \qquad 0 \leq \rho \leq a.$$

Since the range of each variable is not influenced by other variables, the three restraints can be given in any order. ◇

The Volume Swept Out by $\Delta\rho$, $\Delta\phi$, and $\Delta\theta$

In the next section we will need an estimate of the volume of the little box-like region bounded by the spheres with radii ρ and $\rho + \Delta\rho$, the half-planes with angles θ and $\theta + \Delta\theta$, and the cones with half-angles ϕ and $\phi + \Delta\phi$. It is shown in Figure 17.5.17. Two of its surfaces are flat, two are parts of spheres, and two are parts of cones. Segment AD is part of a circle of radius ρ and segment AB is part of a circle of radius $\rho\sin(\phi)$.

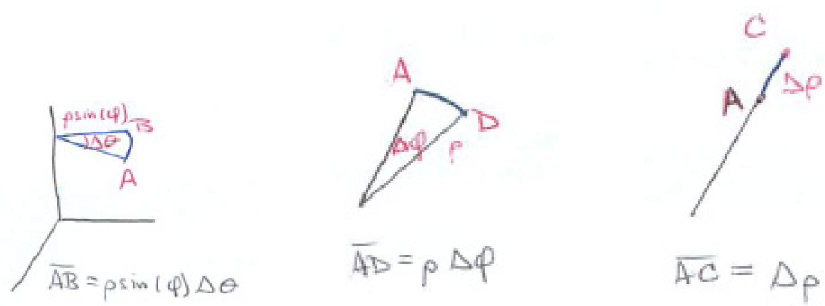


(a) (b)

Figure 17.5.17

AB and AD are arcs of circles, while AC is straight

The product of the lengths of AB, AC, and AD is an estimate of the volume of the little box. Figure 17.5.18 shows how to find the lengths.



(a) (b) (c)

Figure 17.5.18

Therefore the volume of the small box is approximately $(\rho\sin(\phi)\Delta\theta)(\rho\Delta\phi)(\Delta\rho)$:

$$\boxed{\Delta V \approx \rho^2 \sin(\phi)\Delta\rho\ \Delta\phi\ \Delta\theta}$$

As we added an r to an integrand in polar coordinates, we must add the factor $\rho^2 \sin(\phi)$ to an integrand when using an iterated integral in spherical coordinates.

Summary

This section described cylindrical and spherical coordinates. The volume of the small box corresponding to small changes in the three cylindrical coordinates is approximately $r\ \Delta r\ \Delta\theta\ \Delta z$. Because of the presence of the factor r, we must adjoin an r to the integrand when using an iterated integral in cylindrical coordinates.

Similarly, $\rho^2 \sin(\phi)$ must be added to an integrand when using an iterated integral in spherical coordinates.

EXERCISES for Section 17.5

1. On the region in Example 2 draw the set of points described by (a) $z = 2$, (b) $z = 3$, (c) $z = 4.5$.

2. For the cylinder in Example 1 draw the set of points described by (a) $r = a/2$, (b) $\theta = \pi/4$, (c) $z = h/3$.

3.

(a) In the formula $\Delta V \approx r\Delta r\Delta\theta\Delta z$, which factors have the dimension of length?

(b) Why would you expect three such factors?

4.

(a) In the formula $\Delta V \approx \rho^2 \sin(\phi)\Delta\rho\Delta\theta\Delta\phi$, which factors have the dimension of length?

(b) Why would you expect three such factors?

5. Drawing one clear, large diagram, show how to express rectangular coordinates in terms of cylindrical coordinates.

6. Drawing one clear, large diagram, show how to express rectangular coordinates in terms of spherical coordinates.

7. Find the cylindrical coordinates of $(x, y, z) = (3, 3, 1)$, including a clear diagram.

8. Find the spherical coordinates of $(x, y, z) = (3, 3, 1)$, including a clear diagram.

In Exercises 9 to 11 (a) draw the set of points described, and (b) describe it in words.

9. ρ and ϕ fixed, θ varies.

10. ρ and θ fixed, ϕ varies.

11. θ and ϕ fixed, ρ varies.

12. What is the equation of a sphere of radius a centered at the origin in

(a) spherical,

(b) cylindrical ,

(c) rectangular coordinates?

13. Explain why if $P = (x, y, z) = (\rho, \theta, \phi)$, in spherical coordinates, that $x^2 + y^2 + z^2 = \rho^2$. (Draw a box.)

14. Describe the region in Example 6 in cylindrical coordinates in the order $\alpha \leq \theta \leq \beta$, $r_1(\theta) \leq r \leq r_2(\theta)$, $z_1(r, \theta) \leq z \leq z_2(r, \theta)$.

15. Like Exercise 14, but in the order $a \leq z \leq b$, $\theta_1(z) \leq \theta \leq \theta_2(z)$, $r_1(\theta, z) \leq r \leq r_2(\theta, z)$.

16. Sketch the region in the first octant bounded by the planes $\theta = \frac{\pi}{6}$ and $\theta = \frac{\pi}{3}$ and the sphere $\rho = a$.

17. Estimate the area of the bottom face of the curvy box shown in Figure 17.5.17. It lies on the sphere of radius ρ.

18. A cone of half-angle $\pi/6$ is cut by a plane perpendicular to its axis at a distance 4 from its vertex.

(a) Place it on a cylindrical coordinate system.

(b) Describe it in cylindrical coordinates.

19. Like the preceding exercise, but use spherical coordinates.

20. A solid, infinite cone has its vertex at the origin and its axis along the positive z-axis. It is made by revolving a line through the origin that has an angle A with the z-axis, about the z-axis. Describe it in

(a) spherical coordinates

(b) cylindrical coordinates

(c) rectangular coordinates

21. Use spherical coordinates to describe the surface in Figure 17.5.19. It is part of a cone of half vertex angle ϕ with the z-axis as its axis, situated within a sphere of radius a centered at the origin.

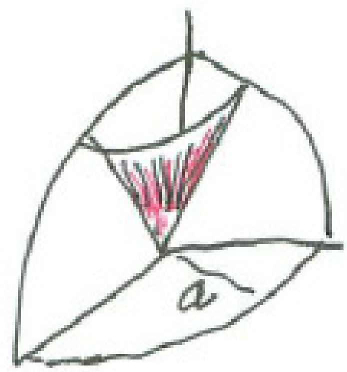


Figure 17.5.19

22. A triangle ABC is inscribed in a circle, with AB a diameter of the circle.

(a) Using geometry, show that angle ACB is a right angle.

(b) Instead, using the equation of a circle in rectangular coordinates, show that AC and BC are perpendicular.

(c) Use (a) or (b) to show that the graph in the plane of $r = b\cos(\theta)$ is a circle of diameter b.

(d) In view of the preceding exercise, show that the equation of the circle in Figure 17.5.19 is $r = 2a\cos(\theta)$.

23. (See Exercise 22.) A ball of radius a has a diameter coinciding with the interval $[0, 2a]$ on the x-axis. Describe the ball in spherical coordinates.

24. The ray described in spherical coordinates by $\theta = \frac{\pi}{6}$ and $\phi = \frac{\pi}{4}$ makes an angle α with the x-axis.

(a) Draw a picture that shows the three direction angles of the ray.

(b) Find $\cos(\alpha)$.

25.

(a) If the region in Example 2 is described in the order $0 \leq \theta \leq 2\pi$, $z_1(\theta) \leq z \leq z_2(\theta)$, $r_1(\theta, z) \leq r \leq r_2(\theta, z)$, what complication arises?

(b) Describe the region using the order given in (a).

26. What is the distance between $P_1 = (\rho_1, \theta_1, \phi_1)$ and $P_2 = (\rho_2, \theta_2, \phi_2)$?

27. The points $P_1 = (\rho_1, \theta_1, \phi_1)$ and $P_2 = (\rho_1, \theta_2, \phi_2)$ both lie on a sphere of radius ρ_1. Assuming that both are in the first octant, find the great circle distance between them. If the sphere is Earth's surface, ρ is approximately 3960 miles, north of the equator ϕ is the complement of the latitude, and θ is related to longitude.

28. At time t a particle moving along a curve is at the point $(\rho(t), \theta(t), \phi(t))$ in spherical coordinates. What is its speed?

29. How far apart are the points (r_1, θ_1, z_1) and (r_2, θ_2, z_2) in the first octant?

(a) Draw a large clear diagram.

(b) Find the distance.

30. The surface of a cylinder is described by $0 \le \theta \le 2\pi, 0 \le r \le 3, 0 \le z \le 2$. One path from $(3, 0, 2)$ to $(3, \pi, 0)$ is straight to $(3, 0, 0)$ followed by a straight path on the base along a diameter. Is that the shortest path on the surface? If not, what is?

By differentiating, verify the equations in Exercises 31 and 32.

31. $\int \frac{dx}{x^3\sqrt{z^2+x^2}} = -\frac{\sqrt{a^2+x^2}}{2a^2+x^2} + \frac{1}{2a^3} \ln \left| \frac{a+\sqrt{a^2+x^2}}{x} \right|$.

32. $\int \frac{x^2\ dx}{a^4-x^4} = \frac{1}{4a} \ln \left|\frac{a+x}{a-x}\right| - \frac{1}{2a} \arctan \frac{x}{a}$.

17.6 Iterated integrals for $\int_R f(P)\, dV$ in Cylindrical or Spherical Coordinates

In Section 17.2 we evaluated an integral of the form $\int_R f(P)\, dA$ by an iterated integral in polar coordinates. In this method it is necessary to multiply the integrand by an r because the area of the small piece determined by small increments in r and θ is not $\Delta r\ \Delta\theta$ but $r\ \Delta r\ \Delta\theta$. Similarly, when developing iterated integrals using cylindrical coordinates, an extra r must be adjoined to the integrand. For spherical coordinates we adjoin $\rho^2 \sin(\phi)$. These adjustments are based on the estimates of the volumes of the small curvy boxes made in the previous section.

A few examples will illustrate the method, which is: Describe the solid R and the integrand in the most convenient coordinate system. Then use the description to set up an iterated integral, being sure to include the appropriate extra factor in the integrand.

Iterated Integrals in Cylindrical Coordinates

To evaluate $\int_R f(P)\ dV$ in cylindrical coordinates express the integrand in cylindrical coordinates and describe the region R in cylindrical coordinates with dV replaced by $r\ dz\ dr\ d\theta$. There are six possible orders of integration, but the most common one has z varying first, then r, and finally θ:

Evaluating $\int_R f(P)\ dV$ in Cylindrical Coordinates

$$\int_R f(P)\ dV = \int_\alpha^\beta \int_{r_1(\theta)}^{r_2(\theta)} \left(\int_{z_1(r,\theta)}^{z_2(r,\theta)} f(r,\theta,z) r\ dz \right) dr\ d\theta.$$

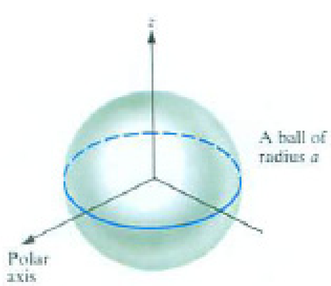


Figure 17.6.1

EXAMPLE 1 Find the volume of a ball R of radius a using cylindrical coordinates.

SOLUTION Place the origin of a cylindrical coordinate system at the center of the ball, as in Figure 17.6.1.

The volume of the ball is $\int_R 1\ dV$. The description of R in cylindrical coordinates is

$$0 \le \theta \le 2\pi, \qquad 0 \le r \le a, \qquad -\sqrt{a^2 - r^2} \le z \le \sqrt{a^2 - r^2}.$$

The iterated integral for the volume is thus

$$\int_R 1\ dV = \int_0^{2\pi}\left(\int_0^a\left(\int_{-\sqrt{a^2-r^s}}^{\sqrt{a^2-r^2}} 1\cdot r\ dz\right)\ dr\right)\ d\theta.$$

Note the factor r.

Evaluation of the first integral yields

$$\int_{-\sqrt{a^2-r^2}}^{\sqrt{a^2-r^2}} r\ dz \quad = \quad rz|_{z=-\sqrt{a^2-r^2}}^{z=\sqrt{a^2-r^2}} = 2r\sqrt{a^2-r^2}.$$

Evaluation of the second integral yields

The order of integration is determined by the order of the variables in describing R.

$$\int_0^a 2r\sqrt{a^2-r^2}\ dr = \left.\frac{-2(a^2-r^2)^{3/2}}{3}\right|_{r=0}^{r=a} = \frac{2a^3}{3}.$$

Finally, evaluation of the third integral gives

$$\int_0^{2\pi}\frac{2a^3}{3}\ d\theta \quad = \quad \frac{2a^3}{3}\int_0^{2\pi} d\theta = \frac{2a^3}{3}\cdot 2\pi = \frac{4}{3}\pi a^3.$$

◇

EXAMPLE 2 Find the volume of the region R inside the cylinder $x^2+y^2=9$, above the xy-plane, and below the plane $z=x+2y+9$. Use cylindrical coordinates.

SOLUTION We wish to evaluate $\int_R 1\ dV$ over the region R described in cylindrical coordinates R by

$$0\le\theta\le 2\pi,\qquad 0\le r\le 3,\qquad 0\le z\le r\cos(\theta)+2r\sin(\theta)+9.$$

(We replace $z=x+2y+9$ by $z=r\cos(\theta)+2r\sin(\theta)+9$.)

The iterated integral takes the form

Note the factor r.

$$\int_0^{2\pi}\left(\int_0^3\left(\int_0^{r\cos(\theta)+2r\sin(\theta)+9} 1\cdot r\ dz\right)\ dr\right)\ d\theta.$$

Integration with respect to z gives

r and θ are constant

$$\int_0^{r\cos(\theta)+2r\sin(\theta)+9} r\ dz \quad = \quad r\int_0^{r\cos(\theta)+2r\sin(\theta)+9} dz = r^2\cos(\theta)+2r^2\sin(\theta)+9r.$$

Then comes integration with respect to r, with θ constant:

$$\begin{aligned}\int_0^3 \left(r^2\cos(\theta)+2r^2\sin(\theta)+9r\right)\,dr &= \left.\frac{r^3}{3}\cos(\theta)+\frac{2r^3}{3}\sin(\theta)+\frac{9r^2}{2}\right|_{r=0}^{r=3}\\ &= 9\cos(\theta)+18\sin(\theta)+\frac{81}{2}.\end{aligned}$$

Finally, integration with respect to θ gives

$$\int_0^{2\pi}\left(9\cos(\theta)+18\sin(\theta)+\frac{81}{2}\right)\,d\theta. \tag{17.6.1}$$

Why is $\int_0^{2\pi}\cos(x)\,dx = \int_0^{2\pi}\sin(x)\,dx = 0$?

Because $\int_0^{2\pi}\cos(\theta)\,d\theta = 0 = \int_0^{2\pi}\sin(\theta)\,d\theta$, (17.6.1) reduces to $\int_0^{2\pi}\frac{81}{2}\,d\theta = 81\pi$. The volume is 81π. ◇

Computing $\int_R f(P)\,dV$ in Spherical Coordinates

To evaluate a triple integral $\int_R f(P)\,dV$ in spherical coordinates, first describe R in spherical coordinates. Usually this will be in the order

$$\alpha\le\theta\le\beta,\qquad \phi_1(\theta)\le\phi\le\phi_2(\theta),\qquad \rho_1(\theta,\phi)\le\rho\le\rho_2(\theta,\phi).$$

Sometimes the order of ρ and ϕ is switched:

$$\alpha\le\theta\le\beta,\qquad \rho_1(\theta)\le\rho\le\rho_2(\theta),\qquad \phi_1(\rho,\theta)\le\phi\le\phi_2(\rho,\theta).$$

Then set up an iterated integral, expressing dV as $\rho^2\sin(\phi)\,d\rho\,d\phi\,d\theta$ or $\rho^2\sin(\phi)\,d\phi\,d\rho\,d\theta$.

EXAMPLE 3 Find the volume of a ball of radius a, using spherical coordinates.

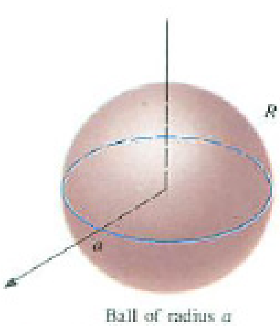


Figure 17.6.2

SOLUTION Place the origin of spherical coordinates at the center of the ball, as in Figure 17.6.2. The ball is described by

$$0\le\theta\le 2\pi,\qquad 0\le\phi\le\pi,\qquad 0\le\rho\le a.$$

Hence

$$\text{Volume of ball} = \int_R 1\,dV = \int_0^{2\pi}\int_0^{\pi}\int_0^{a}\rho^2\sin(\phi)\,d\rho\,d\phi\,d\theta.$$

The inner integral is

$$\int_0^a \rho^2 \sin(\phi)\ d\rho \quad = \quad \sin(\phi) \int_0^a \rho^2\ d\rho = \frac{a^3 \sin(\phi)}{3}.$$

The next integral is

$$\int_0^\pi \frac{a^3 \sin(\phi)}{3}\ d\phi \quad = \quad \left.\frac{-a^3 \cos(\phi)}{3}\right|_0^\pi = \frac{-a^3(-1)}{3} - \frac{-a^3(1)}{3} = \frac{2a^3}{3}.$$

The final integral is

$$\int_0^{2\pi} \frac{2a^3}{3}\ d\theta \quad = \quad \frac{2a^3}{3} \int_0^{2\pi} d\theta = \frac{2a^3}{3} 2\pi = \frac{4\pi a^3}{3}.$$

◇

An Integral in Gravity

The next example is of importance in the theory of gravitational attraction. It implies that a homogeneous ball attracts a particle as if all the mass of the ball is at its center.

EXAMPLE 4 Consider a homogeneous ball of mass M and radius a occupying a region R. Let A be a point at a distance H from the center of the ball, $H > a$. Compute $\int_R (\delta/q(P))\ dV$, where δ is the density of the ball and $q(P)$ is the distance from point P in R to A. (See Figure 17.6.3.)

SOLUTION Express $q(P)$ in spherical coordinates. To do so, choose a spherical coordinate system whose origin is at the center of the sphere and such that the ϕ coordinate of A is 0. (See Figure 17.6.3(b).)

Let $P = (\rho, \theta, \phi)$ be a point in the ball. Applying the law of cosines to triangle AOP, we find that

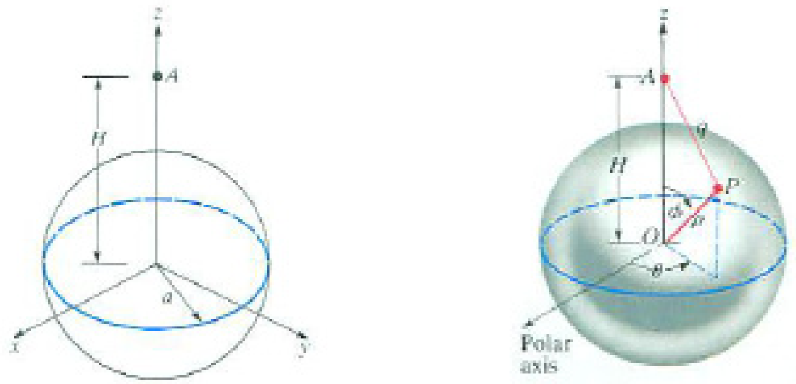
$$q^2 = H^2 + \rho^2 - 2\rho H \cos(\phi).$$

Hence

$$q = \sqrt{H^2 + \rho^2 - 2\rho H \cos(\phi)}.$$

Since the ball is homogeneous, if its total mass is M, then

$$\delta = \frac{M}{\frac{4}{3}\pi a^3} = \frac{3M}{4\pi a^3}.$$



(a) (b)

Figure 17.6.3

Hence

$$\int_R \frac{\delta}{q(P)}\, dV = \int_R \frac{3M}{4\pi a^3 q(P)}\, dV = \frac{3M}{4\pi a^3}\int_R \frac{1}{q(P)}\, dV. \tag{17.6.2}$$

Now evaluate

$$\int_R \frac{1}{q(P)}\, dV$$

An integral where integration with respect to ρ is not first.

by an iterated integral in spherical coordinates:

$$\int_R \frac{1}{q(P)}\, dV = \int_0^{2\pi}\left(\int_0^a\left(\int_0^\pi \frac{\rho^2\sin(\phi)}{\sqrt{H^2+\rho^2-2\rho H\cos(\phi)}}\, d\phi\right)\, d\rho\right)\, d\theta.$$

We integrate with respect to ϕ first, rather than ρ, because it is easier.

Evaluation of the first integral, where ρ and θ are constants, is accomplished with the aid of the Fundamental Theorem of Calculus:

$$\begin{aligned}\int_0^\pi \frac{\rho^2\sin\phi}{\sqrt{H^2+\rho^2-2\rho H\cos(\phi)}}\, d\phi &= \left.\frac{\rho\sqrt{H^2+\rho^2-2\rho H\cos(\phi)}}{H}\right|_{\phi=0}^{\phi=\pi} \\ &= \frac{\rho}{H}\left(\sqrt{H^2+\rho^2+2\rho H}-\sqrt{H^2+\rho^2-2\rho H}\right).\end{aligned}$$

Now, $\sqrt{H^2+\rho^2+2\rho H} = H+\rho$. Since $\rho \le a < H$, $H-\rho$ is positive and $\sqrt{H^2+\rho^2-2\rho H} = H-\rho$.

Thus the first integral equals

$$\frac{\rho}{H}(H+\rho)-(H-\rho)) = \frac{2\rho^2}{H}.$$

Evaluation of the second integral yields

$$\int_0^a \frac{2\rho^2}{H}\, d\rho = \frac{2a^3}{3H}.$$

Evaluation of the third integral yields

$$\int_0^{2\pi} \frac{2a^3}{3H}\, d\theta = \frac{4\pi a^3}{3H}.$$

Hence

$$\int_R \frac{1}{q(P)}\, dV = \frac{4\pi a^3}{3H}.$$

By (17.6.2)

$$\int_R \frac{\delta}{q(P)}\, dV = \frac{3M}{4\pi a^3}\frac{4\pi a^3}{3H} = \frac{M}{H}.$$

This result, M/H, is what we would get if all the mass were located at the center of the ball. ⋄

Newton obtained this remarkable result in 1687.

The Moment of Inertia about a Line

In the study of rotation of an object about an axis its moment of inertia, I, is used. It is defined as follows. The object occupies a region R. Its density at a point P is $\delta(P)$, so its mass is $M = \int_R \delta(P)\, dV$. Usually the density is constant, in which case it is M divided by the volume of R (or M divided by the area of R if R is planar). Let $r(P)$ be the distance from P to a line L. Then, by definition,

$$I = \text{Moment of Inertia about the line } L = \int_R (r(P))^2 \delta(P)\, dV.$$

A similar definition holds for objects distributed on a planar region, with dV replaced by dA.

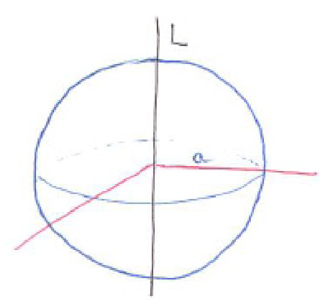


Figure 17.6.4

EXAMPLE 5 Compute the moment of inertia of a uniform mass M in the form of a ball of radius a around a diameter L.

SOLUTION The density $\delta(P)$, being constant, is $M/(\frac{4}{3}\pi a^3)$. We place the diameter L along the z-axis, as in Figure 17.6.4

We will compute the moment of inertia two ways. Because the distance $r(P)$ is r in cylindrical coordinates, we will first use those coordinates. Then we will calculate the moment of inertia in spherical coordinates.

As a check on our answer we note in advance that I must be less than Ma^2, since the maximum of $r(P)$ is a.

One description of the ball in cylindrical coordinates is

$$0 \le \theta \le 2\pi, \qquad -a \le z \le a, \qquad 0 \le r \le \sqrt{a^2 - z^2}.$$

Then

$$\begin{aligned} I &= \int_R \frac{M}{\frac{4}{3}\pi a^3} r^2 \, dV = \frac{3M}{4\pi a^3} \int_R r^2 \, dV \\ &= \frac{3M}{4\pi a^3} \int_0^{2\pi} \int_{-a}^{a} \int_0^{\sqrt{a^2-z^2}} r^3 \, dr \, dz \, d\theta. \end{aligned}$$

Exercise 27 shows that I plays the same role in a rotating body (such as a spinning skater) as mass does in an object moving along a line.

Note the introduction of the extra factor of r in the integrand.

The first integration is

$$\int_0^{\sqrt{a^2-z^2}} r^3 \, dr = \left. \frac{r^4}{4} \right|_0^{\sqrt{a^2-z^2}} = \frac{(a^2 - z^2)^2}{4}.$$

The second is

$$\begin{aligned} \int_{-a}^{a} \frac{(a^2 - z^2)^2}{4} \, dz &= \int_{-a}^{a} \frac{a^4 - 2a^2 z^2 + z^4}{4} \, dz = \frac{1}{4} \left(a^4 z - \frac{2a^2 z^3}{3} + \frac{z^5}{5} \right) \Bigg|_{-a}^{a} \\ &= \frac{1}{4} \left(a^5 - \frac{2a^5}{3} + \frac{a^5}{5} \right) - \frac{1}{4} \left(-a^5 + \frac{2a^5}{3} - \frac{a^5}{5} \right) = \frac{4}{15} a^5. \end{aligned}$$

And the third is

$$\int_0^{2\pi} \frac{4}{15} a^5 \, d\theta = \frac{8\pi}{15} a^5.$$

Then remembering to include the factor $3M/4\pi a^3$, we have

$$I = \frac{3M}{4\pi a^3} \cdot \frac{8\pi}{15} a^5 = \frac{2}{5} M a^2.$$

Because spherical coordinates provide a simple description of the ball, we will use them to find I to see if the computations are easier. The distance $r(P)$ has a more complicated form, $\delta(P) = \delta(\rho, \theta, \phi) = \rho \sin(\phi)$. The integral for the moment of inertia is

$$I = \frac{3M}{4\pi a^3} \int_R (\rho \sin(\phi))^2 \, dV.$$

The iterated integral for it is

$$\int_0^{2\pi}\left(\int_0^{\pi}\left(\int_0^{a}(\rho\sin(\phi))^2\,\rho^2\sin(\phi)\;d\rho\right)\;d\phi\right)\;d\theta.$$

The first integration is

$$\int_0^a \rho^4\sin^3(\phi)\;d\rho = \frac{\rho^5}{5}\sin^3(\phi)\Big|_{\rho=0}^{\rho=a} = \frac{a^5\sin^3(\phi)}{5}.$$

The second is

$$\int_0^{\pi}\frac{a^5}{5}\sin^3(\phi)\;d\phi = \frac{a^5}{5}\int_0^{\pi}\sin^3(\phi)\;d\phi.$$

Since the exponent, 3, is odd, we write $\sin^3(\phi)$ as $(1-\cos^2(\phi))\sin(\phi)$ and have

$$\begin{aligned}\int_0^{\pi}\sin^3(\phi)\;d\phi &= \int_0^{\pi}(\sin(\phi)-\cos^2(\phi)\sin(\phi))\;d\phi = \left(-\cos(\phi)+\frac{\cos^3(\phi)}{3}\right)\Big|_0^{\pi}\\ &= \left(-(-1)+\frac{(-1)^3}{3}\right)-\left(-1+\frac{1}{3}\right)=\frac{4}{3}.\end{aligned}$$

For the final integration, remember the coefficient of $a^5/5$:

$$\frac{a^5}{5}\int_0^{2\pi}\frac{4}{3}\;d\theta = \frac{8a^5}{15}\pi.$$

Bringing back the factor $3M/(4\pi a^3)$ gives the same moment of inertia:

$$I = \frac{2}{5}Ma^2.$$

◇

If all the mass were at a distance a from L, the moment of inertia would be Ma^2. So $2/5Ma^2$ is plausible.

Summary

A multiple integral $\int_R f(P)\;dV$ may be evaluated by an iterated integral in cylindrical or spherical coordinates. In cylindrical coordinates the iterated integral takes the form

$$\int_{\theta_1}^{\theta_2}\left(\int_{r_1(\theta)}^{r_2(\theta)}\left(\int_{z_1(r,\theta)}^{z_2(r,\theta)}f(r,\theta,z)r\;dz\right)\;dr\right)\;d\theta.$$

The description of the region determines the range of integration on the integrals over intervals. (Changing the order of the description of R changes the order of the integrations.) The factor r must be inserted into the integrand.

In spherical coordinates the iterated integral usually takes the form

$$\int_{\theta_1}^{\theta_2} \left(\int_{\phi_1(\theta)}^{\phi_2(\theta)} \left(\int_{\rho_1(\theta,\phi)}^{\rho_2(\theta,\phi)} f(r,\theta,\phi)\rho^2 \sin(\phi)\ d\rho \right)\ d\phi \right)\ d\theta.$$

In this form, integration with respect to ρ is first, but as Example 4 illustrates, it may be convenient to integrate first with respect to ϕ. The factor $\rho^2 \sin(\phi)$ must be inserted in the integrand.

EXERCISES for Section 17.6

In Exercises 1 to 4 (a) draw the region, (b) set up an iterated integral in cylindrical coordinates for the multiple integrals, and (c) evaluate the iterated integral.

1. $\int_R r^2\ dV$, R is bounded by the cylinder $r = 3$ and the planes $z = 2x$ and $z = 3x$.

2. $\int_R z\ dV$, R is bounded by the sphere $z^2 + r^2 = 25$, the plane $z = 0$, and the plane $z = 2$.

3. $\int_R rz\ dV$, R is the part of the ball bounded by $r^2 + z^2 = 16$ in the first octant.

4. $\int_R \cos\theta\ dV$, R is bounded by the cylinder $r = 2\cos(\theta)$ and the paraboloid $z = r^2$.

5. Compute the volume of a right circular cone of height h and radius r using (a) spherical coordinates, (b) cylindrical coordinates, and (c) rectangular coordinates.

6. Find the volume of the region above the xy plane and below the paraboloid $z = 9 - r^2$ using cylindrical coordinates.

7. A right circular cone of radius a and height h has a density at point P equal to the distance from P to the base of the cone. Find its mass, using spherical coordinates.

In Exercises 8 to 9 draw the region R and give a formula for the integrand $f(P)$ such that $\int_R f(P)\ dV$ is described by the iterated integrals.

8. $\int_0^{\pi/2}[\int_0^{\pi/4}(\int_0^{\cos\phi} \rho^3 \sin^2(\theta)\sin(\phi)\ d\rho)\ d\phi]\ d\theta$.

9. $\int_0^{\pi/4}[\int_{\pi/6}^{\pi/2}(\int_0^{\sec\theta} \rho^3 \sin(\theta)\cos(\phi)\ d\rho)\ d\phi]\ d\theta$.

10. Let R be the solid region inside both the sphere $x^2 + y^2 + z^2 = 1$ and the cone $z = \sqrt{x^2 + y^2}$. Let the density at (x, y, z) be z. Set up iterated integrals for the mass in R using (a) rectangular coordinates, (b) cylindrical coordinates, (c) spherical coordinates. (d) Evaluate the iterated integral in (c).

11. Find the average temperature in a ball of radius a if the temperature is the square of the distance from a fixed equatorial plane.

In Exercises 12 and 13 evaluate

12. $\int_0^{2\pi}\left(\int_0^1\left(\int_r^1 zr^3\cos^2\theta\ dz\right)\ dr\right)\ d\theta$

13. $\int_0^{2\pi}\left(\int_0^1\left(\int_{-\sqrt{a^2-r^2}}^{\sqrt{a^2-r^2}} z^2 r\ dz\right)\ dr\right)\ d\theta$

14. Using cylindrical coordinates, find the volume of the region below the plane $z = y + 1$ and above the circle in the xy plane whose center is $(0, 1, 0)$ and whose radius is 1. Include a drawing of the region. (What is the equation of the circle in polar coordinates when the polar axis is along the positive x-axis?)

15. Find the average distance from the center of a ball of radius a to other points of the ball by setting up iterated integrals in the three types of coordinate systems and evaluating the easiest.

16. A solid consists of that part of a ball of radius a that lies within a cone of half-vertex angle $\phi = \pi/6$, the vertex being at the center of the ball. Set up iterated integrals for $\int_R z\, dV$ in three coordinate systems and evaluate the simplest.

In Exercises 17 to 22 evaluate the multiple integrals over a ball of radius a with center at the origin, without using an iterated integral.

17. $\int_R \cos(\theta)\; dV$

18. $\int_R \cos^2\theta\; dV$

19. $\int_R z\; dV$

20. $\int_R (3 + 2\sin(\theta))\; dV$

21. $\int_R \sin^2(\phi)\; dV$

22. $\int_R \cos^3(\phi)\; dV$

23. In polar, cylindrical, and spherical coordinates we introduce an extra factor in the integrand when using an iterated integral. Why is that not necessary when using rectangular coordinates?

24. Is $\sqrt{a^2}$ always equal to a? Explain.

25. Using the method of Example 4 find the average value of q for all points P in the ball. The result is not the same as if the entire ball were placed at its center.

26.

(a) By integrating the function $f(P) = 1$, find the exact volume of the little curvy box corresponding to the changes $\Delta\rho$, $\Delta\theta$, $\Delta\phi$.

(b) Show that the ratio between that exact volume and our estimate, $\rho^2\sin(\phi)\Delta\rho\Delta\theta\Delta\rho$ approaches 1 as $\Delta\rho$, $\Delta\theta$, and $\Delta\phi$ approach 0.

(c) Show that the exact volume in (a) can be written as $(\rho^*)^2\sin(\phi^*)\;\Delta\rho\;\Delta\phi\;\Delta\theta$, where ρ^* is between ρ and $\rho + \Delta\rho$ and ϕ^* is between ϕ and $\phi + \Delta\phi$.

27. The kinetic energy of an object with mass m moving at the velocity v is $mv^2/2$. An object moving in a circle of radius r at the angular speed of ω radians per unit time has velocity $r\omega$. (Why?) Thus its kinetic energy is $(mr^2/2)\omega^2$.
Now consider a mass M that occupies the region R in space. Its density is $\delta(P)$, which may vary from point to point. Let $f(P)$ be the distance from P to a fixed line L. If the mass is spinning around the axis L at the angular rate ω, show that its total kinetic energy is

$$\int\limits_R \frac{1}{2}(f(P))^2\ \delta(P)\ \omega^2\ dV.$$

This can be written as

$$\text{Kinetic Energy} = \frac{1}{2}I\omega^2.$$

Thus I plays the same role in rotational motion that mass m plays in linear motion in the formula $\frac{1}{2}mv^2$.
Every spinning ice skater knows this. When spinning with her arms extended she has a certain amount of kinetic energy. If she puts her arms to her sides she decreases her moment of inertia but has not destroyed her kinetic energy. That forces her angular speed to increase. The larger the mass M is, the harder it is to start it moving and to stop it when it is moving. Similarly, the larger I is, the harder it is to stop the mass from spinning and to stop it when it is spinning.

In Exercises 28 to 32 the objects have a homogeneous (constant density) mass M. Find the moment of inertia, I, of the given object when it rotates around the given axis.

28. A rectangular box of dimensions, $a \times b \times c$ rotating around a line through the center of the box and perpendicular to the face of dimensions $a \times b$.

29. A solid cylinder of radius a and height h rotating around its axis.

30. A solid cylinder of radius a and height h rotating around a line on its surface that is parallel to the cylinder's axis.

31. A cylindrical tube of height h, inner radius a, and outer radius b, rotating around its axis.

32. A solid cylinder of radius a and height h rotating around a diameter in its base.

33. In Example 2 the region R was parameterized as $0 \leq 2\pi$, $0 \leq r \leq 3$, $0 \leq z \leq r\cos(\theta) + 2\sin(\theta) + 9$. There are five other possible orderings of the variables for the parameterization of R. Some of these orderings require splitting the region into two or more pieces to write the parameterization. Which, if any, of these orderings can parameterize the entire region in one piece? You do not have to find each parameterization to answer this question.

34. Solve Example 2 using rectangular coordinates.

35. Evaluate the moment of inertia in Example 5 using the description $0 \le \theta \le 2\pi$, $0 \le r \le a$, $-\sqrt{a^2 - r^2} \le z \le \sqrt{a^2 - r^2}$.

36. Let R be a solid ball of radius a with center at the origin of a coordinate system.

(a) Explain why $\int_R x^2\ dV = \frac{1}{3} \int_R (x^2 + y^2 + z^2)\ dV$.

(b) Evaluate the second integral by spherical coordinates.

(c) Use (b) to find $\int_R x^2\ dV$.

37. Let R be a ball whose center is the origin of a rectangular coordinate system. Show that $\int_R (x^3+y^3+z^3)\ dV = 0$. Do not use an iterated integral. Use symmetry.

38. A homogeneous object with mass M occupies the region R between concentric spheres of radii a and b, $a < b$. Let A be a point at a distance H from their center, $H < a$. Evaluate $\int_R \frac{\delta}{q}\ dV$, where δ is the density and $q = q(P)$ is the distance from H to any point P in R. (That the value of the integral does not involve H immplies that a uniform hollow sphere exerts no gravitational force on objects in its interior.)

39. In Example 4, H is greater than a. Solve the same problem for H less than a. For some ρ, $\sqrt{H^2 + \rho^2 A - 2\rho H}$ equals $H - \rho$ and for some it equals $\rho - H$.

40. (See Exercise 39.) Let A be a point in the plane of a disk but outside the disk. Is the average of the reciprocal of the distance from A to points in the disk equal to the reciprocal of the distance from A to the center of the disk?

SHERMAN: In my files, this Exercise has been removed. I am reinserting it here; it was E 25 (p. 1530) in the version dated August 1, 2011.

41. Show that the result of Example 4 holds if the density $\delta(P)$ depends only on ρ, the distance to the center. (This is approximately the case with Earth, which is not homogeneous.) Let $g(\rho)$ denote $\delta(\rho, \theta, \phi)$.

42. A ball of radius a is not homogeneous. However, its density at P depends only on the distance from P to the center of the ball. That is, there is a function $g(\rho)$ such that the density at $P = (\rho, \theta, \phi)$ is $g(\rho)$. Using an iterated integral, show that the mass of the ball is

$$4\pi \int_0^a g(\rho)\rho^2\ d\rho.$$

43. Let R be the part of a ball of radius a removed by a cylindrical drill of diameter a whose edge passes through the center of the sphere.

(a) Sketch R.

(b) Region R consists of four congruent pieces. Find the volume of one of them using cylindrical coordinates. Multiply by four to get the volume of R.

44. Let R be the ball of radius a. For any point P in the ball other than its center, define $f(P)$ to be the reciprocal of the distance from P to the origin. The average value of r over R involves an improper integral, since the function is unbounded near the origin. Does the improper integral converge or diverge? What is the average value of f over R? (Examine the integral over the region between concentric spheres of radii a and t, and let $t \to 0^+$.)

In Exercises 45 to 46 check the equations by differentiation.

45. $\tan\left(\frac{x}{2}\right) = \int \frac{dx}{1+\cos(x)}$

46. $x\tan\left(\frac{x}{2}\right) + 2\ln\left|\cos\left(\frac{x}{2}\right)\right| = \int \frac{x\ dx}{1+\cos(x)}$

17.7 Integrals Over Surfaces

In this section we define integrals over surfaces and then show how to compute then by an iterated integral.

Definition of a Surface Integral

Let $\mathcal{S}$ be a surface such as the surface of a ball or part of the saddle $z = xy$. If f is a numerical function defined on $\mathcal{S}$, we will define the integral $\int_{\mathcal{S}} f(P)\, dS$. The definition is practically identical with the definition of the double integral, which is the special case when the surface is a plane.

We assume that the surfaces we deal with are smooth, or composed of a finite number of smooth pieces, and that the integrals we define exist. In keeping with the definition of the definite integral on an interval, the surface $\mathcal{S}$ will be partitioned into S_1, S_2, $\dots$, S_n. To measure the size of a partition we define the **mesh** of a partition to be the largest diameter of any of the regions in the partition. The **diameter** of a region is the diameter of the smallest ball that contains the region. In other words, the diameter of a region is the greatest distance between two points in the region. We are now ready to define an integral over a surface.

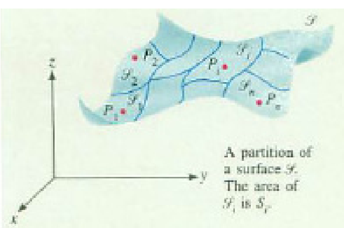


Figure 17.7.1

DEFINITION *(Definite integral of a function f over a surface $\mathcal{S}$.)* Let f be a function that assigns to a point P in a surface $\mathcal{S}$ a number $f(P)$. Consider the sum

$$f(P_1)\Delta S_1 + f(P_2)\Delta S_2 + \cdots + f(P_n)\Delta S_n$$

formed from a partition of $\mathcal{S}$ into S_1, S_2, $\dots$, S_n with P_i in S_i and ΔS_i is the area of S_i. (See Figure 17.7.1.) If the sums approach a number as the mesh of the partition shrinks toward 0, the number is called the **integral of f over $\mathcal{S}$** and is written

$$\int_{\mathcal{S}} f(P)\, dS.$$

Surface integrals are also denoted $\iint_{\mathcal{S}} f(P)\, dS$.

If $f(P)$ is 1 for each point P in $\mathcal{S}$ then $\int_{\mathcal{S}} f(P)\, dS$ is the area of $\mathcal{S}$. If $\mathcal{S}$ is occupied by material of surface density $\sigma(P)$ at P then $\int_{\mathcal{S}} \sigma(P)\, dS$ is the total mass of $\mathcal{S}$.

If matter is distributed on the surface $\mathcal{S}$, its density at a point P in $\mathcal{S}$ is defined much the way density in a lamina was defined in Section 17.1. The only difference is that instead of considering a small disk around P one considers a small patch on $\mathcal{S}$ that contains P. "Small" means that the patch fits in a ball of radius r, and we let r approach 0.

First we show how to integrate over a sphere.

Integrating over a Sphere

If the surface $\mathcal{S}$ is a sphere or part of a sphere, it is often convenient to evaluate an integral over it with the aid of spherical coordinates.

If the center of a spherical coordinate system (ρ, θ, ϕ) is at the center of a sphere of radius a, then ρ is constant on the sphere $\rho = a$. As Figure 17.7.2 suggests, the area of the small region on the sphere corresponding to slight changes $d\theta$ and $d\phi$ is approximately

See Section 17.6 for a similar argument, where ρ was not constant.

$$(a\ d\phi)\,(a\sin(\phi)\ d\theta) = a^2\sin(\phi)\ d\theta\ d\phi.$$

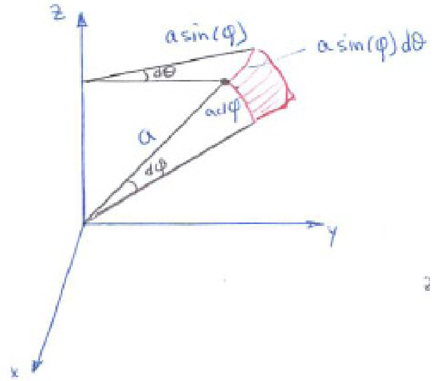


Figure 17.7.2

Thus we may write

$$dS = a^2\sin(\phi)\ d\theta\ d\phi$$

and evaluate

$$\int_{\mathcal{S}} f(P)\ dS$$

as an iterated integral in ϕ and θ. Example 1 illustrates this.

EXAMPLE 1 Let $\mathcal{S}$ be the top half of the sphere with radius a. Evaluate $\int_{\mathcal{S}} z\ dS$.

SOLUTION Since the sphere has radius a, $\rho = a$. The top half of the sphere is described by $0 \le \theta \le 2\pi$ and $0 \le \phi \le \pi/2$. In spherical coordinates $z = \rho\cos(\phi) = a\cos(\phi)$. Thus

$$\int_{\mathcal{S}} z\ dS = \int_{\mathcal{S}} (a\cos(\phi))\ dS = \int_0^{2\pi}\left(\int_0^{\pi/2} (a\cos(\phi))a^2\sin(\phi)\ d\phi\right)\ d\theta.$$

Now,

$$\begin{aligned}\int_0^{\pi/2} (a\cos(\phi))a^2\sin(\phi)\ d\phi &= a^3\int_0^{\pi/2}\cos(\phi)\sin(\phi)\ d\phi = a^3\frac{(-\cos^2(\phi))}{2}\bigg|_0^{\pi/2} \\ &= \frac{a^3}{2}[-0-(-1)] = \frac{a^3}{2}\end{aligned}$$

so that

$$\int_{\mathcal{S}} z\ dS = \int_0^{2\pi}\frac{a^3}{2}\ d\theta = \pi a^3.$$

◇

We can interpret the result in Example 1 in terms of average value. The **average value** of $f(P)$ over a surface $\mathcal{S}$ is defined as

$$\frac{\int_{\mathcal{S}} f(P)\ dS}{\text{Area of } \mathcal{S}}.$$

Geometric interpretation

Example 1 shows that the average value of z over the given hemisphere is

$$\frac{\int_{\mathcal{S}} z\ dS}{\text{Area of } \mathcal{S}} = \frac{\pi a^3}{2\pi a^2} = \frac{a}{2}.$$

The average height above the equator is exactly half the radius.

A General Technique

We evaluated an integral over a curve, $\int_C f\ ds$, by replacing it with $\int_a^b f\ \frac{ds}{dt}\ dt$, an integral over an interval $[a, b]$.

We will do something similar for an integral over a surface: We will replace a surface integral by a double integral over a set in a coordinate plane.

The idea is to replace a small patch on the surface $\mathcal{S}$ by its projection on, say, the xy-plane. The area of the projection is not the same as the area of the patch. With the aid of Figure 17.7.3 we will express the area of the shadow in terms of the tilt of the patch.

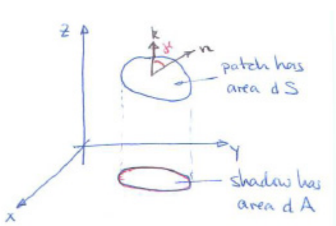


Figure 17.7.3

The unit normal vector to the patch is $\mathbf{n}$. The angle between $\mathbf{n}$ and $\mathbf{k}$ is γ. Call the area of the patch dS, and the area of its projection dA. Then

$$dA \approx |\cos(\gamma)|\ dS.$$

Direction angles and direction cosines were defined in Section 14.4.

The angle γ is one of the direction angles of the unit normal vector, $\mathbf{n}$. For instance, if $\gamma = 0$, then $dA = dS$. If $\gamma = \pi/2$, then $dA = 0$. We use the absolute value of $\cos(\gamma)$, since γ could be larger than $\pi/2$. It follows, if $\cos(\gamma)$ is not 0, that

Approximate Relationship Between dA and dS

$$dS \approx \frac{dA}{|\cos(\gamma)|} \tag{17.7.1}$$

With the aid of (17.7.1), we replace an integral over $\mathcal{S}$ with an integral over its shadow in the xy-plane. The replacement is visible in the approximating sums involved in the integral over a surface. Let $\mathcal{S}$ be a surface that meets each line parallel to the z-axis at most once. Let f be a function whose domain includes $\mathcal{S}$.

An approximating sum for $\int_{\mathcal{S}} f(P)\ dS$ is $\sum_{i=1}^{n} f(P_i)\Delta S_i$. The partition is shown in Figure 17.7.4.

Let $\mathcal{A}$ be the projection of $\mathcal{S}$ in the xy-plane. The patch $\mathcal{S}_i$, with surface area ΔS_i, projects to A_i, of area ΔA_i, and the point P_i on $\mathcal{S}_i$ projects down to Q_i in A_i Let γ_i be the angle between the normal at P_i and $\mathbf{k}$. Then $f(P_i)\Delta S_i$ is approximately $\frac{f(P_i)}{|\cos(\gamma_i)|}\Delta A_i$. Thus an approximation of $\int_{\mathcal{S}} f(P)\ dS$ is

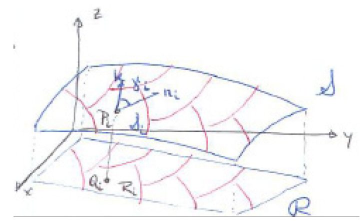


Figure 17.7.4

$$\sum_{i=1}^{n} \frac{f(P_i)}{|\cos(\gamma_i)|}\Delta A_i.$$

Taking the limit as the A_i are chosen smaller and smaller yields the following theorem.

Replacing an integral over a surface with an integral over a planar region.

Theorem 17.7.1. *Let $\mathcal{S}$ be a surface and let $\mathcal{A}$ be its projection on the xy-plane. Assume that for each point Q on $\mathcal{A}$ the line through Q parallel to the z-axis meets $\mathcal{S}$ in exactly one point P. Let f be a function defined on $\mathcal{S}$. Define h on $\mathcal{A}$ by*

$$h(Q) = f(P).$$

Then

$$\int_{\mathcal{S}} f(P)\ dS = \int_{\mathcal{A}} \frac{h(Q)}{|\cos(\gamma)|}\ dA.$$

In this equation γ denotes the angle between $\mathbf{k}$ and a vector normal to the surface of $\mathcal{S}$ at P. (See Figure 17.7.5.)

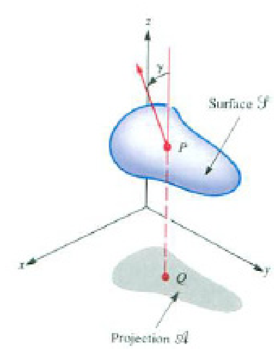


Figure 17.7.5

To apply this, we need to compute $\cos(\gamma)$.

Computing $\cos(\gamma)$

We find a vector perpendicular to the surface in order to compute $\cos(\gamma)$. If $\mathcal{S}$ is the level surface of $g(x, y, z)$, that is, $g(x, y, z) = c$ for some constant c, then the gradient ∇g is such a vector.

If the surface $\mathcal{S}$ is given as $z = f(x, y)$ so it is a level surface of $g(x, y, z) = z - f(x, y)$, Theorem 17.7.2 shows what the formulas for $\cos(\gamma)$ look like. However, it is unnecessary to memorize them. Just remember that a gradient provides a normal to a level surface.

Theorem 17.7.2. *(a) If the surface $\mathcal{S}$ is part of the level surface $g(x, y, z) = c$, then*

$$|\cos(\gamma)| = \frac{|\frac{\partial g}{\partial z}|}{\sqrt{(\frac{\partial g}{\partial x})^2 + (\frac{\partial g}{\partial y})^2 + (\frac{\partial g}{\partial z})^2}}.$$

(b) If the surface $\mathcal{S}$ is given in the form $z = f(x, y)$, then

$$|\cos(\gamma)| = \frac{1}{\sqrt{(\frac{\partial f}{\partial x})^2 + (\frac{\partial f}{\partial y})^2 + 1}}.$$

Proof

(a) A normal vector to $\mathcal{S}$ at a point is provided by the gradient

$$\nabla g = \frac{\partial g}{\partial x}\mathbf{i} + \frac{\partial g}{\partial y}\mathbf{j} + \frac{\partial g}{\partial z}\mathbf{k}.$$

The cosine of the angle between $\mathbf{k}$ and ∇g is

$$\frac{\mathbf{k} \cdot \nabla g}{|\mathbf{k}||\nabla g|} = \frac{\mathbf{k} \cdot (\frac{\partial g}{\partial x}\mathbf{i} + \frac{\partial g}{\partial y}\mathbf{j} + \frac{\partial g}{\partial z}\mathbf{k})}{(1)\left(\cdot\sqrt{(\frac{\partial f}{\partial x})^2 + (\frac{\partial g}{\partial y})^2 + (\frac{\partial g}{\partial z})^2}\right)}$$

so

$$|\cos(\gamma)| = \frac{|\frac{\partial g}{\partial z}|}{\sqrt{(\frac{\partial g}{\partial x})^2 + (\frac{\partial g}{\partial y})^2 + (\frac{\partial g}{\partial z})^2}}.$$

(b) Rewrite $z = f(x, y)$ as $z - f(x, y) = 0$. The surface $z = f(x, y)$ is thus the level surface $g(x, y, z) = 0$ of $g(x, y, z) = z - f(x, y)$. Because

$$\frac{\partial g}{\partial x} = -\frac{\partial f}{\partial x}, \qquad \frac{\partial g}{\partial y} = -\frac{\partial f}{\partial y}, \qquad \text{and} \qquad \frac{\partial g}{\partial z} = 1$$

the formula in (a) gives

$$|\cos(\gamma)| = \frac{1}{\sqrt{(\frac{\partial f}{\partial x})^2 + (\frac{\partial f}{\partial y})^2 + 1}}.$$

Theorem 17.7.2 is stated for projections on the xy-plane. Similar theorems hold for projections on the xz- or yz-planes. The direction angle γ is then replaced by the corresponding direction angle, β or α, and the normal vector is dotted with $\mathbf{j}$ or $\mathbf{i}$. Draw a picture in each case; there is no point in trying to memorize formulas.

EXAMPLE 2 Find the area of the part of the saddle $z = xy$ inside the cylinder $x^2 + y^2 = a^2$.

SOLUTION Let $\mathcal{S}$ be the part of the surface $z = xy$ inside $x^2 + y^2 = a^2$. Then

$$\text{Area of } \mathcal{S} = \int_{\mathcal{S}} 1 \; dS.$$

The projection of $\mathcal{S}$ on the xy-plane is a disk of radius a and center $(0,0)$. Call it $\mathcal{A}$, as in Figure 17.7.6. Then

$$\text{Area of } \mathcal{S} = \int_{\mathcal{S}} 1 \; dS = \int_{\mathcal{A}} \frac{1}{|\cos(\gamma)|} \; dA. \qquad (17.7.2)$$

To find the normal to $\mathcal{S}$ rewrite $z = xy$ as $z - xy = 0$. Thus $\mathcal{S}$ is a level surface of $g(x,y,z) = z - xy$. A normal to $\mathcal{S}$ is therefore

$$\begin{aligned} \nabla g &= \frac{\partial g}{\partial x}\mathbf{i} + \frac{\partial g}{\partial y}\mathbf{j} + \frac{\partial g}{\partial z}\mathbf{k} \\ &= -y\mathbf{i} - x\mathbf{j} + \mathbf{k}. \end{aligned}$$

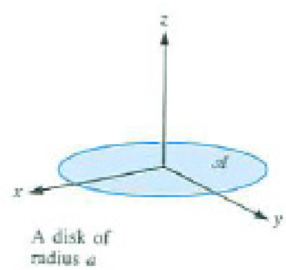


Figure 17.7.6

Then

$$\cos(\gamma) = \frac{\mathbf{k}\cdot\nabla g}{\|\mathbf{k}\|\|\nabla g\|} = \frac{\mathbf{k}\cdot(-y\mathbf{i} - x\mathbf{j} + \mathbf{k})}{\sqrt{y^2+x^2+1}} = \frac{1}{\sqrt{y^2+x^2+1}}.$$

By (17.7.2),

$$\text{Area of } \mathcal{S} = \int_{\mathcal{A}} \sqrt{y^2+x^2+1} \; dA. \qquad (17.7.3)$$

The area of $\mathcal{S}$ is $\int_{\mathcal{A}} \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + 1} \; dA.$

Use polar coordinates to evaluate the integral in (17.7.3):

$$\int_{\mathcal{A}} \sqrt{y^2+x^2+1} \; dA = \int_0^{2\pi}\int_0^{a} \sqrt{r^2+1} \; r \; dr \; d\theta.$$

The inner integration gives

$$\int_0^a \sqrt{r^2+1}r\ dr \quad = \quad \left.\frac{(r^2+1)^{3/2}}{3}\right|_0^a = \frac{(1+a^2)^{3/2}-1}{3}.$$

The second integration gives

$$\int_0^{2\pi} \frac{(1+a^2)^{3/2}-1}{3}\ d\theta = \frac{2\pi}{3}\left((1+a^2)^{3/2}-1\right).$$

◇

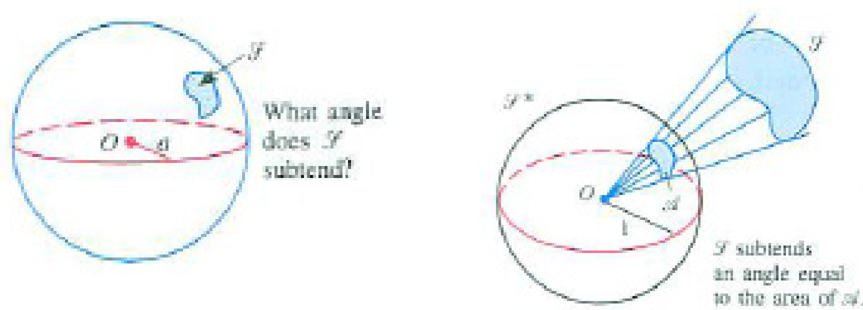
SHERMAN: The definition of steradian measure needs to be moved here. Otherwise, we use it in the paragraphs before Example 3. Figure 17.7.7(a) is misleading; $\mathcal{S}$ is NOT on the sphere with radius a. Should (a) just have $\mathcal{S}$ and the rays, and no sphere at all? Maybe this is already in your notes on the figures? Please scrutinize the current organization.

A Geometric Application: Steradians

Recall the definition of radian measure, reviewed in Section 15.4. An angle with vertex at the origin intercepts an arc of length l on a circle of radius a. The quotient, l/a, is the radian measure of the angle.

Spheres are used to measure a solid angle subtended by a surface $\mathcal{S}$. (See Figure 17.7.7(a).) A solid angle consists of the rays from the origin that meet $\mathcal{S}$. These rays intersect the sphere with radius a centered at the origin in a patch $\mathcal{A}$ with area A. The **steradian measure** of the solid angle subtended by $\mathcal{S}$ at the origin is A/a^2.

Steradians comes from *stereo*, the Greek word for space, and *radians*.



(a) (b)

Figure 17.7.7

The entire sphere subtends a solid angle of $4\pi a^2/a^2 = 4\pi$ steradians.

To visualize a solid angle introduce a sphere of radius a centered at the origin. For ease of drawing, choose the sphere to be large, so it contains $\mathcal{S}$, as in Figure 17.7.7(b).

EXAMPLE 3 Find how large is the angle subtended by one face of a cube at its center.

SOLUTION Imagine a large sphere containing the cube and having center at the center of the cube. The entire surface of the cube subtends an angle of 4π steradians. Because there are six identical faces, each face subtends $4\pi/6 = 2\pi/3$ steradians, which is about 2.09 steradians. ⋄

SHERMAN: Delete the 2.09 steradians?

Let $\mathcal{S}$ be a closed surface and P a point in the convex solid it bounds. Then similar reasoning shows that $\mathcal{S}$ subtends 4π steradians at P. This is the analog of the fact that a closed convex curve subtends an angle of measure 2π radians at any point in the region it bounds.

In Section 15.4 the radians subtended by a curve $\mathcal{C}$ was expressed as a line integral over $\mathcal{C}$, namely $\int_{\mathcal{C}} \widehat{\mathbf{r}} \cdot \mathbf{n}/r ds$. Almost identical reasoning will be used in Example 5 in Section 18.4 to obtain $\int_{\mathcal{S}} \widehat{\mathbf{r}} \cdot \mathbf{n}/r^2 \, dS$ as an expression for the steradians measure of the angle subtended by a surface $\mathcal{S}$ as an integral over $\mathcal{S}$. For now we accept this formula and show its use in the following example.

EXAMPLE 4 One corner C of a cube of side b is at the origin of xyz-space. Find $\int_{\mathcal{S}} \frac{\widehat{\mathbf{r}} \cdot \mathbf{n}}{r^2} \, dS$ where $\mathcal{S}$ is one face of the cube that does not contain the origin.
SOLUTION The surface integral $\int_{\mathcal{S}} \frac{\widehat{\mathbf{r}} \cdot \mathbf{n}}{r^2} \, dS$ equals the steradian measure of the angle subtended by that face at C. Eight identical cubes of side b, all having C as a corner, fill up the space around C and form one large cube. The surface of the large cube consists of 24 congruent squares, each of which subtends the same angle at C. Because the origin is contained within the large cube, the angle subtended by it is 4π. Thus one face subtends $4\pi/24$ steradians at C. Consequently, the value of $\int_{\mathcal{S}} \frac{\widehat{\mathbf{r}} \cdot \mathbf{n}}{r^2} \, dS$ is $4\pi/24 = \pi/6$, which is about 0.52. ⋄

Summary

After defining $\int_{\mathcal{S}} f(P) \, dS$, an integral over a surface, we showed how to compute it when the surface is part of a sphere or can be projected onto a coordinate plane. For a sphere of radius a, $\int_{\mathcal{S}} f(P) \, dS$ equals

$$\int_0^{2\pi} \int_0^{\pi} f(a, \phi, \theta) a^2 \sin(\phi) d\phi d\theta.$$

If each line parallel to the z-axis meets a surface $\mathcal{S}$ in at most one point, an integral over $\mathcal{S}$ can be replaced by an integral over $\mathcal{A}$, the projection of $\mathcal{S}$ on the xy-plane:

$$\int_{\mathcal{S}} f(P) \, dS = \int_{\mathcal{A}} \frac{h(Q)}{|\cos(\gamma)|} \, dA.$$

To find $\cos(\gamma)$, use a gradient. If the surface is a level surface of, $g(x, y, z) = c$, use ∇g. If it has the equation $z = f(x, y)$, rewrite it as $z - f(x, y) = 0$.

We also defined the steradian measure of a solid angle and related it to an integral of the vector field $\widehat{\mathbf{r}}/r^2$. In particular, for a surface $\mathcal{S}$ that encloses the origin, $\int_{\mathcal{S}} \widehat{\mathbf{r}}/r^2 \; dS$, equals 4π, a fact that will be needed in the next chapter.

EXERCISES for Section 17.7

1. A small patch of a surface makes an angle of $\pi/4$ with the xy-plane. Its projection on that plane has area 0.05. Estimate the area of the patch.

2. A small patch of a surface makes an angle of $25°$ with the yz-plane. Its projection on that plane has area 0.03. Estimate the area of the patch.

3.

(a) Draw a diagram of the part of the plane $x + 2y + 3z = 12$ that lies inside the cylinder $x^2 + y^2 = 9$.

(b) Find that area using integration.

(c) Find that area using vectors.

4.

(a) Draw a diagram of the part of the plane $z = x+3y$ that lies inside the cylinder $r = 1 + \cos(\theta)$.

(b) Find its area.

5. Let $f(P)$ be the square of the distance from P to a diameter of a sphere of radius a. Find the average value of $f(P)$ for points on the sphere.

6. Find the area of that part of the sphere of radius a that lies within a cone of half-vertex angle $\pi/4$ and vertex at the center of the sphere, as in Figure 17.7.8.

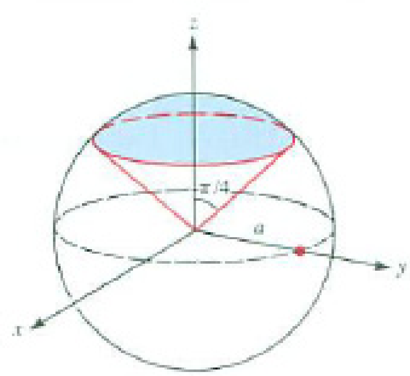


Figure 17.7.8

In Exercises 7 and 8 evaluate $\int_{\mathcal{S}} \mathbf{F}\cdot\mathbf{n}\, dS$ for the sphere and vector field ($\mathbf{n}$ is the outward unit normal.)

7. The sphere $x^2+y^2+z^2=9$ and $\mathbf{F}=x^2\mathbf{i}+y^2\mathbf{j}+z^2\mathbf{k}$.

8. The sphere $x^2+y^2+z^2=1$ and $\mathbf{F}=x^3\mathbf{i}+y^2\mathbf{j}$.

9. Find the area of the part of the spherical surface $x^2+y^2+z^2=1$ that lies within the vertical cylinder erected on the circle $r=\cos\theta$ and above the xy plane.

10. Find the area of that portion of the parabolic cylinder $z=\frac{1}{2}x^2$ between the three planes $y=0$, $y=x$, and $x=2$.

11. Evaluate $\int_{\mathcal{S}} x^2y\, dS$, where $\mathcal{S}$ is the portion in the first octant of a sphere with radius a and center at the origin, as follows:

(a) Set up an integral using x and y as parameters.

(b) Set up an integral using ϕ and θ as parameters.

(c) Evaluate the easier of (a) and (b).

12. A triangle in the plane $z=x+y$ is directly above the triangle in the xy-plane whose vertices are $(1,2)$, $(3,4)$, and $(2,5)$. Find the area of

(a) the triangle in the xy-plane,

(b) the triangle in the plane $z=x+y$.

13. Let $\mathcal{S}$ be the triangle with vertices $(1,1,1)$, $(2,3,4)$, and $(3,4,5)$. Find the area of $\mathcal{S}$

(a) using vectors, find the area of $\mathcal{S}$.

(b) using the formula Area of $\mathcal{S}=\int_{\mathcal{S}} 1\, dS$.

14. Find the area of the portion of the cone $z^2=x^2+y^2$ that lies above one loop of the curve $r=\sqrt{\cos(2\theta)}$.

15. Let $\mathcal{S}$ be the triangle whose vertices are $(1,0,0)$, $(0,2,0)$, and $(0,0,3)$. Let $f(x,y,z)=3x+2y+2z$. Evaluate $\int_{\mathcal{S}} f(P)\, dS$.

16. Let $\mathcal{S}$ be a sphere of radius a with center at the origin of a rectangular coordinate system. Evaluate the integral with a minimum amount of labor.

(a) $\int_{\mathcal{S}} x \; dS$

(b) $\int_{\mathcal{S}} x^3 \; dS$

(c) $\int_{\mathcal{S}} \frac{2x+4y^5}{\sqrt{2+x^2+3y^2}} \; dS$

17. Let $\mathcal{S}$ be a sphere of radius a with center at the origin of a rectangular coordinate system.

(a) Why is $\int_{\mathcal{S}} x^2 \; dS = \int_{\mathcal{S}} y^2 \; dS$?

(b) Evaluate $\int_{\mathcal{S}} (x^2 + y^2 + z^2) \; dS$ with a minimum amount of labor.

(c) In view of (a) and (b), evaluate $\int_{\mathcal{S}} x^2 \; dS$.

(d) Evaluate $\int_{\mathcal{S}} (2x^2 + 3y^3) \; dS$.

18. An electric field radiates power at the rate of $k \sin^2(\phi)/\rho^2$ units per square meter to the point $P = (\rho, \theta, \phi)$. Find the total power radiated to the sphere $\rho = a$.

19. A sphere of radius $2a$ has its center at the origin of a rectangular coordinate system. A circular cylinder of radius a has its axis parallel to the z-axis and passes through the z-axis. Find the area of that part of the sphere that lies within the cylinder and is above the xy-plane.

Consider a distribution of mass M on the surface $\mathcal{S}$. Let its density at P be $\sigma(P)$. The **moment of inertia** of the mass around the z-axis is defined as $\int_{\mathcal{S}} (x^2 + y^2)\sigma(P) \; dS$. Exercises 20 and 21 concern this integral.

20. Find the moment of inertia of a homogeneous distribution of mass on the sphere of radius a around a diameter.

21. Let a, b, and c be positive numbers. Find the moment of inertia about the z-axis of a homogeneous distribution of mass, M, of the triangle whose vertices are $(a, 0, 0)$, $(0, b, 0)$, and $(0, 0, c)$.

22. Let $\mathcal{S}$ be a sphere of radius a. Let $\mathcal{A}$ be a point at distance $b > a$ from the center of $\mathcal{S}$. For P in $\mathcal{S}$ let $f(P)$ be $1/q$, where q is the distance from P to A. Show that the average of $f(P)$ over $\mathcal{S}$ is $1/b$.

23. The data are the same as in Exercise 22 but $b < a$. Show that in this case the average of $1/q$ is $1/a$. The average does not depend on b in this case.

Exercises 24 to 27 deal with steradians. They are intended to be answered using the definition and basic properties of geometry. See also Exercise 37.

24. A right circular cone has angle 2ϕ radians at its vertex. What is the steradian measure of the solid angle its base subtends at the vertex

(a) when ϕ is $\pi/2$?

(b) when ϕ is less than $\pi/2$?

25. Let $\mathcal{C}$ be a convex body bounded by the smooth surface $\mathcal{S}$. Show that if the origin is outside of $\mathcal{C}$, then the integral of $\frac{\widehat{\mathbf{r}}\cdot\mathbf{n}}{r^2}$ over $\mathcal{S}$ is 0. (Note where $\widehat{\mathbf{r}}\cdot\mathbf{n}$ is positive and when it is negative.)

26. Let $\mathcal{C}$ be a convex body bounded by the smooth surface $\mathcal{S}$. (A surface is **smooth** if it has a continuous unit normal vector and no planar parts.) Let $\mathbf{n}$ denote the external normal to $\mathcal{S}$. Assume that the origin is on the surface $\mathcal{S}$. Use steradians to show that $\int_{\mathcal{S}} \mathbf{r}\cdot\mathbf{n}/r^3\,dS = 2\pi$.

27. Let $\mathcal{C}$ be a cube.

(a) How many steradians are there in the solid angle subtended at a corner of $\mathcal{C}$ by the surface of $\mathcal{C}$?

(b) Does this contradict Exercise 26?

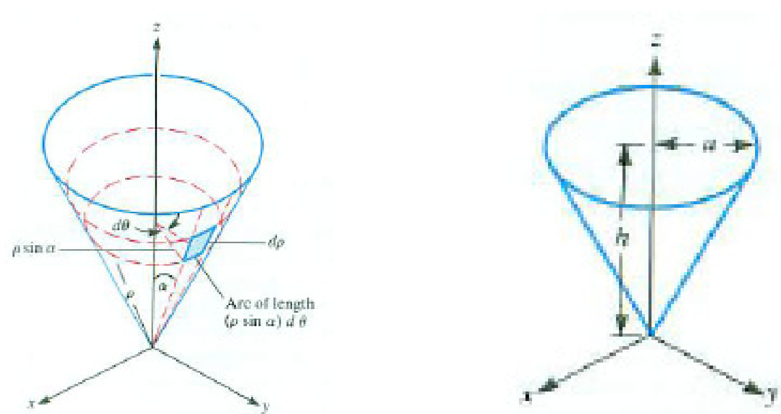
Spherical coordinates are also useful for integrating over a right circular cone as in Exercises 28 to 30. Place the origin at the vertex of the cone and the $\phi = 0$ ray along the axis of the cone, as shown in Figure 17.7.9(a). Let α be the half-vertex angle of the cone.

On the surface of the cone ϕ is constant, $\phi = \alpha$, but ρ and θ vary. A small patch on the surface of the cone corresponding to slight changes $d\theta$ and $d\rho$ has area approximately

$$(\rho\sin(\alpha)\ d\theta)\ d\rho = \rho\sin(\alpha)\ d\rho\ d\theta.$$

(Why? See Figure 17.7.9.) So we may write

$$dS = \rho\sin\alpha\ d\rho\ d\theta.$$



(a) (b)

Figure 17.7.9

28. Find the average distance from points on the curved surface of a cone of radius a and height h to its axis.

29. Evaluate $\int_{\mathcal{S}} z^2\ dS$, where $\mathcal{S}$ is the entire surface of the cone shown in Figure 17.7.9(b), including its base.

30. Evaluate $\int_{\mathcal{S}} x^2\ dS$, where $\mathcal{S}$ is the curved surface of the right circular cone of radius 1 and height 1 with axis along the z-axis.

Integration over the curved surface of a right circular cylinder is easiest in cylindrical coordinates. Given the cylinder of radius a and axis on the z-axis, a patch on the cylinder corresponding to dz and $d\theta$ has area approximately $dS = a\ dz\ d\theta$. (Why?) Exercises 31 and 32 illustrate the use of these coordinates.

31. Let $\mathcal{S}$ be the entire surface of a solid cylinder of radius a and height h. For P in $\mathcal{S}$ let $f(P)$ be the square of the distance from P to one base. Find $\int_{\mathcal{S}} f(P)\ dS$. Be sure to include the two bases in the integration.

32. Let $\mathcal{S}$ be the curved part of the cylinder in Exercise 31. Let $f(P)$ be the square of the distance from P to a fixed diameter in a base. Find the average value of $f(P)$ for points in $\mathcal{S}$.

33. The areas of the projections of a small flat surface patch on the three coordinate planes are 0.01, 0.02, and 0.03. Is that enough information to find the area of the patch? If so, find it. If not, explain why not.

34. Let $\mathbf{F}$ describe the flow of a fluid in space. (See Section 16.3 for fluid flow in a planar region.) $\mathbf{F}(P) = \delta(P)\mathbf{v}(P)$, where $\delta(P)$ is the density of the fluid at P and $\mathbf{v}(P)$ is the velocity of the fluid at P. Making clear, large diagrams, explain why the rate at which the fluid is leaving the solid region enclosed by a surface $\mathcal{S}$ is $\int_{\mathcal{S}} \mathbf{F}\cdot\mathbf{n}\ dS$, where $\mathbf{n}$ denotes the unit outward normal to $\mathcal{S}$.

35. Let $\mathcal{S}$ be the smooth surface of a convex body. Show that $\int_{\mathcal{S}} z\cos(\gamma)\ dS$ is equal to the volume of the solid bounded by $\mathcal{S}$. (Break $\mathcal{S}$ into two parts. In one part $\cos(\gamma)$ is positive; and the other it is negative.)

36. Let $R(x, y, z)$ be a scalar function defined over a closed surface $\mathcal{S}$. (See Figure 17.7.10.)

(a) Show that

$$\int\limits_{\mathcal{S}} R(x,y,z)\cos(\gamma)\ dS = \int\limits_{\mathcal{A}} \left(R(x,y,z_2) - R(x,y,z_1)\right)\ dA,$$

where $\mathcal{A}$ is the projection of $\mathcal{S}$ on the xy-plane and the line through $(x, y, 0)$ parallel to the z-axis meets $\mathcal{S}$ at (x, y, z_1) and (x, y, z_2), with $z_1 \le z_2$.

(b) Let $\mathcal{S}$ be a surface of the type in (a). Evaluate $\int_{\mathcal{S}} z\cos(\gamma)\ dS$.

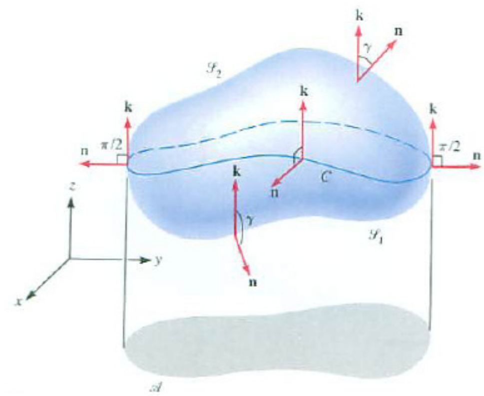


Figure 17.7.10

37. Thie Exercise may be of use in Exercise 38.

(a) Let g be a differentiable function such that $g((x+y)/2) = ((g(x)+g(y))/2$ for all x and y. Show that $g(x) = kx + c$ for some k and c. (Differentiate the first equation.)

(b) Let f be a differentiable function such that $(x+y)f(x+y)+(x-y)f(x-y) = 2xf(x)$ for all x and y. Deduce that there are constants k and c such that $f(x) = k + c/x$.

38. (Suggested by Exercise 22.) Let $d(P)$ be the distance from to a point at a distance b from the center of a sphere of radius a, $a < b$. Let $f(x)$ have the property that the the average value of $f(d(P))$ on the sphere is equal to $f(b)$. This condition holds for all a and b, with $0 < a < b$. Show that f must have the form $f(x) = k + c/x$. (Exercise 37 may be useful.)

17.8 Moments, Centers of Mass, and Centroids

Now that we can integrate over planar regions, surfaces, and solid regions, we can define and calculate the center of mass of a physical object. The center of mass is important for a naval architect, who do not want ships to tip over easily. Pole vaulters hope that as they clear the bar their centers of mass go under the bar. Archimedes, the first person to study the center of mass, was interested in the stability of floating paraboloids.

The Center of Mass

A small boy on one side of a seesaw (which we regard as weightless) can balance a bigger boy on the other side. For example, the two boys in Figure 17.8.1 balance. Each boy exerts a force on the seesaw, due to gravitational attraction, proportional to his mass. The small mass with the long lever arm balances the large mass with the small lever arm. Each boy contributes the same tendency to turn but in opposite directions.

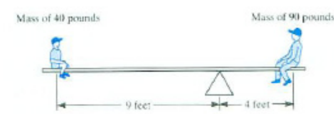


Figure 17.8.1

This tendency is called the **moment**:

$$\text{Moment} = (\text{Mass}) \cdot (\text{Lever arm}),$$

where the lever arm can be positive or negative. To be more precise, introduce on the seesaw an x-axis with its origin 0 at the fulcrum, the point on which the seesaw rests. Define the moment about 0 of a mass m located at the point x on the x-axis to be mx. Then the bigger boy has a moment (90)(4), while the smaller boy has a moment (40)(−9). The total moment of the lever-mass system is 0, and the masses balance. (See Figure 17.8.2.)

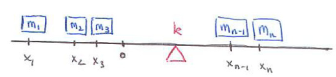


Figure 17.8.2

If a mass m is located on a line with coordinate x, we define its moment about the point having coordinate k as the product $m(x-k)$.

For several point masses m_1, m_2, …, m_n. if mass m_i is located at x_i, for $i = 1, 2, \ldots, n$, then $\sum_{i=1}^{n} m_i(x_i - k)$ is the total moment of the masses about the point k. If a fulcrum is placed at k, then the seesaw rotates clockwise if the total moment is greater that 0, rotates counterclockwise if it is less than 0, and is in equilibrium if the total moment is 0. See Figure 17.8.3.

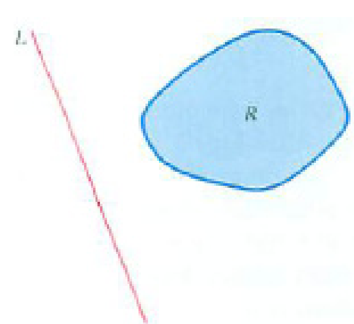


Figure 17.8.3

To find where to place the fulcrum so that the total tendency to turn is 0, we find k such that

$$\sum_{i=1}^{n} m_i(x_i - k) = 0.$$

Writing this as

$$k \sum_{i=1}^{n} m_i = \sum_{i=1}^{n} m_i x_i,$$

we see that

$$k = \frac{\sum_{i=1}^{n} m_i x_i}{\sum_{i=1}^{n} m_i}. \qquad (17.8.1)$$

The number k given by (17.8.1) is called the **center of mass** or center of gravity of the system of masses. It is the point about which all the masses balance. The x coordinate of the center of mass is found by dividing the total moment about 0 by the total mass. It is usually denoted $\bar{x}$.

$\bar{x}$ is pronounced "x bar".

(a) (b)

Figure 17.8.4

Finding the center of mass of a finite number of point masses involves only arithmetic. For example, suppose three masses are placed on a seesaw as in Figure 17.8.4(a). Introduce an x-axis with origin at mass $m_1 = 20$ pounds. Two additional masses are located at $x_2 = 4$ feet and $x_3 = 14$ feet with masses $m_2 = 10$ pounds and $m_3 = 50$ pounds, respectively. The total moment about $x = k$ is

$$M = 20(0-k) + 10(4-k) + 50(14-k) = 740 - 80k.$$

This moment vanishes when $M = 0$, that is, when $k = 740/80 = 9.25$.

This is consistent with the formula for the center of mass:

$$\bar{x} = \frac{m_1x_1 + m_2x_2 + m_3x_3}{m_1 + m_2 + m_3} = \frac{0 + 40 + 700}{10 + 20 + 50} = \frac{740}{80} = 9.25.$$

The seesaw balances when the fulcrum is placed 9.25 feet from the first mass. (See Figure 17.8.4(b).)

Calculus is needed to find the center of mass of a one-dimensional rod that occupies the interval $a \le x \le b$ with density $\lambda(x)$ at x. To apply the previous ideas, divide the rod into n pieces of width $\Delta x = (b-a)/n$. Let $x_i = a + i\Delta x$, $i = 0, \ldots, n$. For the piece of the rod for $x_{i-1} \le x \le x_i$, select a point c_i in it. The mass of the piece is approximately $\lambda(c_i)\Delta x$. An approximation to the total moment about $x = k$ is

$$M \approx \sum_{i=1}^{n} \underbrace{\lambda(c_i)\Delta x}_{\text{mass}}\,\underbrace{(c_i - k)}_{\text{lever}}.$$

As n increases without bound these Riemann sums converge to a definite integral for the total moment

$$M = \int_a^b \lambda(x)(x-k)\ dx.$$

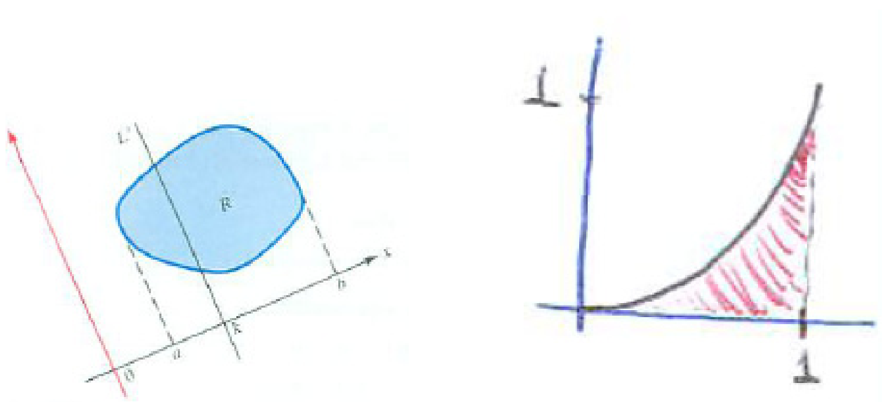
The total moment vanishes when

$$k = \frac{\int_a^b x\lambda(x)\ dx}{\int_a^b \lambda(x)\ dx}.$$

The denominator is the mass of the rod.

Tto find the center of mass of a continuous distribution of matter in a plane region, we use a double integral.

Let R be a region in the plane occupied by a thin piece of metal whose density at P is $\sigma(P)$. Let L be a line in the plane, as shown in Figure 17.8.5(a). We will find a formula for the unique line parallel to L, around which the mass



(a) (b)

Figure 17.8.5

in R balances.

Let L' be any line parallel to L. We will compute the moment about L' and then see how to choose L' to make it 0. To compute the moment of R about L', introduce an x-axis perpendicular to L with its origin at its intersection with L. Assume that L' passes through the x-axis at the point $x = k$, as in Figure 17.8.5(b). In addition, assume that line parallel to L meets R either in a line segment or at a point on the boundary of R. The lever arm of the mass distributed throughout R varies from point to point.

We partition R into n small regions R_1, R_2, …, R_n. Let A_i be the area of R_i. In each region the lever arm around L' varies only a little. So, if we pick a point P_1 in R_1, P_2 in R_2, …, P_n in R_n, and the x-coordinate of P_i is x_i, then

$$\underbrace{(x_i - k)}_{\text{lever arm}}\underbrace{\sigma(P_i)A_i}_{\text{mass in } R_i}$$

is a local estimate of the turning tendency.

Thus

$$\sum_{i=1}^{n}(x_i - k)\sigma(P_i)A_i \tag{17.8.2}$$

would be a good estimate of the total turning tendency around L'. Taking the limit of (17.8.2) as all R_i are chosen smaller and smaller, we expect

$$\int_R (x - k)\sigma(P)\ dA \tag{17.8.3}$$

to represent the turning tendency of the total mass around L'. The quantity (17.8.3) is called the **moment** of the mass distribution around L'.

EXAMPLE 1 Let R be the region under $y = x^2$ and above $[0, 1]$ with the density $\sigma(x, y) = xy$. Find its moment around the line $x = 1/2$.
SOLUTION R is shown in Figure 17.8.6. The moment (17.8.3) equals

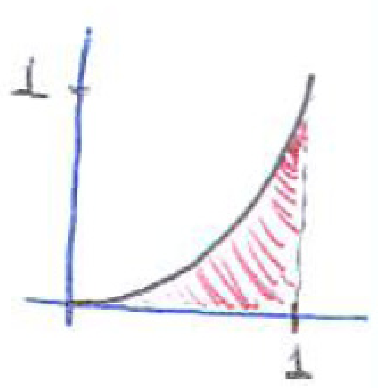


Figure 17.8.6

$$\int_R \left(x - \frac{1}{2}\right) xy\ dA. \tag{17.8.4}$$

We evaluate the double integral by the iterated integral

$$\int_0^1 \left(\int_0^{x^2} \left(x - \frac{1}{2} \right) xy\ dy \right)\ dx.$$

The first integration gives

See Exercise 2.

$$\int_0^{x^2} \left(x - \frac{1}{2} \right) xy\ dy \quad = \quad (x - \frac{1}{2})x \int_0^{x^2} y\ dy = \frac{\left(x - \frac{1}{2}\right) x^5}{2}.$$

The second integration is

$$\int_0^1 \frac{\left(x - \frac{1}{2}\right) x^5}{2} = \int_0^1 \frac{2x^6 - x^5}{4}\ dx = \frac{5}{168}.$$

Since the total moment (17.8.4) is positive, the object would rotate clockwise around the line $x = \frac{1}{2}$. ◇

Now that we have a way to find the moment around any line parallel to the y-axis we can find the line around which the moment is zero, the balancing line. We solve for k in the equation

$$\int_R (x - k)\sigma(P)\ dA = 0.$$

Thus

$$\int_R x\sigma(P)\ dA = k \int_R \sigma(P)\ dA,$$

from which we find that

$$\boxed{k = \frac{\int_R x\sigma(P)\ dA}{\int_R \sigma(P)\ dA}.}$$

The denominator is the total mass. The numerator is the total moment. So we can think of k as the average lever arm as weighted by the density.

There is therefore a unique balancing line parallel to the y-axis. Call its x-coordinate $\overline{x}$. Similarly, there is a unique balancing line parallel to the x-axis. Call its y-coordinate $\overline{y}$. The point $(\overline{x}, \overline{y})$ is called the **center of mass** of the region R. We have:

Center of Mass of Region R with Density σ

The center of mass of a region R with density $\sigma(P)$ has coordinates $(\bar{x}, \bar{y})$ where

$$\overline{x} = \frac{\int_R x\sigma(P)\ dA}{\int_R \sigma(P)\ dA} \qquad \text{and} \qquad \overline{y} = \frac{\int_R y\sigma(P)\ dA}{\int_R \sigma(P)\ dA}.$$

The integral $\int_R x\sigma(P)\ dA$ is called the **moment of R around the y-axis**, and is denoted M_y. Similarly, $M_x = \int_R y\sigma(P)\ dA$ is the **moment of R around the x-axis**.

If the density $\sigma(P)$ is constant, say, equal to 1 everywhere in R, then the two equations reduce to

$$\overline{x} = \frac{\int_R x\ dA}{\int_R\ dA} \qquad \text{and} \qquad \overline{y} = \frac{\int_R y\ dA}{\int_R\ dA}.$$

In this case the center of mass R is also called the **centroid** of the region, a purely geometric concept:

Centroid of R

The centroid of the plane region R has the coordinates $(\bar{x}, \bar{y})$ where

$$\overline{x} = \frac{\int_R x\ dA}{\text{Area of } R} \qquad \text{and} \qquad \overline{y} = \frac{\int_R y\ dA}{\text{Area of } R}. \tag{17.8.5}$$

EXAMPLE 2 Find the center of mass of the region in Example 1.

SOLUTION The density at (x, y) in R is given by $\sigma = xy$. We compute three double integrals: the mass $\int_R xy\ dA$ and the moments $M_y = \int_R x(xy)\ dA$ and $M_x = \int_R y(xy)\ dA$.

We have

$$M_y = \int\limits_R x^2y\ dA \quad = \quad \int\limits_0^1 \left(\int\limits_0^{x^2} x^2y\ dy \right)\ dx = \int\limits_0^1 \frac{x^6}{2}\ dx = \frac{1}{14}.$$

Then

$$M = \int\limits_R xy\ dA \quad = \quad \int\limits_0^1 \left(\int\limits_0^{x^2} xy\ dy \right)\ dx = \int\limits_0^1 \frac{x^5}{2}\ dx = \frac{1}{12}.$$

Finally,

$$M_x = \int_R xy^2 \, dA \;=\; \int_0^1 \left(\int_{x^2}^0 xy^2 \, dy \right) dx = \int_0^1 \frac{x^7}{3} \, dx = \frac{1}{24}.$$

Thus

$$\overline{x} = \frac{\frac{1}{14}}{\frac{1}{12}} = \frac{6}{7} \qquad \text{and} \qquad \overline{y} = \frac{\frac{1}{24}}{\frac{1}{12}} = \frac{1}{2}.$$

It is not surprising that $\overline{x}$ is greater than 1/2, since in Example 17.8.1 we found that the object rotates clockwise around the line $x = 1/2$. ◇

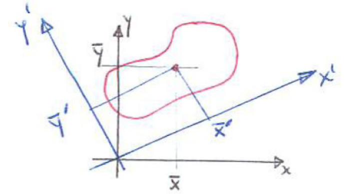


Figure 17.8.7

An Important Point About an Important Point

We defined the center of mass $(\overline{x}, \overline{y})$ by first choosing an xy-coordinate system. What if we choose an $x'y'$ coordinate system at an angle to the xy-coordinate system? Would the center of mass computed in this system, $(\overline{x}', \overline{y}')$ be the same point as $(\overline{x}, \overline{y})$? See Figure 17.8.7. Fortunately, it is, as Exercise 38 shows.

Shortcuts for Computing Centroids

Assume that f is a non-negative function and let R be the region under $y = f(x)$ for x in $[a, b]$. Then, in computing the centroid of R, we encounter the moment about the x-axis

$$M_x = \int_R y \, dA.$$

Thus

$$M_y \;=\; \int_a^b \Big(\int_0^{f(x)} y \, dy \Big) \, dx = \int_a^b \frac{(f(x))^2}{2} \, dx = \frac{1}{2} \int_a^b (f(x))^2 \, dx.$$

Thus, by (17.8.5)

$$\overline{y} = \frac{\frac{1}{2} \int_a^b (f(x))^2 \, dx}{\text{Area of } R}. \tag{17.8.6}$$

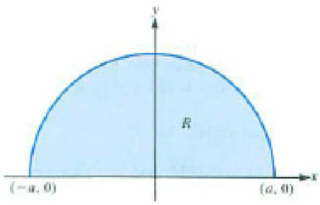


Figure 17.8.8

EXAMPLE 3 Find the centroid of the semicircular region of radius a shown in Figure 17.8.8.

SOLUTION By symmetry, $\overline{x} = 0$.

To find $\overline{y}$, use (17.8.6). The function f is given by $f(x) = \sqrt{a^2 - x^2}$, an even function. The moment of R about the x-axis is

$$\begin{aligned}\int_{-a}^{a} \frac{(\sqrt{a^2-x^2})^2}{2}\,dx &= \int_{-a}^{a} \frac{a^2-x^2}{2}\,dx = 2\int_{0}^{a} \frac{a^2-x^2}{2}\,dx \\ &= \int_{0}^{a} (a^2-x^2)\,dx = \left(a^2x - \frac{x^3}{3}\right)\Big|_0^a \\ &= (a^3 - \frac{a^3}{3}) - 0 = \frac{2}{3}a^3.\end{aligned}$$

Since $4/(3\pi) \approx 0.42$, the centroid of R is at a height of about $0.42a$.

Thus

$$\overline{y} = \frac{\frac{2}{3}a^3}{\text{Area of } R} = \frac{\frac{2}{3}a^3}{\frac{1}{2}\pi a^2} = \frac{4a}{3\pi}.$$

◇

Centers of Other Masses

We defined for masses situated in a plane. They generalize to masses distributed on a curve (such as a wire) or in space (such as an ellipsoid).

For a curve, the mass has a linear density $\lambda(P)$. A short piece around P of length Δs would have mass approximately $\lambda(P)\Delta s$. Thus, the mass and moments of the curve would be

$$M = \int_C \lambda(P)\,ds, \qquad M_y = \int_C x\lambda(P)\,ds, \qquad \text{and} \qquad M_x = \int_C y\lambda(P)\,ds.$$

We state the definition for of a solid object of density $\delta(P)$ occupying the region R. We assume an xyz-coordinate system. The total mass is

$$M = \int_R \delta(P)\,dV.$$

There are three moments, one around each of the coordinate planes:

$$M_{yz} = \int_R x\delta(P)\,dV, \qquad M_{xz} = \int_R y\delta(P)\,dV, \qquad M_{xy} = \int_R z\delta(P)\,dV.$$

The center of mass is $(\overline{x}, \overline{y}, \overline{z})$, where

$$\overline{x} = \frac{\int_R x\delta(P)\ dV}{M}, \qquad \overline{y} = \frac{\int_R y\delta(P)\ dv}{M}, \qquad \overline{z} = \frac{\int_R z\delta(P)\ dV}{M}.$$

If $\delta(P) = 1$ for all P in R, then mass is the same as the volume and the center of mass is called the centroid.

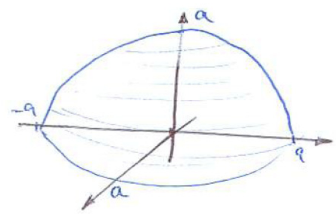


Figure 17.8.9

EXAMPLE 4 Find the centroid of the hemisphere of radius a shown in Figure 17.8.9.

SOLUTION We place the origin of an xyz-coordinate system at the center of the hemisphere, as in Figure 17.8.9.

By symmetry, the centroid is on the z-axis. (If you spin the hemisphere about the z-axis you get the same hemisphere back, which must have the same centroid.) If the centroid were not on the z-axis, you would get more than one centroid for the same object.

So $\overline{x} = \overline{y} = 0$. Calling the hemisphere R, we have

$$\overline{z} = \frac{\int_R z\ dV}{\text{Volume of } R}.$$

The volume of the hemisphere is half that of a ball, $(2/3)\pi a^3$. To evaluate the moment $\int_R z\ dV$, we use an iterated integral in spherical coordinates. Because $z = \cos(\phi)$, we have:

See Exercise 3.

$$\int\limits_R z\ dV = \int\limits_0^{2\pi}\int\limits_0^{\pi/2}\int\limits_0^{a} (\rho\cos(\phi))\rho^2\sin(\phi)\ d\rho\ d\phi\ d\theta.$$

Straightforward computations show that

$$\int\limits_R z\ dV = \frac{\pi a^4}{4}.$$

Thus

$$\overline{z} = \frac{\frac{\pi a^4}{4}}{\frac{2}{3}\pi a^3} = \frac{3a}{8}.$$

The centroid is $(0, 0, \frac{3a}{8})$. ◇

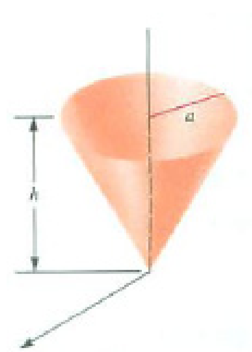


Figure 17.8.10

EXAMPLE 5 Find the centroid of a homogeneous cone of height h and radius a.

SOLUTION As we saw for the sphere in Example 4, symmetry tells us the centroid lies on the axis of the cone.

SHERMAN: half-vertex angle? vertex half-angle? ???

Introduce a spherical coordinate system with the origin at the vertex of the cone and with the axis of the cone lying on the ray $\phi = 0$, as in Figure 17.8.10. The half-vertex angle is $\arctan(a/h)$. The plane of the base of the cone is $z = h$ (in rectangular coordinates), hence

$$\rho\cos(\phi) = h.$$

In spherical coordinates, the cone's description is

$$0 \leq \theta \leq 2\pi, \qquad 0 \leq \phi \leq \arctan(a/h), \qquad 0 \leq \rho \leq h/\cos(\phi).$$

To find the centroid we compute $\int_R z\ dV$ and divide the result by the volume of the cone, which is $\frac{1}{3}\pi a^2 h$.

Now

$$\int\limits_R z\ dV = \int\limits_0^{2\pi} \int\limits_0^{\arctan(a/h)} \int\limits_0^{h/\cos(\phi)} \rho\cos(\phi)(\rho^2\sin(\phi))\ d\rho\ d\phi\ d\theta.$$

See Exercise 4

For the first integration, ϕ and θ are constant so

$$\int\limits_0^{h/\cos(\phi)} \rho\cos(\phi)\rho^2\sin(\phi)\ d\rho \quad = \quad \cos(\phi)\sin(\phi) \int\limits_0^{h/\cos(\phi)} \rho^3\ d\rho = \frac{h^4\sin(\phi)}{4\cos^3(\phi)}.$$

The second integration is

$$\int\limits_0^{\arctan(a/h)} \frac{h^4\sin(\phi)}{4\cos^3(\phi)}\ d\phi \quad = \quad \frac{h^4}{4} \int\limits_0^{\arctan(a/h)} \frac{\sin(\phi)}{\cos^3(\phi)}\ d\phi = \frac{a^2h^2}{8}.$$

The final integral is

$$\int\limits_0^{2\pi} \frac{a^2h^2}{8}\ d\theta = \frac{a^2h^2}{8} 2\pi = \frac{\pi a^2 h^2}{4}.$$

Thus,

$$\overline{z} = \frac{\int_R z\ dV}{\text{Volume of } R} = \frac{\left(\frac{\pi a^2 h^2}{4}\right)}{\left(\frac{\pi a^2 h}{3}\right)} = \frac{3h}{4}.$$

The centroid of a cone is three-fourths of the way from the vertex to the base. ◇

Archimedes and Centroids

How did Archimedes find centroids? Integral calculus was not invented until 1684, some 1900 years after his death. In one approach, he used axioms, in the style of Euclid's geometry text, written a generation or two before him. These are the axioms:

1. The centroids of similar figures are similarly situated.
2. The centroid of a convex region lies within the region.
3. If an object is cut into two pieces, its centroid C lies on the line segment joining the centroids of the two pieces. Moreover, if the pieces are R with centroid A and S with centroid B, then CA times the area of R equals CB times the area of S.

The book cited in Exercise 6 describes how he used them to find the centroid of a triangle.

Summary

We defined the moment about a line and used it to define the center of mass for a plane distribution of mass. The moment of a mass about a line L measures the tendency of the mass to rotate about the line L. The center of mass for a region R in the xy-plane is the point in the region where the region balances.

The moment about the y-axis is

$$M_y = \int\limits_R x\delta(P)\ dA.$$

The moment about the x-axis is

$$M_x = \int\limits_R y\delta(P)\ dA.$$

The center of mass is $(\bar{x}, \bar{y})$ where

$$\overline{x} = \frac{M_y}{\text{Mass}}, \overline{y} = \frac{M_x}{\text{Mass}}.$$

If the density is constant, we have a purely geometric concept,

$$\overline{x} = \frac{\int_R x\ dA}{\text{Area of } R}, \overline{y} = \frac{\int_R y\ dA}{\text{Area of } R}.$$

The definitions, which generalize to curves and solids, are listed in Table 17.8.1.

	curve (C)	solid (R)
density	$\lambda(P)$	$\delta(P)$
M	$\int_C \lambda(P)\ ds$	$\int_S \delta(P)\ dV$
M_{yz}	$\int_C x\lambda(P)\ ds$	$\int_S x\delta(P)\ dV$
M_{xz}	$\int_C y\lambda(P)\ ds$	$\int_S y\delta(P)\ dV$
M_{xy}	$\int_C z\lambda(P)\ ds$	$\int_S z\delta(P)\ dV$

Table 17.8.1 Generalization of mass and moment to curves snd solids.

EXERCISES for Section 17.8

1.

(a) How would you define the centroid of a curve? Call its linear density $\lambda(P)$.

(b) Find the centroid of a semicircle of radius a.

2. Carry out the integrations in Example 1.

3. Carry out the calculations in Example 4.

4. Evaluate the integrals in Example 5.

5. Example 4 showed that the centroid of a hemisphere is less than halfway from the center to its surface. Why is that to be expected?

6. Find the centroid of a solid paraboloid of revolution. This is the region above $z = x^2 + y^2$ and below the plane $z = c$. Archimedes found the centroid without calculus and used the result to analyze the equilibrium of a floating paraboloid. (If it is slightly tilted, will it come back to the vertical or topple over?) To see how he did this 2200 years ago see S. Stein, *Archimedes: What Did He Do Besides Cry Eureka?*, Mathematical Association of America, 1999.

7. Using cylindrical coordinates, find $\overline{z}$ for the region below the paraboloid $z = x^2 + y^2$ and above the disk $r = 2\cos(\theta)$ in the $r\theta$-plane. Include a drawing of the region.

8. Find the z coordinate, $\overline{z}$, of the centroid of the part of the saddle $z = xy$ that lies above the portion of the disk bounded by the circle $x^2 + y^2 = a^2$ in the first quadrant.

In Exercises 9 to 16 find the centroid of R. (Exercises 13 to 16 require integral tables or techniques of Chapter 8.)

9. R is bounded by $y = x^2$ and $y = 4$.

10. R is bounded by $y = x^4$ and $y = 1$.

11. R is bounded by $y = 4x - x^2$ and the x-axis.

12. R is bounded by $y = x$, $x + y = 1$, and the x-axis.

13. R is the region bounded by $y = e^x$ and the x-axis, between the lines $x = 1$ and $x = 2$.

14. R is the region bounded by $y = \sin(2x)$ and the x-axis, between the lines

$x = 0$ and $x = \pi/2$.

15. R is the region bounded by $y = \sqrt{1+x}$ and the x-axis, between the lines $x = 0$ and $x = 3$.

16. R is the region bounded by $y = \ln(x)$ and the x-axis between the lines $x = 1$ and $x = e$.

In Exercises 17 to 24 find the center of mass of the lamina.

17. The triangle with vertices $(0, 0)$, $(1, 0)$, $(0, 1)$; density at (x, y) is $x + y$.

18. The triangle with vertices $(0, 0)$, $(2, 0)$, $(1, 1)$; density at (x, y) is y.

19. The square with vertices $(0, 0)$, $(1, 0)$, $(1, 1)$, $(0, 1)$; density at (x, y) is $y \arctan(x)$.

20. The finite region bounded by $y = 1 + x$ and $y = 2^x$; density at (x, y) is $x + y$.

21. The triangle with vertices $(0, 0)$, $(1, 2)$, $(1, 3)$; density at (x, y) is xy.

22. The finite region bounded by $y = x^2$, the x-axis, and $x = 2$; density at (x, y) is e^x

23. The finite region bounded by $y = x^2$ and $y = x + 6$, situated to the right of the y-axis; density at (x, y) is $2x$.

24. The trapezoid with vertices $(0, 0)$, $(3, 0)$, $(2, 1)$, $(0, 1)$; density at (x, y) is $\sin(x)$.

25. In a letter of 1680 Leibniz wrote:

> Huygens, as soon as he had published his book on the pendulum, gave me a copy of it; and at that time I was quite ignorant of Cartesian algebra and also of the method of indivisibles, indeed I did not know the correct definition of the center of gravity. For, when by chance I spoke of it to Huygens, I let him know that I thought that a straight line drawn through the center of gravity always cut a figure into two equal parts; since that clearly happened in the case of a square, or a circle, an ellipse, and other figures that have a center of magnitude. I imagine that it was the same for all other figures. Huygens laughed when he heard this, and told me that nothing was further from the truth.
>
> [Quoted in C.H. Edwards, *The Historical Development of the Calculus*, p. 239, Springer-Verlag, New York, 1979.]

Give an example showing that Huygen'swas right.

26. Cut an irregular shape out of cardboard and find three balancing lines for it experimentally. Are they concurrent? That is, do they pass through a common point?

27. Let f and g be continuous functions such that $f(x) \geq g(x) \geq 0$ for x in $[a, b]$. Let R be the region above $[a, b]$ that is bounded by the curves $y = f(x)$ and $y = g(x)$.

(a) Set up a definite integral in terms of f and g for the moment of R about the y-axis.

(b) Set up a definite integral with respect to x in terms of f and g for the moment of R about the x-axis.

In Exercises 28 to 31 find (a) the moment of R about the y-axis, (b) the moment of R about the x-axis, (c) the area of R, (d) $\overline{x}$, (e) $\overline{y}$. Assume the density is 1. (See Exercise 27.)

28. R is bounded by the curves $y = x^2$, $y = x^3$.

29. R is bounded by $y = x$, $y = 2x$, $x = 1$, and $x = 2$.

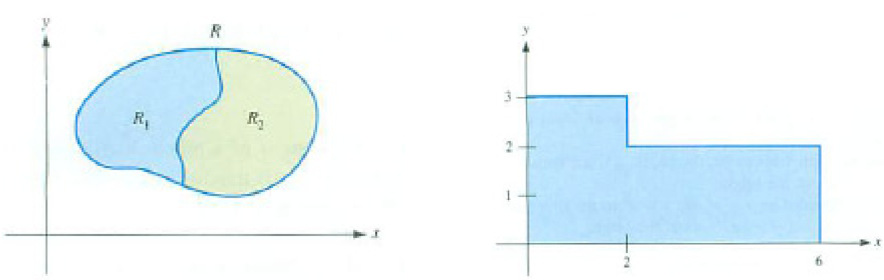
30. R is bounded by the curves $y = 3^x$ and $y = 2^x$ between $x = 1$ and $x = e$.

31. (Use a table of integrals or techniques from Chapter 8.) R is bounded by the curves $y = x - 1$ and $y = \ln(x)$, between $x = 1$ and $x = e$.

32. If R is the region below $y = f(x)$ and above $[a, b]$, show that

$$\overline{x} = \frac{\int_a^b x f(x)\, dx}{\text{Area of } R}.$$

33. A plane distribution of matter occupies the region R. It is cut into two pieces, occupying regions R_1 and R_2, as in Figure 17.8.11(a). The part in R_1 has mass M_1 and centroid $(\overline{x}_1, \overline{y}_1)$. The part in R_2 has mass M_2 and centroid $(\overline{x}_2, \overline{y}_2)$. Find the centroid $(\overline{x}, \overline{y})$ of the entire mass. Express it in terms of M_1, M_2, $\overline{x}_1$, $\overline{x}_2$, $\overline{y}_1$, and $\overline{y}_2$.]



(a) (b)

Figure 17.8.11

34. Let R, R_1, and R_2 be as in Exercise 33. Show that the centroid of R lies on the line segment joining the centroids of R_1 and R_2.

35. Use the formula in Exercise 33 to find the center of mass of the homogeneous lamina shown in Figure 17.8.11(b).

In Exercises 36 and 37 find $\overline{z}$ for the surfaces.

36. The portion of the paraboloid $2z = x^2 + y^2$ below the plane $z = 9$.

37. The portion of the plane $x + 2y + 3z = 6$ above the triangle in the xy plane whose vertices are $(0,0)$, $(4,0)$, and $(0,1)$.

38. This exercise shows that the medians of a triangle meet at the centroid of the triangle. (A **median** of a triangle is a line that passes through a vertex and the midpoint of the opposite edge.)

Let R be a triangle with vertices A, B, and C. Introduce an xy-coordinate system such that the origin is at A and B lies on the x-axis, as in Figure 17.8.12.

(a) Compute $(\overline{x}, \overline{y})$.

(b) Find the equation of the median through C and M, the midpoint of AB.

(c) Verify that the centroid lies on the median computed in (b).

(d) Why would you expect the centroid to lie on each median? (Use physical intuition.)

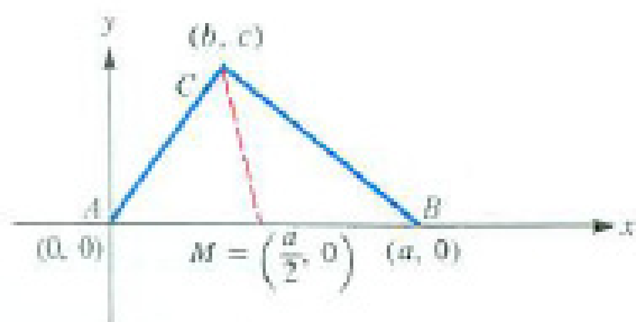
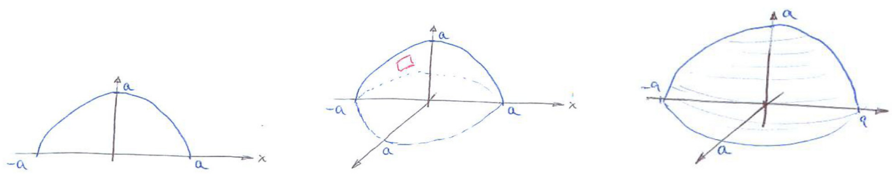


Figure 17.8.12

39. Which do you think would have the highest centroid? The semicircular wire of radius a, shown in Figure 17.8.13(a); the top half of the surface of a ball of radius a, shown in Figure 17.8.13(b); or the top half of a ball of radius a, shown in Figure 17.8.13(c). Give your opinion, with reason.



(a) (b) (c)

Figure 17.8.13

40. Using calculus, determine the highest centroid in Exercise 39.

41. The corners of a homogeneous triangular piece of metal are $(0, 0)$, $(1, 0)$, and $(0, 2)$.

(a) Is the line $y = \frac{11}{5}x$ a balancing line?

(b) If not, if the metal rests on the line which way would it rotate?

DEFINITION *(Section of a region)* Let R be a convex set in the plane. A **section** of R is a part of R that is bounded by a chord and part of the the boundary, as shown as Figure 17.8.14

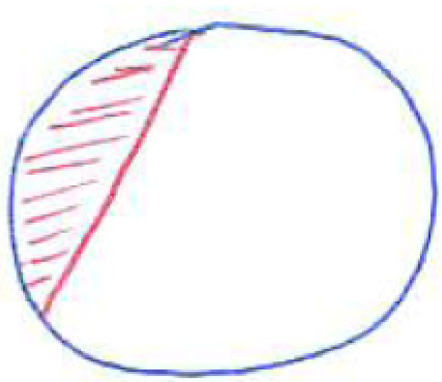


Figure 17.8.14

42. If R is a convex set in the plane, show that different sections have different centers of gravity.

43. (See Exercise 42.) Do you think every point in R that is not on the boundary is the center of mass of some section of R? Why?

44. Archimedes (287-212 B.C.) investigated the centroid of a section of a parabola. A section of the parabola $y = x^2$ is shown in Figure 17.8.15. M is the midpoint of the chord and N is the point on the parabola directly below M.

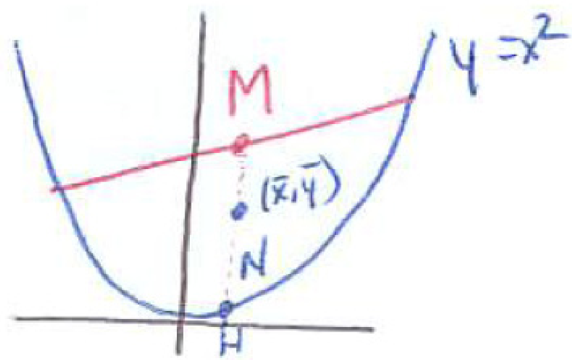


Figure 17.8.15

He showed, without calculus, that the centroid is on the line MN, three-fifths of the way from N and M. Obtain his result with calculus.

45. (See Exercise 44.) Is every point in the region bounded by the parabola the centroid of some section?

46. (See Exercise 6.) The plane $z = c$ in Exercise 6 is perpendicular to the axis of the paraboloid. Archimedes was also interested in the case when the plane is not perpendicular to the axis. Find the centroid of the region below the tilted plane $z = ky$ and above the paraboloid $z = x^2 + y^2$.

Exercises 47 to 49 concern Pappus's Theorem, which relates the volume of a solid of revolution to the centroid of the planar region R that is revolved to form the solid.

Theorem 17.8.1 (Pappus)**.** *Let R be a region in the plane and L a line in the plane that does not cross R, though it can touch R at its border. Then the volume of the solid formed by revolving R about L is equal to the product*

$$(\textit{Distance the centroid of } R \textit{ is rotated}) \cdot (\textit{Area of } R)\,.$$

47.

(a) Prove Pappus's Theorem

(b) Use Pappus's Theorem to find the volume of the torus formed by revolving a disk of radius 3 inches about a line in the plane of the disk and 5 inches from its center.

48. Use Pappus's Theorem to find the centroid of the half disk R of radius a.

49. Use Pappus's Theorem to find the centroid of the right triangle in Figure 17.8.16.

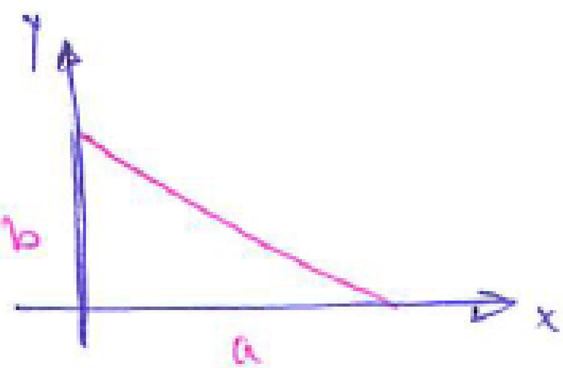


Figure 17.8.16

50. This exercise concerns hydrostatic pressure. (See Section 7.6.)

(a) Show that the pressure of water against a submerged, vertical planar surface occupying the plane region R equals the pressure at the centroid of R times the area of R.

(b) Is the assertion in (a) correct if R is not vertical?

51. Let R be a region in a plane and P a point a distance $h > 0$ from the plane. P and R determine a cone with base R and vertex P, as shown in Figure 17.8.17. Let the area of R be A. What can be said about the distance from the centroid of the cone to the plane of R?

(a) What is the distance in the case of a right circular cone?

(b) Experiment with another cone with a convenient base of your choice.

(c) Make a conjecture.

(d) Explain why it is true.

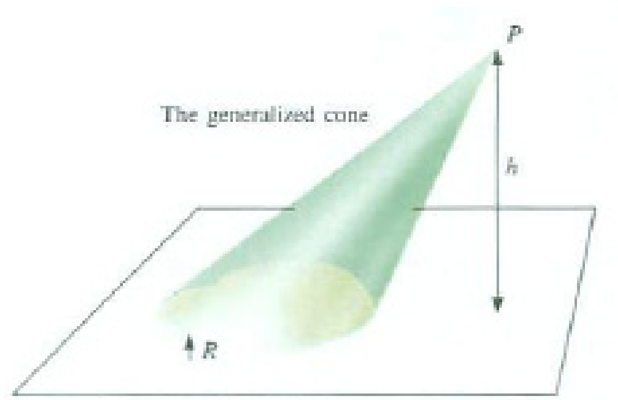


Figure 17.8.17

52. (Contributed by Jeff Lichtman.) Let f and g be two continuous functions such that $f(x) \geq g(x) \geq 0$ for x in $[0, 1]$. Let R be the region under $y = f(x)$ and above $[0, 1]$ and let R^* be the region under $y = g(x)$ and above $[0, 1]$.

(a) Do you think the center of mass of R is at least as high as the center of mass of R^*? (Give your opinion, without any supporting calculations.)

(b) Let $g(x) = x$. Define $f(x)$ to be $\frac{1}{3}$ for $0 \leq x \leq \frac{1}{3}$ and to be x if $\frac{1}{3} \leq x \leq 1$. ($f$ is continuous.) Find $\bar{y}$ for R and also for R^*. (Which is larger?)

(c) Let a be a constant, $0 \leq a \leq 1$. Let $f(x) = a$ for $0 \leq x \leq a$, and let $f(x) = x$ for $a \leq x \leq 1$. Find $\bar{y}$ for R.

(d) Show that the number a for which $\bar{y}$ defined in (c) is a minimum is a root of $x^3 + 3x - 1 = 0$.

(e) Show that the equation in (d) has only one real root q.

(f) Find q to four decimal places.

(g) Show that $\bar{y} = q$

53. (Contributed by Jeff Lichtman) Let f and g be two continuous functions such that $f(x) \geq g(x) \geq 0$ for x in $[0, 1]$. Let R be the region under $y = f(x)$ and above $[0, 1]$; let R^* be the region under $y = g(x)$ and above $[0, 1]$.

(a) Do you think the centroid of R is at least as high as the center of mass of R^*? (An opinion only.)

(b) Let $g(x) = x$. Define $f(x)$ to be $\frac{1}{3}$ for $0 \leq x \leq \frac{1}{3}$ and $f(x)$ to be x if $\frac{1}{3} \leq x \leq 1$. (Note that f is continuous.) Find $\overline{y}$ for R and also for R^*. (Which is larger?)

(c) Let a be a constant, $0 \le a \le 1$. Let $f(x) = a$ for $0 \le x \le a$ and let $f(x) = x$ for $a \le x \le 1$. Find $\overline{y}$ for R.

(d) Show that the number a for which $\overline{y}$ defined in part (c) is a minimum is a root of the equation $x^3 + 3x - 1 = 0$.

(e) Show that the equation in (d) has only one real root q.

(f) Find q to four decimal places.

54. Consider the parabolic surface $z = x^2 + y^2$ below the plane $z = a^2$.

(a) Set up a double integral in the xy-plane for the moment about the xy-plane.

(b) Express it as an iterated integral in polar coordinates.

(c) Evaluate the integral.

(d) Find the centroid of the surface.

17.9 The Jacobian and Multiple Integrals

SHERMAN: Woody did not see this section intro. Please read closely.

In this section the Jacobian, developed in Section 16.8, is used to evaluate multiple integrals by replacing the domain of integration by a simpler domain. This is similar to the way we replaced an integral over a curve with an integral over an interval (replacing ds by $\frac{ds}{dx}\,dx$, in Section 9.4) and an integral over a surface by an integral over a planar region (replacing dS by $\frac{1}{|\cos(\gamma)|}$ in Section 18.7).

The Jacobian Enters the Integral

Let F be a mapping from a region $\mathcal{R}$ in uv-space to the region $\mathcal{S}$ in xy-space and let $f(P)$ be a scalar function defined on $\mathcal{S}$. We will express the integral $\int_{\mathcal{S}} f(P)\,dS$ as a multiple integral over $\mathcal{R}$. If $\mathcal{R}$ is simpler than $\mathcal{S}$, it may be easier to compute the integral over $\mathcal{R}$ than the integral over $\mathcal{S}$.

For a point P in $\mathcal{S}$, let Q be the point in $\mathcal{R}$ such that $F(Q) = P$, as shown in Figure 17.9.1. We can form an approximating sum for $\int_{\mathcal{S}} f(P)\,dS$ indirectly,

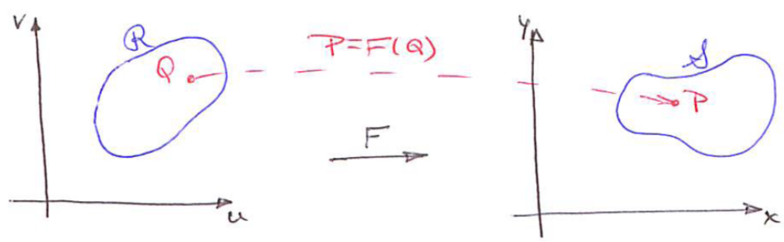


Figure 17.9.1

as follows.

Partition $\mathcal{R}$ into n small patches $\mathcal{R}_1$, $\mathcal{R}_2$, …, $\mathcal{R}_n$. Then $\mathcal{S}_1 = F(\mathcal{R}_1)$, $\mathcal{S}_2 = F(\mathcal{R}_1)$, …, $\mathcal{S}_n = F(\mathcal{R}_n)$ is a partition of $\mathcal{S}$. Pick points Q_1 in $\mathcal{R}_1$, Q_2 in $\mathcal{R}_2$, …, Q_n in $\mathcal{R}_n$. Let $P_1 = F(Q_1)$, $P_2 = F(Q_2)$, …, $P_n = F(Q_n)$, as shown in Figure 17.9.2. Let the area of $\mathcal{S}_i$ be S_i.

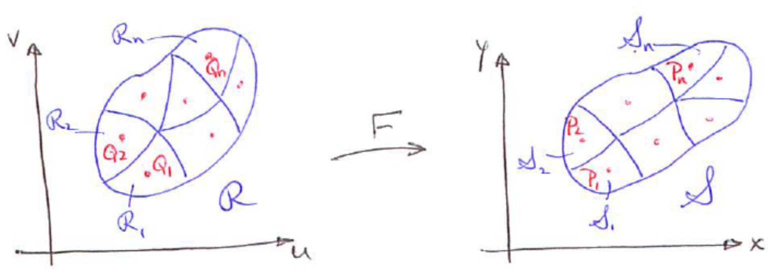


Figure 17.9.2

Then

$$\sum_{i=1}^{n} f(P_i)S_i = \sum_{i=1}^{m} f(F(Q_i))\ S_i \tag{17.9.1}$$

is an approximation of $\int_{\mathcal{S}} f(P)\ dS$.

Let the area of $\mathcal{R}_i$ be A_i. Let $J[F](Q)$ be the Jacobian of F evaluated at Q. Since the Jacobian records local magnification, $|J[F](Q_i)|A_i$ is an approximation of S_i. In view of (17.9.1),

SHERMAN: There are too many () and [] here. Instead of $J[F](Q)$ or $J(F)(Q)$, why not $J_F(Q)$? Stewart uses $\frac{\partial(x,y,z)}{\partial(u,v,w)}$, which I have to admit is more suggestive of the computation.

$$\sum_{i=1}^{n} f(F(Q_i))|J[F](Q_i)|A_i \tag{17.9.2}$$

is an approximation of $\int_{\mathcal{S}} f(P)\ dS$. But (17.9.2) is also an approximation of

$$\int_{\mathcal{R}} f(F(Q))\ |J[F](Q)|\ dA.$$

Taking limits as all $\mathcal{R}_i$ are chosen smaller, we conclude that

Change of Domain

$$\int_{\mathcal{S}} f(P)\ dS = \int_{\mathcal{R}} f(F(Q))|J[F](Q)|\ dA \tag{17.9.3}$$

Equation (17.9.3) says we can replace an integral over $\mathcal{S}$ by an integral over $\mathcal{R}$. The equation remains valid even if F is not one-to-one on the boundary of $\mathcal{R}$.

The notation in (17.9.3) is precise but forbidding. In shorthand, it is summarized in the equation $dS = J(F)\ dA$.

Applying the Idea

EXAMPLE 1 Let $\mathcal{S}$ be the parallelogram bounded by the lines $x + y = 1$, $x + y = 4$, $y - 2x = 2$, and $y - 2x = 3$. Evaluate $\int_{\mathcal{S}} x^2\ dS$.

SOLUTION The set $\mathcal{S}$ is shown in Figure 17.9.3.

Evaluating $\int_{\mathcal{S}} x^2\ dS$ by iterated integrals would require breaking $\mathcal{S}$ into two triangles and a parallelogram. Instead, let us change the domain.

By Example 2 in Section 16.8, $\mathcal{S}$ is the image of the rectangle $\mathcal{R}$ in uv-space described by $1 \leq u \leq 4$, $2 \leq v \leq 3$ by the mapping

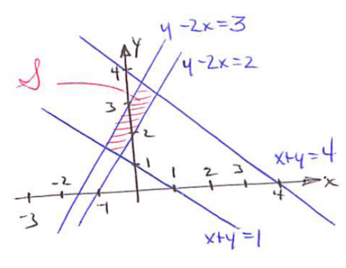


Figure 17.9.3

$$F(u,v) = \left(\frac{u-v}{3}, \frac{2u+v}{3}\right),$$

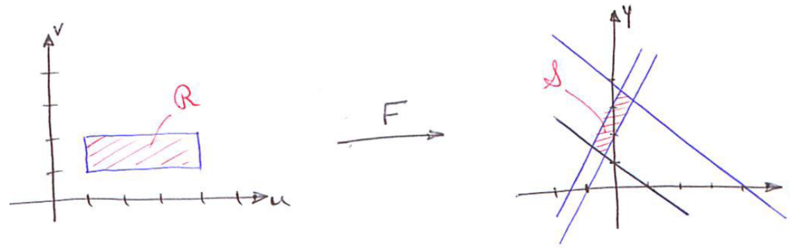


Figure 17.9.4

as shown in Figure 17.9.4. If $F(u,v) = (x,y)$, we have $x = (u-v)/3$ and $y = (2u+v)/3$. The Jacobian of F is

SHERMAN: Woody suggests not mixing $\det(\ )$ and $|\ |$. I have chosen to use $|\ |$, but don't want this to be confused with absolute values, which are needed around the Jacobian. Here, Stewart has some problems because he Jacobian is defined as a determinant, for which he uses vertical bars to delimit the matrix. Would it be enough to include a remark alerting the reader to this possible confusion?

$$\begin{vmatrix} \frac{\partial\left(\frac{u-v}{3}\right)}{\partial u} & \frac{\partial\left(\frac{2u+v}{3}\right)}{\partial u} \\ \frac{\partial\left(\frac{u-v}{3}\right)}{\partial v} & \frac{\partial\left(\frac{2u+v}{3}\right)}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{-1}{3} & \frac{1}{3} \end{vmatrix} = \frac{1}{9} + \frac{2}{9} = \frac{1}{3}.$$

In this case, the Jacobian is constant.

Then $\int_{\mathcal{S}} x^2 \, dS = \int_{\mathcal{R}} \left(\frac{u-v}{3}\right)^2 \left(\frac{1}{3}\right) \, dA$.

The latter integral is easily evaluated as an iterated integral in which u and v are the variables:

$$\int\limits_{\mathcal{R}} \left(\frac{u-v}{3}\right)^2 \left(\frac{1}{3}\right) \, dA = \frac{1}{27} \int\limits_{\mathcal{R}} (u-v)^2 \, dA = \frac{1}{27} \int\limits_1^4 \left(\int\limits_2^3 (u-v)^2 \, dv \right) \, du.$$

The first integration is

$$\int\limits_2^3 (u-v)^2 \, dv = \left. \frac{-(u-v)^3}{3} \right|_{v=2}^{v=3} = \frac{-(u-3)^3}{3} - \frac{-(u-2)^3}{3} = \frac{-(u-3)^3}{3} + \frac{(u-2)^3}{3}.$$

The second integration is

$$\int\limits_1^4 \left(\frac{-(u-3)^3}{3} + \frac{(u-2)^3}{3} \right) \, du = \left. \frac{-(u-3)^4}{12} + \frac{(u-2)^4}{12} \right|_{u=1}^{u=4}$$

$$= \left(\frac{-(1)^4}{12} + \frac{2^4}{12} \right) - \left(\frac{-(-2)^4}{12} + \frac{(-1)^4}{12} \right) = \left(-\frac{1^4}{12} + \frac{2^4}{12} \right) - \left(-\frac{2^4}{12} + \frac{1^4}{12} \right) = \frac{30}{12} =$$

Thus $\int_{\mathcal{S}} x^2 \, dS = \frac{1}{27} \cdot \frac{5}{2} = \frac{5}{54}$. ◇

EXAMPLE 2 Let $\mathcal{S}$ be the region in xy-space bounded by the circles of radii 1 and 2 with centers at the origin, and by the lines $y = x$ and $y = x/2$.

Find $\int_{\mathcal{S}}(x^2+y^2)\ dS$.
SOLUTION The region $\mathcal{S}$ appeared in Example 3 of Section 16.8. As we saw, it is the image of the rectangle $\mathcal{R}$ in the uv-plane described by $1 \le u \le 2$, $\pi/6 \le v \le \pi/4$ by the mapping $F(u,v) = (u\cos(v), u\sin(v))$. Then

$$\int_{\mathcal{S}} (x^2+y^2)\ dS = \int_{\mathcal{R}} \left((u\cos(v))^2 + (u\sin(v))^2\right) |J[F](u,v)|\ dA.$$

The Jacobian of F is

$$\begin{vmatrix} \frac{\partial(u\cos(v))}{\partial u} & \frac{\partial(u\sin(v))}{\partial u} \\ \frac{\partial(u\cos(v))}{\partial v} & \frac{\partial(u\sin(v))}{\partial v} \end{vmatrix} = \begin{vmatrix} \cos(v) & \sin(v) \\ -u\sin(v) & u\cos(v) \end{vmatrix} = u.$$

Also, $(u\cos(v))^2 + (u\sin(v))^2 = u^2$, so we have

$$\int_{\mathcal{S}} (x^2+y^2)\ dS = \int_{\mathcal{R}} \underbrace{u^2}_{\text{function}} \cdot \underbrace{u}_{\text{Jacobian}}\ dA = \int_{\mathcal{R}} u^3\ dA.$$

We see:

$$\int_{\mathcal{R}} u^3\ dA = \int_{\pi/6}^{\pi/4} \left(\int_1^2 u^3\ du \right)\ dv = \int_{\pi/6}^{\pi/4} \frac{15}{4}\ dv = \frac{15}{4}\left(\frac{\pi}{4} - \frac{\pi}{6}\right) = \frac{15}{48}\pi.$$

◇

In this example the Jacobian introduced the extra factor u in the integrand. But u and v are the same as the r and θ of polar coordinates. The extra factor r was introduced in Section 17.3 because the area of a small patch is $r\Delta r\Delta\theta$, not simply $\Delta r\Delta\theta$.

Exercise 23 in Section 16.8 develops the Jacobian for a mapping from uvw-space to xyz-space. We can check that the Jacobian associated with cylindrical coordinates (r,θ,z) where

$$x = r\cos(\theta), \qquad y = r\sin(\theta), \qquad z = z$$

is r, which then must be introduced as a factor in the integrand.

The Jacobian associated with spherical coordinates (ρ,θ,ϕ), where

$$x = \rho\sin(\phi)\cos(\theta), \qquad y = \rho\sin(\phi)\sin(\theta), \qquad z = \rho\cos(\phi),$$

is $\rho^2\sin(\phi)$, which is consistent with what we found in Section 17.6 by considering the volume of a small patch corresponding to changes $\Delta\rho$, $\Delta\phi$, $\Delta\theta$ in the coordinates.

Summary

A mapping F from $\mathcal{R}$ to $\mathcal{S}$ enables us to replace an integral over $\mathcal{S}$ by an integral over $\mathcal{R}$:

$$\int_{\mathcal{S}} f(P)\ dS = \int_{\mathcal{R}} f(F(Q))\ |J(F)(Q)|\ dA.$$

The Jacobian appears because it tells by how much F magnifies the area of a small patch in $\mathcal{R}$. The factors that we introduced in earlier sections into integrands, r for polar and cylindrical coordinates and $\rho^2 \sin(\phi)$ for spherical coordinates, are instances of Jacobians.

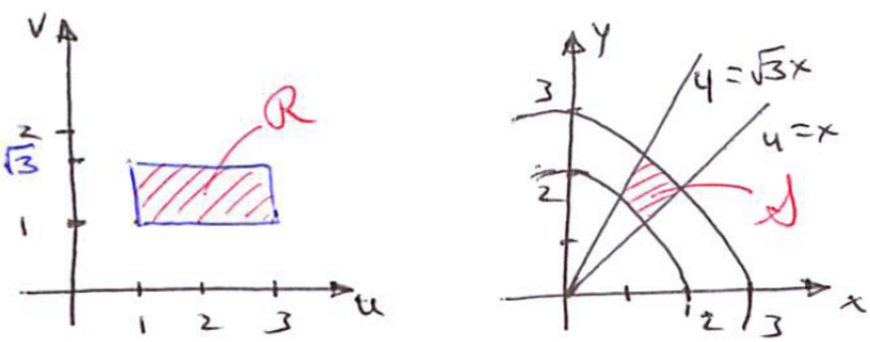
EXERCISES for Section 17.9

1. State the relation between an integral over $\mathcal{R}$ and an integral over its image $\mathcal{S}$ by a mapping F, $\mathcal{S} = F(\mathcal{R})$. Use as few symbols as you can.

2. State, using as few symbols as you can, why the Jacobian appears in the integrand on the right-hand side of $\int_{\mathcal{S}} f(P)\ dS = \int_{\mathcal{R}} f(F(Q))|J[F](Q)|\ dA$. Start your explanation from an approximating sum.

In Exercises 3 to 6, construct a mapping F from $\mathcal{R}$ to $\mathcal{S}$, and use it to evaluate the given integral.

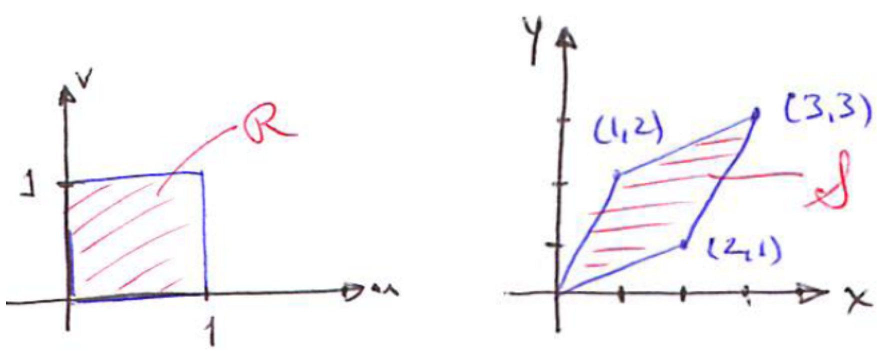
3. $\int_{\mathcal{S}} x^2\ dS$.



(a) (b)

Figure 17.9.5 $\mathcal{S} = F(\mathcal{R})$ for Example 3.

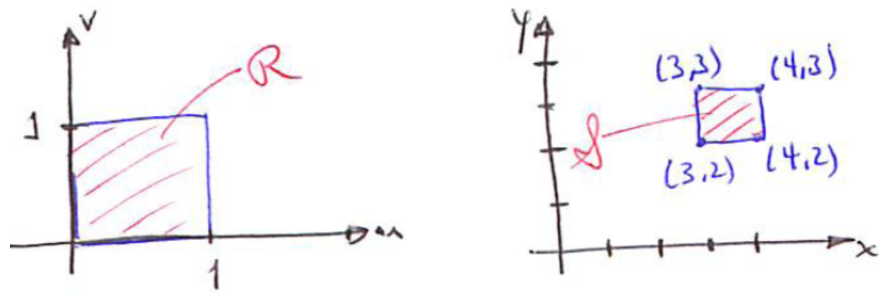
4. $\int_{\mathcal{S}} (x + y)\ dS$.



(a) (b)

Figure 17.9.6 $\mathcal{S} = F(\mathcal{R})$ for Example 4.

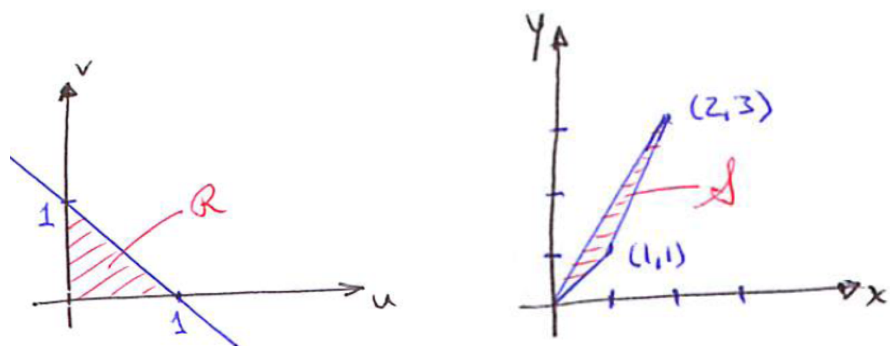
5. $\int_{\mathcal{S}} xy\ dS$.



(a) (b)

Figure 17.9.7 $\mathcal{S} = F(\mathcal{R})$ for Example 5.

6. $\int_{\mathcal{S}} \cos(x)\ dS$.



(a) (b)

Figure 17.9.8 $\mathcal{S} = F(\mathcal{R})$ for Example 6.

7. Let $\mathcal{S}$ be the elliptical region $x^2/25 + y^2/16 \leq 1$.

(a) Sketch $\mathcal{S}$.

(b) Find a mapping F from the disk $\mathcal{R}$ described by $u^2 + v^2 \leq 1$ to $\mathcal{S}$.

(c) Use it to evaluate $\int_{\mathcal{S}} \sin^2(x)\ dS$.

(d) Do enough of the direct calculation of $\int_{\mathcal{S}} \sin^2(x)\ dS$ without using a mapping to see that it is more complicated than the method in (c).

8.

(a) Find a linear mapping F from the uv-plane to the xy-plane such that the image of the parallelogram with vertices $(0,0)$, $(1,0)$, $(1,1)$, and $(0,1)$ is the square $\mathcal{S}$ with vertices $(0,0)$, $(3,2)$, $(5,-1)$, and $(2,-3)$.

(b) Use it to evaluate $\int_{\mathcal{S}} xy\ dS$.

9. It is plausible that if the local magnification of a mapping has the constant value k, then it would magnify all areas by that factor. However, there is nothing in the definition of the local magnification that assures us that this is so. After all, the definition involves a limit of quotients that need not equal k. However, a mapping that has a constant local magnification k does magnify all areas by the factor k. Justify this claim.

10.

(a) Use the mapping in Example 1 to find the area of the parallelogram in that example.

(b) Find the area of the parallelogram by using the cross product.

11. The Jacobian of a mapping F at (u, v) is u^2v. Find the area of the image of the rectangle $1 \leq u \leq 2$, $3 \leq v \leq 5$.

12. (See Exercise 14 in Section 16.8.) Show that an integral of the form $\int_a^b f(x)\ dx$ can be replaced by an integral of the form $\int_0^1 g(u)\ du$.

13. Mappings $F(u, v) = (au + bv + e, cu + dv + f)$, with $ad - bc \neq 0$, are **affine mappings**. Note that $F(0, 0) = (e, f)$. The linear mappings are the affine mappings that map the origin in the uv-plane to the origin in the xy-plane. Show that if $\mathcal{S}$ is a triangle in the xy-plane there is an affine mapping from the triangle $\mathcal{R}$ in the uv-plane whose vertices are $(0, 0)$, $(1, 0)$, and $(0, 1)$ onto $\mathcal{S}$. That implies that an integral over a triangle can be replaced by an integral over the fixed triangle $\mathcal{R}$.

14.

SHERMAN: The last parts of the E 14 have changed since Woody reviewed it. The old one pointed out that the new integral would be improper, with an unbounded integrand. This version does not give so many hints. Which do you want to use?

Sam: I can even use a mapping to get rid of improper integrals.

Jane: Another of your tricks.

Sam: Say I had $\int_0^\infty e^{-x^2}\ dx$. The mapping $x = \tan(u)$ sends the interval $[0, \pi/2)$ onto the infinite interval $[0, \infty)$. Since $dx = \sec^2(u)\ du$, it follows that that integral equals $\int_0^{\pi/2} e^{-\tan^2(u)} \sec^2(u)\ du$.

Jane: Very impressive. Surely something is wrong.

Is Sam's claim correct for a change?

15. Let F be described by $x = u^2 - v^2$, $y = uv$.

(a) Find the Jacobian of F.

(b) Let $\mathcal{R}$ be the square with vertices $(0,0)$, $(1,0)$, $(1,1)$, and $(0,1)$ and let $\mathcal{S}$ be $F(R)$. Sketch $\mathcal{S}$. (Sketch the images of the edges of $\mathcal{R}$.)

(c) Evaluate $\int_{\mathcal{S}} y^2\, dS$ using the Jacobian.

16.

(a) Sketch the set $\mathcal{S}$ in xyz-space given by $x^2 + y^2/9 + z^2/16 \leq 1$. (It is called an ellipsoid, and resembles a football.)

(b) Find its volume by using a mapping from a sphere $\mathcal{R}$ of radius 1, whose volume is $4\pi/3$.

17. This exercise uses the Jacobian to find the formula for the area of a surface of revolution. In Section 9.5 we developed such a formula. Let f be a positive function defined for x in $[a,b]$. The graph of f is rotated around the x-axis to produce a surface of revolution, $\mathcal{S}$.

(a) Sketch $\mathcal{S}$.

(b) Show on the sketch that each point (x,y,z) on $\mathcal{S}$ is determined by x and an angle θ, $0 \leq \theta \leq 2\pi$.

(c) Express (x,y,z) in terms of x and θ.

(d) Part (c) describes a mapping from the rectangle $a \leq x \leq b$, $0 \leq \theta \leq 2\pi$. Find its Jacobian.

(e) Use the Jacobian to show that the area of $\mathcal{S}$ is $2\pi \int_a^b \sqrt{1 + (f'(x))^2}\, dx$.

We have concentrated on using mappings to simplify the domain of integration. They may also simplify an integrand, as the next two exercises show.

18. Consider $\int_{\mathcal{S}} \exp((y-x)/(y+x))\, dS$ where $\mathcal{S}$ is the triangle bounded by the lines $x=0$, $y=0$, and $x+y=1$. The substitution $u = x+y$, $v = x-y$ simplifies the integrand.

(a) Sketch $\mathcal{S}$ and the set $\mathcal{R}$ in uv-space that is the image of $\mathcal{S}$ under the mapping $u = x+y$, $v = y-x$.

(b) Find the mapping F from $\mathcal{R}$ to $\mathcal{S}$ that is the inverse of the mapping in (a).

(c) Use F to evaluate the integral. Choose the iterated integral wisely.

19. Evaluate $\int_{\mathcal{S}}(x+y)^2\cos^2(x-y)\,dS$, where $\mathcal{S}$ is the square with vertices $(\pi/2, 0)$, $(\pi, \pi/2)$, $(\pi/2, \pi)$, and $(0, \pi/2)$.

17.S Chapter Summary

SHERMAN: Do these sums need limits? At least i? Generic $i = 0$ (or 1) and n? Or, are the OK as is, just as the generic "limits" is used. Woody did not object, but I just want to be sure we agree.
SHERMAN: OK to omit interior dividers?

This chapter generalized the notion of a definite integral over an interval to integrals over plane sets, surfaces, and solids. The definitions of multiple integrals are almost the same, the integral of $f(P)$ over a set being the limit of sums of the form $\sum f(P_i)\ \Delta A_i$, $\sum f(P_i)\ \Delta S_i$, or $\sum f(P_i)\ \Delta V_i$ for integrals over plane sets, surfaces, or solids, respectively.

Three different interpretations of double integrals are given in Table 17.S.1. The only difference is the interpretation of the integrand.

Integral	Integrand	Interpretation
$\int_R 1\ dA$	1	Area of R
$\int_R \sigma(P)\ dA$	$\sigma(P) =$ density per unit area	Mass of R
$\int_R c(P)\ dA$	$c(P) =$ length of cross section of solid	Volume of R

Table 17.S.1 Three interpretations of

Average value extends easily to functions of several variables. For instance, if $f(P)$ is defined on a plane region R, its average value over R is defined as

$$\frac{1}{\text{area of } R}\int\limits_R f(P)\ dA.$$

Some multiple integrals (also known as double or triple integrals) can be calculated by repeated integrations over intervals, that is, as iterated integrals. This requires a description of the region in a coordinate system and replaces dA or dV by an expression based on the area or volume of a small patch swept out by small changes in the coordinates, as recorded in Table 17.S.2.

The final section showed the role of the Jacobian determinant in replacing an integral over one set by an integral over another. The absolute value of the Jacobian is the magnification factor that is inserted into an integrand if the computation uses coordinates other than rectangular: r for polar coordinates and $\rho^2\sin(\phi)$ for spherical coordinates. That is, if F is a mapping from $\mathcal{R}$ to $\mathcal{S}$, $\int_{\mathcal{S}} f(P)\ dS$ equals $\int_{\mathcal{R}} f(F(Q))|J[F](Q)|\ dA$.

The equation holds whether $\mathcal{R}$ and $\mathcal{S}$ are solids, surfaces, or intervals. The last case was called integration by substitution in Section 8.2.

An integral over a surface S, $\int_S f(P)\ dS$, can often be replaced by an integral over the projection of S onto a plane R, replacing dS by $dA/\cos(\gamma)$ where γ is the angle between a normal to S and a normal to R.

Coordinate System	Substitution
Rectangular (2-D)	$dA = dx\ dy$
Rectangular (3-D)	$dV = dx\ dy\ dz$
Polar	$dA = r\ dr\ d\theta$
Cylindrical	$dV = r\ dr\ d\theta\ dz$
Cylindrical (surface)	$dS = r\ d\theta\ dz$
Spherical	$dV = \rho^2 \sin(\phi)\ d\phi\ d\rho\ d\theta$
Spherical (surface)	$dS = \rho^2 \sin(\phi)\ d\phi\ d\theta$

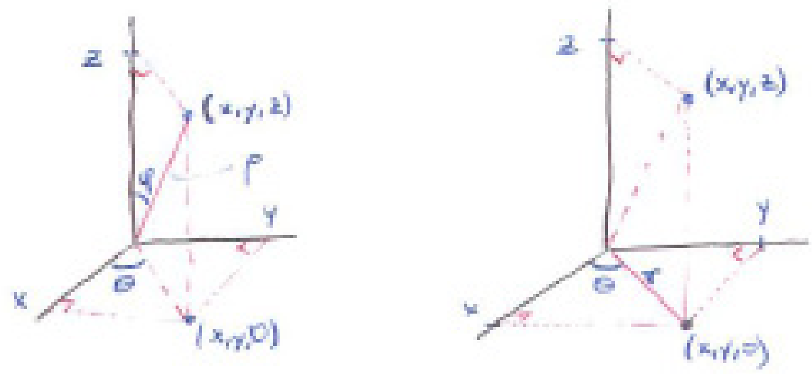
Table 17.S.2

Key Facts

If density is 1, the center of mass is called the **centroid**.

Formula	Significance
$\int_R 1\ dA$	Area of R
$\int_R 1\ dV$	Volume of R
$\frac{\int_R f(P)\ dA}{\text{Area of } R}$ or $\frac{\int_R f(P)\ dV}{\text{Volume of } R}$	Average value of f over R
$\int_R \sigma(P)\ dA$ or $\int_R \delta(P)\ dV$	Total mass of R, M (σ and δ denote density)
$\int_R y\sigma(P)\ dA$, $\int_R x\sigma(P)\ dA$	Moments, M_x and M_y about x and y axes, respectively. (A moment can be computed around any line in the plane.)
$\int_R f(P)\sigma(P)\ dA$, $\int_R f(P)\sigma(P)\ dV$ where $f(P)$ is the square of the distance from P to some fixed line L	Moment of inertia around L for planar and solid regions, respectively.
$\int_R x^2\sigma(P)\ dA$, $\int_R y^2\sigma(P)\ dA$	Second moments, M_{xx} and M_{yy} about x and y axes, respectively.
$\left(\frac{M_y}{M}, \frac{M_x}{M}\right)$	Center of mass, $(\bar{x}, \bar{y})$
$\int_R z\delta(P)\ dV$	Moment M_{xy}
$\int_R y\delta(P)\ dV$	Moment M_{xz}
$\int_R x\delta(P)\ dV$	Moment M_{yz}
$\left(\frac{M_{yz}}{M}, \frac{M_{xz}}{M}, \frac{M_{xy}}{M}\right)$	Center of mass of solid, $(\bar{x}, \bar{y}, \bar{z})$

Table 17.S.3



(a) (b)

Figure 17.S.1

EXERCISES for 17.S

1. In each of the following conversions from one integral to another, what is the "local magnification" that is involved?

(a) Using substitution to replace an integral over $[a, b]$ by an integral over $[c, d]$. (Section 8.2)

(b) Replacing an integral over a curve by an integral over an interval. (Section 15.3)

(c) Replacing an integral over a surface that is the graph of $z = f(x, y)$ by an integral over its projection on the xy-plane. (Section 17.7)

(d) Replacing a double integral in rectangular coordinates by one in polar coordinates. (Section 17.3)

(e) Using spherical coordinates to integrate over the a ball or solid cone. (Section 17.6)

(f) Using spherical coordinates to integrate over surface of a sphere or cone. (Section 17.7)

2. The temperature at (x, y) at time t is $T(x, y, t) = e^{-tx}\sin(x + 3y)$. Let $f(t)$ be the average temperature in the rectangle $0 \le x \le \pi$, $0 \le y \le \pi/2$ at time t. Find df/dt.

3. Let f be a function such that $f(-x, y) = -f(x, y)$.

(a) Give some examples of such functions.

(b) For what type regions R in the xy plane is $\int_R f(x,y)\ dA$ certainly equal to 0?

4. Find $\int_R (2x^3y^2 + 7)\ dA$ where R is the square with vertices $(1,1)$, $(-1,1)$, $(-1,-1)$, and $(1,-1)$. Do this with as little work as possible.

5. Let $f(x,y)$ be a continuous function. Define $g(x)$ to be $\int_R f(P)\ dA$, where R is the rectangle with vertices $(3,0)$, $(3,5)$, $(x,0)$, and $(x,5)$, $x > 3$. Express dg/dx as a suitable integral.

6. Let R be a plane lamina in the shape of the region bounded by the graph of the function $r = 2a\sin(\theta)$ $(a > 0)$. If the variable density of the lamina is given by $\sigma(r,\theta) = \sin(\theta)$, find the centroid R.

In Exercises 7 and 8, use iterated integrals in polar coordinates to find the point.

7. The centroid of the region within the cardioid $r = 1 + \cos(\theta)$.

8. The centroid of the region within the leaf of $r = \cos 3(\theta)$ that lies along the polar axis.

In Exercises 9 to 12 find the moment of inertia of a homogeneous lamina of mass M of the given shape, around the given line.

9. A disk of radius a, about the line perpendicular to it through its center.

10. A disk of radius a, about a line perpendicular to it through a point on the circumference.

11. A disk of radius a, about a diameter.

12. A disk of radius a, about a tangent.

13.

(a) In a diagram much larger than Figure 16.8.5 in Section 16.8, show $\mathcal{C}$ and the parallelogram that approximates it. Include the vectors $(\partial\mathbf{r}/\partial u)\Delta u$ and $(\partial\mathbf{r}/\partial v)\Delta v$.

(b) Why are the vectors in (a) tangent to the curves that meet at $F(u_0, v_0)$?

14. Let $F(u,v,w) = (u\sin(v)\cos(w), u\sin(v)\sin(w), u\cos(v))$. Let $\mathcal{R}$ in uvw-space be described by $1 \le u \le 2$, $0 \le v \le \pi/4$, $0 \le w \le \pi/2$.

(a) Sketch $\mathcal{R}$.

(b) Sketch $F(\mathcal{R})$.

(c) Find the magnification of F.

15. Let $\mathcal{S}$ be the sphere with radius a and center at the origin. We want to find $\int_{\mathcal{S}}(xz + y^2)dS$.

(a) Why is $\int_{\mathcal{S}} xzdS = 0$?

(b) Why is $\int_{\mathcal{S}} x^2dS = \int_{\mathcal{S}} y^2dS = \int_{\mathcal{S}} z^2dS$?

(c) Why is $\int_{\mathcal{S}} y^2dS = \int_{\mathcal{S}}(a^2/3)dS$?

(d) Show that $\int_{\mathcal{S}}(xz + y^2)dS = 4\pi a^2/3$.

16. Let a be a positive number and $\mathcal{R}$ the region bounded by $y = x^a$, the x-axis, and the line $x = 1$.

(a) Show that the centroid of $\mathcal{R}$ is $\left(\frac{a+1}{a+2}, \left(\frac{a+1}{a+2}\right)^a\right)$.

(b) Find $\lim_{a\to\infty} \overline{x}$ and $\lim_{a\to\infty} \overline{y}$.

(c) Show that the centroid of $\mathcal{R}$ lies in $\mathcal{R}$ for all large values of a.

It is true that the centroid lies in $\mathcal{R}$ for all positive values of a, but the proof is more difficult.

17. Define the **moment of a curve in the xy-plane around the x-axis** to be $\int_{s_1}^{s_2} y\ ds$, where s_1 and s_2 refer to the range of the arc length s. The **moment of the curve around the y-axis** is defined as $\int_{s_1}^{s_2} x\ ds$. The **centroid** of the curve, $(\overline{x}, \overline{y})$, is defined by setting

$$\overline{x} = \frac{\int_{s_1}^{s_2} x\ ds}{\text{length of curve}} \quad \text{and} \quad \overline{y} = \frac{\int_{s_1}^{s_2} y\ ds}{\text{length of curve}}.$$

Find the centroid of the top half of the circle $x^2 + y^2 = a^2$.

18. Show that the area of the surface obtained by revolving about the x-axis a curve that lies above it is equal to the length of the curve times the distance that the centroid of the curve moves. See Exercise 17.

19. Use Exercise 18 to find the surface area of the torus formed by revolving a circle of radius a around a line a distance b from its center, $b \geq a$.

20. Use Exercise 18 to find the area of the curved part of a cone of radius a and height h.

21. Let $f(P)$ and $g(P)$ be continuous functions defined on a plane region R.

(a) Show that

$$\left(\int_R f(P)g(P)\,dA\right)^2 \le \left(\int_R f(P)^2\,dA\right)\left(\int_R g(P)^2\,dA\right).$$

(Review the proof of the Cauchy-Schwarz inequality presented in the CIE on Average Speed and Class Size on page 680.)

(b) Show that if equality occurs in the inequality in (a), then f is a constant times g.

22. (Contributed by G. D. Chakerian.) A solid region $\mathcal{R}$ is bounded below by the xy-plane, above by the surface $z = f(P)$, and the sides by the surface of a cylinder, as shown in Figure 17.S.2.

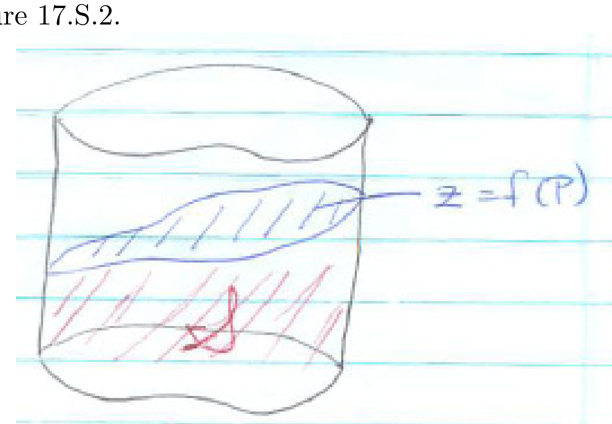


Figure 17.S.2

The volume of $\mathcal{R}$ is V. If V is fixed, show that the top surface that minimizes the height of the centroid of $\mathcal{R}$ is a horizontal plane. Water in a glass illustrates this, for nature minimizes the height of the centroid of the water. (See Exercise 21.)

Exercises 23 to 31 explore the average distance for all points on a curve or in a region. The distance from a point to a curve is the shortest distance from the point to the curve.

23. Find the average distance from points in a disk of radius a to the center of the disk.

(a) Set up the pertinent definite integral in rectangular coordinates.

(b) Set it up in polar coordinates.

(c) Evaluate the easier integral in (a) and (b).

24. Find the average distance from points in a square of side a to the center of the square.

(a) Set up the pertinent definite integral in rectangular coordinates.

(b) Set it up in polar coordinates.

(c) Evaluate the easier integral in (a) and (b).

25. Find the average distance from points in a ball of radius a to the center of the ball.

(a) Set up the pertinent definite integral in rectangular coordinates.

(b) Set it up in spherical coordinates.

(c) Evaluate the easier integral in (a) and (b).

26. Find the average distance from points in a cube of side a to the center of the cube.

(a) Set up the pertinent definite integral in rectangular coordinates.

(b) Set it up in polar coordinates.

(c) Evaluate the easier integral in (a) and (b).

27. Find the average distance from points in a square of side a to the border of the square.

(a) Set up the pertinent definite integral in rectangular coordinates.

(b) Set it up in polar coordinates.

(c) Evaluate the easier integral in (a) and (b).

28. Find the average distance from the points in a disk of radius a to the circular border.

(a) Before doing any calculations, decide whether the average distance is greater than $a/2$ or less than $a/2$. Explain how you made this decision.

(b) Carry out the calculation using a convenient coordinate system.

29.

(a) Show that a region of diameter d can always fit into a disk of diameter $2d$.

(b) Can it always fit into a disk of diameter d?

30. If a region has diameter d,

(a) how small can its area be?

(b) show that area is less than or equal to $\pi d^2/2$.

31. Let A and B be two points in the xy-plane. A curve (in the xy-plane) consists of all points P such that the sum of the distances from P to A and P to B is constant, say $2a$. Consider the distance from P to A as a function of arclength on the curve. Find the average of that distance.

32. A surface is called **closed** when it is the boundary of a region R, as a balloon surrounds the air within it. A surface is called **smooth** when it has a continuous outward unit normal vector at each point of the surface. Let S be a smooth closed surface bounding R. Show that for any point P_0 in R, there are at least two points on S such that $\overrightarrow{P_0P}$ is normal to S. It is conjectured that if P_0 is the centroid of R, then there are at least four points on S such that P_0P is normal to S.

Exercises 33 to 36 concern the moment of inertia. Note that if the object is homogeneous, has mass M and volume V, its density $\delta(P)$ is M/V.

33. A homogeneous rectangular solid box has mass M and sides of lengths a, b, and c. Find its moment of inertia about an edge of length a.

34. A rectangular homogeneous box of mass M has dimensions a, b and c. Show that the moment of inertia of the box about a line through its center and parallel to the side of length a is $M(b^2+c^2)/12$.

35. A right solid circular cone has altitude h, radius a, constant density, and mass M.

(a) Why is its moment of inertia about its axis less that Ma^2?

(b) Show that its moment of inertia about its axis is $3Ma^2/10$.

36. Let P_0 be a point in a solid of mass M. Show that for all choices of three mutually perpendicular lines that meet at P_0 the sum of the moments of inertia of the solid about the lines is the same.

37. Let $\mathcal{S}_a$ be the sphere with radius a centered at the origin. Let $\mathcal{S}$ be a part of the sphere $\mathcal{S}_a$. Use the definition of steradian measure of the angle subtended by a surface to show that

$$\text{Angle } \mathcal{S} \text{ subtends} = \frac{\text{Area of } \mathcal{S}}{a^2}.$$

(The angle subtended by $\mathcal{S}_a$ is 4π.)

SHERMAN: This was added after the version reviewed by Woody. The next 2 E's are the ones you provided on an extra handwritten page. Please decide how you want these to appear.

38. Let $(\bar{x}, \bar{y})$ be the centroid of a region R computed using the moments of R around the x- and y-axes.

(a) Compute the moment of R around the line through $(\bar{x}, \bar{y})$ with slope 1.

(b) Show that the moment of R around any line through $(\bar{x}, \bar{y})$ is 0.

(c) Does the centroid depend upon the choice of axes? That is, if the moments were computed around a different pair of perpendicular axes, would the centroid still be $(\bar{x}, \bar{y})$?

39. Matter in a planar region R has density $\delta(P)$. Relative to the xy-axes its center of mass is $P = (\bar{x}, \bar{y})$. Introduce a second rectangular coordinate system with $x'y'$-axes parallel to the original system. The $x'y'$-axes are a translation of the xy-axes. The origin of the $x'y'$-axes is at the point (h, k) relative to the xy-axes. The center of gravity is $Q = (\bar{x'}, \bar{y'})$ when computed using the $x'y'$-axes. Show that $P = Q$. This shows that the center of gravity does not depend on the choice of axes, as long as they are parallel systems.

40. (This continues Exercise 39.) Assume that the center of mass is $(0, 0)$ when computed relative to the xy-axes. These axes are rotated around $(0, 0)$ by an angle θ to produce a second rectangular coordinate system $x'y'$.

(a) Show that the center of mass computed with the $x'y'$-axes is the same as that computed with the xy-axes. (Show that the x'-axis and the y'-axis are balancing lines.)

(b) From (a) and the preceding Exercise, show that the center of mass does not depend on the particular coordinate system chosen.

41. A solid of varying density $\delta(P)$ occupies the region R in space. Let L_1 be the line through its center of mass and L_2 a line parallel to L_1 and at a distance r from it. Let I_1 be the moment of inertia of the solid around L_1 and I_2 the moment of inertia around L_2.

(a) Show that $I_1 = I_2 + r^2 M$ where M is the mass of the solid.

(b) Which choice of r leads to the smallest value of I_2?

SHERMAN: The key in (b) is that $I_2 \geq 0$. Is this obvious, or is that the point? Could the wording of (b) be better?

42. Let $z = g(y)$ be a decreasing function of y such that $g(1) = 0$. Let R be the solid of revolution formed by revolving about the z-axis the region in the yz-plane bounded by $y = 0$, $z = 0$, and $z = g(y)$. Using repeated integrals in cylindrical coordinates, show that $\int_R z\, dV = \int_0^1 \pi y \left(g(y)\right)^2 \, dy$ and $\int_R z\, dV = \int_0^{g(0)} \pi z \left(g^{-1}(z)\right)^2 \, dz$.

43. (See Exercise 42.)

(a) Show that the z-coordinate of the centroid of the solid described in Exercise 42 is

$$\frac{\int_{)}^1 \frac{x}{2} \left(g(x)\right)^2 \, dx}{\int_0^1 x g(x) \, dx},$$

while the z-coordinate of the centroid of the plane region that was revolved is

$$\frac{\int_{)}^1 \frac{1}{2} \left(g(x)\right)^2 \, dx}{\int_0^1 g(x) \, dx}.$$

(b) By considering

$$\int_0^1 \int_0^1 g(x)g(y)(x-y)(g(x)-g(y)) \, dx \, dy,$$

show that the centroid of the solid of revolutions is below that of the plane region. (Why is the repeated integral less that or equal to 0?)

Calculus is Everywhere # 24
Solving the Wave Equation

The Calculus is Everywhere section in the previous chapter (The Wave in a Rope) introduced the partial differential equation

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 y}{\partial t^2}. \tag{C.24.1}$$

We will solve it to find y as a function of x and t. First, we solve some simpler equations, which will help us solve (C.24.1).

EXAMPLE 1 Let $u(x, y)$ satisfy $\partial u/\partial x = 0$. Find the form of $u(x, y)$.
SOLUTION Since $\partial u/\partial x$ is 0, for a fixed value of y, $u(x, y)$ is constant. Thus, $u(x, y)$ depends only on y, and can be written in the form $h(y)$ for some function h of a single variable.

On the other hand, any function $u(x, y)$ that can be written in the form $h(y)$ has the property that $\partial u/\partial x = 0$. ◇

EXAMPLE 2 Let $u(x, y)$ satisfy

$$\frac{\partial^2 u}{\partial x \partial y} = 0. \tag{C.24.2}$$

Find the form of $u(x, y)$.
SOLUTION We know that

$$\frac{\partial\left(\frac{\partial u}{\partial y}\right)}{\partial x} = \frac{\partial^2 u}{\partial x \partial y} = 0.$$

By Example 1, $\frac{\partial u}{\partial y} = h(y)$ or some function $h(y)$. By the Fundamental Theorem of Calculus, for any number b,

$$u(x, b) - u(x, 0) = \int\limits_0^b \frac{\partial u}{\partial y} dy = \int\limits_0^b h(y) dy.$$

Let H be an antiderivative of h. Then

$$u(x, b) - u(x, 0) = H(b) - H(0).$$

Replacing b by y shows that

$$u(x, y) = u(x, 0) + H(y) - H(0).$$

So $u(x, y)$ is a sum of a function of x and a function of y,

$$u(x, y) = f(x) + g(y). \tag{C.24.3}$$

A quick calculation shows that any function of this form satisfies (C.24.2). $\diamond$

We will solve the wave equation (C.24.1) by using a change of variables that transforms it into the one solved in Example 2.

The new variables are

$$p = x + ct \qquad \text{and} \qquad q = x - ct.$$

so

$$x = \frac{1}{2}(p + q) \qquad \text{and} \qquad t = \frac{1}{2c}(p - q).$$

We will apply the chain rule, where y is a function of p and q and p and q are functions of x and t, as indicated in Figure C.24.1. Thus $y(x, t) = u(p, q)$.

Because

$$\frac{\partial p}{\partial x} = 1, \qquad \frac{\partial p}{\partial t} = c, \qquad \frac{\partial q}{\partial x} = 1, \qquad \text{and} \qquad \frac{\partial q}{\partial t} = -c,$$

we have

$$\frac{\partial y}{\partial x} = \frac{\partial u}{\partial p}\frac{\partial p}{\partial x} + \frac{\partial u}{\partial q}\frac{\partial q}{\partial x} = \frac{\partial u}{\partial p} + \frac{\partial u}{\partial q}.$$

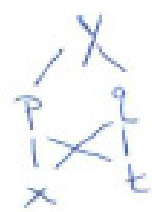


Figure C.24.1

Then

$$\begin{aligned}
\frac{\partial^2 y}{\partial x^2} &= \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial p} + \frac{\partial u}{\partial q}\right) \\
&= \frac{\partial}{\partial p}\left(\frac{\partial u}{\partial p} + \frac{\partial u}{\partial q}\right)\frac{\partial p}{\partial x} + \frac{\partial}{\partial q}\left(\frac{\partial u}{\partial p} + \frac{\partial u}{\partial q}\right)\frac{\partial q}{\partial x} \\
&= \left(\frac{\partial^2 u}{\partial p^2} + \frac{\partial^2 u}{\partial p \partial q}\right)\cdot 1 + \left(\frac{\partial^2 u}{\partial q \partial p} + \frac{\partial^2 u}{\partial q^2}\right)\cdot 1.
\end{aligned}$$

Thus

$$\frac{\partial^2 y}{\partial x^2} = \frac{\partial^2 u}{\partial p^2} + 2\frac{\partial^2 u}{\partial p \partial q} + \frac{\partial^2 u}{\partial q^2}. \tag{C.24.4}$$

A similar calculation shows that

$$\frac{\partial^2 y}{\partial t^2} = c^2\left(\frac{\partial^2 u}{\partial p^2} - 2\frac{\partial^2 u}{\partial p \partial q} + \frac{\partial^2 u}{\partial q^2}\right). \tag{C.24.5}$$

Substituting (C.24.4) and (C.24.5) in (C.24.1) leads to

$$\frac{\partial^2 u}{\partial p^2} + 2\frac{\partial^2 u}{\partial p \partial q} + \frac{\partial^2 u}{\partial q^2}. = \frac{1}{c^2}\left(c^2\right)\left(\frac{\partial^2 u}{\partial p^2} - 2\frac{\partial^2 u}{\partial p \partial q} + \frac{\partial^2 u}{\partial q^2}\right),$$

which reduces to

$$4\frac{\partial^2 u}{\partial p \partial q} = 0.$$

By Example 2, there are function $f(p)$ and $g(q)$ such that

$$y(x,t) = u(p,q) = f(p) + g(q)$$

which can be written as

$$y(x,t) = f(x+ct) + g(x-ct). \tag{C.24.6}$$

The expression (C.24.6) is the most general solution of the wave equation (C.24.1).

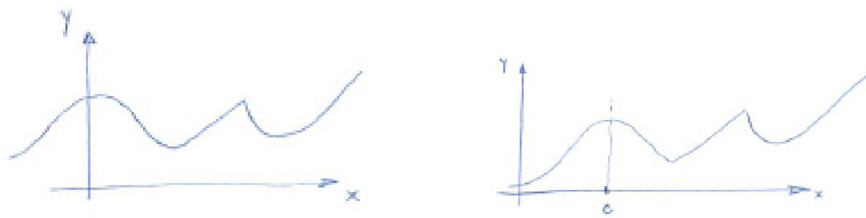
What does a solution (C.24.6) look like? What does the constant c tell us? To answer these questions, suppose

$$y(x,t) = g(x-ct). \tag{C.24.7}$$

Here t represents time. For each t, $y(x,t) = g(x-ct)$ is a function of x and we can graph it in the xy plane. For $t = 0$, (C.24.7) becomes

$$y(x,0) = g(x).$$

That is the graph of $y = g(x)$, whatever g is, as shown in Figure C.24.2(a).



(a) (b)

Figure C.24.2 (a) $t = 0$, (b) $t = 1$.

One unit of time later, when $t = 1$, Then

$$y = y(x,1) = g(x - c \cdot 1) = g(x-c).$$

The value of $y(x,1)$ is the same as the value of g at $x - c$, c units to the left of x. So the graph at $t = 1$ is the graph of g in Figure C.24.2(a) shifted to the right c units, as in Figure C.24.2(b).

As t increases, the initial 'wave shown in Figure C.24.2(a) moves to the right at the constant speed, c. Thus c tells us its velocity. That will play a role in Maxwell's prediction that electromagnetic waves travel at the speed of light, as we will see in the Calculus is Everywhere at the end of Chapter 18.

EXERCISES

1. Which functions $u(x, y)$ have both $\partial u/\partial x$ and $\partial u/\partial y$ equal to 0 for all x and y?

2. Let $u(x, y)$ satisfy $\partial^2 u/\partial x^2 = 0$. Find the form of $u(x, y)$.

3. Show that any function of the form (C.24.3) satisfies (C.24.2).

4. Verify that any function of the form (C.24.6) satisfies the wave equation.

5. We interpreted $y(x, t) = g(x - ct)$ as the description of a wave moving with speed c to the right. Interpret $y(x, t) = f(x + ct)$.

6. Carry out the similar calculations to verify (C.24.5).

7. Let k be a positive constant.

(a) What are the solutions to the equation

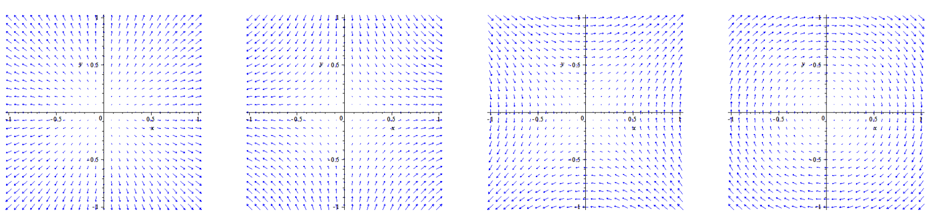
$$\frac{\partial^2 y}{\partial x^2} = k\frac{\partial^2 y}{\partial t^2}?$$

(b) What is the speed of the waves?

Chapter 18

The Theorems of Green, Stokes, and Gauss

Imagine a fluid or gas moving through space or on a plane. Its density may vary from point to point as may its velocity vector may vary from point to point. Figure 18.0.1 shows four situations. The diagrams show flows in the plane because they are easier to visualize.



(a) (b) (c) (d)

Figure 18.0.1 Four vector fields in the plane.

The plots in Figure 18.0.1 resemble the slope fields of Section 3.6 but instead of short segments, we have vectors, which may be short or long. Two questions are:

For a fixed region of the plane (or in space), is the amount of fluid in it increasing, decreasing, or not changing?

At a point, does the field create a tendency for the fluid to rotate? If we put a little propeller in the fluid would it turn? If so, in which direction, and how fast?

These questions arise in several areas, such as fluid flow, electromagnetism, thermodynamics, and gravity. This chapter provides techniques for answering them.

Throughout we assume that all partial derivatives of the first and second orders exist and are continuous.

18.1 Conservative Vector Fields

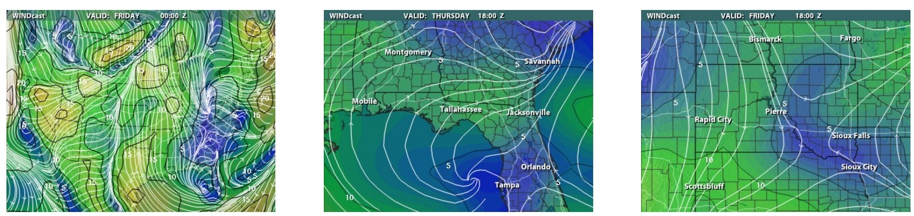
In Section 15.3 we defined integrals of the form

$$\int_C (P\ dx + Q\ dy + R\ dz) \tag{18.1.1}$$

where P, Q, and R are scalar functions of x, y, and z and C is a curve in space. Similarly, in the xy-plane, for scalar functions of x and y, P and Q, we have

$$\int_C (P\ dx + Q\ dy).$$

Instead of three scalar fields, P, Q, and R, we could think of a single vector function $\mathbf{F}(x,y,z) = P(x,y,z)\mathbf{i} + Q(x,y,z)\mathbf{j} + R(x,y,z)\mathbf{k}$, called a **vector field**, in contrast to a scalar field.



(a) (b) (c)

Figure 18.1.1 Wind maps showing the direction and speed of the winds for (a) the United States, (b) near Pierre, SD and (c) near Tallahassee, FL on April 24, 2009. [Source: `www.intellicast.com/National/Wind/Windcast.aspx`]

In Chapter 15 the formal vector $d\mathbf{r} = dx\mathbf{i} + dy\mathbf{j} + dz\mathbf{k}$ was introduced as a way to rewrite (18.1.1) as

$$\int_C \mathbf{F} \cdot\ d\mathbf{r}.$$

The vector notation is compact, is the same in the plane and in space, and emphasizes the idea of a vector field. More important, it frees us from referring to any specific coordinate system. The longer notations

$$\int_C (P\ dx + Q\ dy + R\ dz) \qquad \text{and} \qquad \int_C (P(x,y,z)\ dx + Q(x,y,z)\ dy + R(x,y,z)\ dz)$$

are used to prove theorems and to carry out calculations.

Return to Central Vector Fields

In Section 15.3 we made this definition:

> **DEFINITION** *(Conservative Field)* A vector field $\mathbf{F}$ defined in a planar or spatial region is called **conservative** if
>
> $$\int_{C_1} \mathbf{F}\cdot\ dr = \int_{C_2} \mathbf{F}\cdot\ dr$$
>
> whenever C_1 and C_2 are any two simple curves in the region with the same initial and terminal points.

An equivalent definition of a conservative vector field $\mathbf{F}$ is that for a simple closed curve C in the region $\oint_C \mathbf{F}\cdot\ d\mathbf{r} = 0$, as Theorem 18.1.1 implies. A closed curve is a curve that begins and ends at the same point, forming a loop. It is simple if it passes through no point more than once other than its start and finish points. A curve that starts at one point and ends at a different point is simple if it never intersects itself. Figure 18.1.2 shows some curves that are simple and some that are not.

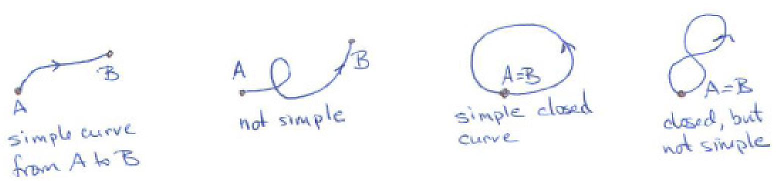


Figure 18.1.2

The next theorem is the first step towards a simple test to identify a conservative field.

Theorem 18.1.1. *A vector field* $\mathbf{F}$ *is conservative if and only if* $\oint_C \mathbf{F}\cdot\ dr = 0$ *for every simple closed curve in the region where* $\mathbf{F}$ *is defined.*

Proof

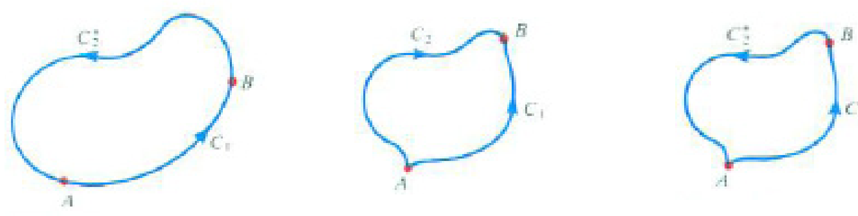
Assume that $\mathbf{F}$ is a conservative and let C be simple closed curve that starts and ends at the point A. Pick a point B on the curve and break C into two curves: C_1 from A to B and C_2^* from B to A, as in Figure 18.1.3(a).

Let C_2 be the curve C_2^* traversed in the opposite direction, from A to B. Then, since $\mathbf{F}$ is conservative,

Note the sign change.

$$\oint_C \mathbf{F}\cdot\ d\mathbf{r} \ = \ \int_{C_1} \mathbf{F}\cdot\ d\mathbf{r} + \int_{C_2^*} \mathbf{F}\ d\mathbf{r} = \int_{C_1} \mathbf{F}\cdot\ d\mathbf{r} - \int_{C_2} \mathbf{F}\cdot\ d\mathbf{r} = 0.$$

To prove the converse, assume that $\oint_C \mathbf{F}\cdot\ d\mathbf{r} = 0$ for any simple closed curve C in the region. Let C_1 and C_2 be two simple curves in the region, starting at A and ending at B. Let $-C_2$ be C_2 taken in the reverse direction. (See Figures 18.1.3(b) and (c).) Then C_1 followed by $-C_2$ is a closed curve C from A back to A. Thus



(a) (b) (c)

Figure 18.1.3

$$0 = \oint\limits_C \mathbf{F}\cdot\ dr \ = \int\limits_{C_1} \mathbf{F}\cdot\ dr + \int\limits_{-C_2} \mathbf{F}\cdot\ dr = \int\limits_{C_1} \mathbf{F}\cdot\ dr - \int\limits_{C_2} \mathbf{F}\cdot\ dr.$$

Consequently,

$$\int\limits_{C_1} \mathbf{F}\cdot\ dr = \int\limits_{C_2} \mathbf{F}\cdot\ dr.$$

•

In this proof we assumed that C_1 and C_2 overlap only at their endpoints, A and B. Exercise 25 treats the case when the curves intersect elsewhere.

Every Gradient Field is Conservative

Whether a vector field is conservative is important in the study of gravity, electromagnetism, and thermodynamics. In the rest of this section we describe ways to determine whether a vector field $\mathbf{F}$ is conservative.

It is impossible to evaluate $\oint \mathbf{F}\cdot d\mathbf{r}$ for every simple closed curve and see if it is always 0 because there are infinitely many of them. If you find a closed curve where it is not 0, then $\mathbf{F}$ is not conservative. The first test involves gradients.

The fundamental theorem of calculus asserts that $\int_a^b f'(x)\ dx = f(b) - f(a)$. Theorem 18.1.2 asserts that $\int_C \nabla f\cdot d\mathbf{r} = f(B) - f(A)$, where f is a function of two or three variables and C is a curve from A to B. Because of its resemblance to the fundamental theorem of calculus, it is sometimes called the **fundamental theorem of vector fields**.

A vector field that is the gradient of a scalar field is conservative. That is the substance of Theorem 18.1.2. It says that the circulation of a gradient field of a scalar function f along a curve is the difference in values of f at the end points.

Theorem 18.1.2. *Let f be a scalar field defined in a region in the plane or in space. Then the gradient field $\mathbf{F} = \nabla f$ is conservative. For any points A and B in the region,*

$$\int_C \nabla f \cdot \, d\mathbf{r} = f(B) - f(A).$$

Proof

For simplicity take the planar case. Let C be given by the parameterization $\mathbf{r} = \mathbf{G}(t)$ for t in $[a, b]$. Let $\mathbf{G}(t) = x(t)\mathbf{i} + y(t)\mathbf{j}$. Then

$$\int_C \nabla f \cdot d\mathbf{r} \;=\; \int_C \left(\frac{\partial f}{\partial x}\, dx + \frac{\partial f}{\partial y}\, dy\right) = \int_a^b \left(\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}\right)\, dt.$$

The integrand $(\partial f/\partial x)(dx/dt) + (\partial f/\partial y)(dy/dt)$ is reminiscent of the chain rule in Section 16.3. If we define H to be

$$H(t) = f(x(t), y(t)),$$

then the chain rule asserts that

$$\frac{dH}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

Thus

$$\int_a^b \left(\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}\right)\, dt = \int_a^b \frac{dH}{dt}\, dt = H(b) - H(a)$$

by the fundamental theorem of calculus. Because

$$H(b) = f(x(b), y(b)) = f(B)$$

and

$$H(a) = f(x(a), y(a)) = f(A)$$

we have

$$\int_C \nabla f \cdot d\mathbf{r} = f(B) - f(A), \qquad (18.1.2)$$

and the theorem is proved. •

In differential form Theorem 18.1.2 reads

SHERMAN: Is this a suitable title?

Differential Form of Theorem 18.1.2

If f is defined as the xy-plane, and C starts at A and ends at B,

$$\int_C \left(\frac{\partial f}{\partial x}\, dx + \frac{\partial f}{\partial y}\, dy \right) = f(B) - f(A) \tag{18.1.3}$$

If f is defined in space, then,

$$\int_C \left(\frac{\partial f}{\partial x}\, dx + \frac{\partial f}{\partial y}\, dy + \frac{\partial f}{\partial z}\, dz \right) = f(B) - f(A). \tag{18.1.4}$$

SHERMAN: If we prefer the notation in the theorem, shouldn't that be boxed instead of (18.1.3) and (18.1.4)?

One vector equation (18.1.2) covers (18.1.3) and (18.1.4). This illustrates an advantage of vector notation.

It is a much easier to evaluate $f(B) - f(A)$ than to compute a line integral.

EXAMPLE 1 Let $f(x, y, z) = \frac{1}{\sqrt{x^2+y^2+z^2}}$, which is defined everywhere except at the origin. (a) Find the gradient field $\mathbf{F} = \nabla f$, (b) Compute $\int_C \mathbf{F} \cdot d\mathbf{r}$ where C is a curve from $(1, 2, 2)$ to $(3, 4, 0)$.

SOLUTION (a) Straightforward computations show that

$$\frac{\partial f}{\partial x} = \frac{-x}{(x^2+y^2+z^2)^{3/2}}, \frac{\partial f}{\partial y} = \frac{-y}{(x^2+y^2+z^2)^{3/2}}, \frac{\partial f}{\partial z} = \frac{-z}{(x^2+y^2+z^2)^{3/2}}.$$

So

$$\nabla f = \frac{-z\mathbf{i} - y\mathbf{j} - z\mathbf{k}}{(x^2+y^2+z^2)^{3/2}}. \tag{18.1.5}$$

If we let $\mathbf{r}(x, y, z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$, $r = |\mathbf{r}|$, and $\widehat{\mathbf{r}} = \mathbf{r}/r$, then (18.1.5) can be written as

$$\mathbf{F} = \nabla f = \frac{-\mathbf{r}}{r^3} = \frac{-\widehat{\mathbf{r}}}{r^2}.$$

(b) For a curve C from $(1, 2, 2)$ to $(3, 4, 0)$,

$$\begin{aligned} \int_C \nabla f \cdot d\mathbf{r} &= f(3,4,0) - f(1,2,2) = \frac{1}{\sqrt{3^2+4^2+0^2}} - \frac{1}{\sqrt{1^2+2^2+2^2}} \\ &= \frac{1}{5} - \frac{1}{3} = -\frac{2}{15}. \end{aligned}$$

◇

For a constant k, a vector field, $\mathbf{F} = k\widehat{\mathbf{r}}/r^2$, is called an **inverse square central field**. They play an important role in gravity and electromagnetism.

In Example 1, $|\nabla f| = \frac{|-\mathbf{r}|}{r^3} = \frac{r}{r^3} = \frac{1}{r^2}$ and $f(x,y,z) = \frac{1}{r}$. In the study of gravity, ∇f measures gravitational attraction, and f measures potential.

EXAMPLE 2 Evaluate $\oint_C (y\ dx + x\ dy)$ around a closed curve C taken counterclockwise.

SOLUTION In Section 15.3 it was shown that if the area enclosed by a curve C is A, then $\oint_C x\ dy = A$ and $\oint_C y\ dx = -A$. Thus,

$$\oint_C (y\ dx + x\ dy) = -A + A = 0.$$

A second solution uses Theorem 18.1.2. The gradient of xy is

$$\nabla(xy) = \frac{\partial(xy)}{\partial x}\mathbf{i} + \frac{\partial(xy)}{\partial y}\mathbf{j} = y\mathbf{i} + x\mathbf{j}.$$

Hence, by Theorem 18.1.2, if the endpoints of C are A and B

$$\oint_C (y\ dx + x\ dy) = \oint_C \nabla(xy)\cdot d\mathbf{r} = xy|_A^B.$$

Because C is a closed curve, $A = B$ and so the integral is 0. ◇

A differential form $P(x,y,z)\ dx + Q(x,y,z)\ dy + R(x,y,z)\ dz$ is called **exact** if there is a scalar function f such that $P(x,y,z) = \partial f/\partial x$, $Q(x,y,z) = \partial f/\partial y$, and $R(x,y,z) = \partial f/\partial z$. Then the expression takes the form

$$\frac{\partial f}{\partial x}\ dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z}\ dz.$$

That is the same thing as saying that the vector field $\mathbf{F} = P(x,y,z)\mathbf{i} + Q(x,y,z)\mathbf{j} + R(x,y,z)\mathbf{k}$ is a gradient field: $\mathbf{F} = \nabla f$.

If F is Conservative Must It Be a Gradient Field?

The proof of the next theorem is similar to the proof of the second part of the fundamental theorem of calculus. We suggest you review that proof (page 534) before reading the following proof.

FTC II states that every continuous function has an antiderivative.

If $\mathbf{F}$ is conservative, is it necessarily the gradient of some scalar function? The answer is yes. That is the substance of the next theorem. First we introduce some terminology about regions.

Recall that a region $\mathcal{R}$ in the plane is open if for each point P in $\mathcal{R}$ there is a disk with center at P that lies entirely in $\mathcal{R}$. For instance, a square without its edges is open. and a square with its edges is not open.

Open is defined in Section 16.2.

An open region in space is defined similarly, with "disk" replaced by "ball."

An open region $\mathcal{R}$ is **arcwise-connected** if two points in it can be joined by a curve that lies completely in $\mathcal{R}$. An arcwise-connected region has only one piece.

Theorem 18.1.3. *Let* $\mathbf{F}$ *be a conservative vector field defined in an arcwise-connected region* $\mathcal{R}$ *in the plane (or in space). Then there is a scalar function* f *defined there such that* $\mathbf{F} = \nabla f$*.*

Proof

Suppose $\mathbf{F}$ is planar, $\mathbf{F} = P(x,y)\mathbf{i} + Q(x,y)\mathbf{j}$. (If $\mathbf{F}$ is defined in space the proof is similar.) Define the scalar function f as follows. Let (a,b) and (x,y) be points in $\mathcal{R}$. Select a curve C in $\mathcal{R}$ that starts at (a,b) and ends at (x,y).

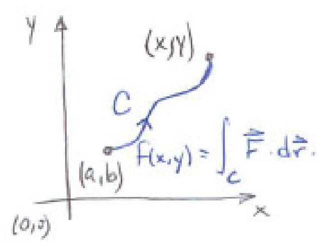


Figure 18.1.4

Define $f(x,y)$ to be $\int_C \mathbf{F}\cdot d\mathbf{r}$. Since $\mathbf{F}$ is conservative, $f(x,y)$ depends only on the point (x,y) and not on the choice of C. (See Figure 18.1.4.)

All that remains is to show that $\nabla f = \mathbf{F}$; that is, $\partial f/\partial x = P$ and $\partial f/\partial y = Q$. We will go through the details for the first case, $\partial f/\partial x = P$. The other is similar.

Let (x_0, y_0) be a point in $\mathcal{R}$ and form the difference quotient whose limit is $\partial f/\partial x(x_0,y_0)$, namely,

$$\frac{f(x_0+h,y_0) - f(x_0,y_0)}{h},$$

for h small enough so that (x_0+h, y_0) is also in the region.

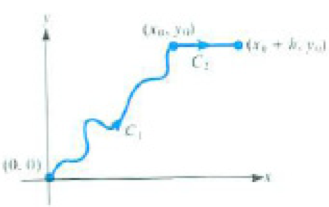


Figure 18.1.5

Let C_1 be a curve from (a,b) to (x_0,y_0) and let C_2 be the straight path from (x_0,y_0) to (x_0+h,y_0). (See Figure 18.1.5.) Let C be the curve from $(0,0)$ to (x_0+h,y_0) formed by taking C_1 first and continuing on C_2. Then

$$f(x_0,y_0) = \int_{C_1} \mathbf{F}\cdot d\mathbf{r},$$

and

$$f(x_0+h,y_0) = \int_{C} \mathbf{F}\cdot d\mathbf{r} = \int_{C_1} \mathbf{F}\cdot d\mathbf{r} + \int_{C_2} \mathbf{F}\cdot d\mathbf{r}.$$

Thus

$$\frac{f(x_0+h,y_0) - f(x_0,y_0)}{h} = \frac{\int_{C_2} \mathbf{F}\cdot d\mathbf{r}}{h} = \frac{\int_{C_2}(P(x,y)\ dx + Q(x,y)\ dy)}{h}.$$

On C_2, y is constant, $y = y_0$ so $dy = 0$. Thus $\int_{C_2} Q(x,y)\ dy = 0$. Also,

$$\int_{C_2} P(x,y)\ dx = \int_{x_0}^{x_0+h} P(x,y)\ dx.$$

By the mean-value theorem for definite integrals, there is a number x^* between x_0 and $x_0 + h$ such that

See Section 6.3 for the MVT for Definite Integrals.

$$\int_{x}^{x+h} P(x,y)\ dx = P(x^*, y_0)h.$$

Hence

$$\begin{aligned}\frac{\partial f}{\partial x}(x_0,y_0) &= \lim_{h\to 0}\frac{f(x_0+h,y_0)-f(x_0,y_0)}{h}\\ &= \lim_{h\to 0}\frac{1}{h}\int_{x_0}^{x_0+h} P(x,y_0)\ dx = \lim_{h\to 0} P(x^*,y_0) = P(x_0,y_0).\end{aligned}$$

Consequently,

$$\frac{\partial f}{\partial x}(x_0,y_0) = P(x_0,y_0),$$

as was to be shown.

Similarly, we can show that

$$\frac{\partial f}{\partial y}(x_0,y_0) = Q(x_0,y_0).$$

•

For a vector field **F** defined throughout some region in the plane (or space) the three properties in Figure 18.1.6 are therefore equivalent because any one of them describes a conservative field. We used property (3) as the definition. Figure 18.1.6 tells us that any one of the three properties, (1), (2), or (3),

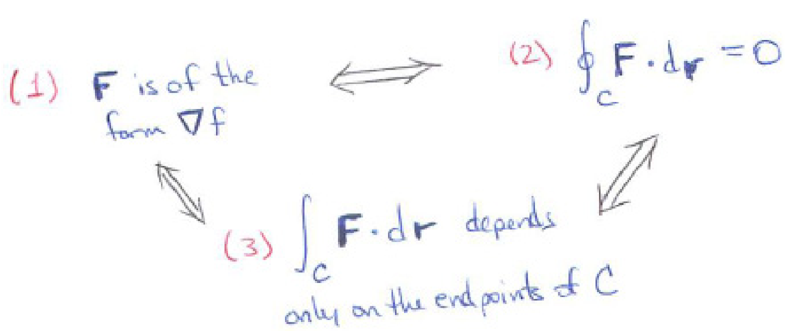


Figure 18.1.6 Double-headed arrows (⇔) mean "if and only if" or "is equivalent to." (Single-headed arrows (⇒) mean "implies.")

describes

Almost A Test For Being Conservative

Figure 18.1.6 describes three ways of deciding whether a vector field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ is conservative. Now we give a simple way to tell that it is not conservative. It is simpler than finding a line integral $\int_C \mathbf{F} \cdot d\mathbf{r}$ that is not 0.

The test depends on the two orders in which we may compute a second-order mixed partial derivative giving the same result.

If $P\ dx + Q\ dy + R\ dz$ (or equivalently a vector field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$) is anexact form, then $\mathbf{F}$ is a gradient and there is a scalar function f such that

$$\frac{\partial f}{\partial x} = P, \qquad \frac{\partial f}{\partial y} = Q, \qquad \frac{\partial f}{\partial z} = R.$$

Since

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)$$

we have

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}.$$

Similarly we find

$$\frac{\partial Q}{\partial z} = \frac{\partial R}{\partial y} \qquad \text{and} \qquad \frac{\partial P}{\partial z} = \frac{\partial R}{\partial x}.$$

To summarize,

Test for a Conservative Field

If the vector field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ is conservative, then

$$\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = 0, \qquad \frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} = 0, \qquad \frac{\partial R}{\partial x} - \frac{\partial P}{\partial z} = 0. \tag{18.1.6}$$

If one of the equations in (18.1.6) does not hold, then $P\ dx + Q\ dy + R\ dz$ is *not* exact and $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ is not conservative.

EXAMPLE 3 Show that $\cos(y)\ dx + \sin(xy)\ dy + \ln(1+x)\ dz$ is not exact.
SOLUTION Checking (18.1.6) we compute

$$\frac{\partial(\sin(xy))}{\partial x} - \frac{\partial(\cos(y))}{\partial y},$$

which equals

$$y\cos(xy) + \sin(y),$$

which is not 0. There is no need to check the remaining equations in (18.1.6). The expression $\sin(xy)\ dx + \cos(y)\ dy + \ln(1+x)\ dz$ is not exact. Thus the vector field $\sin(xy)\mathbf{i}+\cos(y)\mathbf{j}+\ln(1+x)\mathbf{k}$ is not a gradient field and hence not conservative. ◇

To restate (18.1.6) as a vector equation introduce a 3×3 determinant

$$\det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial_x} & \frac{\partial}{\partial_y} & \frac{\partial}{\partial_z} \\ P & Q & R \end{pmatrix} \tag{18.1.7}$$

Expanding it as though its entries were numbers, we get

$$\mathbf{i}\left(\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}\right)-\mathbf{j}\left(\frac{\partial R}{\partial x}-\frac{\partial P}{\partial z}\right)+\mathbf{k}\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right). \tag{18.1.8}$$

If the scalar equations in (18.1.6) hold, then (18.1.8) is the **0**-vector. It is given a name.

> **DEFINITION** *(Curl of a Vector Field)* The **curl** of the vector field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ is the vector field given by (18.1.7) or (18.1.8). It is denoted **curl F**.

The determinant (18.1.7) is like the one for the cross product of two vectors. For this reason, it is also denoted $\nabla\times\mathbf{F}$ (read as "del cross F"). That is easier to write than (18.1.8), which refers to components.

The definition also applies to a vector field $\mathbf{F} = P(x,y)\mathbf{i} + Q(x,y)\mathbf{j}$ in the plane. Writing $\mathbf{F}$ as $P(x,y)\mathbf{i} + Q(x,y)\mathbf{j} + 0\mathbf{k}$ and observing that $\partial Q/\partial z = 0$ and $\partial P/\partial z = 0$, we find that

$$\nabla\times\mathbf{F} = \left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)\mathbf{k}.$$

EXAMPLE 4 Compute the curl of $\mathbf{F} = xyz\mathbf{i} + x^2\mathbf{j} - xy\mathbf{k}$.
SOLUTION The curl of $\mathbf{F}$ is given by

$$\det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ xyz & x^2 & -xy, \end{pmatrix}$$

which is short for

$$\begin{aligned} &\left(\frac{\partial}{\partial y}(-xy)-\frac{\partial}{\partial z}(x^2)\right)\mathbf{i}-\left(\frac{\partial}{\partial x}(-xy)-\frac{\partial}{\partial z}(xyz)\right)\mathbf{j}+\left(\frac{\partial}{\partial x}(x^2)-\frac{\partial}{\partial y}(xyz)\right)\mathbf{k} \\ &= (-x-0)\mathbf{i}-(-y-xy)\mathbf{j}+(2x-xz)\mathbf{k} \\ &= -x\mathbf{i}+(y+xy)\mathbf{j}+(2x-xz)\mathbf{k}. \end{aligned}$$

◇

From (18.1.6), for vector fields in space or in the xy-plane we have

Theorem 18.1.4. *If* $\mathbf{F}$ *is a conservative vector field, then* $\nabla \times \mathbf{F} = \mathbf{0}$.

You may wonder why the vector field $\mathbf{curl}\,\mathbf{F}$ is called the curl of $\mathbf{F}$. It is because of its physical significance as we will see in Section 18.6. If $\mathbf{F}$ describes a fluid flow, $\mathbf{curl}\,\mathbf{F}$ describes the tendency of the fluid to rotate and form whirlpools — that is, to curl.

The Converse of Theorem 18.1.4 Is False

The converse of Theorem 18.1.4 is not true. There are vector fields $\mathbf{F}$ whose curls are $\mathbf{0}$ that are not conservative. Example 5 provides one such $\mathbf{F}$ in the xy-plane. Its curl is $\mathbf{0}$ but it is not conservative.

EXAMPLE 5 Let $\mathbf{F} = \frac{-y\mathbf{i}}{x^2+y^2} + \frac{x\mathbf{j}}{x^2+y^2}$. Show that (a) $\nabla \times \mathbf{F} = \mathbf{0}$, but (b) $\mathbf{F}$ is not conservative.

SOLUTION (a) We compute

$$\det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \frac{-y}{x^2+y^2} & \frac{x}{x^2+y^2} & 0 \end{pmatrix},$$

which equals

$$\left(\frac{\partial(0)}{\partial y} - \frac{\partial}{\partial z}\left(\frac{x}{x^2+y^2}\right)\right)\mathbf{i} - \left(\frac{\partial(0)}{\partial x} - \frac{\partial}{\partial z}\left(\frac{-y}{x^2+y^2}\right)\right)\mathbf{j}$$
$$+ \left(\frac{\partial}{\partial x}\left(\frac{x}{x^2+y^2}\right) - \frac{\partial}{\partial y}\left(\frac{-y}{x^2+y^2}\right)\right)\mathbf{k}.$$

The $\mathbf{i}$ and $\mathbf{j}$ components are clearly 0, and a computation shows that the $\mathbf{k}$ component is

$$\frac{y^2-x^2}{(x^2+y^2)^2} - \frac{y^2-x^2}{(x^2+y^2)^2} = 0.$$

Thus the curl of $\mathbf{F}$ is $\mathbf{0}$.

(b) To show that $\mathbf{F}$ is not conservative, it suffices to exhibit a closed curve C such that $\oint_C \mathbf{F} \cdot\, d\mathbf{r}$ is not 0. One such C is the unit circle parameterized counterclockwise by

$$x = \cos(\theta), \qquad y = \sin(\theta), \qquad 0 \le \theta \le 2\pi.$$

On it $x^2 + y^2 = 1$. Figure 18.1.7 shows a few values of $\mathbf{F}$ at points on C. It appears that $\int_C \mathbf{F} \cdot\, d\mathbf{r}$, which measures circulation, is positive, not 0. Its exact

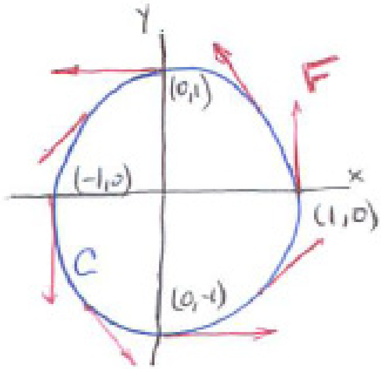


Figure 18.1.7

value is $\int_C \mathbf{F} \cdot\, d\mathbf{r}$:

Recall that, on C, $x^2 + y^2 = 1$.

$$\begin{aligned}
\oint_C \mathbf{F} \cdot\, d\mathbf{r} &= \oint_C \left(\frac{-y\ dx}{x^2 + y^2} + \frac{x\ dy}{x^2 + y^2} \right) \\
&= \int_0^{2\pi} (-\sin\theta\ d(\cos\theta) + \cos\theta\ d(\sin\theta)) \\
&= \int_0^{2\pi} (\sin^2\theta + \cos^2\theta)\ d\theta = \int_0^{2\pi} d\theta = 2\pi.
\end{aligned}$$

This establishes (b), **F** is not conservative. ◇

The curl of **F** being **0** is not enough to assure us that a vector field **F** is conservative. An extra condition must be satisfied by **F**. This condition concerns the domain of **F**. This extra assumption will be developed for planar fields in Section 18.2 and for spatial fields in Section 18.6. Then we will have a test for determining whether a vector field is conservative.

Summary

We showed that a vector field is conservative is equivalent if and only if it is the gradient of a scalar field. Then we defined the curl of a vector field. If the field is denoted **F**, the curl of **F** is a new vector field denoted **curl F** or $\nabla \times \mathbf{F}$. If **F** is conservative, then $\nabla \times \mathbf{F}$ is **0**. However, if the curl of **F** is **0**, it does not follow that **F** is conservative. An extra assumption on the domain of **F** must be added. That assumption will be described in the next section.

EXERCISES for Section 18.1

In Exercises 1 to 4 answer true or false and explain.

1. $\mathbf{F}$ is conservative, then $\nabla \times \mathbf{F} = \mathbf{0}$.

2. $\nabla \times \mathbf{F} = \mathbf{0}$, then $\mathbf{F}$ is conservative.

3. $\mathbf{F}$ is a gradient field, then $\nabla \times \mathbf{F} = \mathbf{0}$.

4. $\nabla \times \mathbf{F} = \mathbf{0}$, then $\mathbf{F}$ is a gradient field.

5. Using information in this section, describe various ways of showing a vector field $\mathbf{F}$ is not conservative.

6. Using information in this section, describe various ways of showing a vector field $\mathbf{F}$ is conservative.

7. Decide if the sets are open, closed, neither open nor closed, or both open and closed.

(a) unit disk with its boundary

(b) unit disk without any of its boundary points

(c) the x-axis

(d) the xy-plane

(e) the xy-plane with the x-axis removed

(f) a square with its edges and corners

(g) a square with its edges but with its corners removed

(h) a square with none of its edges

8. In Example 1 we computed a line integral by using the fact that the vector field $(-x\mathbf{i} - y\mathbf{j} - z\mathbf{k})/(x^2 + y^2 + z^2)^{3/2}$ is a gradient field. Compute the integral directly.

9. Let $\mathbf{F} = y\cos(x)\mathbf{i} + (\sin(x) + 2y)\mathbf{j}$.

(a) Show that $\mathbf{curl}\,\mathbf{F}$ is $\mathbf{0}$ and $\mathbf{F}$ is defined in an arcwise-connected region of the plane.

(b) Construct a potential function f whose gradient is $\mathbf{F}$.

10. Let $f(x,y,z) = e^{3x}\ln(z+y^2)$. Compute $\int_C \nabla f \cdot\ d\mathbf{r}$, where C is the straight path from $(1,1,1)$ to $(4,3,1)$.

11. We obtained the first equation in (18.1.6). Derive the other two.

12. Find the curl of $\mathbf{F}(x,y,z) = e^{x^2}yz\mathbf{i} + x^3\cos^2 3y\mathbf{j} + (1+x^6)\mathbf{k}$.

13. Find the curl of $\mathbf{F}(x,y) = \tan^2(3x)\mathbf{i} + e^{3x}\ln(1+x^2)\mathbf{j}$.

14. Using theorems of this section, explain why the curl of a gradient is $\mathbf{0}$, that is, $\mathbf{curl}(\nabla f) = \mathbf{0}$ ($\nabla \times \nabla f = \mathbf{0}$) for a scalar function $f(x,y,z)$. (No computations are needed.)

15. By a computation using components, show that for the scalar function $f(x,y,z)$, $\mathbf{curl}(\nabla f) = \mathbf{0}$.

16. Let $f(x,y) = \cos(x+y)$. Evaluate $\int_C \nabla f \cdot\ d\mathbf{r}$, where C is the part of the parabola $y = x^2$ that goes from $(0,0)$ to $(2,4)$.

17. In Example 5 we computed $\oint_C \mathbf{F}\cdot d\mathbf{r}$, where $\mathbf{F} = \frac{-y\mathbf{i}+x\mathbf{j}}{x^2+y^2}$ and C is the unit circle with center at the origin. Compute the integral when C is the circle of radius 5 with center at the origin.

18. Give an example of two conservative vector fields $\mathbf{F}$ and $\mathbf{G}$ such that $\mathbf{F}+\mathbf{G}$ is not conservative.

SHERMAN: Does this exercise need something more?

19. Show that $\mathbf{curl}(f\mathbf{F}) = \nabla f \times \mathbf{F} + f\,\mathbf{curl}\,\mathbf{F}$.

20. Show that $\mathbf{curl}(\mathbf{F}\times\mathbf{G}) = (\mathbf{G}\cdot\nabla)\mathbf{F} - (\mathbf{F}\cdot\nabla)\mathbf{G} + \mathbf{F}(\nabla\cdot\mathbf{G}) - \mathbf{G}(\nabla\cdot\mathbf{F})$. The first two terms have a form not seen before now. If $\mathbf{F} = F_1\mathbf{i} + F_2\mathbf{j} + F_3\mathbf{k}$ and $\mathbf{G} = G_1\mathbf{i} + G_2\mathbf{j} + G_3\mathbf{k}$, then

$$(\mathbf{G}\cdot\nabla)\mathbf{F} = G_1\frac{\partial F_1}{\partial x} + G_2\frac{\partial F_2}{\partial y} + G_3\frac{\partial F_3}{\partial z}.$$

21. If $\mathbf{F}$ and $\mathbf{G}$ are conservative, is $\mathbf{F}\times\mathbf{G}$?

22. Assume that $\mathbf{F}(x,y)$ is conservative. Let C_1 be the straight path from $(0,0,0)$ to $(1,0,0)$ and C_2 the straight path from $(1,0,0)$ to $(1,1,1)$. If $\int_{C_1}\mathbf{F}\ d\mathbf{r} = 3$ and

$\int_{C_2} \mathbf{F}\, d\mathbf{r} = 4$, what can be said about $\int_C \mathbf{F}\, d\mathbf{r}$, where C is the straight path from $(0,0,0)$ to $(1,1,1)$?

23. Let $\mathbf{F}(x,y)$ be the field

$$\mathbf{F}(x,y) = g(\sqrt{x^2+y^2})\frac{x\mathbf{i}+y\mathbf{j}}{\sqrt{x^2+y^2}}$$

where g is a scalar function. If we denote $x\mathbf{i}+y\mathbf{j}$ as $\mathbf{r}$, then $\mathbf{F}(x,y) = g(r)\widehat{\mathbf{r}}$, where $r = |\mathbf{r}|$ and $\widehat{\mathbf{r}} = |\mathbf{r}|/r$. Show that $\oint_C \mathbf{F}\cdot\ d\mathbf{r} = 0$ for any path $ABCDA$ of the form shown in Figure 18.1.8. The path consists of two circular arcs and parts of two rays from the origin.

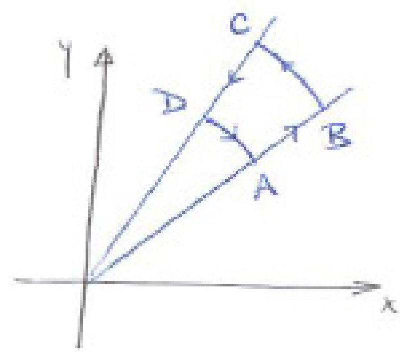


Figure 18.1.8

24. In view of the previous exercise, we may expect $\mathbf{F}(x,y) = g(\sqrt{x^2+y^2})\frac{x\mathbf{i}+y\mathbf{j}}{\sqrt{x^2+y^2}}$ to be conservative. Show that it is by showing that $\mathbf{F}$ is the gradient of $G(x,y) = H(\sqrt{x^2+y^2})$, where H is an antiderivative of g, that is, $H' = g$.

25. In Theorem 18.1.1 we proved that $\partial f/\partial x = P$. Prove that $\partial f/\partial y = Q$.

26. The domain of a vector field $\mathbf{F}$ is all of the xy-plane. Assume that there are two points A and B such that $\int_C \mathbf{F}\, d\mathbf{r}$ is the same for all curves C from A to B. Deduce that $\mathbf{F}$ is conservative.

27. A gas at temperature T_0 and pressure P_0 is brought to temperature $T_1 > T_0$ and pressure $P_1 > P_0$. The work done in this process is given by the line integral in the TP- plane

$$\int\limits_C \left(\frac{RT\ dP}{P} - R\ dT\right),$$

where R is a constant and C is the curve that records the various combinations of T and P during the process. Evaluate the integral over the paths shown in Figure 18.1.9.

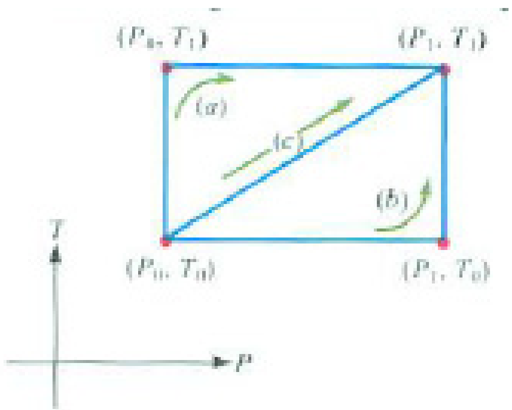


Figure 18.1.9

(a) The pressure is kept constant at P_0 while the temperature is raised from T_0 to T_1 and then the temperature is kept constant at T_1 while the pressure is raised from P_0 to P_1.

(b) The temperature is kept constant at T_0 while the pressure is raised from P_0 to P_1 and then the temperature is raised from T_0 to T_1 while the pressure is kept constant at P_1.

(c) Both pressure and temperature are raised so that the path from (P_0, T_0) to (P_1, T_1) is straight.

Because the integrals are path dependent, the differential expression $RT\ dP/P - R\ dT$ defines a thermodynamic quantity that depends on the process, not only on the state. The vector field $(RT/P)\mathbf{i} - R\mathbf{j}$ is not conservative.

28. Assume that $\mathbf{F}(x, y)$ is defined throughout the xy-plane and that $\oint_C \mathbf{F}(x, y) \cdot d\mathbf{r} = 0$ for every closed curve that can fit inside a disk of diameter 0.01. Show that $\mathbf{F}$ is conservative.

29. This exercise completes the proof of Theorem 18.1.1 when C_1 and C_2 overlap outside of their endpoints A and B. Introduce a third simple curve from A to B that overlaps C_1 and C_2 only at A and B. Make an argument similar to that in the proof of Theorem 18.1.1 to dispose of this case.

30. We proved that $\lim \frac{\int_{x_0}^{x_0+h} P(x,y_0)\ dx}{h}$ equals $P(x_0, y_0)$, by using the mean-value theorem for definite integrals. Find a proof that uses a part of the fundamental theorem of calculus.

18.2 Green's Theorem and Circulation

In this section we discuss a theorem that relates an integral of a vector field over a closed curve C in a plane to an integral of a related scalar function over the region $\mathcal{R}$ whose boundary is C. We will also see what this means in terms of the circulation of a vector field. In thte next section we relate it to find flux across a curve.

Statement of Green's Theorem

We begin by stating Green's Theorem and explaining each term in it. Then we will see several applications of the theorem. Its proof is at the end of the next section.

Green's Theorem

Let C be a simple, closed counterclockwise curve in the xy-plane, bounding a region $\mathcal{R}$. Let P and Q be scalar functions defined at least on an open set containing $\mathcal{R}$. Assume P and Q have continuous first partial derivatives. Then

$$\oint\limits_C (P\ dx + Q\ dy) = \int\limits_{\mathcal{R}} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\ dA.$$

Since P and Q are independent of each other, Green's Theorem really consists of two theorems:

$$\int\limits_C P\ dx = -\int\limits_R \frac{\partial P}{\partial y}\ dA \qquad \text{and} \qquad \oint\limits_C Q\ dy = \int\limits_R \frac{\partial Q}{\partial x}\ dA. \tag{18.2.1}$$

EXAMPLE 1 In Section 15.3 we showed that if the counterclockwise curve C bounds a region $\mathcal{R}$, then $\oint_C y\ dx$ is the negative of the area of $\mathcal{R}$. Obtain this result with the aid of Green's Theorem.

SOLUTION Let $P(x,y) = y$, and $Q(x,y) = 0$. Then Green's Theorem says that

$$\oint\limits_C y\ dx = -\int\limits_R \frac{\partial y}{\partial y}\ dA.$$

Since $\partial y/\partial y = 1$, it follows that $\oint y\ dx$ is $-\int_{\mathcal{R}} 1\ dA$, the negative of the area of $\mathcal{R}$. ◇

Green's Theorem and Circulation

What does Green's Theorem say about a vector field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$? First of all, $\oint_C (P\ dx + Q\ dy)$ now becomes simply $\oint_C \mathbf{F} \cdot\ d\mathbf{r}$.

The right hand side of Green's Theorem looks a bit like the curl of a vector field in the plane. To be specific, we compute the curl of $\mathbf{F}$:

$$\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \partial_x & \partial_y & \partial_z \\ P(x,y) & Q(x,y) & 0 \end{pmatrix} = 0\mathbf{i} - 0\mathbf{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathbf{k}$$

Thus the curl of $\mathbf{F}$ equals the vector function

$$\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathbf{k}. \tag{18.2.2}$$

To obtain the (scalar) integrand on the right-hand side of (18.2.2), we "dot (18.2.2) with $\mathbf{k}$,"

$$\left(\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathbf{k}\right) \cdot \mathbf{k} = \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}.$$

We can now express Green's Theorem using vectors. In particular, circulation around a closed curve can be expressed in terms of a double integral of the curl over a region.

Green's Theorem in Vector Notation

If the counterclockwise closed curve C bounds the region $\mathcal{R}$, then

$$\oint\limits_C \mathbf{F} \cdot\ d\mathbf{r} = \int\limits_{\mathcal{R}} (\nabla \times \mathbf{F}) \cdot \mathbf{k}\ dA.$$

Recall that if $\mathbf{F}$ describes the flow of a fluid in the xy-plane, then $\oint_C \mathbf{F} \cdot d\mathbf{r}$ represents its circulation, or tendency to form whirlpools. This theorem tells us that the magnitude of the curl of $\mathbf{F}$ represents the tendency of the fluid to rotate. If the curl of $\mathbf{F}$ is $\mathbf{0}$ everywhere, then $\mathbf{F}$ is called **irrotational** — there is no rotational tendency.

This form of Green's theorem provides an easy way to show that a vector field $\mathbf{F}$ is conservative. It uses the idea of a simply-connected region. Informally "a simply-connected region in the xy-plane comes in one piece and has no holes." More precisely, an arcwise-connected region $\mathcal{R}$ in the plane or in space is **simply-connected** if each closed curve in $\mathcal{R}$ can be shrunk gradually to a point while remaining in $\mathcal{R}$.

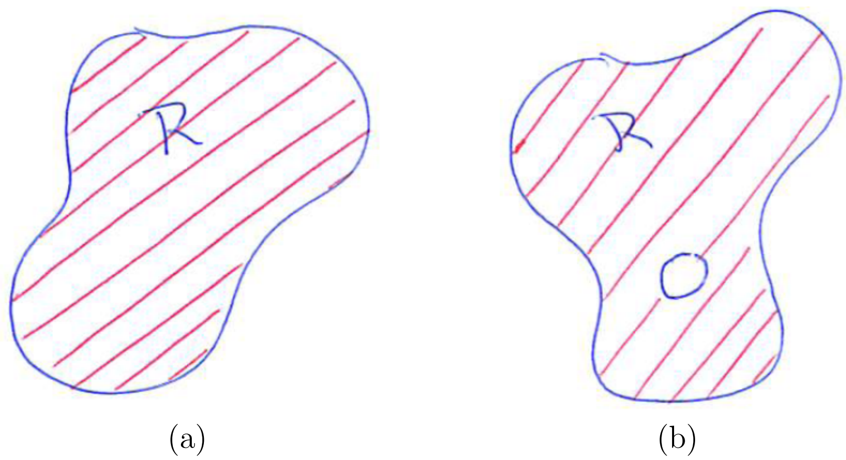


Figure 18.2.1 Regions in the plane that are (a) simply connected and (b) not simply connected.

Figure 18.2.1 shows two regions in the plane. The one on the left is simply-connected, while the one on the right is not simply connected. For instance, the xy-plane is simply connected. So is the xy-plane without its positive x-axis. However, the xy-plane, without the origin is *not* simply connected, because a circular path around the origin cannot be shrunk to a point while staying within the region.

If the origin is removed from xyz-space, what is left *is* simply connected. However, if we remove the z-axis, what is left is *not* simply connected.

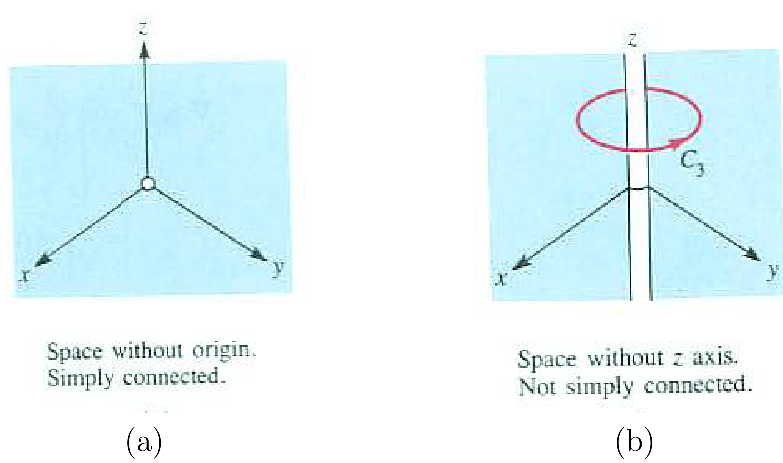


Figure 18.2.2 (a) xyz-space with the origin removed is simply connected. (b) xyz-space with the z-axis removed is not simply connected.

Figure 18.2.2(b) shows a curve that cannot be shrunk to a point while avoiding the z-axis.

Now we can state an easy way to tell whether a vector field is conservative.

Theorem 18.2.1. *If a vector field* $\mathbf{F}$ *is defined in a simply-connected region in the* xy*-plane and* $\nabla \times \mathbf{F} = \mathbf{0}$ *throughout that region, then* $\mathbf{F}$ *is conservative.*

Proof

Let C be any simple closed curve in the region and $\mathcal{R}$ the region it bounds. We wish to prove that the circulation of $\mathbf{F}$ around C is $\mathbf{0}$. We have

$$\oint\limits_C \mathbf{F} \cdot \ d\mathbf{r} = \int\limits_{\mathcal{R}} (\mathbf{curl}\,\mathbf{F}) \cdot \mathbf{k}\ dA.$$

Since $\mathbf{curl}\,\mathbf{F}$ is $\mathbf{0}$ throughout $\mathcal{R}$, it follows that $\oint_C \mathbf{F} \cdot \ d\mathbf{r} = 0$. •

In Example 5 in Section 18.1, there is a vector field whose curl is $\mathbf{0}$ but is not conservative. In view of the theorem just proved, its domain must not be simply connected. Indeed, the domain of the vector field in that example is the xy-plane with the origin deleted.

EXAMPLE 2 Let $\mathbf{F}(x, y, z) = e^x y\mathbf{i} + (e^x + 2y)\mathbf{j}$.

1. Show that $\mathbf{F}$ is conservative.
2. Exhibit a scalar function f whose gradient is $\mathbf{F}$.

SOLUTION

1. A straightforward calculation shows that $\nabla \times \mathbf{F} = \mathbf{0}$. Since $\mathbf{F}$ is defined throughout the xy-plane, a simply-connected region, Theorem 18.2.1 tells us that $\mathbf{F}$ is conservative.

2. By Section 18.1, we know that there is a scalar function f such that $\nabla f = \mathbf{F}$. There are several ways to find f. We show one of these methods here. Additional approaches are pursued in Exercises 7 and 8.

   The approach chosen here follows the construction in the proof of Theorem 18.1.3. For a point (a, b), define $f(a, b)$ to equal $\int_C \mathbf{F} \cdot d\mathbf{r}$, where C is any curve from $(0, 0)$ to (a, b). Any curve with the prescribed endpoints will do. For simplicity, choose C to be the curve that goes from $(0, 0)$ to (a, b) in a straight line. (See Figure 18.2.3.) When a is not zero, we can use x as a parameter and write this segment as: $x = t$, $y = (b/a)t$ for $0 \le t \le a$. (If $a = 0$, we would use y as a parameter.) Then

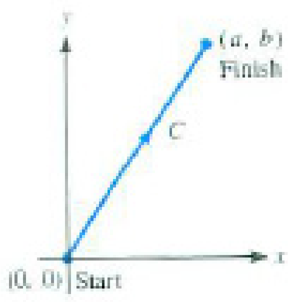


Figure 18.2.3

$$
\begin{aligned}
f(a,b) &= \int\limits_C (e^x y\,dx + (e^x + 2y)\,dy) = \int\limits_0^a \left(e^t \frac{b}{a} t\,dt + \left(e^t + 2\frac{b}{a}t \right) \frac{b}{a}\,dt \right) \\
&= \frac{b}{a} \int\limits_0^a \left(te^t + e^t + 2\frac{b}{a}t \right)\,dt = \frac{b}{a} \left((t-1)e^t + e^t + \frac{b}{a}t^2 \right) \Bigg|_0^a \\
&= \frac{b}{a} \left(te^t + \frac{b}{a}t^2 \right) \Bigg|_0^a = be^a + b^2.
\end{aligned}
$$

$ye^x + y^2 + k$ for any constant k, also would be a potential.

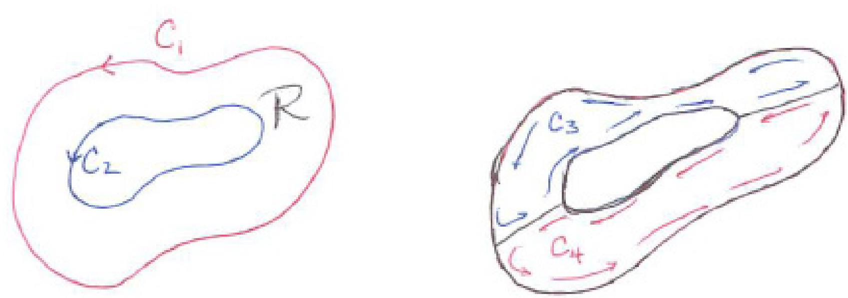
Since $f(a,b) = be^a + b^2$, we see that $f(x,y) = ye^x + y^2$ is the desired function. One could check this by showing that the gradient of f is indeed $ye^x\mathbf{i} + (e^x + 2y)\mathbf{j}$. Other suitable potential functions f are $ye^x + y^2 + k$ for any constant k.

$\diamond$

The next example uses the *cancellation principle*, which is based on the fact that the sum of two line integrals in opposite direction on a curve is zero. This idea is used here to develop the two-curve version of Green's Theorem and then several more times before the end of this chapter.

Green's Theorem — The Two-Curve Case

EXAMPLE 3 Figure 18.2.4(a) shows two closed counterclockwise curves C_1, and C_2 that enclose a ring-shaped region $\mathcal{R}$ in which $\nabla \times \mathbf{F}$ is $\mathbf{0}$. Show that the circulation of $\mathbf{F}$ over C_1 equals the circulation of $\mathbf{F}$ over C_2.
SOLUTION Cut $\mathcal{R}$ into two regions, each bounded by a simple curve, to



(a) (b)

Figure 18.2.4

which we can apply Theorem 18.2.1. Let C_3 bound one of the regions and C_4 bound the other, with the usual counterclockwise orientation. On the cuts, C_3

and C_4 go in opposite directions. On the outer curve C_3 and C_4 have the same orientation as C_1. On the inner curve they are the opposite orientation of C_2. (See Figure 18.1.2(b).) Thus

$$\int_{C_3} \mathbf{F}\cdot\ dr + \int_{C_4} \mathbf{F}\cdot\ dr = \int_{C_1} \mathbf{F}\cdot\ dr - \int_{C_2} \mathbf{F}\cdot\ dr. \qquad (18.2.3)$$

By Theorem 18.2.1 each integral on the left side of (18.2.3) is 0. Thus

$$\int_{C_1} \mathbf{F}\cdot d\mathbf{r} = \int_{C_2} \mathbf{F}\cdot d\mathbf{r} \qquad (18.2.4)$$

◇

Example 3 justifies the "two-curve" variation of Green's Theorem:

Two-Curve Version of Green's Theorem
Assume two nonoverlapping curves C_1 and C_2 lie in a region where **curl F** is **0** and form the border of a ring. Then, if C_1 and C_2 both have the same orientation,

$$\oint_{C_1} \mathbf{F}\cdot d\mathbf{r} = \oint_{C_2} \mathbf{F}\cdot d\mathbf{r}.$$

This theorem tells us "as you move a closed curve within a region of zero-curl, you don't change the circulation." The next Example illustrates this point.

EXAMPLE 4 Let $\mathbf{F} = \frac{-y\mathbf{i}+x\mathbf{j}}{x^2+y^2}$ and C be the closed counterclockwise curve bounding the square whose vertices are $(-2,-2)$, $(2,-2)$, $(2,2)$, and $(-2,2)$. Evaluate the circulation of $\mathbf{F}$ around C as easily as possible.
SOLUTION This vector field appeared in Example 5 of Section 18.1. Since its curl is $\mathbf{0}$, at all points except the origin, where $\mathbf{F}$ is not defined, we may use the two-curve version of Green's Theorem. Thus $\oint_C \mathbf{F}\cdot d\mathbf{r}$ equals the circulation of $\mathbf{F}$ over the unit circle in Example 5, hence equals 2π.

This is a lot easier than integrating $\mathbf{F}$ directly over each of the four edges of the square. ◇

How to Draw $\nabla \times \mathbf{F}$

For the planar vector field $\mathbf{F}$, its curl, $\nabla\times\mathbf{F}$, is of the form $z(x,y)\mathbf{k}$. If $z(x,y)$ is positive, the curl points directly up from the page. Indicate this by the symbol $\odot$, which suggests the point of an arrow or the nose of a rocket. If $z(x, y)$ is negative, the curl points down from the page. To show this, use the symbol $\oplus$, which suggests the feathers of an arrow or the fins of a rocket. Figure 18.2.5 illustrates their use, is standard notation in physics.

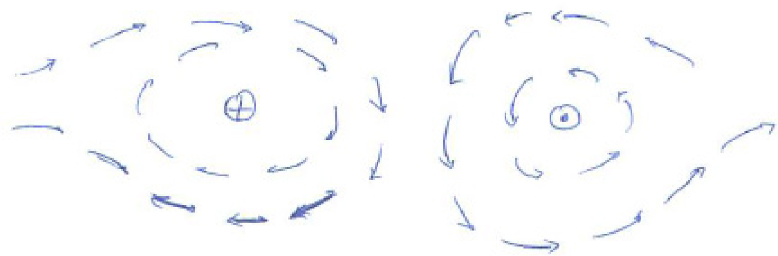


Figure 18.2.5

Summary

We first expressed Green's theorem in terms of scalar functions

$$\oint_C (P\ dx + Q\ dy) = \int_{\mathcal{R}} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right)\ dA.$$

We then translated it into a statement about the circulation of a vector field;

$$\oint_C \mathbf{F} \cdot\ d\mathbf{r} = \int_{\mathcal{R}} (\nabla \times \mathbf{F}) \cdot \mathbf{k}\ dA.$$

In this theorem the closed curve C is oriented counterclockwise.

With the aid of this theorem we were able to show the following important result:

If the curl of $\mathbf{F}$ is $\mathbf{0}$ *and if the domain of* $\mathbf{F}$ *is simply connected*, then $\mathbf{F}$ is conservative.

Also, in a region in which $\nabla \times \mathbf{F} = \mathbf{0}$, the value of $\oint_C \mathbf{F} \cdot\ d\mathbf{r}$ does not change as you gradually change C to other curves in the region.

EXERCISES for Section 18.2

In Exercises 1 through 4 verify Green's Theorem for the given functions P and Q and curve C.

1. $P = xy$, $Q = y^2$ and C is the border of the square whose vertices are $(0,0)$, $(1,0)$, $(1,1)$ and $(0,1)$.

2. $P = x^2$, $Q = 0$ and C is the boundary of the unit circle with center $(0,0)$.

3. $P = e^y$, $Q = e^x$ and C is the triangle with vertices $(0,0)$, $(1,0)$, and $(0,1)$.

4. $P = \sin(y)$, $Q = 0$ and C is the boundary of the portion of the unit disk with center $(0,0)$ in the first quadrant.

5. Figure 18.2.6 shows a vector field for a fluid flow $\mathbf{F}$. At the indicated points A, B, C, and D tell when the curl of $\mathbf{F}$ is pointed up, down or is $\mathbf{0}$. (Use the $\odot$ and $\oplus$ notation.) (When the fingers of your right hand copy the direction of the flow, your thumb points in the direction of the curl, up or down.)

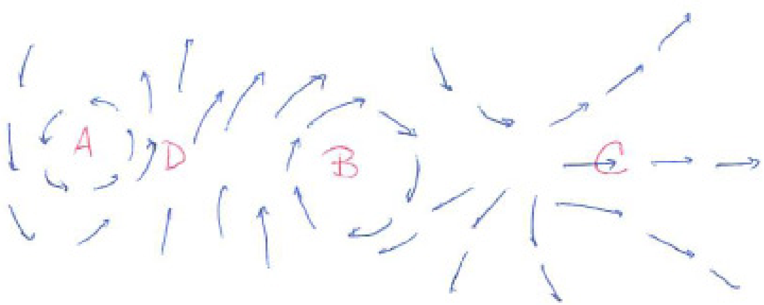


Figure 18.2.6

6. Assume that $\mathbf{F}$ describes a fluid flow. Let P be a point in the domain of $\mathbf{F}$ and C a small circular path around P.

(a) If the curl of $\mathbf{F}$ points upward, in what direction is the fluid tending to turn near P, clockwise or counterclockwise?

(b) If C is oriented clockwise, would $\oint_C \mathbf{F} \cdot d\mathbf{r}$ to be positive or negative?

7. In Example 2 we constructed a function f by using a straight path from $(0,0)$ to (a,b). Instead, construct f by using a path that consists of two line segments, the first from $(0,0)$ to $(a,0)$, and the second, from $(a,0)$ to (a,b).

8. In Example 2 we constructed a function f by using a straight path from $(0,0)$ to (a,b). Instead, construct f by using a path that consists of two line segments, the first from $(0,0)$ to $(0,b)$, and the second from $(0,b)$ to (a,b).

9. Another way to construct a potential function f for a vector field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$

is to work directly with the requirement that $\nabla f = \mathbf{F}$. That is, with the equattions

$$\frac{\partial f}{\partial x} = P(x,y) \qquad \text{and} \qquad \frac{\partial f}{\partial y} = Q(x,y).$$

(a) Integrate $\frac{\partial f}{\partial x} = e^x y$ with respect to x to conclude that $f(x,y) = e^x y + C(y)$. Note that the "constant of integration" can be any function of y, which we call $C(y)$. (Why?)

(b) Next, differentiate the result found in (a) with respect to y. This gives two formulas for $\frac{\partial f}{\partial y}$: $e^x + C'(y)$ and $e^x + 2y$. Use this fact to explain why $C'(y) = 2y$.

(c) Solve the equation for C found in (b).

(d) Combine the results of (a) and (c) to obtain the general form for a potential function for this vector field.

In Exercises 10 through 13

(a) check that $\mathbf{F}$ is conservative in the given domain, that is $\nabla \times \mathbf{F} = \mathbf{0}$, and the domain of $\mathbf{F}$ is simply connected

(b) construct f such that $\nabla f = \mathbf{F}$, using integrals on curves

(c) construct f such that $\nabla f = \mathbf{F}$, using antiderivatives, as in Exercise 9.

10. $\mathbf{F} = 3x^2y\mathbf{i} + x^3\mathbf{j}$, domain the xy-plane

11. $\mathbf{F} = y\cos(xy)\mathbf{i} + (x\cos(xy) + 2y)\mathbf{j}$, domain the xy-plane

12. $\mathbf{F} = (ye^{xy} + 1/x)\mathbf{i} + xe^{xy}\mathbf{j}$, domain all (x,y) with $x > 0$

13. $\mathbf{F} = \frac{2y\ln(x)}{x}\mathbf{i} + (\ln(x))^2\mathbf{j}$, domain all (x,y) with $x > 0$

14. Verify Green's Theorem when $\mathbf{F}(x,y) = x\mathbf{i} + y\mathbf{j}$ and $\mathcal{R}$ is the disk of radius a and center at the origin.

15. In Example 1 we used Green's Theorem to show that $\oint_C y\ dx$ is the negative of the area that C encloses. Use Green's Theorem to show that $\oint_C x\ dy$ equals that area. (We obtained this result in Section 15.3 without Green's Theorem.)

16. Let A be a plane region with boundary C a simple closed curve swept out counterclockwise. Use Green's theorem to show that the area of A equals

$$\frac{1}{2}\oint(-y\ dx + x\ dy).$$

17. Use Exercise 16 to find the area of the region bounded by the line $y = x$ and the curve

$$\begin{cases} x = t^6 + t^4 \\ y = t^3 + t \end{cases} \quad \text{for } t \text{ in } [0, 1].$$

18. Assume that $\mathbf{curl}\,\mathbf{F}$ at $(0, 0)$ is -3. Let C sweep out counterclockwise the boundary of a circle of radius a, center at $(0, 0)$. When a is small, estimate the circulation $\int_C \mathbf{F} \cdot d\mathbf{r}$.

DOUG: Did we assume all circulation is computed with counterclockwise oriented curves? Check this!

19. Which of these fields are conservative:

(a) $x\mathbf{i} - y\mathbf{j}$

(b) $\frac{x\mathbf{i} - y\mathbf{j}}{x^2 + y^2}$

(c) $3\mathbf{i} + 4\mathbf{j}$

(d) $(6xy - y^3)\mathbf{i} + (4y + 3x^2 - 3xy^2)\mathbf{j}$

(e) $\frac{y\mathbf{i} - x\mathbf{j}}{1 + x^2y^2}$

(f) $\frac{x\mathbf{i} + y\mathbf{j}}{x^2 + y^2}$

20. Figure 18.2.7 shows a fluid flow $\mathbf{F}$. All the vectors are parallel, but their magnitudes increase from bottom to top. A small simple curve C is placed in the flow.

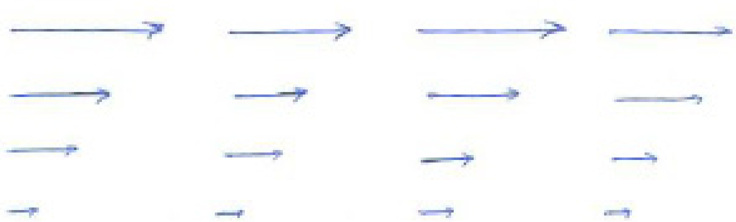


Figure 18.2.7

(a) Assume C has a counterclockwise orientation. Is the circulation around C positive, negative, or 0? Justify your opinion.

(b) Assume that a wheel with small blades is free to rotate around its axis, which is perpendicular to the page. When it is inserted into this flow, which way would it turn, or would it not turn at all? (Don't just say, "It would get wet.")

21. Let $\mathbf{F}(x, y) = y^2\mathbf{i}$.

(a) Sketch the field.

(b) Without computing it, predict when $(\nabla \times \mathbf{F}) \cdot \mathbf{k}$ is positive, negative or zero.

(c) Compute $(\nabla \times \mathbf{F}) \cdot \mathbf{k}$.

(d) What would happen if you dipped a wheel with small blades free to rotate around its axis, which is perpendicular to the page, into this flow.

22. Check that the curl of the vector field in Example 2 is $\mathbf{0}$, as asserted.

23. Explain in words, without explicit calculations, why the circulation of the field $f(r)\widehat{\mathbf{r}}$ around the curve $PQRSP$ in Figure 18.2.8 is zero. As usual, f is a scalar function, $r = |\mathbf{r}|$, and $\widehat{r} = \mathbf{r}/r$.

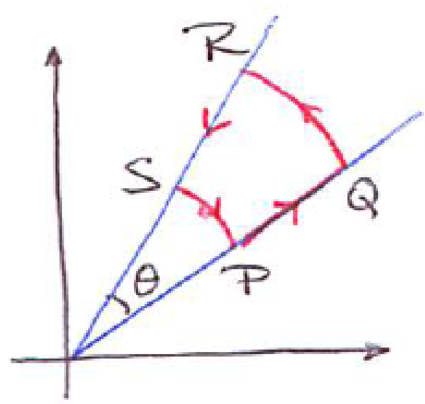


Figure 18.2.8

In Exercises 24 to 27 let $\mathbf{F}$ be a vector field defined everywhere in the plane except at the point P shown in Figure 18.2.9. Assume that $\nabla \times \mathbf{F} = \mathbf{0}$ and that $\int_{C_1} \mathbf{F} \cdot d\mathbf{r} = 5$.

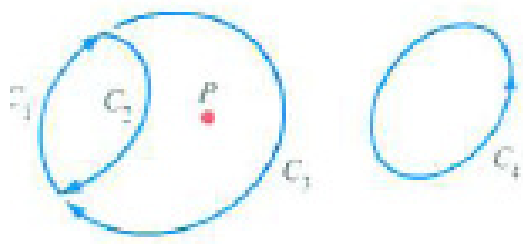


Figure 18.2.9

24. What, if anything, can be said about $\int_{C_2} \mathbf{F} \cdot d\mathbf{r}$?

25. What, if anything, can be said about $\int_{C_3} \mathbf{F} \cdot d\mathbf{r}$?

26. What, if anything, can be said about $\int_{C_4} \mathbf{F} \cdot d\mathbf{r}$?

27. What, if anything, can be said about $\int_{C} \mathbf{F} \cdot d\mathbf{r}$, where C is the curve formed by C_1 followed by C_3?

In Exercises 28 to 31 show that the vector field is conservative and then construct a scalar function of which it is the gradient. Use the method in Example 2.

28. $2xy\mathbf{i} + x^2\mathbf{j}$

29. $\sin(y)\mathbf{i} + (x\cos(y) + 3)\mathbf{j}$

30. $(y+1)\mathbf{i} + (x+1)\mathbf{j}$

31. $3y\sin^2(xy)\cos(xy)\mathbf{i} + (1 + 3x\sin^2(xy)\cos(xy))\mathbf{j}$

32. Show that

(a) $3x^2y\ dx + x^3\ dy$ is exact.

(b) $3xy\ dx + x^2\ dy$ is not exact.

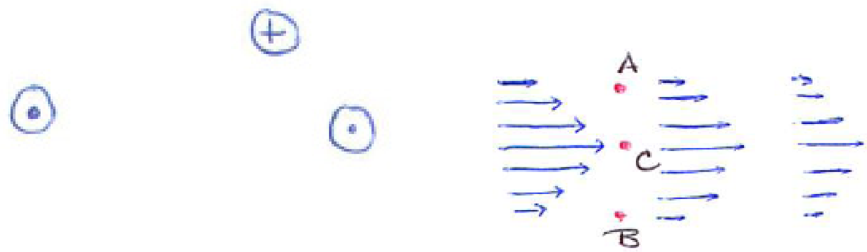
33. Show that $(x\ dx + y\ dy)/(x^2+y^2)$ is exact and exhibit a function f such that df equals the given expression. (That is, find f such that $\nabla f \cdot d\mathbf{r}$ agrees with the given differential form.)

34. Let $\mathbf{F} = \widehat{\mathbf{r}}/|\mathbf{r}|$ in the xy plane and let C be the circle of radius a and center $(0,0)$.

(a) Evaluate $\oint_C \mathbf{F} \cdot \mathbf{n}\ ds$ without using Green's theorem.

(b) Let C now be the circle of radius 3 and center $(4,0)$. Evaluate $\oint_C \mathbf{F}\cdot\mathbf{n}\ ds$, doing as little work as possible.

35. Figure 18.2.10(a) shows the direction of a vector field at three points. Draw a vector field compatible with these values. (No zero-vectors, please.)



(a) (b)

Figure 18.2.10

36. Consider the vector field in Figure 18.2.10(b). Will a paddle wheel turn at A? At B? At C? If so, in which direction?

37. Use Exercise 16 to obtain the formula for area in polar coordinates:

$$\text{Area} = \frac{1}{2}\int_{\alpha}^{\beta} r^2\ d\theta.$$

(Assume C is given parametrically as $x = r(\theta)\cos(\theta)$, $y = r(\theta)\sin(\theta)$, for $\alpha \le \theta \le \beta$.)

38. A curve is given parametrically by $x = t(1-t^2)$, $y = t^2(1-t^3)$, for t in $[0,1]$.

(a) Sketch the points corresponding to $t = 0$, 0.2, 0.4, 0.6, 0.8, and 1.0, and use them to sketch the curve.

(b) Let $\mathcal{R}$ be the region enclosed by the curve. What difficulty arises when you try to compute the area of $\mathcal{R}$ by a definite integral involving vertical or horizontal cross sections?

(c) Use Exercise 16 to find the area of $\mathcal{R}$.

39. Repeat Exercise 38 for $x = \sin(\pi t)$ and $y = t - t^2$, for t in $[0, 1]$. In (a), let $t = 0$, $1/4$, $1/2$, $3/4$, and 1.

40. Assume that you know that Green's Theorem is true when $\mathcal{R}$ is a triangle and C its boundary.

(a) Deduce that it therefore holds for quadrilaterals.

(b) Deduce that it holds for polygons.

41. Assume that $\nabla \times \mathbf{F} = \mathbf{0}$ in the region $\mathcal{R}$ bounded by an exterior curve C_1 and two interior curves C_2 and C_3, as in Figure 18.2.11. Show that $\int_{C_1} \mathbf{F} \cdot \, d\mathbf{r} = \int_{C_2} \mathbf{F} \cdot \, d\mathbf{r} + \int_{C_3} \mathbf{F} \cdot \, d\mathbf{r}$. All three curves have a counterclockwise orientation.

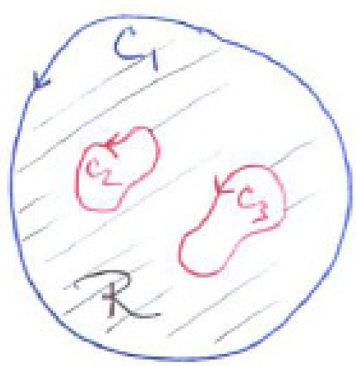


Figure 18.2.11

18.3 Green's Theorem, Flux, and Divergence

In the previous section we introduced Green's Theorem and applied it to discover a theorem about circulation and curl. That concerned the line integral of $\mathbf{F}\cdot\mathbf{T}$, the tangential component of $\mathbf{F}$, since $\mathbf{F}\cdot d\mathbf{r}$ is short for $(\mathbf{F}\cdot\mathbf{T})\ ds$. Now we will translate Green's Theorem into a theorem about the line integral of $\mathbf{F}\cdot\mathbf{n}$, the normal component of $\mathbf{F}$, $\oint \mathbf{F}\cdot\mathbf{n}\ ds$. Thus Green's Theorem will provide information about the flow of the vector field $\mathbf{F}$ across a closed curve C.

Green's Theorem Expressed in Terms of Flux

Let $\mathbf{F} = M\mathbf{i} + N\mathbf{j}$ and C be a counterclockwise closed curve. At a point on a closed curve the **unit exterior normal vector** (or **unit outward normal vector**) $\mathbf{n}$ is perpendicular to the curve and points outward from the region enclosed by the curve. To compute $\mathbf{F}\cdot\mathbf{n}$ in terms of M and N, we first express $\mathbf{n}$ in terms of $\mathbf{i}$ and $\mathbf{j}$.

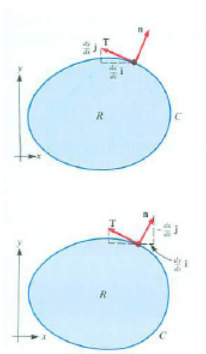


Figure 18.3.1

The vector
$$\mathbf{T} = \frac{dx}{ds}\mathbf{i} + \frac{dy}{ds}\mathbf{j}$$
is tangent to the curve, has length 1, and points in the direction in which the curve is swept out, as shown in Figure 18.3.1. This figure shows the exterior unit normal $\mathbf{n}$ has its x component equal to the y component of $\mathbf{T}$ and its y component equal to the negative of the x component of $\mathbf{T}$. Thus
$$\mathbf{n} = \frac{dy}{ds}\mathbf{i} - \frac{dx}{ds}\mathbf{j}.$$
Consequently, if $\mathbf{F} = M\mathbf{i} + N\mathbf{j}$, then
$$\begin{aligned}\oint_C \mathbf{F}\cdot\mathbf{n} &= \oint_C (M\mathbf{i}+N\mathbf{j})\cdot\left(\frac{dy}{ds}\mathbf{i} - \frac{dx}{ds}\mathbf{j}\right)\ ds = \oint_C \left(M\frac{dy}{ds} - N\frac{dx}{ds}\right)\ ds \\ &= \oint_C (M\ dy - N\ dx) = \oint_C (-N\ dx + M\ dy). \qquad (18.3.1)\end{aligned}$$
In (18.3.1), $-N$ plays the role of P and M plays the role of Q in Green's Theorem. Since Green's Theorem states that
$$\oint_C (P\ dx + Q\ dy) = \int_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\ dA$$
we have
$$\oint_C (-N\ dx + M\ dy) = \int_R \left(\frac{\partial M}{\partial x} - \frac{\partial(-N)}{\partial y}\right)\ dA$$

or, if $\mathbf{F} = M\mathbf{i} + N\mathbf{j}$, then

$$\oint_C \mathbf{F} \cdot \mathbf{n}\, ds = \int_R \left(\frac{\partial M}{\partial x} + \frac{\partial N}{\partial y} \right)\, dA.$$

In our customary notation, we have

Green's Theorem Expressed in Terms of Flux

Theorem 18.3.1. *If* $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$, *then*

$$\oint_C \mathbf{F} \cdot \mathbf{n}\, ds = \int_{\mathcal{R}} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} \right)\, dA$$

where C *is the boundary of* $\mathcal{R}$.

The expression

$$\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y},$$

the sum of two partial derivatives, is called the **divergence** of $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$. It is written $\operatorname{div} \mathbf{F}$ or $\nabla \cdot \mathbf{F}$. The latter notation is suggested by the symbolic dot product

$$\left(\frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} \right) \cdot (P\mathbf{i} + Q\mathbf{j}) = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}.$$

It is pronounced "del dot eff". Theorem 18.3.1 is called the divergence theorem in the plane. It can be written as

Divergence Theorem in the Plane

Theorem 18.3.2.

$$\oint_C \mathbf{F} \cdot \mathbf{n}\, ds = \int_{\mathcal{R}} \operatorname{div} \mathbf{F}\, dA$$

where C *is the boundary of* $\mathcal{R}$.

EXAMPLE 1 Compute the divergence of (a) $\mathbf{F} = e^{xy}\mathbf{i} + \arctan(3x)\mathbf{j}$ and (b) $\mathbf{F} = -x^2\mathbf{i} + 2xy\mathbf{j}$.
SOLUTION

(a) $\frac{\partial}{\partial x} e^{xy} + \frac{\partial}{\partial y} \arctan(3x) = ye^{xy} + 0 = ye^{xy}$

(b) $\frac{\partial}{\partial x}(-x^2) + \frac{\partial}{\partial y}(2xy) = -2x + 2x = 0$

◇

The double integral of the divergence of $\mathbf{F}$ over a region describes the amount of flow across the border of the region. It tells how rapidly the fluid is leaving (diverging) or entering the region (converging). Hence the name "divergence".

In the next section we will be using the divergence of a vector field defined in space, $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$, where P, Q and R are functions of x, y, and z. It is defined as the sum of three partial derivatives

$$\nabla \cdot \mathbf{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}.$$

It will play a role in measuring flux across a surface.

EXAMPLE 2 Verify that $\oint_C \mathbf{F} \cdot \mathbf{n}\ ds$ equals $\int_R \nabla \cdot \mathbf{F}\ dA$, when $\mathbf{F}(x, y) = x\mathbf{i} + y\mathbf{j}$, R is the disk of radius a and center at the origin and C is the boundary curve of R.

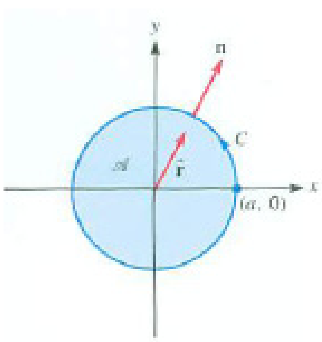


Figure 18.3.2

SOLUTION We compute $\oint_C \mathbf{F} \cdot \mathbf{n}\ ds$, where C is the circle bounding $\mathcal{R}$. See Figure 18.3.2.

Since C is a circle centered at $(0, 0)$, the unit exterior normal $\mathbf{n}$ is $\widehat{\mathbf{r}}$:

$$\mathbf{n} = \widehat{\mathbf{r}} = \frac{x\mathbf{i} + y\mathbf{j}}{|x\mathbf{i} + y\mathbf{j}|} = \frac{x\mathbf{i} + y\mathbf{j}}{a}.$$

Thus, because $\oint_C\ ds$ is the arclength of C,

$$\begin{aligned} \oint_C \mathbf{F} \cdot \mathbf{n}\ ds &= \oint_C (x\mathbf{i} + y\mathbf{j}) \cdot \left(\frac{x\mathbf{i} + y\mathbf{j}}{a}\right)\ ds = \oint_C \frac{x^2 + y^2}{a}\ ds \\ &= \oint_C \frac{a^2}{a}\ ds = a \oint_C\ ds = a(2\pi a) = 2\pi a^2. \end{aligned} \qquad (18.3.2)$$

Next we compute $\int_\mathcal{R} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right)\ dA$. Since $P = x$ and $Q = y$, $\partial P/\partial x + \partial Q/\partial y = 1 + 1 = 2$. Then

$$\int_\mathcal{R} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right)\ dA = \int_\mathcal{R} 2\ dA,$$

which is twice the area of the disk $\mathcal{R}$, and hence is $2\pi a^2$. This agrees with (18.3.2). ◇

As the next example shows, a double integral can provide a way to compute the flux $\oint \mathbf{F}\cdot\mathbf{n}\;ds$.

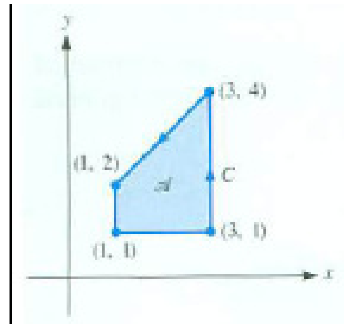


Figure 18.3.3

EXAMPLE 3 Let $\mathbf{F} = x^2\mathbf{i} + xy\mathbf{j}$. Evaluate $\oint \mathbf{F}\cdot\mathbf{n}\;ds$ over the curve that bounds the quadrilateral with vertices $(1,1)$, $(3,1)$, $(3,4)$, and $(1,2)$ shown in Figure 18.3.3.

SOLUTION The line integral could be evaluated directly, but would require parameterizing each of the four edges of C. With Green's theorem we can instead evaluate an integral over a single plane region.

Let $\mathcal{R}$ be the region that C bounds. By Green's theorem

$$\begin{aligned}\oint\limits_{C} \mathbf{F}\cdot\mathbf{n}\;ds &= \int\limits_{\mathcal{R}} \nabla\cdot\mathbf{F}\;dA = \int\limits_{\mathcal{R}} \left(\frac{\partial(x^2)}{\partial x} + \frac{\partial(xy)}{\partial y}\right)\;dA \\ &= \int\limits_{\mathcal{R}} (2x+x)\;dA = \int\limits_{\mathcal{R}} 3x\;dA.\end{aligned}$$

See Exercise 15.

Then

$$\int\limits_{\mathcal{R}} 3x\;dA = \int\limits_{1}^{3}\int\limits_{1}^{y(x)} 3x\;dy\;dx,$$

where $y(x)$ is determined by the equation of the line that provides the top edge of $\mathcal{R}$. The line through $(1,2)$ and $(3,4)$ has the equation $y = x+1$. Therefore,

$$\int\limits_{\mathcal{R}} 3x\;dA = \int\limits_{1}^{3}\int\limits_{1}^{x+1} 3x\;dy\;dx.$$

The inner integration gives

$$\int\limits_{1}^{x+1} 3x\;dy = 3xy\big|_{y=1}^{y=x+1} = 3x(x+1) - 3x = 3x^2.$$

The second integration gives

$$\int\limits_{1}^{3} 3x^2\;dx = x^3\big|_1^3 = 27 - 1 = 26.$$

◇

A Local View of div F

We have presented a global view of div $\mathbf{F}$, integrating it over a region $\mathcal{R}$ to get the total divergence across the boundary of $\mathcal{R}$. There is a way of viewing div $\mathbf{F}$, locally. It uses an extension of the Permanence Principle of Section 2.5 to the plane and to space.

Let $P = (a, b)$ be a point in the plane and $\mathbf{F}$ a vector field describing fluid flow. Choose a small region $\mathcal{R}$ around P, and let C be its boundary. See Figure 18.3.4. Then the net flow out of $\mathcal{R}$ is

$$\oint_C \mathbf{F}\cdot\mathbf{n}\; ds.$$

By Green's theorem, the net flow is also

$$\int_{\mathcal{R}} \operatorname{div}\mathbf{F}\; dA.$$

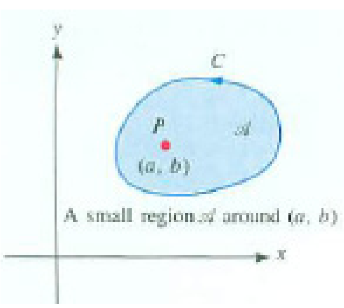


Figure 18.3.4

Since div $\mathbf{F}$ is continuous and $\mathcal{R}$ is small, div $\mathbf{F}$ is almost constant throughout $\mathcal{R}$, staying close to the divergence of $\mathbf{F}$ at (a, b). Thus

$$\int_{\mathcal{R}} \operatorname{div}\mathbf{F}\; dA \approx \operatorname{div}\mathbf{F}(a, b)\ \text{Area}(\mathcal{R}).$$

or, equivalently,

$$\frac{\text{Net flow out of } R}{\text{Area of } \mathcal{R}} \approx \operatorname{div}\mathbf{F}(a, b). \qquad (18.3.3)$$

This means that

$$\operatorname{div}\mathbf{F} \text{ at } P$$

is a measure of the rate at which fluid tends to leave a small region around P, hence another reason for the name "divergence." If div $\mathbf{F}$ is positive, fluid near P tends to get less dense (diverge). If div $\mathbf{F}$ is negative, fluid near P tends to accumulate (converge). Physicists also refer to div $\mathbf{F}$ as "flux density," for if it is multiplied by the area of a small region around it, the product approximates the flux out of the region.

Moreover, (18.3.3) suggests a different definition of the divergence div $\mathbf{F}$ at (a, b), namely

Diameter is defined in Section 17.1.

Local Definition of div $\mathbf{F}(a, b)$

$$\operatorname{div}\mathbf{F}(a, b) = \lim_{\text{Diameter of } \mathcal{R}\to 0} \frac{\oint_C \mathbf{F}\cdot\mathbf{n}\; ds}{\text{Area of } \mathcal{R}}$$

where $\mathcal{R}$ is a region enclosing (a, b) whose boundary C is a simple closed curve.

The definition appeals to our physical intuition. We began by defining $\operatorname{div}\mathbf{F}$ mathematically, as $\partial P/\partial x + \partial Q/\partial y$. We now see its physical meaning, which is independent of a coordinate system. The coordinate-free definition is the basis for Section 18.9.

EXAMPLE 4 Estimate the flux of $\mathbf{F}$ across a small circle C of radius a if $\operatorname{div}\mathbf{F}$ at the center of the circle is 3.

SOLUTION The flux of $\mathbf{F}$ across C is $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$, which equals $\int_{\mathcal{R}} \operatorname{div}\mathbf{F}\,dA$, where $\mathcal{R}$ is the disk that C bounds. Since $\operatorname{div}\mathbf{F}$ is continuous, it changes little in a small enough disk, and we treat it as almost constant. Then $\int_{\mathcal{R}} \operatorname{div}\mathbf{F}\,dA$ is approximately $(3)(\text{Area of }\mathcal{R}) = 3(\pi a^2) = 3\pi a^2$. ◇

Proof of Green's Theorem

The proof is not here just to show that Green's theorem is true. It has been known for over 150 years, and no one has said it is false. Studying a proof strengthens one's understanding of the fundamentals.

As Steve Whitaker of the chemical engineering department at the University of California at Davis has observed, "The concepts that one must understand to prove a theorem are frequently the concepts one must understand to apply the theorem."

In the proof we will use the concepts of a double integral, an iterated integral, a line integral, and the Fundamental Theorem of Calculus. So it provides a review of four basic ideas.

We prove that $\oint_{\mathcal{R}} Q\,dy = \int_{\mathcal{R}} \frac{\partial Q}{\partial x}\,dA$. The proof that $\oint_C P\,dx = -\int \frac{\partial P}{\partial y}\,dA$ is similar.

To avoid distracting details we assume that $\mathcal{R}$ is **strictly convex**. It has no dents and its border has no straight line segments. The ideas of the proof show up clearly in this special case. Thus $\mathcal{R}$ has the description $a \le x \le b$, $y_1(x) \le y \le y_2(x)$, as shown in Figure 18.3.5. We will express $\int_{\mathcal{R}} \frac{\partial Q}{\partial y}\,dA$ and

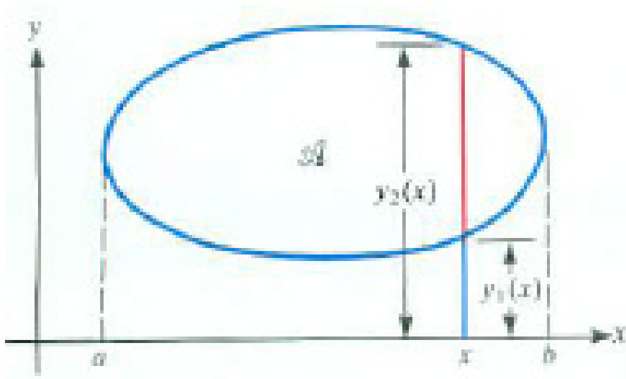


Figure 18.3.5

$\int_C Q\,dy$ as definite integrals over the interval $[a, b]$.

We have

$$\int_{\mathcal{R}} \frac{\partial Q}{\partial y}\, dA = \int_a^b \int_{y_1(x)}^{y_2(x)} \frac{\partial Q}{\partial y}\, dy\, dx.$$

By the Fundamental Theorem of Calculus,

$$\int_{y_1(x)}^{y_2(x)} \frac{\partial Q}{\partial y}\, dy = Q(x, y_2(x)) - Q(x, y_1(x)).$$

Hence

$$\int_{\mathcal{R}} \frac{\partial Q}{\partial y}\, dA = \int_a^b \left(Q(x, y_2(x)) - Q(x, y_1(x))\right)\, dx. \tag{18.3.4}$$

To express $\int_C -Q\, dx$ as an integral over $[a.b]$, break C into two paths, one along the bottom part of $\mathcal{R}$, described by $y = y_1(x)$, the other along the top part of $\mathcal{R}$, described by $y = y_2(x)$. Denote the bottom path C_1 and the top path C_2. See Figure 18.3.6.

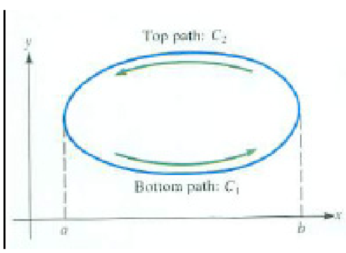


Figure 18.3.6

Then

$$\oint_C (-Q)\, dx = \int_{C_1} (-Q)\, dx + \int_{C_2} (-Q)\, dx. \tag{18.3.5}$$

But

$$\int_{C_1} (-Q)\, dx = \int_{C_1} \left(-Q(x, y_1(x))\right)\, dx = \int_a^b \left(-Q(x, y_1(x))\right)\, dx,$$

and

$$\int_{C_2} (-Q)\, dx = \int_{C_2} \left(-Q(x, y_2(x))\right)\, dx = \int_b^a \left(-Q(x, y_2(x))\right)\, dx = \int_a^b Q(x, y_2(x))\, dx.$$

Thus by (18.3.5),

$$\begin{aligned}\oint_C (-Q)\, dx &= \int_a^b -Q(x, y_1(x))\, dx + \int_a^b Q(x, y_2(x))\, dx \\ &= \int_a^b \left(Q(x, y_2(x)) - Q(x, y_1(x))\right) dx.\end{aligned}$$

This is the right side of (18.3.4) and concludes the proof.

Summary

We introduced the divergence of a vector field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$, namely the scalar field $\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}$ denoted $\operatorname{div}\mathbf{F}$ or $\nabla \cdot \mathbf{F}$.

We translated Green's theorem into a theorem about the flux of a vector field in the xy-plane,

$$\oint\limits_{C} \mathbf{F} \cdot \mathbf{n}\, ds = \int\limits_{\mathcal{R}} \operatorname{div}\mathbf{F}\, dA.$$

In words, it says that the integral of the normal component of $\mathbf{F}$ around a simple closed curve equals the integral of the divergence of $\mathbf{F}$ over the region which the curve bounds.

From this it follows that

$$\operatorname{div}\mathbf{F}(P) = \lim_{\text{diameter of } \mathcal{R} \to 0} \frac{\oint_C \mathbf{F} \cdot \mathbf{n}\, ds}{\text{Area of } \mathcal{R}} = \lim_{\text{diameter of } \mathcal{R} \to 0} \frac{\text{Flux across } C}{\text{Area of } \mathcal{R}}$$

where C is the boundary of the region $\mathcal{R}$, which contains P.

We concluded with a proof of Green's theorem.

EXERCISES for Section 18.3

1. State the divergence form of Green's theorem in symbols.

2. State the divergence form of Green's theorem in words, using no symbols.

In Exercises 3 to 6 compute the divergence of

3. $\mathbf{F} = x^3y\mathbf{i} + x^2y^3\mathbf{j}$

4. $\mathbf{F} = \arctan(3xy)\mathbf{i} + (e^{y/x})\mathbf{j}$

5. $\mathbf{F} = \ln(x+y)\mathbf{i} + xy(\arcsin y)^2\mathbf{j}$

6. $\mathbf{F} = y\sqrt{1+x^2}\mathbf{i} + \ln((x+1)^3(\sin(y))^{3/5}e^{x+y})\mathbf{j}$

In Exercises 7 to 10 compute $\int_{\mathcal{R}} \operatorname{div}\mathbf{F}\ dA$ and $\oint_C \mathbf{F}\cdot\mathbf{n}\ ds$ and check that they are equal.

7. $\mathbf{F} = 3x\mathbf{i} + 2y\mathbf{j}$, and $\mathcal{R}$ is the disk of radius 1 with center $(0,0)$.

8. $\mathbf{F} = 5y^3\mathbf{i} - 6x^2\mathbf{j}$, and $\mathcal{R}$ is the disk of radius 2 with center $(0,0)$.

9. $\mathbf{F} = xy\mathbf{i} + x^2y\mathbf{j}$, and $\mathcal{R}$ is the square with vertices $(0,0)$, $(a,0)$ (a,b) and $(0,b)$, where a, $b > 0$.

10. $\mathbf{F} = \cos(x+y)\mathbf{i} + \sin(x+y)\mathbf{j}$, and $\mathcal{R}$ is the triangle with vertices $(0,0)$, $(a,0)$ and (a,b), where a, $b > 0$.

In Exercises 11 to 14 use Green's theorem expressed in terms of divergence to evaluate $\oint_C \mathbf{F}\cdot\mathbf{n}\ ds$, where C is the boundary of R.

11. $\mathbf{F} = e^x \sin y\mathbf{i} + e^{2x}\cos(y)\mathbf{j}$, and R is the rectangle with vertices $(0,0)$, $(1,0)$, $(0,\pi/2)$, and $(1,\pi/2)$.

12. $\mathbf{F} = y\tan(x)\mathbf{i} + y^2\mathbf{j}$, and R is the square with vertices $(0,0)$, $(1,0)$, $(1,1)$, and $(0,1)$.

13. $\mathbf{F} = 2x^3y\mathbf{i} - 3x^2y^2\mathbf{j}$, and R is the triangle with vertices $(0,1)$, $(3,4)$, and $(2,7)$.

14. $\mathbf{F} = \frac{-\mathbf{i}}{xy^2} + \frac{\mathbf{j}}{x^2y}$, and R is the triangle with vertices $(1,1)$, $(2,2)$, and $(1,2)$. (Write $\mathbf{F}$ with a common denominator.)

15. In Example 3 we found $\oint_C \mathbf{F}\cdot\mathbf{n}\ ds$ by computing a double integral. Evaluate it directly.

16. Let $\mathbf{F}(x,y) = \mathbf{i}$, a constant field.

(a) Evaluate directly the flux of $\mathbf{F}$ around the triangular path, $(0,0)$ to $(1,0)$, to $(0,1)$, back to $(0,0)$.

(b) Use the divergence of $\mathbf{F}$ to evaluate the flux in (a).

17. Let a be a small number and $\mathcal{R}$ be the square with vertices (a, a), $(-a, a)$, $(-a, -a)$, and $(a, -a)$, and C its boundary. If the divergence of $\mathbf{F}$ at the origin is 3, estimate $\oint_C \mathbf{F} \cdot \mathbf{n}\, ds$.

18. Assume $|\mathbf{F}(P)| \leq 4$ for points P on a curve of length L that bounds a region $\mathcal{R}$ of area A. What can be said about the integral $\int_{\mathcal{R}} \nabla \cdot \mathbf{F}\, dA$?

19. Verify the divergence form of Green's theorem for $\mathbf{F} = 3x\mathbf{i} + 4y\mathbf{j}$ and C is the square whose vertices are $(2, 0)$, $(5, 0)$, $(5, 3)$, and $(2, 3)$.

A vector field $\mathbf{F}$ is said to be **divergence-free** or **incompressible** when $\nabla \cdot \mathbf{F} = 0$ at every point in the field.

20. Figure 18.3.7 shows four vector fields. Two are divergence-free and two are not. Decide which two are not, copy them onto a sheet of drawing paper, and sketch a closed curve C for which $\oint_C \mathbf{F} \cdot \mathbf{n}\, ds$ is not 0.

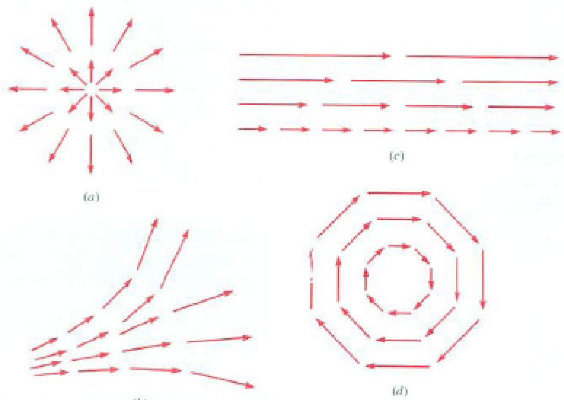


Figure 18.3.7

21. For a vector field $\mathbf{F}$,

(a) Is the curl of the gradient of $\mathbf{F}$ always $\mathbf{0}$?

(b) Is the divergence of the gradient of $\mathbf{F}$ always 0?

(c) Is the divergence of the curl of $\mathbf{F}$ always 0?

(d) Is the gradient of the divergence of $\mathbf{F}$ always $\mathbf{0}$?

22. Figure 18.3.8 shows the flow $\mathbf{F}$ of a fluid. Decide whether $\nabla \cdot \mathbf{F}$ is positive, negative, or zero at A, B, and C.

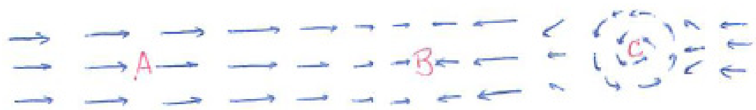


Figure 18.3.8

23. If div $\mathbf{F}$ at $(0.1, 0.1)$ is 3 estimate $\oint_C \mathbf{F}\cdot\mathbf{n}\ ds$, where C is the curve around the square whose vertices are $(0,0)$, $(0.2,0)$, $(0.2,0.2)$, and $(0,0.2)$.

24. Find the area of the region bounded by the line $y = x$ and the curve

$$\begin{aligned} x &= t^6 + t^4 \\ y &= t^3 + t \end{aligned}$$

for t in $[0,1]$. (Use Green's theorem.)

25. Let f be a scalar function. Let $\mathcal{R}$ be a convex region and C its boundary taken counterclockwise. Show that

$$\int\limits_{\mathcal{R}} \left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right)\ dA = \oint\limits_{C} \left(\frac{\partial f}{\partial x}\ dy - \frac{\partial f}{\partial x}\ dx\right).$$

26. Let $\mathbf{F}$ be the vector field whose formula in polar coordinates is $\mathbf{F}(r,\theta) = r^n\widehat{\mathbf{r}}$, where $\mathbf{r} = x\mathbf{i}+y\mathbf{j}$, $r = |\mathbf{r}|$, and $\widehat{\mathbf{r}} = \mathbf{r}/r$. Show that the divergence of $\mathbf{F}$ is $(n+1)r^{n-1}$. (First express $\mathbf{F}$ in rectangular coordinates.)

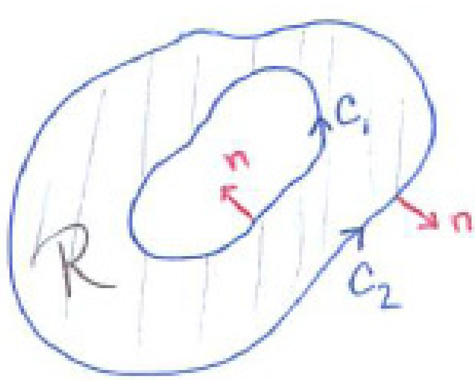


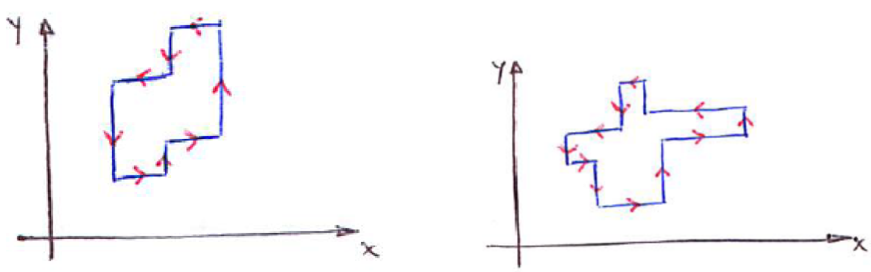
Figure 18.3.9

27. A region R with a hole is bounded by two oriented curves C_1 and C_2, as in Figure 18.3.9, which includes exterior pointing unit normal vectors.

(a) Show that $\oint_{C_1} \mathbf{F}\cdot\mathbf{n}\ ds - \int_{C_2} \mathbf{F}\cdot\mathbf{n}\ ds = \int_R (\nabla\cdot\mathbf{F})\ dA$. (Break R into two regions that have no holes, as in Exercises 34 and 35.)

(b) If $\nabla\cdot\mathbf{F} = 0$ in $\mathcal{R}$, show that $\int_{C_1} \mathbf{F}\cdot\mathbf{n}\ ds = \int_{C_2} \mathbf{F}\cdot\mathbf{n}\ ds$.

28. Let $\mathbf{F}$ be a vector field in the xy-plane whose flux across any rectangle is 0. Show that its flux across the curves in Figure 18.3.10(a) and (b) is also 0.



(a) (b)

Figure 18.3.10

29. Assume that the circulation of $\mathbf{F}$ along every circle in the xy-plane is 0. Must $\mathbf{F}$ be conservative?

30. The line integral for flux, $o\int_C \mathbf{F}\cdot\mathbf{n}\ ds$, and for circulation it is $\oint_C \mathbf{F}\cdot\mathbf{T}\ ds$. Why is the first integral independent of the orientation of the curve but the second one is dependent on the orientation?

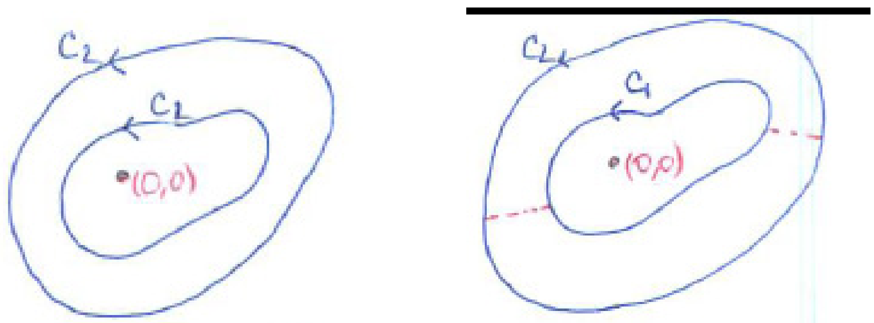
31. The field $\mathbf{F}$ is defined throughout the xy-plane. If the flux of $\mathbf{F}$ across every circle is 0, must the flux of $\mathbf{F}$ across every square be 0? Explain.

32. Let $\mathbf{F}(x,y)$ describe a fluid flow. Assume $\nabla\cdot\mathbf{F}$ is never 0 in a certain region R. Show that none of the streamlines in the region forms a loop within $\mathcal{R}$. (At each point P on a stream line, $\mathbf{F}(P)$ is tangent to the streamline.)

33. Let $\mathcal{R}$ be a region in the xy-plane bounded by the closed curve C. Let $f(x,y)$ be defined on the plane. Show that

$$\int\limits_{\mathcal{R}} \left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial x^2}\right)\ dA = \oint\limits_{C} D_{\mathbf{n}}(f)\ ds.$$

34. Assume that $\mathbf{F}$ is defined everywhere in the xy-plane except at the origin and that the divergence of $\mathbf{F}$ is identically 0. Let C_1 and C_2 be two counterclockwise simple curves circling the origin. C_1 lies inside the region within C_2. Show that $\oint_{C_1} \mathbf{F} \cdot \mathbf{n}\, ds = \int_{C_2} \mathbf{F} \cdot \mathbf{n}\, ds$.
See Figure 18.3.11(a). (Draw the dashed lines in Figure 18.3.11(b) to cut the region between C_1 and C_2 into two regions.)



(a) (b)

Figure 18.3.11

35. (This continues Exercise 34.) Assume that $\mathbf{F}$ is defined everywhere in the xy-plane except at the origin and that the divergence of $\mathbf{F}$ is identically 0. Let C_1 and C_2 be two counterclockwise simple curves circling the origin. They may intersect. Show that $\oint_{C_1} \mathbf{F} \cdot \mathbf{n}\, ds = \oint_{C_2} \mathbf{F} \cdot \mathbf{n}\, ds$. The exercise shows that if the divergence of $\mathbf{F}$ is 0, an integral over a complicated curve can be replaced by an integral over a simpler curve.

36.

(a) Draw enough vectors for the field $\mathbf{F}(x, y) = (x\mathbf{i} + y\mathbf{j})/(x^2 + y^2)$ to show what it looks like.

(b) Compute $\nabla \cdot \mathbf{F}$.

(c) Does your sketch in (a) agree with what you found for $\nabla \cdot \mathbf{F}$. in (b)? If not, redraw the vector field.

37. We proved that $\int_{\mathcal{R}} \frac{\partial Q}{\partial y}\, dA = \int_C Q\, dy$ in a special case. Prove it in this more general case, in which we assume less about the region $\mathcal{R}$. Assume that $\mathcal{R}$ has the description $a \leq x \leq b$, $y_1(x) \leq y \leq y_2(x)$. Figure 18.3.12 shows such a region, which need not be convex. The curved path C breaks up into four paths, two of which are straight (or may be empty), as in Figure 18.3.12.

38. We proved the second part of (18.2.1), namely that $\oint_C Q\ dy = \int_R \partial Q/\partial x\ dA$. Prove the first part, $\oint_C P\ dx = -\int_R \partial p/\partial y\ dA$.

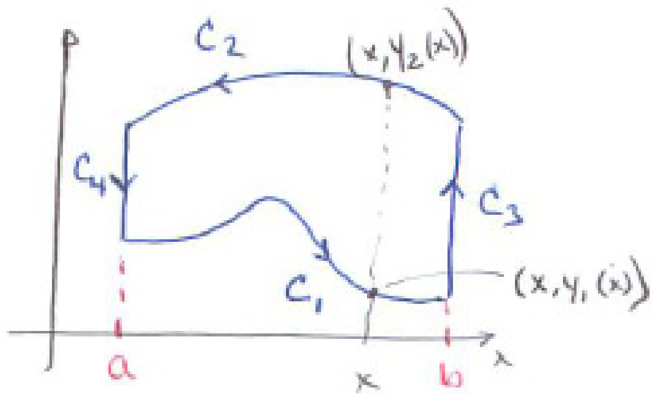


Figure 18.3.12

18.4 Central Vector Fields and Steradians

Central vector fields are a special but important type of vector field that appears in the study of gravity and the attraction or repulsion of electric charges. They radiate from a point mass or point charge. Physicists invented them in order to avoid action at a distance. One particle acts on another directly, through the vector field it creates.

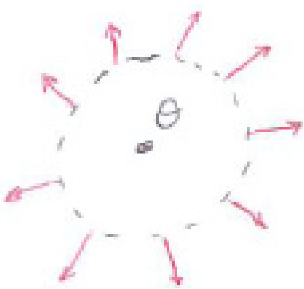


Figure 18.4.1

Central Vector Fields

A **central vector field** is a continuous vector field defined everywhere in the plane (or in space) except, perhaps, at a point $\mathcal{O}$, with the properties:

1. Each vector points towards (or away from) $\mathcal{O}$.

2. The magnitudes of all vectors at a given distance from $\mathcal{O}$ are equal.

The point $\mathcal{O}$ is called the center, or pole, of the field. A central vector field is also called radially symmetric. There are various ways to think of a central vector field. For one in the plane, the vectors at points on a circle with center $\mathcal{O}$ are perpendicular to the circle and have the same length, as shown in Figures 18.4.1 and 18.4.2.

The same holds for central vector fields in space, with "circle" replaced by "sphere."

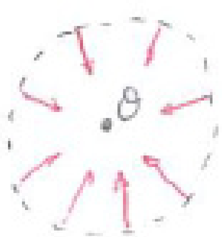


Figure 18.4.2

The formula for a central vector field has a simple form. Let the field be $\mathbf{F}$ and let P be any point other than $\mathcal{O}$. Denote the vector $\overrightarrow{OP}$ by $\mathbf{r}$, its magnitude by r, and the unit vector $\mathbf{r}/r$ by $\widehat{\mathbf{r}}$. Then there is a scalar function f, defined for all positive numbers, such that

$$\mathbf{F}(P) = f(r)\widehat{\mathbf{r}}.$$

The magnitude of $\mathbf{F}(P)$ is $|f(r)|$. If $f(r)$ is positive, $\mathbf{F}(P)$ points away from $\mathcal{O}$. If $f(r)$ is negative, $\mathbf{F}(P)$ points toward $\mathcal{O}$.

A central field is a vector-valued function of more than one variable. Because $\mathbf{r} = \overrightarrow{OP} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ we may denote $\mathbf{F}(P)$ as $\mathbf{F}(x, y, z)$ or $\mathbf{F}(\mathbf{r})$.

Central Vector Fields in the Plane

Using polar coordinates with pole at $\mathcal{O}$, we can express a central field in the xy-plane in the form

$$\mathbf{F}(\mathbf{r}) = f(r)\widehat{\mathbf{r}},$$

where $r = |\mathbf{r}|$ and $\widehat{\mathbf{r}} = \mathbf{r}/r$. The magnitude of $\mathbf{F}(\mathbf{r})$ is $|f(r)|$.

We already met such a field in Section 18.1 in the study of line integrals. There, $f(r) = 1/r$ so the field varied as the inverse first power. When, in Section 15.4, we encountered the line integral for the normal component of the field along a curve we found that it gives the number of radians the curve subtends.

See page 1280.

The vector field $\mathbf{F}(\mathbf{r}) = (1/r)\widehat{\mathbf{r}}$ can also be written as

$$\mathbf{F}(\mathbf{r}) = \frac{\mathbf{r}}{r^2}. \tag{18.4.1}$$

Its magnitude is not inversely proportional to the square of r because the magnitude of $\mathbf{r}/r^2$ is $r/r^2 = 1/r$, the reciprocal of the first power of r.

EXAMPLE 1 Evaluate the flux $\oint_C \mathbf{F}\cdot\mathbf{n}\ ds$ for the central field $\mathbf{F}(x,y) = f(r)\widehat{\mathbf{r}}$, where $\mathbf{r} = x\mathbf{i} + y\mathbf{j}$, over the closed curve shown in Figure 18.4.3. We have $a < b$ and the path goes from $A = (a,0)$ to $B = (b,0)$ to $C = (0,b)$ to $D = (0,a)$ and ends at $A = (a,0)$.

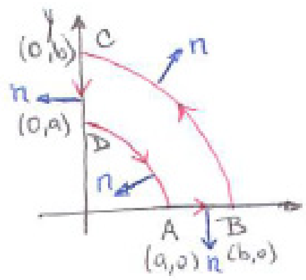


Figure 18.4.3

SOLUTION On the paths from A to B and from C to D the exterior normal, $\mathbf{n}$, is perpendicular to $\mathbf{F}$, so $\mathbf{F}\cdot\mathbf{n} = 0$, and these integrands contribute nothing to the integral. On BC, $\mathbf{F}$ equals $f(b)\widehat{\mathbf{r}}$. There $\widehat{\mathbf{r}} = \mathbf{n}$, so $\mathbf{F}\cdot\mathbf{n} = f(b)$ since $\mathbf{r}\cdot\mathbf{n} = 1$. The length of arc BC is $(2\pi b)/4 = \pi b/2$. Thus

$$\int_B^C \mathbf{F}\cdot\mathbf{n}\ ds = \int_B^C f(b)\ ds = f(b)\int_B^C ds = \frac{\pi}{2}bf(b).$$

On the arc DC, $\widehat{\mathbf{r}} = -\mathbf{n}$. A similar calculation shows that

$$\int_D^C \mathbf{F}\cdot\mathbf{n}\ ds = -\frac{\pi}{2}af(a).$$

Hence

$$\oint_C \mathbf{F}\cdot\mathbf{n}\ ds = 0 + \frac{\pi}{2}bf(b) + 0 - \frac{\pi}{2}af(a) = \frac{\pi}{2}(bf(b) - af(a)).$$

◇

For a central field $f(r)\widehat{\mathbf{r}}$ to have zero flux around all paths of the special type shown in Figure 18.4.3, we must have

$$f(b)b - f(a)a = 0$$

for positive a and b. In particular,

$$f(b)b - f(1)1 = 0 \qquad \text{or} \qquad f(b) = \frac{f(1)}{b}.$$

Thus $f(r)$ is inversely proportional to r and there is a constant c such that

$$f(r) = \frac{c}{r}.$$

If $f(r)$ is not of the form c/r, the vector field $\mathbf{F}(x,y) = f(r)\widehat{\mathbf{r}}$ does not have zero flux across these paths. In Exercise 5 you may compute the divergence of $(c/r)\widehat{\mathbf{r}}$ and show that it is zero.

The only central vector fields with center at the origin in the plane with zero divergence are these whose magnitude is inversely proportional to the distance from the origin.

In space the only central fields with zero flux across closed surfaces have a magnitude inversely proportional to the square of the distance to the pole, as we will see in a moment.

Knowing that the central field $\mathbf{F} = \widehat{\mathbf{r}}/r$ has zero divergence enables us to evaluate line integrals of the form $\oint_C \frac{\widehat{\mathbf{r}}\cdot\mathbf{n}}{r}\,ds$, as the next example shows.

EXAMPLE 2 Let $\mathbf{F}(\mathbf{r}) = \widehat{\mathbf{r}}/r$. Evaluate $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$ where C is the counterclockwise circle of radius 1 and center $(2,0)$, as shown in Figure 18.4.4.

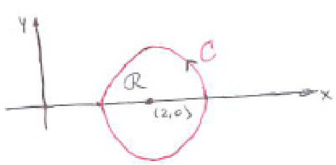


Figure 18.4.4

SOLUTION Exercise 5 shows that the field $\mathbf{F}$ has divergence zero throughout the region R that C bounds. By Green's theorem, the line integral equals the integral of the divergence over R:

$$\oint_C \mathbf{F}\cdot\mathbf{n}\,ds = \int_R \nabla\cdot\mathbf{F}\,dA. \qquad (18.4.2)$$

Since the divergence of $\mathbf{F}$ is 0 throughout R, the right side of (18.4.2) is 0. Therefore $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds = 0$. ◇

The next example involves a curve that surrounds a point where the vector field $\mathbf{F} = \widehat{\mathbf{r}}/r$ is not defined.

EXAMPLE 3 Let C be a simple closed curve enclosing the origin. Evaluate $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$, where $\mathbf{F} = \widehat{\mathbf{r}}/r$.

SOLUTION Figure 18.4.5 shows C and a small circle D centered at the origin and in the region that C bounds. Without a formula for C, we can not compute $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds$ directly. However, since the divergence of $\mathbf{F}$ is 0

See page 1280 in Section 15.4.

throughout the region bounded by C and D, we have, by the two-curve case of Green's theorem,

$$\oint\limits_C \mathbf{F}\cdot\mathbf{n}\,ds = \oint\limits_D \mathbf{F}\cdot\mathbf{n}\,ds. \tag{18.4.3}$$

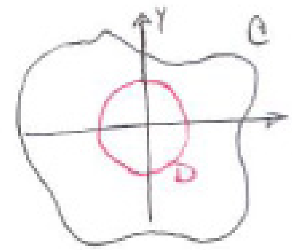


Figure 18.4.5

The integral on the right-hand side of (18.4.3) computed directly. To do so, let the radius of D be a. Then for points P on D, $\mathbf{F}(P) = \widehat{\mathbf{r}}/a$. Because $\widehat{\mathbf{r}}$ and $\mathbf{n}$ are the same unit vector, $\widehat{\mathbf{r}}\cdot\mathbf{n} = 1$. Thus

$$\oint\limits_D \mathbf{F}\cdot\mathbf{n}\,ds = \oint\limits_D \frac{\widehat{\mathbf{r}}\cdot\mathbf{n}}{a}\,ds = \int\limits_D \frac{1}{a}\,ds = \frac{1}{a}2\pi a = 2\pi.$$

Hence $\oint_C \mathbf{F}\cdot\mathbf{n}\,ds = 2\pi$. ◇

This should not be a surprise, for the integral equals the number of radians that C subtends.

Central Vector Fields in Space

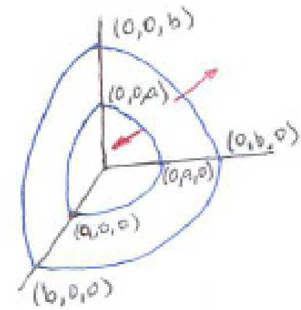


Figure 18.4.6

A central vector field in space with center at the origin has the form $\mathbf{F}(x,y,z) = \mathbf{F}(\mathbf{r}) = f(r)\widehat{\mathbf{r}}$. We show that if the flux of $\mathbf{F}$ over surfaces described below is zero then $f(r)$ must be inversely proportional to the square of r.

The surface S shown in Figure 18.4.6 consists of an octant of two concentric spheres, one of radius a, the other of radius b, $a < b$, together with the flat surfaces on the coordinate planes. Let $\mathcal{R}$ be the region bounded by S. On its three flat sides $\mathbf{F}$ is perpendicular to the exterior normal. On the outer sphere $\mathbf{F}(x,y,z)\cdot\mathbf{n} = f(b)$. On the inner sphere $\mathbf{F}(x,y,z)\cdot\mathbf{n} = -f(a)$. Thus

Surface area of a sphere of radius r is $4\pi r^2$.

$$\oint\limits_S \mathbf{F}\cdot\mathbf{n}\,dS = f(b)(\frac{1}{8})(4\pi b^2) - f(a)(\frac{1}{8})(4\pi a^2) = \frac{\pi}{2}(f(b)b^2 - f(a)a^2).$$

Since this is 0 for positive a and b, it follows that there is a constant c such that

Compare with Example 1.

$$f(r) = \frac{c}{r^2}.$$

The magnitude must be proportional to the inverse square.

The following assertion is justified in Exercise 28:

The only central vector field with center at the origin in space with zero divergence are these whose magnitude is inversely proportional to the square of the distance from the origin.

A Geometric Application: Steradians (Revisited)

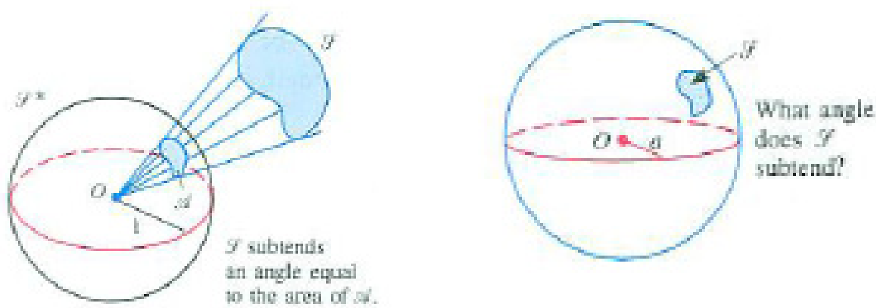
SHERMAN: We changed this significantly, so I cannot completely follow Woody's suggestions. Please read this particularly carefully.

In Section 17.7 we saw that the inverse square central field plays an important role in geometry. The purpose of the current discussion is to derive the formula for the steradian measure of the angle subtended by a surface as a surface integral involving the inverse square central field.

In Section 15.4 we showed how radian measure could be expressed in terms of the line integral $\int_C (\widehat{\mathbf{r}}/r) \cdot \mathbf{n}\, ds$, that is, in terms of the central field whose magnitude is inversely proportional to the first power of the distance from the center. That was based on circular arcs in a plane. Measuring solid angles involves patches on surfaces of spheres.

Let $\mathcal{O}$ be a point and $\mathcal{S}$ a surface such that each ray from $\mathcal{O}$ meets $\mathcal{S}$ in at most one point. Let $\mathcal{S}^*$ be the unit sphere with center at $\mathcal{O}$. The rays from $\mathcal{O}$ that meet $\mathcal{S}$ intersect $\mathcal{S}^*$ in a set that we call $\mathcal{R}$, as shown in Figure 18.4.7(a). Let the area of $\mathcal{R}$ be A. The solid angle subtended by $\mathcal{S}$ at $\mathcal{O}$ is said to have a measure of A steradians

For instance, a closed surface $\mathcal{S}$ that encloses $\mathcal{O}$ subtends a solid angle of 4π steradians because the area of the unit sphere is 4π.



(a) (b)

Figure 18.4.7

EXAMPLE 4 Let $\mathcal{S}$ be part of the surface of a sphere of radius a, $\mathcal{S}_a$, whose center is $\mathcal{O}$. Find the angle subtended by $\mathcal{S}$ at $\mathcal{O}$. See Figure 18.4.7(b).

SOLUTION The entire sphere $\mathcal{S}_a$ subtends an angle of 4π steradians because it has area $4\pi a^2$. We therefore have the proportion

$$\frac{\text{Angle } \mathcal{S} \text{ subtends}}{\text{Angle } \mathcal{S}_a \text{ subtends}} = \frac{\text{Area of } \mathcal{S}}{\text{Area of } \mathcal{S}_a},$$

or

$$\frac{\text{Angle } \mathcal{S} \text{ subtends}}{4\pi} = \frac{\text{Area of } \mathcal{S}}{4\pi a^2}.$$

Hence

$$\text{Angle } \mathcal{S} \text{ subtends} = \frac{\text{Area of } \mathcal{S}}{a^2} \quad \text{steradians.}$$

◇

EXAMPLE 5 Let $\mathcal{S}$ be a surface such that each ray from $\mathcal{O}$ meets $\mathcal{S}$ in at most one point. Find an integral that represents in steradians the solid angle that $\mathcal{S}$ subtends at $\mathcal{O}$.

SOLUTION Consider a small patch of $\mathcal{S}$. Call it $d\mathcal{S}$ and let its area be dA. If we can estimate the angle that it subtends at $\mathcal{O}$, then we will have the local approximation that will tell us what integral represents the total solid angle subtended by $\mathcal{S}$.

Let $\mathbf{n}$ be a unit normal at a point in the patch, which we regard as essentially flat, as in Figure 18.4.8. Let $d\mathcal{A}$ be the projection $d\mathcal{S}$ on a plane perpendicular to $\mathbf{r}$, as shown in Figure 18.4.8. The area of $d\mathcal{A}$ is approximately dA, where

$$dA = \widehat{\mathbf{r}} \cdot \mathbf{n}\ dS.$$

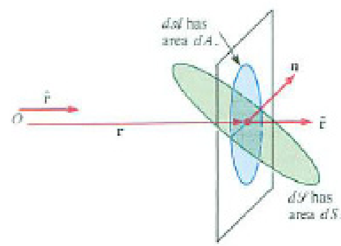


Figure 18.4.8

Both $d\mathcal{S}$ and $d\mathcal{A}$ subtend approximately the same solid angle, which according to Example 4 is about

$$\frac{\widehat{\mathbf{r}} \cdot \mathbf{n}}{|\mathbf{r}|^2}\ dS \quad \text{steradians.}$$

Consequently $\mathcal{S}$ subtends a solid angle of

$$\int\limits_{\mathcal{S}} \frac{\widehat{\mathbf{r}} \cdot \mathbf{n}}{|\mathbf{r}|^2}\ dS \quad \text{steradians.}$$

◇

The following special case will be used in Section 18.5.

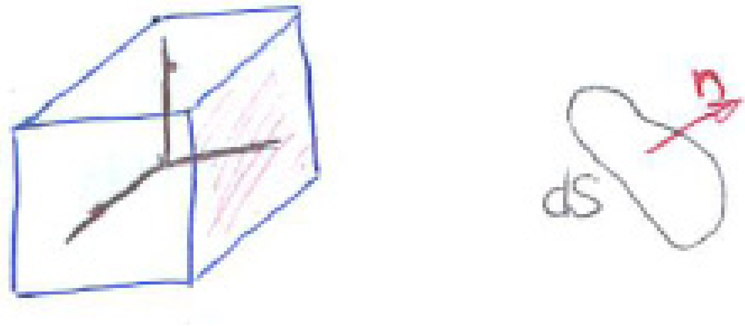
SHERMAN: Woody didn't want this in a box. I tried adding a title, but I can easily take this back to plain text. Your thoughts?

A Closed Surface Subtends a Solid Angle of 4π Steradians

Let $\mathcal{O}$ be a point in the region bounded by the closed surface $\mathcal{S}$. Assume each ray from $\mathcal{O}$ meets $\mathcal{S}$ in exactly one point, and let $\mathbf{r}$ denote the position vector from $\mathcal{O}$ to that point. Then

$$\int\limits_{\mathcal{S}} \frac{\widehat{\mathbf{r}} \cdot \mathbf{n}}{r^2}\ dS = 4\pi. \tag{18.4.4}$$

When $\mathcal{S}$ is a sphere of radius a and center at the origin, $\widehat{\mathbf{r}} = \mathbf{n}$, so $\widehat{\mathbf{r}} \cdot \mathbf{n} = 1$. Also, $r = a$. Then (18.4.4) becomes $\int_{\mathcal{S}} (1/a^2)\ dS = (1/a^2)4\pi a^2 = 4\pi$. However,



(a) (b)

Figure 18.4.9

it is not obvious that (18.4.4) holds when $\mathcal{S}$ is a sphere and the origin is not its center, or when $\mathcal{S}$ is not a sphere. **EXAMPLE 6** Let $\mathcal{S}$ be the cube of side

2 bounded by the planes $x = \pm 1$, $y = \pm 1$, $z = \pm 1$, shown in Figure 18.4.9(a). Find $\oint_{\mathcal{S}} \frac{\widehat{\mathbf{r}} \cdot \mathbf{n}}{r^2}\, dS$, where $\mathcal{S}$ is one of the faces of the cube.
SOLUTION Each of the faces subtends the same solid angle at the origin. Since the entire surface subtends 4π steradians, each face subtends $4\pi/6 = 2\pi/3$ steradians. Then the flux over each face is

$$\int_{\mathcal{S}} \frac{\widehat{\mathbf{r}} \cdot \mathbf{n}}{r^2}\, dS = \frac{2\pi}{3}.$$

$\diamond$

In physics books the integral $\int_{\mathcal{S}} \frac{\widehat{\mathbf{r}} \cdot \mathbf{n}}{r^2}\, dS$ is also written as

$$\int_{\mathcal{S}} \frac{\mathbf{r} \cdot \mathbf{n}}{r^3}\, dS, \qquad \int_{\mathcal{S}} \frac{\widehat{\mathbf{r}} \cdot d\mathbf{S}}{r^2}, \qquad \int_{\mathcal{S}} \frac{\mathbf{r} \cdot d\mathbf{S}}{r^3}, \quad \text{or} \quad \int_{\mathcal{S}} \frac{\cos(\mathbf{r}, \mathbf{n})}{r^2}\, dS.$$

The symbol $d\mathbf{S}$ is short for $\mathbf{n}\, dS$, and calls to mind Figure 18.4.9(b), which shows a small patch on the surface, together with an exterior normal unit vector.

Recall that $\cos(\mathbf{r}, \mathbf{n})$ denotes the cosine of the angle between $\mathbf{r}$ and $\mathbf{n}$; see also Section 14.2.

Summary

We investigated central vector fields. Vector fields whose divergence is always zero are called divergence-free or incompressible. In the plane the only divergence-free central fields are of the form $(c/r)\widehat{\mathbf{r}}$ where c is a constant, an inverse first power. In space the only incompressible central fields are of the form $(c/r^2)\widehat{\mathbf{r}}$, an inverse second power. The field $\widehat{\mathbf{r}}/r^2$ can be used to express the size of a solid angle of a surface $\mathcal{S}$ in steradians as an integral, $\int_{\mathcal{S}} \widehat{\mathbf{r}}\cdot\mathbf{n}/r^2\, dS$. In particular, if $\mathcal{S}$ encloses the center of the field, then $\int_{\mathcal{S}} \widehat{\mathbf{r}} \cdot \mathbf{n}/r^2\, dS = 4\pi$. Divergence-free vector fields are discussed again in Section 18.6.

EXERCISES for Section 18.4

1. Define a central field in words, using no symbols.

2. Define a central field with center at $\mathcal{O}$ in symbols.

3. Give an example of a non-constant central field in the plane that

(a) does not have zero divergence,

(b) has zero divergence.

4. Give an example of a non-constant central field in space that

(a) is not divergence-free,

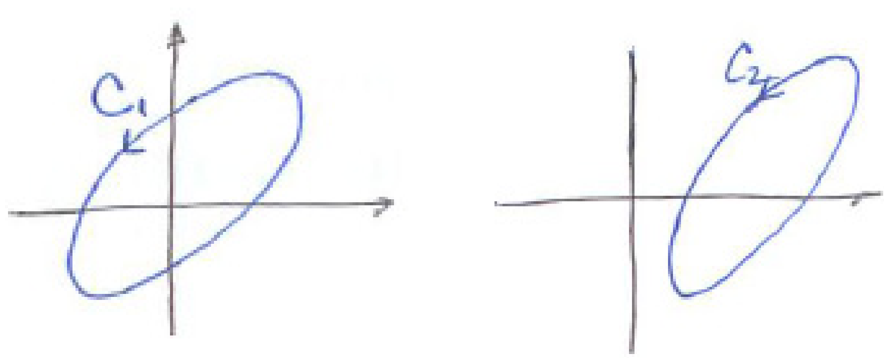
(b) is divergence-free.

5. Let $\mathbf{F}(x, y)$ be an inverse-first-power central field in the plane $\mathbf{F}(x, y) = (c/r)\widehat{\mathbf{r}}$, where $\mathbf{r} = x\mathbf{i} + y\mathbf{j}$. Compute the divergence of $\mathbf{F}$. (First write $\mathbf{F}(x, y)$ as $\frac{cx\mathbf{i}+cy\mathbf{j}}{x^2+y^2}$.)

6. Show that the curl of a central vector field in the plane is $\mathbf{0}$.

7. Show that the curl of a central vector field in space is $\mathbf{0}$.

8. Let $\mathbf{F}(\mathbf{r}) = \widehat{\mathbf{r}}/r$. Evaluate the flux $\oint_C \mathbf{F} \cdot \mathbf{n}\, ds$ as simply as you can for the ellipses in Figure 18.4.10.



(a) (b)

Figure 18.4.10

9. Figure 18.4.11 shows a cube of side 2 with one corner at the origin.

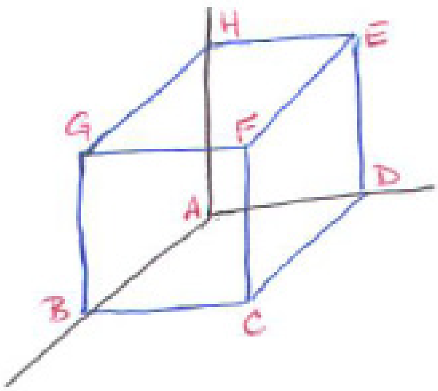


Figure 18.4.11

Evaluate the integral of the function $\widehat{\mathbf{r}} \cdot \mathbf{n}/r^2$ over

(a) the square $EFGH$

(b) the square $ABCD$

(c) the entire surface of the cube.

10. Let $\mathbf{F}(\mathbf{r}) = \widehat{\mathbf{r}}/r^3$. Evaluate the flux of $\mathbf{F}$ over the sphere of radius 2 and center at the origin.

11. A pyramid is made of four congruent equilateral triangles. Find the steradians subtended by one face at the centroid of the pyramid. No integration is necessary.

12. How many steradians does one face of a cube subtend at

(a) One of the four vertices not on that face?

(b) The center of the cube? No integration is necessary.

13. In Example 2 the integral $\oint_C \widehat{\mathbf{r}} \cdot \mathbf{n}/r \; ds$ was 0. How would you explain this in terms of subtended angles and steradians?

14. Let $\mathbf{F}$ and $\mathbf{G}$ be central vector fields in the plane with different centers.

(a) Show that $\mathbf{F} + \mathbf{G}$ is not a central field.

(b) Show that the divergence of $\mathbf{F} + \mathbf{G}$ is 0.

15. In Example 6, we evaluated a surface integral by interpreting it in terms of the size of a subtended solid angle. Evaluate the integral directly, without that knowledge.

16. Let $\mathcal{S}$ be the triangle whose vertices are $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$. Evaluate $\int_{\mathcal{S}} \frac{\widehat{\mathbf{r}} \cdot \mathbf{n}}{r^2}\, dS$ by using steradians.

17. Evaluate the integral in Exercise 16 directly.

18. Let $\mathbf{F}(x, y, z) = \frac{x\mathbf{i}+y\mathbf{j}+0\mathbf{k}}{x^2+y^2}$ be a vector field in space.

(a) What is the domain of $\mathbf{F}$?

(b) Sketch $\mathbf{F}(1, 1, 0)$ and $\mathbf{F}(1, 1, 2)$ with tails at the given points.

(c) Show $\mathbf{F}$ is not a central field.

(d) Show its divergence is 0.

Exercises 19 to 26 are related.

19. Let $\mathbf{F}$ be a planar central field. Show that $\nabla \times \mathbf{F}$ is $\mathbf{0}$. ($\mathbf{F}(x, y) = \frac{g(\sqrt{x^2+y^2}(x\mathbf{i}+y\mathbf{j}))}{\sqrt{x^2+y^2}}$ for some scalar function g.)

20. (This continues Exercise 19.) Show that $\mathbf{F}$ is a gradient field; to be specific, $\mathbf{F} = \nabla g(\sqrt{x^2 + y^2})$.

21. Carry out the computation to show that the only central fields in space that have zero divergence have the form $\mathbf{F}(\mathbf{r}) = c\widehat{\mathbf{r}}/r^2$ if the origin of the coordinates is at the center of the field.

22. If we worked in four-dimensional space instead of the two-dimensional plane or three-dimensional space, which central fields do you think would have zero divergence? Carry out a calculation to confirm your conjecture.

23. Let $\mathbf{F} = \widehat{\mathbf{r}}/r^2$ and S be the surface of the lopsided pyramid with square base, whose vertices are $(0, 0, 0)$, $(1, 1, 0)$, $(0, 1, 0)$, $(0, 1, 1)$, and $(1, 1, 1)$.

(a) Sketch the pyramid.

(b) What is the integral of $\mathbf{F} \cdot \mathbf{n}$ over the square base?

(c) What is the integral of $\mathbf{F} \cdot \mathbf{n}$ over each of the remaining four faces?

(d) Evaluate $\oint_S \mathbf{F} \cdot \mathbf{n}\, dS$.

24. Let C be the circle $x^2 + y^2 = 4$ in the xy-plane. For each point Q in the disk bounded by C consider the central field with center Q, $\mathbf{F}(P) = \overrightarrow{PQ}/|PQ|^2$. Its magnitude is inversely proportional to the first power of the distance from P to Q. Evaluate the flux of $\mathbf{F}$ across C

(a) when Q is the origin $(0, 0)$

(b) when Q is not the origin

(c) when Q lies on C.

25. Let $\mathbf{F}$ be the central field in the plane, with center at $(1, 0)$ and with magnitude inversely proportional to the first power of the distance to $(1, 0)$: $\mathbf{F}(x, y) = \frac{(x-1)\mathbf{i}+y\mathbf{j}}{|(x-1)\mathbf{i}+y\mathbf{j}|^2}$. Let C be the circle of radius 2 and center at $(0, 0)$.

(a) By thinking in terms of subtended angle, evaluate the flux $\oint_C \mathbf{F} \cdot \mathbf{n}\, ds$.

(b) Evaluate it by carrying out the integration.

26. This exercise gives a geometric way to see why a central force is conservative. Let $\mathbf{F}(x, y) = f(r)\widehat{\mathbf{r}}$. Figure 18.4.12 shows $\mathbf{F}(x, y)$ and a short vector $\overrightarrow{d\mathbf{r}}$ and two circles.

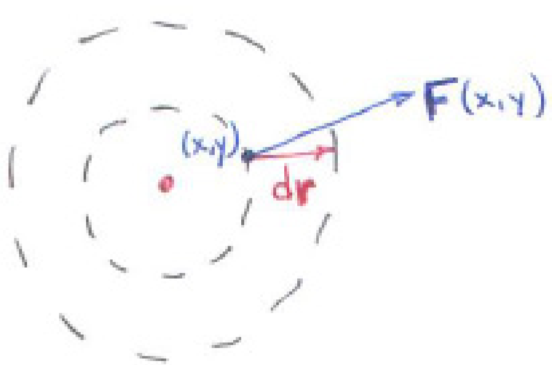


Figure 18.4.12

(a) Why is $\mathbf{F}(x, y) \cdot\ d\mathbf{r}$ approximately $f(r)\ dr$, where dr is the difference in the radii of the two circles?

(b) Let C be a curve from A to B, where $A = (a, \alpha)$ and $B = (b, \beta)$ in polar coordinates. Why is $\int_C \mathbf{F} \cdot \, d\mathbf{r} = \int_a^b f(r)\, dr$?

(c) Why is $\mathbf{F}$ conservative?

27. Show that the derivative of $\frac{1}{3}\tan^3(x) - \tan(x) + x$ is $\tan^4(x)$.

28. Use integration by parts to show that

$$\int \tan^n(x)\, dx = \frac{\tan^{n-1}(x)}{n-1} - \int \tan^{n-2}(x)\, dx.$$

29. Entry 16 in the table of antiderivatives in the front of this book is

$$\int \frac{dx}{x(ax+b)} = \frac{1}{b} \ln \left| \frac{x}{ax+b} \right|.$$

(a) Use a partial fraction expansion to evaluate the antiderivative.

(b) Use differentiation to check that the formula is correct.

30. Repeat Exercise 29 for entry 17 in the Table of Antiderivatives:

$$\int \frac{dx}{x^2(ax+b)} = \frac{-1}{bx} + \frac{a}{b^2} \ln \left| \frac{ax+b}{x} \right|.$$

31. Show that $x \arccos(x) - \sqrt{1-x^2}$ is an antiderivative of $\arccos(x)$.

32. Find $\int \arctan(x)$.

33.

(a) Find $\int xe^{ax}\, dx$.

(b) Use integration by parts to show that

$$\int x^m e^{ax}\, dx = \frac{x^m e^{ax}}{a} - \frac{m}{a} \int x^{m-1} e^{ax}\, dx.$$

(c) Verify the equation in (b) by differentiating the right hand side.

Radiation, light, or sound comes from a point source at a distance h from a plane. In applications it is important to know the fraction of the signal that strikes the plane hits a disk located in that plane. There are tables that list that fraction as a function of h and the distance the center of the disk is from the point in the plane closest to the source (with the radius of the disk taken as 1). Exercises 34 to 37 concern the special case where the center of the disk is on the line through the source perpendicular to the plane of the disk.

34. Let the radius of the disk be a. Let $s(a)$ be the number of steradians subtended by the disk. Explain why the fraction of interest equals $s(a)/(2\pi)$.

35. Recall that the surface area of a sphere between two parallel planes that intersect the sphere is proportional to the distance between the planes. Use this information (and an integration) to show that

$$s(a) = 2\pi\left(1 - \frac{1}{\sqrt{z^2 + h^2}}\right).$$

36. Use the integral for the flux of the field $\widehat{\mathbf{r}}/r^2$ to find $s(a)$. (The integration is much easier in polar coordinates.)

37. For $h = 0.8$ and $a = 1$ a table lists $s(a) = 2.35811$.

(a) Does that agree with the formula obtained in Exercise 35?

(b) What fraction of the radiation that strikes the plane hits the disk?

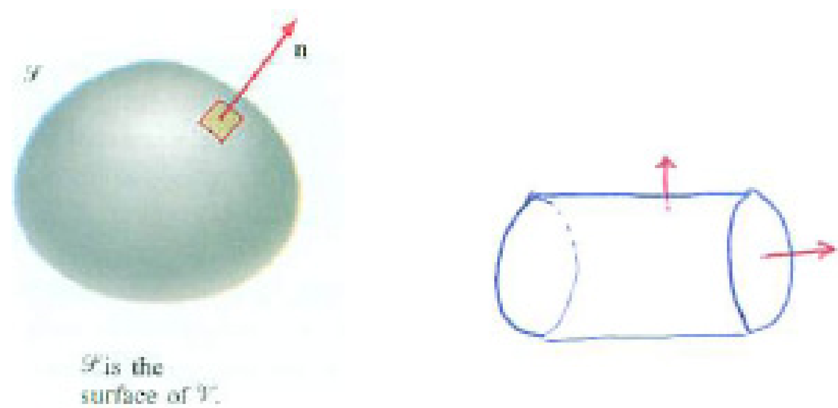
(c) Why does the table list only disks of radius 1? How would you use the table if $h = 3$ and $a = 2$?

18.5 The Divergence Theorem in Space (Gauss's Law)

In Sections 18.2 and 18.3 we developed Green's theorem and applied it in two forms for a vector field $\mathbf{F}$ in the plane. One concerned the line integral of the tangential component of $\mathbf{F}$, $\oint_C \mathbf{F} \cdot \mathbf{T}\, ds$, also written as $\oint_C \mathbf{F} \cdot d\mathbf{r}$. The other concerned the integral of the normal component of $\mathbf{F}$, $\oint_C \mathbf{F} \cdot \mathbf{n}\, ds$. In this section we develop the **Divergence theorem**, an extension of the second form from the plane to space. The extension of the first form to space is the subject of Section 18.6. In Section 18.7 the Divergence Theorem will be applied to electromagnetism.

The Divergence (or Gauss') Theorem

Consider a region $\mathcal{R}$ in space bounded by a surface $\mathcal{S}$. For instance, $\mathcal{R}$ may be a ball and $\mathcal{S}$ its surface, a case encountered in the elementary theory of electromagnetism. In another case, $\mathcal{R}$ is a right circular cylinder and $\mathcal{S}$ is its surface, which consists of two disks and its curved side. See Figure 18.5.1(a). Both figures show unit exterior normals, perpendicular to the surface. The



(a) (b)

Figure 18.5.1 Normal vectors to surfaces.

Divergence theorem relates an integral over the surface to an integral over the region it bounds. It is assumed that all surfaces of interest have a continuous exterior normal (such as a sphere) or are made up of a finite number of such surfaces (such as the surface of a cube).

Theorem 18.5.1. *Divergence Theorem —One-Surface Case. Let $\mathcal{V}$ be the region in space bounded by the surface $\mathcal{S}$. Let $\mathbf{n}$ denote the exterior unit normal* *of $\mathcal{V}$ along the boundary $\mathcal{S}$. Then*

$$\int\limits_{\mathcal{S}} \mathbf{F}\cdot\mathbf{n}\; ds = \int\limits_{\mathcal{V}} \nabla\cdot\mathbf{F}\; dV$$

for any vector field $\mathbf{F}$ defined on $\mathcal{V}$.

In words, the integral of the normal component of $\mathbf{F}$ over a surface equals the integral of the divergence of $\mathbf{F}$ over the region the surface bounds.

The integral $\int_{\mathcal{S}} \mathbf{F}\cdot\mathbf{n}\; dS$ is called the **flux** of the field $\mathbf{F}$ across the surface $\mathcal{S}$.

If $\mathbf{F} = P\mathbf{i}+Q\mathbf{j}+R\mathbf{k}$ and $\cos(\alpha)$, $\cos(\beta)$, and $\cos(\gamma)$ are the direction cosines of the exterior normal, then the Divergence theorem reads

$$\int\limits_{\mathcal{S}} (P\mathbf{i}+Q\mathbf{j}+R\mathbf{k})\cdot(\cos(\alpha)\mathbf{i}+\cos(\beta)\mathbf{j}+\cos(\gamma)\mathbf{k})\; dS = \int\limits_{\mathcal{V}} \left(\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}+\frac{\partial R}{\partial z}\right)\; dV.$$

Evaluating the dot product puts the Divergence theorem in the form

Direction cosines are defined in Section 14.4.

$$\int\limits_{\mathcal{S}} (P\cos(\alpha)+Q\cos(\beta)+R\cos(\gamma))\; dS = \int\limits_{\mathcal{V}} \left(\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}+\frac{\partial R}{\partial z}\right)\; dV.$$

When the Divergence theorem is expressed in this form, we see that it amounts to three scalar theorems:

$$\int\limits_{\mathcal{S}} P\cos(\alpha)\; dS = \int\limits_{\mathcal{V}} \frac{\partial P}{\partial x}\; dV, \quad \int\limits_{\mathcal{S}} Q\cos(\beta)\; dS = \int\limits_{\mathcal{V}} \frac{\partial Q}{\partial y}\; dV, \tag{18.5.1}$$

$$\text{and} \quad \int\limits_{\mathcal{S}} R\cos(\gamma)\; dS = \int\limits_{\mathcal{V}} \frac{\partial R}{\partial z}\; dV.$$

Establishing these equations proves the Divergence theorem. We delay the proof to the end of this section, after we have shown how the Divergence theorem is applied.

Two-Surface Version of the Divergence Theorem

The Divergence theorem also holds if the solid region has holes it it. Then the boundary consists of several separate closed surfaces. The most important case is when there is just one hole and hence an inner surface $\mathcal{S}_1$ and an outer surface $\mathcal{S}_2$, as shown in Figure 18.5.2.

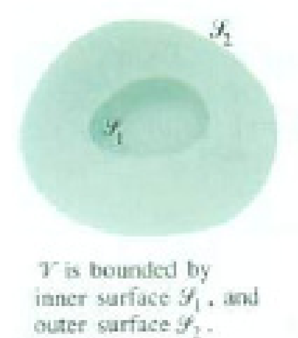


Figure 18.5.2

Theorem 18.5.2 (Divergence Theorem — Two-Surface Case.)**.** *Let $\mathcal{V}$ be a region in space bounded by the surfaces $\mathcal{S}_1$ and $\mathcal{S}_2$. Let $\mathbf{n}^*$ denote the exterior normal along the boundary. Then*

$$\int\limits_{\mathcal{S}_1} \mathbf{F} \cdot \mathbf{n}^* \, dS + \int\limits_{\mathcal{S}_2} \mathbf{F} \cdot \mathbf{n}^* \, dS = \int\limits_{\mathcal{V}} \operatorname{div} \mathbf{F} \, dV$$

for a vector field defined on $\mathcal{V}$.

Compare with (18.2.4) in Exercise 3 in Section 18.2.

The importance of this form of the Divergence Theorem is that it allows us to conclude that the flux across the surfaces are the same provided they form the boundary of a solid where $\operatorname{div} \mathbf{F} = 0$.

Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two closed surfaces that form the boundary of the region $\mathcal{V}$. Let $\mathbf{F}$ be a vector field defined on $\mathcal{V}$ such that the divergence of $\mathbf{F}$, $\nabla \cdot \mathbf{F}$, is 0 throughout $\mathcal{V}$. Then

$$\int\limits_{\mathcal{S}_1} \mathbf{F} \cdot \mathbf{n} \, dS = \int\limits_{\mathcal{S}_2} \mathbf{F} \cdot \mathbf{n} \, dS \qquad (18.5.2)$$

The proof of this result parallels the derivation of (18.2.4) for curves in Section 18.2 so we omit it.

The next example illustrates (18.5.2), which enables us, if the divergence of $\mathbf{F}$ is 0, to replace the integral of $\mathbf{F} \cdot \mathbf{n}$ over a surface by an integral of $\mathbf{F} \cdot \mathbf{n}$ over a more convenient surface.

EXAMPLE 1 Let $\mathbf{F}(\mathbf{r}) = \widehat{\mathbf{r}}/r^2$, the inverse square vector field with center at the origin. Let $\mathcal{S}$ be a convex surface that encloses the origin. Find the flux of $\mathbf{F}$ over the surface, $\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n} \, dS$.

SOLUTION Select a sphere with center at the origin that does not intersect $\mathcal{S}$. It should be small in order to miss $\mathcal{S}$. Call it $\mathcal{S}_1$ and its radius a, then, by (18.5.2),

$$\int\limits_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n} \, dS = \int\limits_{\mathcal{S}_1} \mathbf{F} \cdot \mathbf{n} \, dS$$

Because $(\widehat{\mathbf{r}}/r^2) \cdot \mathbf{n}$ equals $\frac{\mathbf{r} \cdot \mathbf{n}}{r^2}$, then, $\mathbf{r} \cdot \mathbf{n}$ is just 1. Thus:

$$\int\limits_{\mathcal{S}_1} \mathbf{F} \cdot \mathbf{n} \, dS = \int\limits_{\mathcal{S}_1} \frac{1}{a^2} \, dS = \frac{1}{a^2} \int\limits_{\mathcal{S}_1} dS = \frac{1}{a^2} 4\pi a^2 = 4\pi.$$

◇

Example 1 agrees with the fact that a convex surface subtends an angle of 4π of 4π steradians at any point in the region it bounds.

A **uniform** or **constant** vector field is a vector field whose vectors are identical. We use one in the next example.

EXAMPLE 2 Verify the Divergence theorem for the constant field $\mathbf{F}(x, y, z) = 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$ and the surface $\mathcal{S}$ of the cube whose sides have length 5 in Figure 18.5.3.

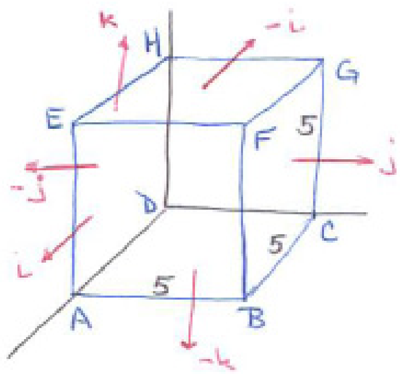


Figure 18.5.3

SOLUTION To find $\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n}\; dS$ we consider the integral of $\mathbf{F} \cdot \mathbf{n}$ over the faces.

On the bottom face, $ABCD$ the unit exterior normal is $-\mathbf{k}$. Thus

$$\mathbf{F} \cdot \mathbf{n} = (2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}) \cdot (-\mathbf{k}) = -4.$$

So

$$\int_{ABCD} \mathbf{F} \cdot \mathbf{n}\; dS = \int_{ABCD} (-4)\; dS = -4 \int_{ABCD} dS = (-4)(25) = -100.$$

The integral over the top face involves the exterior unit normal $\mathbf{k}$ instead of $-\mathbf{k}$. Then $\int_{EFGH} \mathbf{F} \cdot \mathbf{n}\; dS = 100$. The sum of the two integrals is 0. Similar computations show that the flux of $\mathbf{F}$ over the entire surface is 0.

The Divergence theorem says that the flux equals $\int_{\mathcal{R}} \operatorname{div} \mathbf{F}\; dV$, where $\mathcal{R}$ is the solid cube. Now, $\operatorname{div} \mathbf{F} = \partial(2)/\partial x + \partial(3)/\partial y + \partial(4)/\partial z = 0 + 0 + 0 = 0$. So the integral of $\operatorname{div} \mathbf{F}$ over $\mathcal{R}$ is 0, verifying the divergence theorem. $\diamond$

Why $\operatorname{div} \mathbf{F}$ is Called the Divergence

Let $\mathbf{F}(x, y, z)$ be the vector field describing the flow for a gas. That is, $\mathbf{F}(x, y, z)$ is the product of the density of the gas at (x, y, z) and the velocity vector of the gas there.

The integral $\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n}\, dS$ over a closed surface $\mathcal{S}$ represents the tendency of the gas to leave the region $\mathcal{R}$ that $\mathcal{S}$ bounds. If it is positive the gas is tending to escape or diverge. If negative, the effect is for the amount of gas in $\mathcal{R}$ to increase and become denser.

Let $\rho(x, y, z, t)$ be the density of the gas at time t at the point $P = (x, y, z)$, with units mass per unit volume. Then $\int_{\mathcal{R}} \rho\, dV$ is the total mass of gas in $\mathcal{R}$ at a given time. So the rate at which the mass in $\mathcal{R}$ changes is given by the derivative

$$\frac{d}{dt} \int\limits_{\mathcal{R}} \rho\, dV.$$

If ρ is sufficiently well-behaved, we may differentiate past the integral sign. Then

$$\frac{d}{dt} \int\limits_{\mathcal{R}} \rho\, dV = \int\limits_{\mathcal{R}} \frac{\partial p}{\partial t}\, dV.$$

Therefore

$$\int\limits_{\mathcal{R}} \frac{\partial p}{\partial t}\, dV = \int\limits_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n}\, dS$$

since both represent the rate at which gas accumulates in or escapes from $\mathcal{R}$. But by the Divergence theorem, $\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n}\, dS = \int_{\mathcal{R}} \nabla \cdot \mathbf{F}\, dV$, and so

$$\int\limits_{\mathcal{R}} \nabla \cdot \mathbf{F}\, dV = \int\limits_{\mathcal{R}} \frac{\partial p}{\partial t}\, dV$$

or

$$\int\limits_{\mathcal{R}} (\nabla \cdot \mathbf{F} - \frac{\partial p}{\partial t})\, dV = 0. \qquad (18.5.3)$$

From this it is possible to conclude that $\nabla \cdot \mathbf{F} - \frac{\partial p}{\partial t} = 0$, as follws.

The Zero-Integral Principle (Section 6.3) says: If a continuous function f on an interval $[a, b]$ has the property that $\int_c^d f(x)\, dx = 0$ for every subinterval $[c, d]$ then $f(x) = 0$ on $[a, b]$. An extension of the Zero-Integral Principle (Exercise 27) is:

Zero-Integral Principle in Space

Let $\mathcal{R}$ be a region in space bounded by a surface, and let f be a continuous function on $\mathcal{R}$. Assume that for every region $\mathcal{S}$ in $\mathcal{R}$, $\int_{\mathcal{S}} f(P)\, dS = 0$. Then $f(P) = 0$ for all P in $\mathcal{R}$.

Equation 18.5.3 holds not just for the solid $\mathcal{R}$ but for any solid region within $\mathcal{R}$. By the zero-integral principle, the integrand must be zero throughout $\mathcal{R}$, and we conclude that

$$\nabla \cdot \mathbf{F} = \frac{\partial p}{\partial t}.$$

This tells us that div $\mathbf{F}$ at a point P represents the rate gas is getting denser or lighter near P. That is why div $\mathbf{F}$ is called the divergence of $\mathbf{F}$. Where div $\mathbf{F}$ is positive, the gas is dissipating. Where div $\mathbf{F}$ is negative, the gas is collecting.

See Exercise 20 in Section 18.3.

For this reason a vector field for which the divergence is zero is called **incompressible**. An incompressible gas is also called divergence free.

We conclude this section with a proof of the Divergence theorem.

Proof of the Divergence Theorem

We prove the theorem for the special case that each line parallel to an axis meets the surface $\mathcal{S}$ in at most two points and $\mathcal{V}$ is convex. We prove the third equation in (18.5.1). The other two are established the same way.

We wish to show that

$$\int\limits_{\mathcal{V}} R\cos(\gamma)\; dS = \int\limits_{\mathcal{V}} \frac{\partial R}{\partial z}\; dV. \tag{18.5.4}$$

Let $\mathcal{A}$ be the projection of $\mathcal{S}$ on the xy-plane. Its description is

$$a \le x \le b, \qquad y_1(x) \le y \le y_2(x).$$

The description of $\mathcal{V}$ is then

$$a \le x \le b, \qquad y_1(x) \le y \le y_2(x), \qquad z_1(x,y) \le z \le z_2(x,y).$$

Then (see Figure 18.5.4)

$$\int\limits_{\mathcal{V}} \frac{\partial R}{\partial z}\; dV = \int\limits_a^b \int\limits_{y_1(x)}^{y_2(x)} \int\limits_{z_1(x,y)}^{z_2(x,y)} \frac{\partial R}{\partial z}\; dz\; dy\; dx. \tag{18.5.5}$$

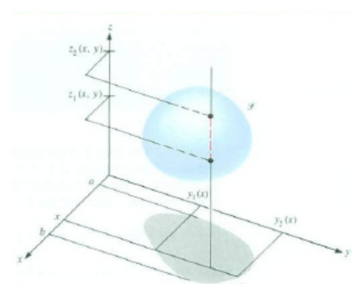


Figure 18.5.4

The first integration gives

$$\int\limits_{z_1(x,y)}^{z_2(x,y)} \frac{\partial R}{\partial z}\; dz = R(x,y,z_2) - R(x,y,z_1), \tag{18.5.6}$$

by the Fundamental Theorem of Calculus. We have, therefore,

$$\int\limits_{\mathcal{V}} \frac{\partial R}{\partial z}\; dV = \int\limits_a^b \int\limits_{y_1(x)}^{y_2(x)} \left(R(x,y,z_2) - R(x,y,z_1)\right)\; dy\; dx,$$

hence
$$\int\limits_{\mathcal{V}} \frac{\partial R}{\partial z}\, dV = \int\limits_{\mathcal{A}} \left(R(x,y,z_2) - R(x,y,z_1)\right)\, dA.$$

This says that, essentially, on the top part of $\mathcal{V}$, where $0 < \gamma < \pi/2$, $dA = \cos(\gamma)\, dS$ is positive. On the bottom part of $\mathcal{V}$, where $\pi/2 < \gamma < \pi$, $dA = -\cos(\gamma)\, dS$ is negative. According to (17.7.1) in Section 17.7, we can replace dA by $\cos(\gamma)\, dS$ and write the last integral as

$$\int\limits_{\mathcal{S}} R(x,y,z)\cos(\gamma)\, dS.$$

Thus
$$\int\limits_{\mathcal{V}} \frac{\partial R}{\partial z}\, dV = \int\limits_{\mathcal{S}} R\cos\gamma\, dS,$$

and (18.5.4) is established.

Summary

We stated the Divergence theorem for a single surface and for two surfaces. It lets us calculate the flux of a vector field $\mathbf{F}$ in terms of an integral of its divergence $\nabla \cdot \mathbf{F}$ over a region. This is useful for fields that are incompressible (divergence free), such as the inverse-square field in space, $\widehat{\mathbf{r}}/r^2$. The flux across a surface of such a field depends on whether its center is inside or outside the surface. If the center is at Q and the field is of the form $c\frac{\overline{QP}}{|\overline{QP}|^3}$, its flux across a surface not enclosing Q is 0. If it encloses Q, its flux is 4π. This is a consequence of the Divergence theorem. It can be explained geometrically in terms of solid angles.

EXERCISES for Section 18.5

1. State the Divergence theorem in symbols.

2. State the Divergence theorem using only words.

3. Explain why $\nabla \cdot \mathbf{F}$ at a point P can be expressed as a coordinate-free limit.

4. What is the two-surface version of Gauss's theorem?

5. Verify the divergence theorem for $\mathbf{F}(x, y, z) = x\mathbf{i} + y\mathbf{j} + 0\mathbf{k}$ and the surface $x^2 + y^2 + z^2 = 9$.

6. Verify the divergence theorem for the field $\mathbf{F}(x, y, z) = x\mathbf{i}$ and the cube whose vertices are $(0, 0, 0)$, $(2, 0, 0)$, $(2, 2, 0)$, $(0, 2, 0)$, $(0, 0, 2)$, $(2, 0, 2)$, $(2, 2, 2)$, and $(0, 2, 2)$.

7. Verify the divergence theorem for $\mathbf{F} = 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$ and the tetrahedron whose four vertices are $(0, 0, 0)$, $(1, 0, 0)$, and $(0, 1, 0)$, and $(0, 0, 1)$.

8. Verify the two-surface version of Gauss's theorem for $\mathbf{F}(x, y, z) = (x^2 + y^2 + z^2)(x\mathbf{i} + y\mathbf{j} + z\mathbf{k})$ where the surfaces are spheres of radii 2 and 3 centered at the origin.

9. Let $\mathbf{F} = 2x\mathbf{i} + 3y\mathbf{j} + (5z + 6x)\mathbf{k}$, and let $\mathbf{G} = (2x + 4z^2)\mathbf{i} + (3y + 5x)\mathbf{j} + 5z\mathbf{k}$. Show that

$$\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n}\, dS = \int_{\mathcal{S}} \mathbf{G} \cdot \mathbf{n}\, dS,$$

where $\mathcal{S}$ is a surface bounding a region in space.

10. Show that the divergence of $\widehat{\mathbf{r}}/r^2$ is 0. ($\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$.)

In Exercises 11 to 18 use the Divergence Theorem.

11. Let $\mathcal{V}$ be the solid region bounded by the xy plane and the paraboloid $z = 9 - x^2 - y^2$. Evaluate $\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n}\, dS$ where $\mathbf{F} = y^3\mathbf{i} + z^3\mathbf{j} + x^3\mathbf{k}$ and $\mathcal{S}$ is the boundary of $\mathcal{V}$.

12. Evaluate $\int_{\mathcal{V}} \nabla \cdot \mathbf{F}\, dV$ for $\mathbf{F} = \sqrt{x^2 + y^2 + z^2}(x\mathbf{i} + y\mathbf{j} + z\mathbf{k})$ and $\mathcal{V}$ the ball of radius 2 and center at $(0, 0, 0)$.

In Exercises 13 and 14 find $\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n}\, dS$.

13. $\mathbf{F} = z\sqrt{x^2 + z^2}\mathbf{i} + (y + 3)\mathbf{j} - x\sqrt{x^2 + z^2}\mathbf{k}$ and $\mathcal{S}$ is the boundary of the solid region between $z = x^2 + y^2$ and the plane $z = 4x$.

14. $\mathbf{F} = x\mathbf{i} + (3y+z)\mathbf{j} + (4x+2z)\mathbf{k}$ and $\mathcal{S}$ is the surface of the cube bounded by the planes $x=1$, $x=3$, $y=2$, $y=4$, $z=3$, and $z=5$.

15. Evaluate $\int_{\mathcal{S}} \mathbf{F}\cdot\mathbf{n}\, dS$, where $\mathbf{F} = 4xz\mathbf{i} - y^2\mathbf{j} + yz\mathbf{k}$ and $\mathcal{S}$ is the surface of the cube bounded by the planes $x=0$, $x=1$, $y=0$, $z=0$, and $z=1$, with the face corresponding to $x=1$ removed.

16. Evaluate $\int_{\mathcal{S}} \mathbf{F}\cdot\mathbf{n}\, dS$, where $\mathbf{F} = x\mathbf{i} + y\mathbf{j} + 2x\mathbf{k}$ and $\mathcal{S}$ is the boundary of the tetrahedron with vertices $(1,2,3)$, $(1,0,1)$ $(2,1,4)$, and $(1,3,5)$.

17. Let $\mathcal{S}$ be a surface of area S that bounds a region $\mathcal{V}$ of volume V. Assume that $|\mathbf{F}(P)| \leq 5$ for all points P on the surface $\mathcal{S}$. What can be said about $\int_{\mathcal{V}} \nabla\cdot\mathbf{F}\, dV$?

18. Evaluate $\int_{\mathcal{S}} \mathbf{F}\cdot\mathbf{n}\, dS$, where $\mathbf{F} = x^3\mathbf{i} + y^3\mathbf{j} + z^3\mathbf{k}$ and $\mathcal{S}$ is the sphere of radius a and center $(0,0,0)$.

In Exercises 19 to 22 evaluate $\int_{\mathcal{S}} \mathbf{F}\cdot\mathbf{n}\, dS$ for $\mathbf{F} = \widehat{\mathbf{r}}/r^2$.

19. $\mathcal{S}$ is the sphere of radius 2 and center $(5,3,1)$.

20. $\mathcal{S}$ is the sphere of radius 3 and center $(1,0,1)$.

21. $\mathcal{S}$ is the surface of the box bounded by the planes $x=-1$, $x=2$, $y=2$, $y=3$, $z=-1$, and $z=6$.

22. $\mathcal{S}$ is the surface of the box bounded by the planes $x=-1$, $x=2$, $y=-1$, $y=3$, $z=-1$, and $z=4$.

23. Assume that the flux of $\mathbf{F}$ across every sphere is 0. Must the flux of $\mathbf{F}$ across the surface of every cube be 0 also?

24. If $\mathbf{F}$ is always tangent to a surface $\mathcal{S}$ what can be said about the integral of $\nabla\cdot\mathbf{F}$ over the region that $\mathcal{S}$ bounds?

25. Let $\mathbf{F}(\mathbf{r}) = f(r)\widehat{\mathbf{r}}$ be a central vector field in space that has zero divergence. Show that $f(r)$ has the form $f(r) = a/r^2$ for some constant a. (Consider the flux of $\mathbf{F}$ across the closed surface bounded by a cone with vertex at the origin and two spheres centered at the origin, see Figure 18.5.5.)

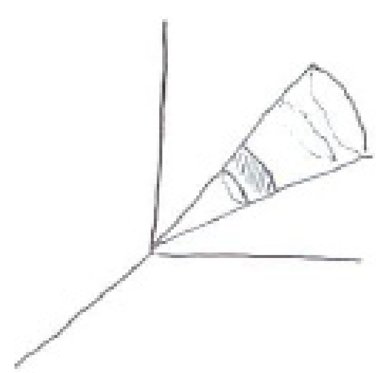


Figure 18.5.5

26. Let $\mathbf{F}$ be defined everywhere except at the origin and be divergence-free. Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two closed surfaces that enclose the origin. Explain why $\int_{\mathcal{S}_1} \mathbf{F} \cdot \mathbf{n}\, dS = \int_{\mathcal{S}_2} \mathbf{F} \cdot \mathbf{n}\, dS$. The two surfaces may intersect.

27. Provide the details for the proof of the Zero-Integral Principle in space. You will need to consider two cases, when $f > 0$ and when $f < 0$.

28. Show that the flux of an inverse-square central field $c\widehat{\mathbf{r}}/r^2$ across any closed surface that bounds a region that does not contain the origin is zero.

29.

(a) Show that the proof in the text of the Divergence Theorem applies to a tetrahedron. (Choose your coordinate system carefully.)

(b) Deduce that if the Divergence Theorem holds for a tetrahedron then it holds for any polyhedron. A polyhedron can be cut into tetrahedra.

30. In Exercise 25 you showed by considering a particular type of surface that the only central fields with zero divergence are the inverse square fields. Show this, instead, by computing the divergence of $\mathbf{F}(x, y, z) = f(r)\widehat{\mathbf{r}}$, where $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$.

31. Let $\mathbf{F}$ be defined everywhere in space except at the origin. Assume that

$$\lim_{|\mathbf{r}| \to \infty} \frac{\mathbf{F}(\mathbf{r})}{|\mathbf{r}|^2} = \mathbf{0}$$

and that $\mathbf{F}$ is defined everywhere except at the origin, and is divergence-free. What can be said about $\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n}\, dS$, where $\mathcal{S}$ is the sphere of radius 2 centered at the origin?

We proved one-third of the Divergence Theorem. Exercises 32 and 33 concern the other two-thirds.

32. Prove that

$$\int\limits_{\mathcal{S}} Q \cos(\beta)\, dS = \int\limits_{\mathcal{V}} \frac{\partial Q}{\partial y}\, dV.$$

33. Prove that

$$\int\limits_{\mathcal{S}} P \cos(\alpha)\, dS = \int\limits_{\mathcal{V}} \frac{\partial P}{\partial x}\, dV.$$

18.6 Stokes' Theorem

In Section 18.1 we learned that Green's theorem in the xy-plane can be written as

$$\oint_C \mathbf{F} \cdot \, d\mathbf{r} = \int_{\mathcal{R}} (\mathbf{curl}\,\mathbf{F}) \cdot \mathbf{k} \; dA,$$

where C is traversed counterclockwise and bounds the region $\mathcal{R}$. The general Stokes' theorem in this section extends this to closed curves in space. It asserts that if the closed curve C bounds a surface $\mathcal{S}$, then

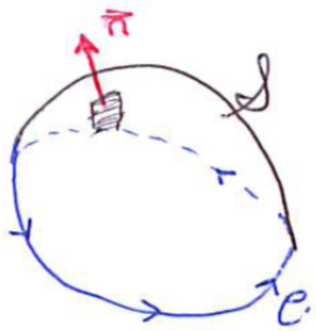


Figure 18.6.1

$$\oint_C \mathbf{F} \cdot \, d\mathbf{r} = \int_{\mathcal{S}} (\mathbf{curl}\,\mathbf{F}) \cdot \mathbf{n} \; dS.$$

As usual, the vector $\mathbf{n}$ is a unit normal to the surface. There are two normals at each point on the surface and we will describe how to decide which unit normal vector to use. The choice depends on the orientation of C.

In words, Stokes' theorem reads says that the circulation of a vector field around a closed curve is equal to the integral of the normal component of the curl of the field over any surface that the curve bounds.

Stokes published his theorem in 1854 (without proof, for it appeared as a question on a Cambridge University examination). By 1870 it was in common use. It is the most recent of the three theorems discussed in this chapter, for Green published his theorem in 1828 and Gauss published the divergence theorem in 1839.

Figure 18.6.2

Choosing the Normal n

In order to state Stokes' theorem precisely, we must describe what kind of surface $\mathcal{S}$ is permitted and which of the two normals $\mathbf{n}$ to choose.

The surfaces $\mathcal{S}$ that we consider it is possible to assign, at each point on $\mathcal{S}$, a unit normal $\mathbf{n}$ in a continuous manner. On the surface shown in Figure 18.6.2, there are two ways to do this, shown in Figure 18.6.3.

(a) (b)

Figure 18.6.3

For the surface shown in Figure 18.6.4, a Möbius band, it is impossible to choose. If you start with choice (1) and move the normal continuously along the surface, by the time you return to the initial point on the surface at stage (9), you have the opposite normal. A surface for which a continuous choice can be made is called **orientable** or **two-sided**. Stokes' theorem holds for orientable surfaces, which include, for instance, any part of the surface of a convex body, such as a ball, cube or cylinder.

Right-hand rule for choosing $\mathbf{n}$.

Consider an orientable surface $\mathcal{S}$, bounded by a parameterized curve C so that the curve is swept out in a definite direction. If the surface is part of a plane, we can use the right-hand rule to choose $\mathbf{n}$: The direction of $\mathbf{n}$ should match the thumb of the right hand if the fingers curl in the direction of C and the thumb and palm are perpendicular to the plane. If the surface is not flat, we still use the right-hand rule to choose a normal at points near C. The choice of one normal determines normals throughout the surface. Figure 18.6.5 illustrates the choice of $\mathbf{n}$. For instance, if C is counterclockwise in the xy-plane, the definition picks out the normal $\mathbf{k}$.

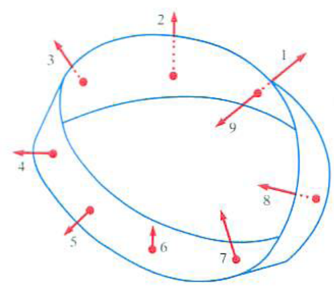
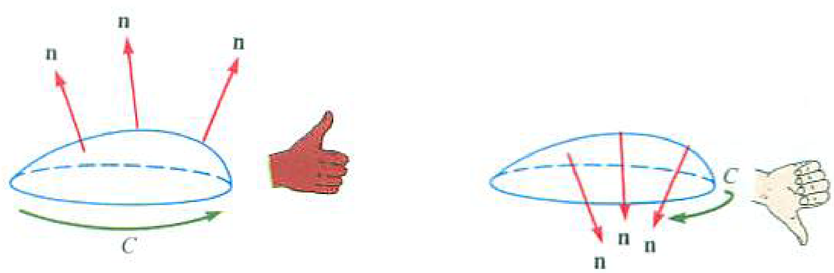


Figure 18.6.4 Follow the choices through all nine stages — there's trouble.



(a) (b)

Figure 18.6.5

Theorem 18.6.1 (Stokes' Theorem)**.** *Let $\mathcal{S}$ be an orientable surface bounded by the parameterized curve C. At each point of $\mathcal{S}$ let $\mathbf{n}$ be the unit normal chosen by the right-hand rule. Let $\mathbf{F}$ be a vector field defined on some region in space including $\mathcal{S}$. Then*

$$\oint_C \mathbf{F}\cdot\, d\mathbf{r} = \int_{\mathcal{S}} (\nabla\times\mathbf{F})\cdot\mathbf{n}\, dS.$$

Some Applications of Stokes' Theorem

Stokes' theorem enables us to replace $\int_{\mathcal{S}}(\mathbf{curl}\,\mathbf{F})\cdot\mathbf{n}\,dS$ by a similar integral over a surface that might be simpler than $\mathcal{S}$. That is the substance of the following special case.

One way to evaluate some surface integrals is to choose a simpler surface.

Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two surfaces bounded by the same curve C and oriented so that they yield the same orientation on C. Let $\mathbf{F}$ be a vector field defined on both $\mathcal{S}_1$ and $\mathcal{S}_2$. Then

$$\int_{\mathcal{S}_1} (\mathbf{curl}\,\mathbf{F}) \cdot \mathbf{n}\ dS = \int_{\mathcal{S}_2} (\mathbf{curl}\,\mathbf{F}) \cdot \mathbf{n}\ dS \qquad (18.6.1)$$

The two integrals in (18.6.1) are equal since both equal $\oint_C \mathbf{F} \cdot\ d\mathbf{r}$.

EXAMPLE 1 Let $\mathbf{F} = xe^z\mathbf{i} + (x + xz)\mathbf{j} + 3e^z\mathbf{k}$ and let $\mathcal{S}$ be the top half of the sphere $x^2 + y^2 + z^2 = 1$. Find $\int_{\mathcal{S}} (\mathbf{curl}\,\mathbf{F}) \cdot \mathbf{n}\ dS$, where $\mathbf{n}$ is the outward normal. See Figure 18.6.6.

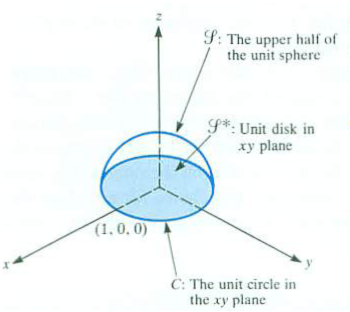


Figure 18.6.6

SOLUTION Let $\mathcal{S}^*$ be the flat base of the hemisphere. By (18.6.1),

$$\int_{\mathcal{S}} (\nabla \times \mathbf{F}) \cdot \mathbf{n}\ dS = \int_{\mathcal{S}^*} (\nabla \times \mathbf{F}) \cdot \mathbf{k}\ dS.$$

(On $\mathcal{S}^*$ $\mathbf{k}$, not $-\mathbf{k}$, is the correct normal to use.)

A calculation shows that

$$\nabla \times \mathbf{F} = -x\mathbf{i} + xe^z\mathbf{j} + (z + 1)\mathbf{k},$$

hence $(\nabla \times \mathbf{F}) \cdot \mathbf{k} = z + 1$. On $\mathcal{S}^*$, $z = 0$, so

$$\int_{\mathcal{S}^*} (\nabla \times \mathbf{F}) \cdot \mathbf{k}\ dS = \int_{\mathcal{S}^*} dS = \pi.$$

Thus the original integral over $\mathcal{S}$ is also π. ◇

Just as there are two-curve versions of Green's theorem and of the Divergence theorem, there is a two-curve version of Stokes' theorem.

Two-curve version of Stokes' theorem

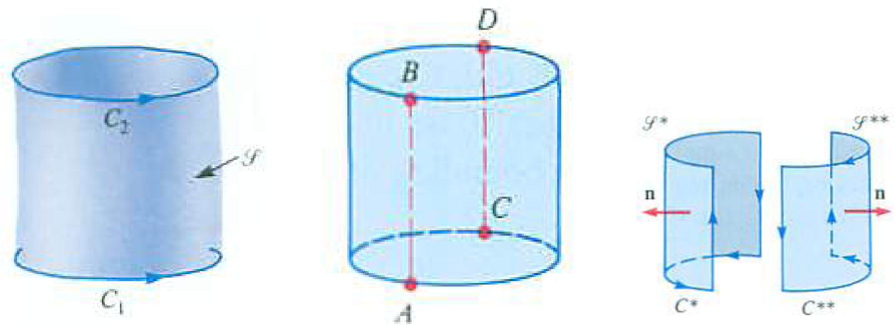
Stokes' Theorem for a Surface Bounded by Two Closed Curves

Theorem 18.6.2. *Let $\mathcal{S}$ be an orientable surface whose boundary consists of the two closed curves C_1 and C_2. Give C_1 an orientation. Orient $\mathcal{S}$ consistent with the right-hand rule as applied to C_1. Give C_2 the same orientation as C_1. (If C_2 is moved on $\mathcal{S}$ to C_1, the orientations will agree.) Then*

$$\oint_{C_1} \mathbf{F} \cdot\ d\mathbf{r} - \oint_{C_2} \mathbf{F} \cdot\ d\mathbf{r} = \int_{\mathcal{S}} (\nabla \times \mathbf{F}) \cdot \mathbf{n}\ dS. \qquad (18.6.2)$$

Proof

Figure 18.6.7(a) shows the typical situation.



(a) (b) (c)

Figure 18.6.7

We will obtain (18.6.2) from Stokes' theorem with the aid of the cancellation principle. Introduce lines AB and CD on $\mathcal{S}$, cutting $\mathcal{S}$ into two surfaces, $\mathcal{S}^*$ and $\mathcal{S}^{**}$. See Figure 18.6.7(c). Apply Stokes' theorem to $\mathcal{S}^*$ and $\mathcal{S}^{**}$. See Figure 18.6.7(c).

The cancellation principle was introduced in Section 18.2.

Let C^* be the curve that bounds $\mathcal{S}^*$, oriented so that where it overlaps C_1 it has the same orientation as C_1. Let C^{**} be the curve that bounds $\mathcal{S}^{**}$, again oriented to match C_1. See Figure 18.6.7(c).

By Stokes' theorem,

$$\oint_{C^*} \mathbf{F} \cdot \, d\mathbf{r} = \int_{\mathcal{S}^*} (\mathbf{curl}\,\mathbf{F}) \cdot \mathbf{n} \; dS \tag{18.6.3}$$

and

$$\oint_{C^{**}} \mathbf{F} \cdot \, d\mathbf{r} = \int_{\mathcal{S}^{**}} (\mathbf{curl}\,\mathbf{F}) \cdot \mathbf{n} \; dS. \tag{18.6.4}$$

Adding (18.6.3) and (18.6.4) and using the cancellation principle gives

$$\oint_{C_1} \mathbf{F} \cdot \, d\mathbf{r} - \oint_{C_2} \mathbf{F} \cdot \, d\mathbf{r} = \int_{\mathcal{S}} (\mathbf{curl}\,\mathbf{F}) \cdot \mathbf{n} \; dS.$$

•

In practice, it is most common to apply (18.6.2) when $\mathbf{curl}\,\mathbf{F} = \mathbf{0}$. This is so important we state it explicitly:

$\mathbf{F}$ is irrotational when $\mathbf{curl}\,\mathbf{F} = \mathbf{0}$.

Let $\mathbf{F}$ be a field such that $\mathbf{curl}\,\mathbf{F} = \mathbf{0}$. Let C_1 and C_2 be two closed curves that together bound an orientable surface $\mathcal{S}$ on which $\mathbf{F}$ is defined. If C_1 and C_2 are similarly oriented, then

$$\oint_{C_1} \mathbf{F}\cdot\ d\mathbf{r} = \oint_{C_2} \mathbf{F}\cdot\ d\mathbf{r}. \tag{18.6.5}$$

This follows directly from (18.6.2) since $\int_{\mathcal{S}} (\mathbf{curl}\,\mathbf{F})\cdot\mathbf{n}\ dS = 0$.

EXAMPLE 2 Assume that $\mathbf{F}$ is irrotational and defined everywhere except on the z-axis. Given that $\oint_{C_1} \mathbf{F}\cdot\ d\mathbf{r} = 3$, find (a) $\oint_{C_2} \mathbf{F}\cdot\ d\mathbf{r}$ and (b) $\oint_{C_3} \mathbf{F}\cdot\ d\mathbf{r}$. (See Figure 18.6.8.)

SOLUTION (a) By (18.6.5), $\oint_{C_2} \mathbf{F}\cdot\ d\mathbf{r} = \oint_{C_1} \mathbf{F}\cdot\ d\mathbf{r} = 3$. (b) By Stokes' theorem, (18.6.1), $\oint_{C_3} \mathbf{F}\cdot\ d\mathbf{r} = 0$. ◇

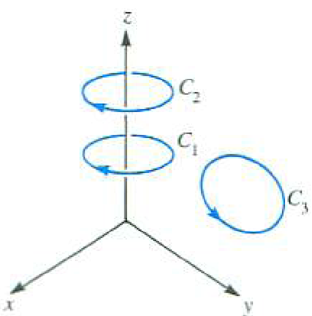


Figure 18.6.8

Curl and Conservative Fields

In Section 18.1 we saw that if $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$ is defined on a simply connected region in the xy-plane and if $\mathbf{curl}\,\mathbf{F} = \mathbf{0}$, then $\mathbf{F}$ is conservative. Now that we have Stokes' theorem, this can be extended to a field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ defined on a simply connected region in space.

Theorem 18.6.3. *Let* $\mathbf{F}$ *be defined on a simply connected region in space. If* $\mathbf{curl}\,\mathbf{F} = \mathbf{0}$*, then* $\mathbf{F}$ *is conservative.*

Proof

We provide only a sketch of the proof. Let C be a simple closed curve situated in the simply connected region. To avoid topological complexities, we assume that it bounds an orientable surface $\mathcal{S}$. To show that $\oint_C \mathbf{F}\cdot d\mathbf{r} = 0$, we use the same argument as in Section 18.2:

$$\oint_C \mathbf{F}\cdot\ d\mathbf{r} = \int_{\mathcal{S}} (\nabla\times\mathbf{F})\cdot\mathbf{n}\ dS = \int_{\mathcal{S}} 0\ dS = 0.$$

•

It follows from Theorem 18.6.3 that every central field $\mathbf{F}$ is conservative because a calculation shows that its curl is $\mathbf{0}$. (See Exercises 6 and 7 in Section 18.4.) Moreover, $\mathbf{F}$ is defined either throughout space or everywhere except at the center of the field. (Observe that in either case the region is simply connected.)

Exercise 26 of Section 18.4 presents a geometric argument for why a central field is conservative.

In Sections 18.7 and 18.9 we will show how Stokes' theorem is applied in the theory of electromagnetism.

Why Curl is Called Curl

Let $\mathbf{F}$ be a vector field describing the flow of a fluid, as in Section 18.1. Stokes' theorem will give a physical interpretation of $\mathbf{curl}\,\mathbf{F}$.

Consider a fixed point P_0 in space. Imagine a small circular disk $\mathcal{S}$ with center P_0. Let C be the boundary of $\mathcal{S}$ oriented in such a way that C and $\mathbf{n}$ fit the right-hand rule. (See Figure 18.6.9)

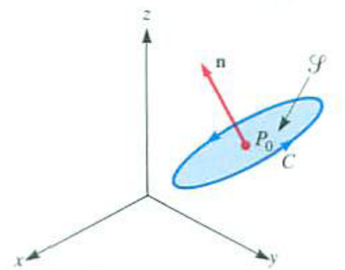


Figure 18.6.9

Now examine the two sides of

$$\int_{\mathcal{S}} (\mathbf{curl}\,\mathbf{F}) \cdot \mathbf{n}\; dS = \oint_C \mathbf{F} \cdot \mathbf{T}\; ds. \tag{18.6.6}$$

The right side measures the tendency of the fluid to move along C (rather than, say, perpendicular to it.) Thus $\oint_C \mathbf{F} \cdot \mathbf{T}\; ds$ might be thought of as the circulation or whirling tendency of the fluid along C. For each tilt of the small disk $\mathcal{S}$ at P_0, or, equivalently, each choice of unit normal vector $\mathbf{n}$, the line integral $\oint_C \mathbf{F} \cdot \mathbf{T}\; ds$ measures a circulation. It records the tendency of a paddle wheel at P_0 with axis along $\mathbf{n}$ to rotate. (See Figure 18.6.10.)

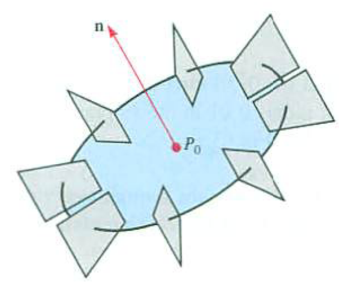


Figure 18.6.10

On the left side of (18.6.6), if $\mathcal{S}$ is small, the integrand is almost constant and the integral is approximately

$$(\mathbf{curl}\,\mathbf{F})_{P_0} \cdot \mathbf{n} \cdot \text{Area of } \mathcal{S}, \tag{18.6.7}$$

where $(\mathbf{curl}\,\mathbf{F})_{P_0}$ denotes the curl of $\mathbf{F}$ evaluated at P_0.

Keeping the center of $\mathcal{S}$ at P_0, vary the vector $\mathbf{n}$ by tilting $\mathcal{S}$. For which choice of $\mathbf{n}$ will (18.6.7) be largest? The answer is that for that $\mathbf{n}$ with the same direction as the fixed vector $(\mathbf{curl}\,\mathbf{F})_{P_0}$. With that choice of $\mathbf{n}$, (18.6.7) becomes

$$|(\mathbf{curl}\,\mathbf{F})_{P_0}| \text{ Area of } \mathcal{S}\ .$$

The physical interpretation of curl

Thus a paddle wheel placed in the fluid at P_0 rotates most quickly when its axis is in the direction of $\mathbf{curl}\,\mathbf{F}$ at P_0. The magnitude of $\mathbf{curl}\,\mathbf{F}$ is a measure of how fast the paddle wheel can rotate when placed at P_0. Thus $\mathbf{curl}\,\mathbf{F}$ records the direction and magnitude of maximum circulation at a given point.

A Vector Definition of Curl

In Section 18.1 $\mathbf{curl\,F}$ was defined in terms of the partial derivatives of the components of $\mathbf{F}$. By Stokes' theorem, $\mathbf{curl\,F}$ is related to the circulation $\oint_C \mathbf{F}\cdot\, d\mathbf{r}$. We exploit it to obtain a new view of $\mathbf{curl\,F}$, free of coordinates.

Let P_0 be a point in space and let $\mathbf{n}$ be a unit vector. Consider a small disk $\mathcal{S}_{\mathbf{n}}(a)$, perpendicular to $\mathbf{n}$, whose center is P_0 and which has radius a. Let $C_{\mathbf{n}}(a)$ be the boundary of $\mathcal{S}_{\mathbf{n}}(a)$, oriented to be compatible with the right-hand rule. Then

$$\int\limits_{\mathcal{S}_{\mathbf{n}}(a)} (\mathbf{curl\,F})\cdot\mathbf{n}\; dS = \oint\limits_{C_{\mathbf{n}}(a)} \mathbf{F}\cdot\, d\mathbf{r}.$$

As in our discussion of the physical meaning of curl, we see that

$$(\mathbf{curl\,F})(P_0)\cdot\mathbf{n}\cdot \text{Area of } \mathcal{S}_{\mathbf{n}}(a) \approx \oint\limits_{C_{\mathbf{n}}(a)} \mathbf{F}\cdot d\mathbf{r}$$

or

$$(\mathbf{curl\,F})(P_0)\cdot\mathbf{n} \approx \frac{\oint_{C_{\mathbf{n}}(a)} \mathbf{F}\cdot\, d\mathbf{r}}{\text{Area of } \mathcal{S}_{\mathbf{n}}(a)}.$$

Thus

$$(\mathbf{curl\,F})(P_0)\cdot\mathbf{n} = \lim_{a\to 0} \frac{\oint_{C_{\mathbf{n}}(a)} \mathbf{F}\cdot\, d\mathbf{r}}{\text{Area of } \mathcal{S}_{\mathbf{n}}(a)}. \tag{18.6.8}$$

This gives meaning to the component of $(\mathbf{curl\,F})(P_0)$ in a direction $\mathbf{n}$. So the magnitude and direction of $\mathbf{curl\,F}$ at P_0 can be described in terms of $\mathbf{F}$, without looking at the components of $\mathbf{F}$.

Local Definition of curl F

The magnitude of $(\mathbf{curl\,F})_{P_0}$ is the maximum value of

$$\lim_{a\to 0} \frac{\oint_{C_{\mathbf{n}}(a)} \mathbf{F}\cdot\, d\mathbf{r}}{\text{Area of } \mathcal{S}_{\mathbf{n}}(a)}, \tag{18.6.9}$$

for all unit vectors $\mathbf{n}$.
The direction of $(\mathbf{curl\,F})_{P_0}$ is given by the vector $\mathbf{n}$ that maximizes the limit (18.6.9).

EXAMPLE 3 Let $\mathbf{F}$ be a vector field such that at the origin $\mathbf{curl\,F} = 2\mathbf{i} + 4\mathbf{j} + 4\mathbf{k}$. Estimate $\oint_C \mathbf{F}\cdot\, d\mathbf{r}$ if C encloses a disk of radius 0.01 in the xy-plane with center $(0,0,0)$. C is swept out clockwise. (See Figure 18.6.11.)
SOLUTION Let $\mathcal{S}$ be the disk whose border is C. Choose the normal to

$\mathcal{S}$ that is consistent with the orientation of C and the right-hand rule. That choice is $-\mathbf{k}$. Thus

$$(\mathbf{curl}\,\mathbf{F})\cdot(-\mathbf{k}) \approx \frac{\oint_C \mathbf{F}\cdot\ d\mathbf{r}}{\text{Area of } \mathcal{S}}.$$

The area of $\mathcal{S}$ is $\pi(0.01)^2$ and $\mathbf{curl}\,\mathbf{F} = 2\mathbf{i}+3\mathbf{j}+4\mathbf{k}$. Thus

$$(2\mathbf{i}+3\mathbf{j}+4\mathbf{k})\cdot(-\mathbf{k}) \approx \frac{\oint_C \mathbf{F}\cdot\ d\mathbf{r}}{\pi(0.01)^2}.$$

From this it follows that

$$\oint_C \mathbf{F}\cdot\ d\mathbf{r} \approx -4\pi(0.01)^2.$$

◇

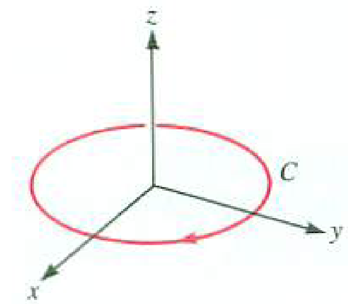


Figure 18.6.11

SHERMAN: Do we need a more complete reference for this quote? Do you have it?

The Origin of the Term curl

In a letter to the mathematician Tait written on November 7, 1870, Maxwell offered some names for $\nabla \times \mathbf{F}$:

> Here are some rough-hewn names. Will you like a good Divinity shape their ends properly so as to make them stick? ...
>
> The vector part $\nabla \times \mathbf{F}$ I would call the twist of the vector function. Here the word twist has nothing to do with a screw or helix. The word *turn* ... would be better than twist, for twist suggests a screw. Twirl is free from the screw motion and is sufficiently racy. Perhaps it is too dynamical for pure mathematicians, so for Cayley's sake I might say Curl (after the fashion of Scroll.)

His last suggestion, curl, has stuck.

Proof of Stokes' Theorem

We include the proof because it reviews several basic ideas. The proof uses Green's theorem, the normal to a surface $z = f(x, y)$, and the expression for an integral over a surface as an integral over its projection on a plane. We write the theorem using components. We will assume that the surface $\mathcal{S}$ meets each line parallel to an axis in at most one point. That there are one-to-one projections of $\mathcal{S}$ onto the coordinate planes .

We write $\mathbf{F}(x,y,z)$ as $P(x,y,z)\mathbf{i}+Q(x,y,z)\mathbf{j}+R(x,y,z)\mathbf{k}$, or $\mathbf{F}=P\mathbf{i}+Q\mathbf{j}+R\mathbf{k}$. We project $\mathcal{S}$ onto the xy-plane, so write the equation for $\mathcal{S}$ as $z-f(x,y)=0$. A unit normal to $\mathcal{S}$ is

$$\mathbf{n}=\frac{-\frac{\partial f}{\partial x}\mathbf{i}-\frac{\partial f}{\partial y}\mathbf{j}+\mathbf{k}}{\sqrt{\left(\frac{\partial f}{\partial x}\right)^2+\left(\frac{\partial f}{\partial y}\right)^2+1}}.$$

(Since the $\mathbf{k}$ component of $\mathbf{n}$ is positive, it is the correct normal, given by the right-hand rule.) Let C^* be the projection of C on the xy-plane, swept out counterclockwise.

See Exercise 41.

A computation shows that Stokes' theorem, expressed in components, reads

$$\int\limits_C P\,dx+Q\,dy+R\,dz$$
$$=\int\limits_{\mathcal{S}}\frac{\left(\frac{\partial R}{\partial x}-\frac{\partial Q}{\partial z}\right)\left(-\frac{\partial f}{\partial x}\right)-\left(\frac{\partial R}{\partial x}-\frac{\partial P}{\partial z}\right)\left(-\frac{\partial f}{\partial y}\right)+\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)(1)}{\sqrt{\left(\frac{\partial f}{\partial x}\right)^2+\left(\frac{\partial f}{\partial y}\right)^2+1}}\,dS.$$

This reduces to three equations, one for P, one for Q, and one for R.

We will establish the result for P, namely

$$\int\limits_C P\,dx=\int\limits_{\mathcal{S}}\frac{\frac{\partial P}{\partial z}(-\frac{\partial f}{\partial y})-\frac{\partial P}{\partial y}(1)}{\sqrt{(\frac{\partial f}{\partial x})^2+(\frac{\partial f}{\partial y})^2+1}}\,dS. \tag{18.6.10}$$

To change the integral over $\mathcal{S}$ to one over its projection, $\mathcal{S}^*$, on the xy-plane, we replace dS by $\sqrt{(\partial f/\partial x)^2+(\partial f/\partial y)^2+1}\,dA$. At the same time we project C onto a counterclockwise curve C^*. The square roots cancel leaving us with this equation in the xy-plane:

$$\int\limits_{C*}P(x,y,f(x,y))\,dx=\int\limits_R\left(-\frac{\partial P}{\partial z}\frac{\partial f}{\partial y}-\frac{\partial P}{\partial y}\right)\,dA. \tag{18.6.11}$$

We apply Green's theorem to the left side of (18.6.11), and obtain

$$\int\limits_{C^*}P(x,y,f(x,y))\,dx=\int\limits_{\mathcal{S}^*}-\frac{\partial P(x,y,f(x,y))}{\partial y}\,dA.$$

Be sure you understand each of the four steps in this proof, and why they are valid.

But

$$\frac{\partial P(x,y,f(x,y))}{\partial y}=\frac{\partial P}{\partial y}+\frac{\partial P}{\partial z}\frac{\partial f}{\partial y}. \tag{18.6.12}$$

Combining (18.6.11) and (18.6.12) completes the proof of (18.6.10).

We assumed that $\mathcal{S}$ has a special form, meeting lines parallel to an axis just once. More general surfaces, such as the surface of a sphere or a polyhedron can be cut into pieces of this type. Exercise 49 shows why Stokes' theorem holds in these cases also.

Summary

Stokes' theorem relates the circulation of a vector field over a closed curve C to the integral over a surface $\mathcal{S}$ with boundary C. The integrand over the surface is the component of the curl of the field perpendicular to the surface,

$$\int\limits_C \mathbf{F}\cdot\ d\mathbf{r} = \int\limits_{\mathcal{S}} (\mathbf{curl}\,\mathbf{F})\cdot\mathbf{n}\ dS.$$

The normal $\mathbf{n}$ to $\mathcal{S}$ is determined by the right-hand rule.

EXERCISES for Section 18.6

SHERMAN: Is the wording of these first two exercises consistent with similar questions in other sections? Please check.

1. State Stokes' theorem (symbols permitted).

2. State Stokes' theorem in words (symbols not permitted).

3. While we are careful to mention the orientation of a curve in the plane (either clockwise or counterclockwise). Why is this distinction not made for curves in space?

4. Let $F(r)$ be an antiderivative of $f(r)$. Show that $f(r)\widehat{\mathbf{r}}$ is the gradient of $F(r)$, hence is conservative. $(f(r)\frac{\mathbf{r}}{r} = f(r)\widehat{\mathbf{r}}.)$

5.

(a) Use the fact that a gradient, ∇f, is conservative, to show that its curl is $\mathbf{0}$.

(b) Compute $\nabla \times \nabla f$ in terms of components to show that the curl of a gradient is $\mathbf{0}$.

In Exercises 6 to 9 verify Stokes' theorem for $\mathbf{F}$ and $\mathcal{S}$.

6. $\mathbf{F} = xy^2\mathbf{i} + y^3\mathbf{j} + y^2z\mathbf{k}$, $\mathcal{S}$ is the top half of the sphere $x^2 + y^2 + z^2 = 1$.

7. $\mathbf{F} = y\mathbf{i} + xz\mathbf{j} + x^2\mathbf{k}$, $\mathcal{S}$ is the triangle with vertices $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$.

8. $\mathbf{F} = y^5\mathbf{i} + x^3\mathbf{j} + z^4\mathbf{k}$, $\mathcal{S}$ is the portion of $z = x^2 + y^2$ below the plane $z = 1$.

9. $\mathbf{F} = -y\mathbf{i} + x\mathbf{j} + z\mathbf{k}$, $\mathcal{S}$ is the portion of the cylinder $z = x^2$ inside the cylinder $x^2 + y^2 = 4$.

10. Assume that $\mathbf{F}$ is defined everywhere except on the z-axis and is irrotational. The curves C_1, C_2, C_3, and C_4 are as shown in Figure 18.6.12. What, if anything, can be said about

$$\oint_{C_1} \mathbf{F}\cdot\, d\mathbf{r}, \qquad \oint_{C_2} \mathbf{F}\cdot\, d\mathbf{r}, \qquad \oint_{C_3} \mathbf{F}\cdot\, d\mathbf{r}, \qquad \text{and} \qquad \oint_{C_4} \mathbf{F}\cdot\, d\mathbf{r}.$$

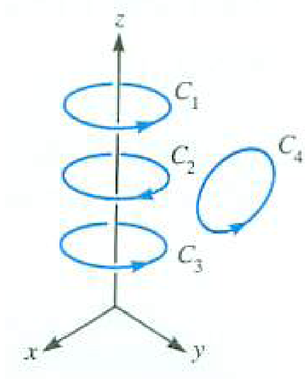


Figure 18.6.12

11. Evaluate $\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n}\, dS$, where $\mathbf{F}(x,y,z) = x\mathbf{i} - y\mathbf{j}$ and $\mathcal{S}$ is the surface of the cube bounded by the three coordinate planes and the planes $x = 1$, $y = 1$, $z = 1$, exclusive of the surface in the plane $x = 1$. Let $\mathbf{n}$ be outward from the cube.

12. Using Stokes' Theorem, evaluate $\int_{\mathcal{S}} (\nabla \times \mathbf{F}) \cdot \mathbf{n}\, dS$, where $\mathbf{F} = (x^2 + y - 4)\mathbf{i} + 3xy\mathbf{j} + (2xz + z^2)\mathbf{k}$, and $\mathcal{S}$ is the portion of the surface $z = 4 - (x^2 + y^2)$ above the xy plane. Let $\mathbf{n}$ be the upward normal.

In Exercises 13 to 16 use Stokes' theorem to evaluate $\oint_C \mathbf{F} \cdot d\mathbf{r}$ for $\mathbf{F}$ and C. Assume that C is oriented counterclockwise when viewed from above.

13. $\mathbf{F} = \sin(xy)\mathbf{i}$, C is the intersection of the plane $x + y + z = 1$ and the cylinder $x^2 + y^2 = 1$.

14. $\mathbf{F} = e^x\mathbf{j}$, C is the triangle with vertices $(2, 0, 0)$, $(0, 3, 0)$, and $(0, 0, 4)$.

15. $\mathbf{F} = xy\mathbf{k}$, C is the intersection of the plane $z = y$ with the cylinder $x^2 - 2x + y^2 = 0$.

16. $\mathbf{F} = \cos(x + z)\mathbf{j}$, C is the boundary of the rectangle with vertices $(1, 0, 0)$, $(1, 1, 1)$, $(0, 1, 1)$, and $(0, 0, 0)$.

17. Let $\mathcal{S}_1$ be the top half and $\mathcal{S}_2$ the bottom half of a sphere of radius a in space. Let $\mathbf{F}$ be a vector field defined on the sphere and let $\mathbf{n}$ denote an exterior normal to the sphere. What relation, if any, is there between $\int_{\mathcal{S}_1} (\nabla \times \mathbf{F}) \cdot \mathbf{n}\, dS$ and $\int_{\mathcal{S}_2} (\nabla \times \mathbf{F}) \cdot \mathbf{n}\, dS$?

18. Let $\mathbf{F}$ be a vector field throughout space such that $\mathbf{F}(P)$ is perpendicular to the curve C at each point P on C, the boundary of a surface $\mathcal{S}$. What can one conclude about

$$\int_{\mathcal{S}} (\nabla \times \mathbf{F}) \cdot \mathbf{n} \, dS?$$

19. Let C_1 and C_2 be two closed curves in the xy-plane that encircle the origin and are similarly oriented, as in Figure 18.6.13.

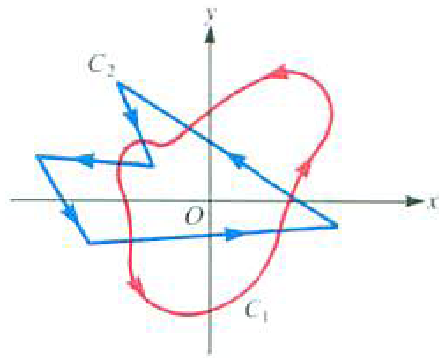


Figure 18.6.13

Let $\mathbf{F}$ be a vector field defined throughout the plane except at the origin. Assume that $\nabla \times \mathbf{F} = \mathbf{0}$.

(a) Must $\oint_C \mathbf{F} \cdot d\mathbf{r} = 0$?

(b) What, it any, relation exists between $\oint_{C_1} \mathbf{F} \cdot d\mathbf{r}$ and $\oint_{C_2} \mathbf{F} \cdot d\mathbf{r}$?

20. Let $\mathbf{F}$ be defined everywhere in space except on the z-axis. Assume also that $\mathbf{F}$ is irrotational, $\oint_{C_1} \mathbf{F} \cdot d\mathbf{r} = 3$, and $\oint_{C_2} \mathbf{F} \cdot d\mathbf{r} = 5$. See Figure 18.6.14. What if, anything, can be said about

(a) $\oint_{C_3} \mathbf{F} \cdot d\mathbf{r}$,

(b) $\oint_{C_4} \mathbf{F} \cdot d\mathbf{r}$?

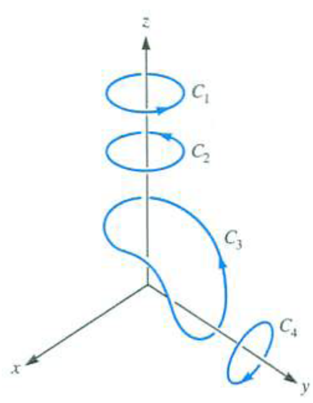


Figure 18.6.14

21. Which of the sets are connected? simply connected?

(a) A circle ($x^2 + y^2 = 1$) in the xy-plane

(b) A disk ($x^2 + y^2 \leq 1$) in the xy-plane

(c) The xy-plane from which a circle is removed

(d) The xy-plane from which a disk is removed

(e) The xy-plane from which one point is removed

(f) xyz-space from which one point is removed

(g) xyz-space from which a sphere is removed

(h) xyz-space from which a ball is removed

(i) A solid torus

(j) xyz-space from which a solid torus is removed

(k) A coffee cup with one handle

(l) xyz-space from which a solid torus is removed

22. Which central fields have curl $\mathbf{0}$?

23. Let $\mathcal{V}$ be the solid bounded by $z = x + 2$, $x^2 + y^2 = 1$, and $z = 0$. Let $\mathcal{S}_1$ be the portion of the plane $z = x + 2$ that lies within the cylinder $x^2 + y^2 = 1$. Let C be the boundary of $\mathcal{S}_1$, with a counterclockwise orientation as viewed from above. Let $\mathbf{F} = y\mathbf{i} + xz\mathbf{j} + (x + 2y)\mathbf{k}$. Use Stokes' Theorem for $\mathcal{S}_1$ to evaluate $\oint_C \mathbf{F} \cdot d\mathbf{r}$.

24. (See Exercise 23.) Let $\mathcal{S}_2$ be the curved surface of $\mathcal{V}$ together with the base of $\mathcal{V}$. Use Stokes' Theorem for $\mathcal{S}_2$ to evaluate $\oint_C \mathbf{F} \cdot d\mathbf{r}$.

25. Verify Stokes' theorem for the special case when $\mathbf{F}$ has the form ∇f, that is, is a gradient field.

26. Let $\mathbf{F}$ be a vector field defined on the surface $\mathcal{S}$ of a convex solid. Show that $\int_{\mathcal{S}} (\nabla \times \mathbf{F}) \cdot \mathbf{n}\, dS = 0$

(a) by the Divergence theorem,

(b) by drawing a closed curve on C on $\mathcal{S}$ and using Stokes' theorem on the parts into which C divides $\mathcal{S}$.

27. Evaluate $\oint_C \mathbf{F}\cdot d\mathbf{r}$ if $\mathbf{F}(x,y,z) = (-y\mathbf{i}+x\mathbf{j})/(x^2+y^2)$ and C is the intersection of the plane $z = 2x+2y$ and the paraboloid $z = 2x^2+3y^2$ oriented counterclockwise as viewed from above.

28. Let $\mathbf{F}(x,y)$ be a vector field defined everywhere in the plane except at the origin. Assume that $\nabla\times\mathbf{F} = \mathbf{0}$. Let C_1 be the circle $x^2+y^2 = 1$ counterclockwise; let C_2 be the circle $x^2+y^2 = 4$ clockwise; let C_3 be the circle $(x-2)^2+y^2 = 1$ counterclockwise, and let C_4 be the circle$(x-1)^2+y^2 = 9$ clockwise. Assuming that $\oint_{C_1}\mathbf{F}\cdot d\mathbf{r}$ is 5, evaluate

(a) $\oint_{C_2}\mathbf{F}\cdot d\mathbf{r}$

(b) $\oint_{C_3}\mathbf{F}\cdot d\mathbf{r}$

(c) $\oint_{C_4}\mathbf{F}\cdot d\mathbf{r}$.

29. Let $\mathbf{F}(x,y,z) = \mathbf{r}/|\mathbf{r}|^a$, where $\mathbf{r} = \mathbf{i}+u\mathbf{j}+z\mathbf{k}$ and a is a real number.

(a) Show that $\nabla\times\mathbf{F} = \mathbf{0}$.

(b) Show that $\mathbf{F}$ is conservative.

(c) Exhibit a scalar function f such that $\mathbf{F} = \nabla f$.

30. Explain why (18.6.4) holds if $\mathcal{S}_1$ and $\mathcal{S}_2$ together form the boundary surface $\mathcal{S}$ of a solid region R. Use the Divergence Theorem, not Stokes' Theorem.

31. Show that a central field $f(r)\widehat{\mathbf{r}}$ is conservative by showing that it is irrotational and defined on a simply connected region. (Express $\widehat{\mathbf{r}}\mathbf{r}$ in terms of x, y and z.) See also Exercise 45.)

Work Exercises 32 to 34 in preparation for Exercise 35.

32. Assume that $\mathbf{G}$ is the curl of a vector field $\mathbf{F}$, $\mathbf{G} = \nabla\times\mathbf{F}$. Let $\mathcal{S}$ be a surface that bounds a solid region V. Let C be a closed curve on the surface $\mathcal{S}$ breaking $\mathcal{S}$ into two pieces $\mathcal{S}_1$ and $\mathcal{S}_2$. Show that $\int_{\mathcal{S}_1}\mathbf{G}\cdot\mathbf{n}\,dS = -\int_{\mathcal{S}_2}\mathbf{G}\cdot\mathbf{n}\,dS$.

33. Using the Divergence theorem, show that $\int_{\mathcal{S}}\mathbf{G}\cdot\mathbf{n}\,dS = 0$.

34. Using Stokes' theorem, show that $\int_{\mathcal{S}}\mathbf{G}\cdot\mathbf{n}\,dS = 0$. (Break the integral into integrals over $\mathcal{S}_1$ and $\mathcal{S}_2$.)

35. Let $\mathbf{F} = e^{xy}\mathbf{i} + \tan(3yz)\mathbf{j} + 5z\mathbf{k}$ and $\mathcal{S}$ be the tetrahedron whose vertices are $(0,0,0)$, $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$. Let $\mathcal{S}_1$ be the base of $\mathcal{S}$ in the xy-plane and $\mathcal{S}_2$ consist of the other three faces. Find $\int_{\mathcal{S}} (\nabla \times \mathbf{F}) \cdot \mathbf{n}\, dS$. (Think about the preceding three exercises.)

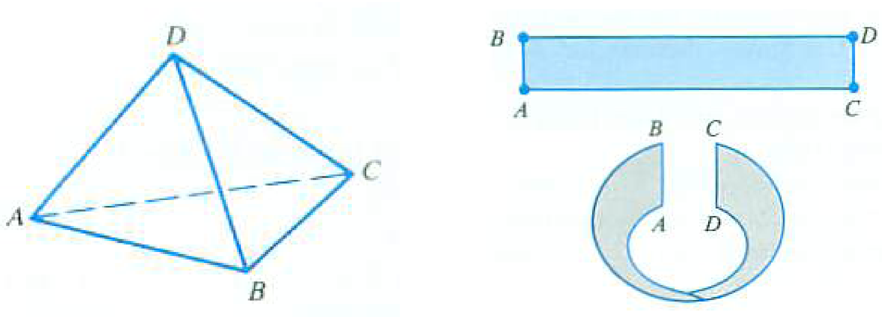
36. Let $\mathbf{F}$ be defined throughout space and have continuous divergence and curl.

(a) For which $\mathbf{F}$ is $\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n}\, dS = 0$ for all spheres $\mathcal{S}$?

(b) For which $\mathbf{F}$ is $\oint_C \mathbf{F} \cdot\, d\mathbf{r} = 0$ for all circles C?

37. Let C be the curve formed by the intersection of the plane $z = x$ and the paraboloid $z = x^2 + y^2$. Orient C to be counterclockwise when viewed from above. Evaluate $\oint_C (xyz\, dx + x^2\, dy + xz\, dz)$.

38. Assume that Stokes' theorem is true for triangles. Deduce that it holds for the surface $\mathcal{S}$ in Figure 18.6.15(a), consisting of the three triangles DAB, DBC, DCA, and the curve $ABCA$.



(a) (b)

Figure 18.6.15

39. Sam has a different way to choose $\mathbf{n}$.

Sam: I think the book's way of choosing $\mathbf{n}$ is too complicated.

Jane: OK. How would you do it?

Sam: Glad you asked. First, I would choose a unit normal $\mathbf{n}$ at one point on the orientable surface.

Jane: That's a good start.

Sam: Then I choose unit normals in a continuous way everywhere on the surface starting at my initial choice.

Jane: And how would you finish?

Sam: My last step is to orient the boundary curve to be compatible with the right-hand rule.

Would this proposal work? If it does, would it agree with the approach in the text?

40. See also Exercises 4 and 31.

Sam: The only conservative fields in space that I know are the inverse square central fields with centers anywhere I please.

Jane: There are a lot more.

Sam: Oh?

Jane: Start with any scalar function $f(x, y, z)$ with continuous partial derivation of the first and second orders. Then its gradient will be a conservative field.

Sam: O.K. But I bet there are still more.

Jane: No. I got them all.

Who is right?

Exercises 41 to 46 concern the proof of Stokes' Theorem.

41. Carry out the calculations in the proof that translated Stokes' Theorem into an equation involving P, Q, and R.

42. Draw a picture of $\mathcal{S}$, $\mathcal{S}^*$, C, and C^* that appear in the proof of Stokes' Theorem.

43. Write the four steps involved in the proof of Stokes' Theorem, giving an explanation for each.

44. In the proof of Stokes' Theorem we used a normal $\mathbf{n}$. Show that it is the correct one, compatible with counterclockwise orientation of C^*.

45.

(a) State Stokes' Theorem for $\int_C Q\ dy$.

(b) Prove Stokes' Theorem for $\int_C Q\ dy$.

(c) State Stokes' Theorem for $\int_C R\ dz$.

(d) Prove Stokes' Theorem for $\int_C R\ dz$.

46. Draw a picture of $\mathcal{S}$, $\mathcal{S}^*$, C and C^* that appear in the proof of Stokes' theorem.

47. A Möbius band can be made by making a half-twist in a narrow rectangular strip, bringing the two ends together, and fastening them with glue or tape. See Figure 18.6.15(b).

(a) Make a Möbius band.

(b) Letting a pencil represent a normal $\mathbf{n}$ to the band, check that it is not orientable.

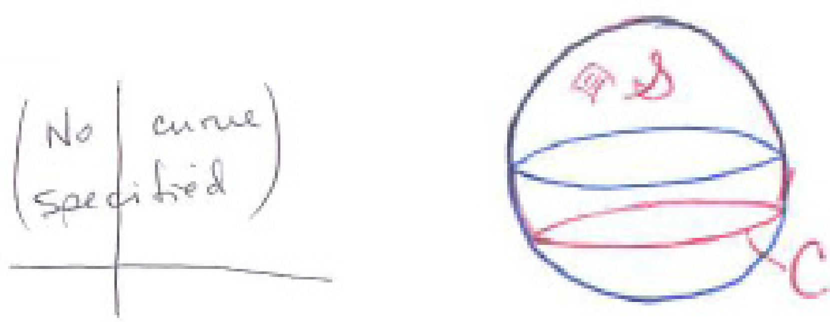
(c) If you form a band by putting in a full twist (360°), is it orientable?

(d) What happens when you cut the bands in (a) and (c) down the middle? one third of the way from one edge to the other?

48.

(a) Explain why the line integral of a central vector field $f(r)\widehat{\mathbf{r}}$ around the path in Figure 18.6.16(a) is 0.

(b) Deduce from (a) and the coordinate-free view of curl that the curl of a central field is $\mathbf{0}$.



(a) (b)

Figure 18.6.16

49.

(a) The proof of Stokes' theorem that we gave would not apply to surfaces that are more complicated, such as the top three fourths of a sphere, shown in Figure 18.6.16(b). How could you cut $\mathcal{S}$ into pieces to each of which the proof applies? Describe them, in words.

(b) How could you use (a) to show that Stokes' Theorem holds for C and $\mathcal{S}$ in Figure 18.6.16(b)

50. We dealt only with the component P. What is the analog of (18.6.10) for Q? Prove it. (The steps would parallel the steps used for P.)

18.7 Connections Between the Electric Field and $\widehat{\mathbf{r}}/|\mathbf{r}|^2$

We now introduce one of the four equations that describe the phenomena of electricity and magnetism. The other three will be encountered later in the chapter. All are expressed in terms of vector fields. Using just these four equations, James Clerk Maxwell predicted that light is an electromagnetic phenomenon. We do not assume a prior knowledge of physics.

The Electric Field Due To a Single Charge

We make some assumptions about the fundamental electrical charges, electrons and protons. An electron has a negative charge and a proton has a positive charge of equal absolute value. Two like charges exert a force of repulsion on each other and unlike charges attract each other.

Let C and P denote the location of charges q and q_0, respectively. Let $\mathbf{r}$ be the vector from C to P, as in Figure 18.7.1, so $r = |\mathbf{r}|$ is the distance between them. The unit vector $\widehat{\mathbf{r}}$ in the direction of $\mathbf{r}$ is $\mathbf{r}/r$.

If both q and q_0 are protons or both are electrons, the force pushes them farther apart. If one is a proton and the other is an electron, the force draws them closer. The magnitude of the force is inversely proportional to r^2.

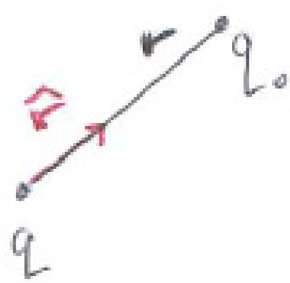


Figure 18.7.1

Assume that q is positive. The magnitude of the force it exerts on charge q_0 is proportional to q and q_0, and it is also inversely proportional to r^2. So, for some constant k, the magnitude of the force is

$$k\frac{q\ q_0}{r^2}.$$

It is directed along the vector $\mathbf{r}$. If q_0 is also positive, it is in the same direction as $\mathbf{r}$. If q_0 is negative, it is in the direction of $-\mathbf{r}$. So we have the vector equation

$$\mathbf{F} = k\frac{q\ q_0}{r^2}\widehat{\mathbf{r}} \qquad (18.7.1)$$

where k is positive.

We will write k as $1/(4\pi\epsilon_0)$. The value of ϵ_0 depends on the units in which charge, distance, and force are measured. Then (18.7.1) is written

Read ϵ_0 as "epsilon zero" or "epsilon null."

$$\mathbf{F} = \frac{q\ q_0}{4\pi\epsilon_0 r^2}\widehat{\mathbf{r}}.$$

Physicists associate a vector field with a charge q. The field exerts a force on other charges.

A positive charge q at point C. creates a central inverse-square vector field $\mathbf{E}$ with center at C. It is defined everywhere except at C. Its value at a point

P is

$$\mathbf{E}(P) = \frac{q\,\widehat{\mathbf{r}}}{4\pi\epsilon_0 r^2}$$

where $\mathbf{r} = \overline{CP}$, as in Figure 18.7.2.

(a) (b) (c)

Figure 18.7.2

The value of $\mathbf{E}$ depends only on q and the vector from C to P.

The force $\mathbf{F}$ exerted by charge q on charge q_0 at P is obtained by multiplying $\mathbf{E}$ by q_0:

$$\mathbf{F} = q_0\mathbf{E}. \tag{18.7.2}$$

The field $\mathbf{E}$ can be calculated in principle by putting a charge q_0 at P, observing the force $\mathbf{F}$ and then dividing $\mathbf{F}$ by q_0. The field $\mathbf{E}$ enables the charge q to act at a distance on other charges. It plays the role of a rubber band or a spring.

The Electric Field Due to a Distribution of Charge

Electrons and protons usually do not exist in isolation. Charge may be distributed on a line, a curve, a surface or in space.

If a total charge Q occupies a region R in space its density varies from point to point. Denote the density at P by $\delta(P)$. Like the density of mass, it is defined as a limit. Let $V(r)$ be a small ball of radius r and center at P. Then we have the definition

$$\delta(P) = \lim_{r\to 0^+} \frac{\text{charge in } V(r)}{\text{volume of } V(r)}.$$

The charge in $V(r)$ is approximately the volume of $V(r)$ times $\delta(P)$. We will be interested only in uniform charges, where the density is constant with the value δ. Thus the charge in a region of volume V is δV.

The field due to a uniform charge Q distributed in a region R is the sum of the fields due to the individual point charges in Q. To describe the field we need the concept of the integral of a vector field. The definition is similar to the definition of the definite integral in Section 6.2. Let $\mathbf{F}(P)$ be a continuous vector field defined on a solid region R. Break R into regions R_1, R_2, $\dots$, R_n and choose a point P_i in R_i, $1 \le i \le n$. Let the volume of R_i be V_i. The sums $\sum_{i=1}^{n} \mathbf{F}(P_i)V_i$ have a limit as all R_i are chosen smaller and smaller. This limit, denoted $\int_R \mathbf{F}(P)\ dV$ is called the **integral of F over** R. TIt can be computed componentwise. For example, if $\mathbf{F} = F_1\mathbf{i} + F_2\mathbf{j} + F_3\mathbf{k}$ then $\int_R \mathbf{F}(P)\ dV = \int_R F_1\ dV\mathbf{i} + \int_R F_2\ dV\mathbf{j} + \int_R F_3\ dV\mathbf{k}$. Similar definitions hold for vector fields defined on surfaces or curves.

To estimate the field due to a distribution of charge, partition R into small regions R_1, R_2, $\dots$, R_n and choose a point P_i in R_i, $i = 1, 2 \dots, n$. The volume of R_i is V_i. The charge in R_i is δV_i, where δ is the density of the charge. Figure 18.7.3 shows the contribution to the field at a point P.

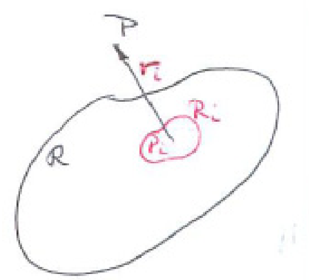


Figure 18.7.3

Let $\mathbf{r}_i$ be the vector from P_i to P, and $r_i = |\mathbf{r}_i|$. Then the field due to the charge in R_i is approximately

$$\frac{\delta\ V_i\ \widehat{\mathbf{r}}_i}{4\pi\epsilon_0\ r_i^2} = \frac{\delta\ \widehat{\mathbf{r}}_i}{4\pi\epsilon_0\ r_i^2} V_i.$$

As an estimate of the field due to Q, we have the sum

$$\sum_{i=1}^{n} \frac{\delta\ \widehat{\mathbf{r}}_i\ V_i}{4\pi\epsilon_0 r_i^2}.$$

Taking limits as all the regions R_i are chosen smaller, we have

$$\mathbf{E}(P) = \text{Field at } P = \int\limits_R \frac{\delta\ \widehat{\mathbf{r}}}{4\pi\epsilon_0 r^2}\ dV$$

Factoring out the constant $\delta/(4\pi\epsilon_0)$, we have

$$\mathbf{E}(P) = \frac{\delta}{4\pi\epsilon_0} \int\limits_R \frac{\widehat{\mathbf{r}}}{r^2}\ dV$$

That is an integral over a solid region. If the charge is on a surface S with uniform surface density σ, the field would be given by

$$\mathbf{E}(P) = \frac{\sigma}{4\pi\epsilon_0} \int\limits_S \frac{\widehat{\mathbf{r}}}{r^2}\ dS.$$

If the charge lies on a line or a curve C, with uniform density λ, then

$$\mathbf{E}(P) = \frac{\lambda}{4\pi\epsilon_0} \int\limits_C \frac{\widehat{\mathbf{r}}}{r^2}\ ds.$$

To illustrate the definition we compute one such field value directly. In Example 2 we solve the same problem more simply.

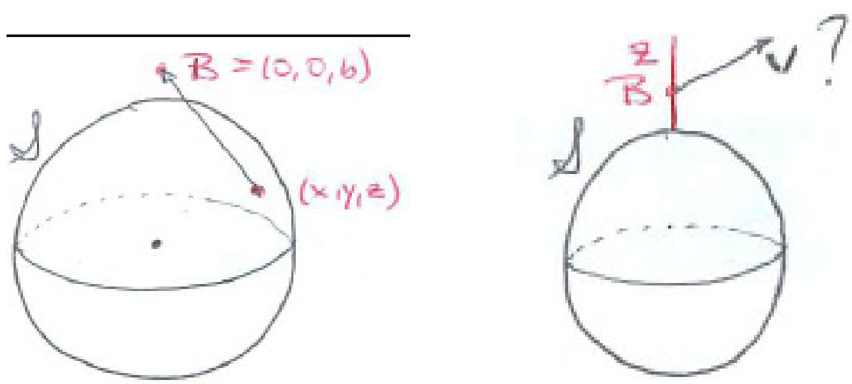
EXAMPLE 1 A charge Q is uniformly distributed on a sphere of radius a, $\mathcal{S}$. Find the electrostatic field $\mathbf{E}$ at a point B a distance $b > a$ from the center of the sphere.

SOLUTION We evaluate

$$\frac{\sigma}{4\pi\epsilon_0}\int\limits_{S}\frac{\widehat{\mathbf{r}}}{r^2}\,dS. \tag{18.7.3}$$

Since the charge is uniform over a region with area $4\pi a^2$, $\sigma = Q/4\pi a^2$.

Place a rectangular coordinate system with its origin at the center of the sphere and the z-axis on B, so that $B = (0,0,b)$, as in Figure 18.7.4(a). We



(a) (b)

Figure 18.7.4

us use the symmetry of the sphere to predict something about the vector $\mathbf{E}(B)$. Could it look like the vector $\mathbf{v}$, which is not parallel to the z-axis, as in Figure 18.7.4(b)?

If you spin the sphere around the z-axis, the vector $\mathbf{v}$ would change. But the sphere is unchanged and so is the charge. So $\mathbf{E}(B)$ must be parallel to the z-axis. That means its x- and y-components are both 0. So we need only find its z-component, which is $\mathbf{E}(B)\cdot\mathbf{k}$.

Let (x, y, z) be a point on the sphere $\mathcal{S}$. Then

$$\mathbf{r} = (0\mathbf{i} + 0\mathbf{j} + b\mathbf{k}) - (x\mathbf{i} + y\mathbf{j} - z\mathbf{k}) = -x\mathbf{i} - y\mathbf{j} + (b - z)\mathbf{k}. \tag{18.7.4}$$

So

$$\frac{\widehat{\mathbf{r}}}{r^2} = \frac{\mathbf{r}}{r^3} = \frac{-x\mathbf{i} - y\mathbf{j} + (b-z)\mathbf{k}}{(\sqrt{x^2 + y^2 + b^2 - 2bz + z^2})^3} = \frac{-x\mathbf{i} - y\mathbf{j} + (b-z)\mathbf{k}}{(a^2 + b^2 - 2bz)^{3/2}}. \tag{18.7.5}$$

We need only its z-component,

$$\frac{b-z}{(a^2+b^2-2bz)^{3/2}}.$$

The magnitude of $\mathbf{E}(B)$ is therefore

$$\frac{\sigma}{4\pi\epsilon_0}\int\limits_S \frac{b-z}{(a^2+b^2-2bz)^{3/2}}\,dS. \tag{18.7.6}$$

We evaluate the integral in (18.7.6) by introducing spherical coordinates in the standard position. We have $dS = a^2\sin(\phi)d\phi\ d\theta$ and $z = a\cos(\phi)$. So (18.7.6) becomes

$$\int\limits_0^{\pi}\int\limits_0^{2\pi} \frac{(b-a\cos\phi)a^2\sin\phi}{(a^2+b^2-2ab\cos\phi)^{(3/2)}}\,d\theta\ d\phi$$

which reduces, after the first integration with respect to θ, to

$$2\pi a^2\int\limits_0^{\pi} \frac{(b-a\cos\phi)\sin\phi\ d\phi}{(a^2+b^2-2ab\cos\phi)^{3/2}}. \tag{18.7.7}$$

Let $u=\cos(\phi)$, hence $du = -\sin(\phi)\ d\phi$. This transforms (18.7.7) into

$$-2\pi a^2\int\limits_1^{-1} \frac{(b-au)\ du}{(a^2+b^2-2abu)^{3/2}}. \tag{18.7.8}$$

Then we make a substitution, $v = a^2+b^2-2abu$, which. changes (18.7.8) into

$$\frac{2\pi a^2}{4ab^2}\int\limits_{(b-a)^2}^{(b+a)^2} \frac{v+b^2-a^2}{v^{3/2}}\,dv \tag{18.7.9}$$

Write the integrand as the sum of $1/\sqrt{v}$ and $(b^2-a^2)/v^{3/2}$ and use the fundamental theorem of calculus to show that (18.7.8) equals $4\pi a^2/b^2$.

Combining this with (18.7.9) shows that

$$\mathbf{E}(B) = \frac{\sigma}{4\pi\epsilon_0}\frac{4\pi a^2}{b^2}\mathbf{k} = \frac{Q}{4\pi\epsilon_0 b^2}\mathbf{k}.$$

◇

The result in this example, $Q/(4\pi\epsilon_0 b^2)\mathbf{k}$, is the same as if all the charge Q were at the center of the sphere. A uniform charge on a sphere acts on external particles as though the whole charge were placed at its center. This was discovered for the gravitational field by Newton and proved geometrically in his *Principia* of 1687.

DOUG: Do we still need to insert mention of earlier calculation for gravity; mention section and page.

Using Flux and Symmetry to Find E

We included Example 1 because it reviews some integration techniques and serves as a contrast to a field **E** due to a charge distribution.

Suppose a charge Q is distributed outside the region bound by a surface S, as in Figure 18.7.5.

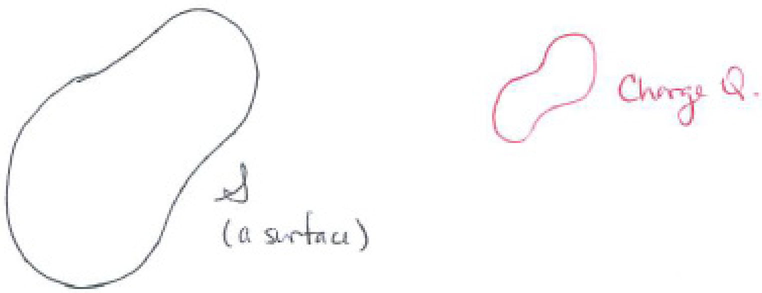


Figure 18.7.5

Let **r** be the vector from q to a typical point in S. The flux of **E** associated with a point charge q over a closed surface $\mathcal{S}$ is

$$\int\limits_{\mathcal{S}} \mathbf{E}(P) \cdot \mathbf{n}\, dS = \int\limits_{\mathcal{S}} \frac{q\widehat{\mathbf{r}} \cdot \mathbf{n}}{4\pi\epsilon_0 r^2}\, dS = \frac{q}{4\pi\epsilon_0} \int\limits_{\mathcal{S}} \frac{\widehat{\mathbf{r}} \cdot \mathbf{n}}{r^2}\, dS.$$

As we saw in Section 18.5, the integral is 4π when the charge is inside the solid bounded by the surface and 0 if the charge is outside. (See Exercise 28 in that section). Thus the total flux is q/ϵ_0 if the charge is inside and 0 if it is outside.

Next we will find the flux of a charge Q distributed in a solid R inside a surface $\mathcal{S}$. (See Exercise 5 for the case when the charge is outside $\mathcal{S}$.)

Divide R into n small regions R_1, R_2, ..., R_n. In region R_i select a point P_i. Let the density of charge at P_i be $\delta(P_i)$. Thus the charge in R_i produces a flux of approximately $\delta(P_i)V_i/\epsilon_0$. Consequently

$$\sum_{i=1}^{n} \frac{\delta(P_i)V_i}{\epsilon_0}$$

estimates the flux produced by Q. Taking limits, we see that

$$\text{Flux across } S \text{ produced by } Q = \int\limits_R \frac{\delta(P)}{\epsilon_0}\, dV.$$

But $\int_R \delta(P)\, dV$ is the total charge Q. Thus we have

$$\text{Flux} = \frac{Q}{\epsilon_0}.$$

Thus we have one of the four fundamental equations of electrostatics:

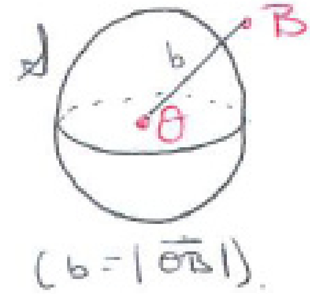


Figure 18.7.6

Gauss's Law
The flux produced by a distribution of charge across a closed surface is the charge Q in the region bounded by the surface divided by ϵ_0.

The charge outside of S produces no flux across S because the negative flux across S cancels the positive flux.

We illustrate the power of Gauss's law by applying it to the case in Example 1.

EXAMPLE 2 A charge Q is distributed uniformly on a sphere of radius a. Find the electrostatic field $\mathbf{E}$ at a point B at a distance b from the center of the sphere with $b > a$.

SOLUTION We do not need to introduce a coordinate system in Figure 18.7.6. By symmetry, the field at a point P outside the sphere is parallel to the vector $\overrightarrow{CP}$. Moreover, its magnitude is the same for all points the same distance from the origin C. Call the magnitude $f(r)$ where r is the distance from C. We want to find $f(b)$.

Figure 18.7.7

To do this, take another sphere S^*, with center C and radius b, as in Figure 18.7.7. The flux of $\mathbf{E}$ across S^* is $\int_{S^*} \mathbf{E}\cdot\mathbf{n}\, dS$.

But $\mathbf{E}\cdot\mathbf{n}$ is just $f(b)$ since $\mathbf{E}$ and $\mathbf{n}$ are parallel and $\mathbf{E}(P)$ has magnitude $f(b)$ for all points P on S^*. Thus $\int_{S^*} \mathbf{E}\cdot\mathbf{n}\, dS = \int_{S^*} f(b)\, dS = f(b)\int_{S^*}\, dS = f(b)4\pi b^2$.

By Gauss's law

$$\frac{Q}{\epsilon_0} = f(b)(4\pi b^2).$$

That tells us that

$$f(b) = \frac{Q}{4\pi\epsilon_0 b^2}.$$

This is the same result as in Example 1, but uses less work. Symmetry and Gauss's law provide a way to find the electrostatic field due to distribution of charge. ◇

The same approach shows that the field $\mathbf{E}$ produced by the spherical charge in Examples 1 and 2 inside the sphere is $\mathbf{0}$. Let $f(r)$ be the magnitude of $\mathbf{E}$ at a distance r from the center of the sphere. For $r > a$, $f(r) = Q/(4\pi\epsilon_0 r^2)$ and for $0 < r < a$, $f(r) = 0$. The graph of f is in Figure 18.7.8.

Summary

The field due to a point charge q at a point C is given by the formula $\mathbf{E}(P) = \frac{1}{4\pi\epsilon_0}\frac{q\widehat{\mathbf{r}}}{r^2}$, where $\mathbf{r} = \overrightarrow{OP}$. This field produces a force $q_0\mathbf{E}(P)$ on a charge q_0

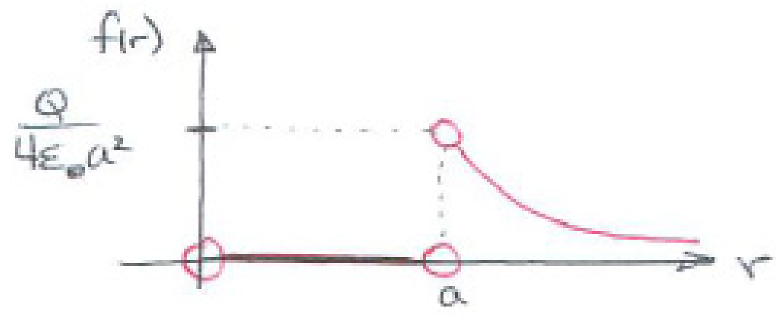


Figure 18.7.8

located at P. The field due to a distribution of charge is obtained by an integration.

We showed that a charge Q outside a surface produces a net flux of zero across the surface. However the flux produced by a charge within the surface is Q/ϵ_0. That is Gauss's law.

We used Gauss's law to find the field produced by a spherical distribution of charge.

EXERCISES for Section 18.7

1. Fill in the omitted details in the calculation in Example 1.

2. Describe to a friend who knows no physics the field $\mathbf{E}$ produced by a point charge q.

3. State Gauss's law using no mathematical symbols.

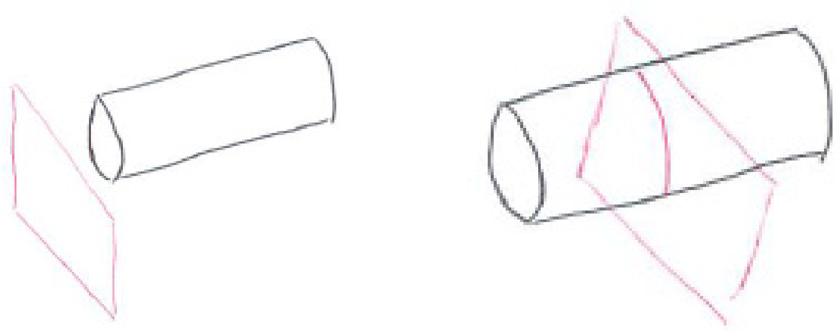
4. Why do you think that the constant k was replaced by $1/4\pi\epsilon_0$? Later we will see why it is convenient to have ϵ_0 in the denominator.

5. Show that a charge Q distributed in a solid region R outside a closed surface $\mathcal{S}$ induces zero flux across $\mathcal{S}$.

6. A charge is distributed uniformly over an infinite plane. For a part of this surface of area A the charge is kA, with density k. Find the field $\mathbf{E}$ due to the charge at any point P not in the plane.

(a) Use symmetry to say as much as you can about it. Be sure to discuss its direction.

(b) Show that the magnitude is constant by applying Gauss's law to a cylinder whose axis is perpendicular to the plane and which does not intersect the plane as shown in Figure 18.7.9(a).



(a) (b)

Figure 18.7.9

(c) Find the magnitude of $\mathbf{E}$ by applying Gauss's law to the cylinder in Figure 18.7.9(b) which intersects the plane of the charge. Let the area of the circular cross section be A and the area of its curved side be B.

7. Find the field **E** of the charge in Example 1 at a point on the surface of the sphere. Why is Gauss's law not applicable here? (Let the point be $(0, 0, a)$.)

8. Find the field **E** of the charge in Example 1 at the center of the sphere. (Use symmetry, don't integrate.)

9. Complete the graph in Figure 18.7.8. That is, fill in the function values corresponding to $r = 0$ and $r = a$.

10. A charge is distributed uniformly along an infinite straight wire. The charge on a section of length l is kl. Find the field **E** due to the charge.

(a) Use symmetry to say as much as you can about the direction and magnitude of **E**.

(b) Find the magnitude by applying Gauss's Law to the cylinder of radius r and height h shown in Figure 18.7.10

(c) Find the force directly by an integral over the line, as in Example 1.

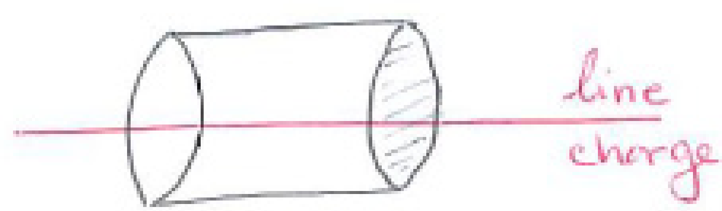


Figure 18.7.10

11. Suppose that there is a uniform distribution of charge Q throughout a ball of radius a. Use Gauss' law to find the electrostatic field **E** produced by the charge

(a) at points outside the ball

(b) at points inside the ball.

12. Let $f(r)$ be the magnitude of the field in Exercise 11 at a distance r from the center of the ball. Graph $f(r)$ for $r \geq 0$.

13. A charge Q lies partly inside a closed surface S and partly outside. Let Q_1 be the amount inside and Q_2 the amount outside, as in Figure 18.7.11. What is the flux across $\mathcal{S}$ of the charge Q?

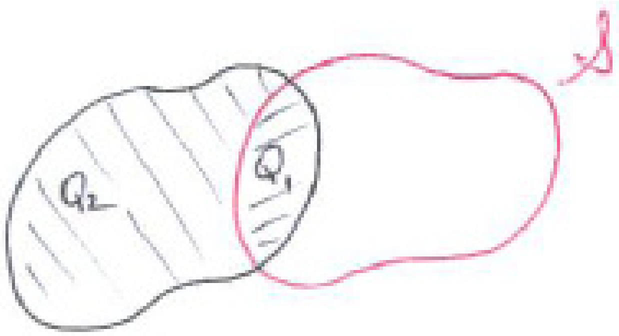


Figure 18.7.11

14. Exercise 10 concerned the field $\mathbf{E}$ due to a charge uniformly spread on an infinite line. If the charge density is λ, $\mathbf{E}$ at a point at a distance a from the line is $(\lambda/(2\pi a\epsilon_0))\,\mathbf{j}$.
Assume instead that the line occupies only the right half of the x-axis, $[0, \infty)$.

(a) Using the result in Exercise 10, show that the $\mathbf{j}$-component of $\mathbf{E}(0, a)$ is $(\lambda/4\pi a\epsilon_0)\mathbf{j}$.

(b) By integrating over $[0, \infty)$, show that the $\mathbf{i}$-component of $\mathbf{E}$ at $(0, a)$ is $\lambda/(4\pi a\epsilon_0)\mathbf{i}$.

(c) What angle does $\mathbf{E}(0, a)$ make with the y-axis?

(d) Why is Gauss' law of no use in determining the $\mathbf{i}$-component of $\mathbf{E}$?

15. We showed that $\mathbf{E}(P) = \frac{\delta}{4\pi\epsilon_0}\int_R \frac{\widehat{\mathbf{r}}}{r^2}\,dV$ if the charge density is constant. Find the integral for $\mathbf{E}(P)$ when the charge density varies.

16. In Example 1, we used an integral to find the electrostatic field outside a uniformly charged sphere. Carry out similar calculation to find the field inside the sphere. (Is the square root of $(b-a)^2$ still $b-a$?)

17. Use the approach in Example 2 to find the electrostatic field inside a uniformly charged sphere.

18. Graph the magnitude of the field in Example 1 as a function of the distance from the center of the sphere. This will need the results of Exercises 16 or 17.

19. Find the field $\mathbf{E}$ in the Exercise 6 by integrating over the whole (infinite) plane. Do not use Gauss's law.

18.8 Expressing Vector Functions in Other Coordinate Systems

We have expressed the gradient, divergence, and curl in rectangular coordinates. In engineering and physics functions are also expressed in polar, cylindrical, and spherical coordinates. This section shows how to compute **grad**, div, and **curl** in such cases.

The Gradient in Polar Coordinates

Let $g(r,\theta)$ be a scalar function expressed in polar coordinates. Its gradient has the form $A(r,\theta)\widehat{\mathbf{r}} + B(r,\theta)\widehat{\boldsymbol{\theta}}$, where $\widehat{\mathbf{r}}$ and $\widehat{\boldsymbol{\theta}}$ are the unit vectors shown in Figure 18.8.1. The unit radial vector $\widehat{\mathbf{r}}$ points in the direction of increasing r. The unit tangential vector $\widehat{\boldsymbol{\theta}}$ points in the direction determined by increasing θ. Note that $\widehat{\boldsymbol{\theta}}$ is tangent to the circle through (r,θ) with center at the pole.

Our goal is to find $A(r,\theta)$ and $B(r,\theta)$, which we denote as A and B.

One might guess, in analogy with rectangular coordinates, that $A(r,\theta)$ would be $\partial g/\partial r$ and $B(r,\theta)$ would be $\partial g/\partial\theta$. That guess is part right and part wrong, for we will show that

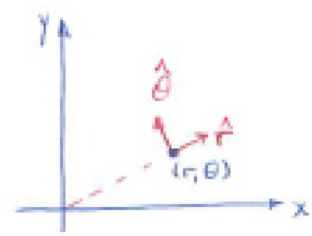


Figure 18.8.1

$$\boxed{\mathbf{grad}\, g = \frac{\partial g}{\partial r}\widehat{\mathbf{r}} + \frac{1}{r}\frac{\partial g}{\partial \theta}\widehat{\boldsymbol{\theta}}}$$

In the $\widehat{\boldsymbol{\theta}}$ component an extra factor of $1/r$ appears.

We reserve the use of ∇ for rectangular coordinates, and use **grad** in other coordinate systems.

One way to obtain (18.8) is long and not illuminating: express g, $\widehat{\mathbf{r}}$, and $\widehat{\boldsymbol{\theta}}$ in terms of x, y, $\mathbf{i}$, and $\mathbf{j}$ and use the formula for gradient in terms of rectangular coordinates. Then translate it to polar coordinates. This approach is outlined in Exercises 17 and 18.

We will use a simpler way, one that generalizes to cylindrical and spherical coordinates. It exploits the connection between a gradient and directional derivative of g at a point P in the direction $\mathbf{u}$. It shows why the coefficient $1/r$ appears in (18.8).

If $\mathbf{u}$ is a unit vector, the directional derivative of g in the direction $\mathbf{u}$ is the dot product of $\mathbf{grad}\, g$ with $\mathbf{u}$:

$$D_{\mathbf{u}}g = \mathbf{grad}\, g \cdot \mathbf{u}, \tag{18.8.1}$$

so

$$D_{\widehat{\mathbf{r}}}g = (A\widehat{\mathbf{r}} + B\widehat{\boldsymbol{\theta}}) \cdot \widehat{\mathbf{r}} = A \tag{18.8.2}$$

and

$$D_{\widehat{\boldsymbol{\theta}}}g = (A\widehat{\mathbf{r}} + B\widehat{\boldsymbol{\theta}}) \cdot \widehat{\boldsymbol{\theta}} = B. \tag{18.8.3}$$

So all we need to do is find $D_{\widehat{\mathbf{r}}}g$ and $D_{\widehat{\boldsymbol{\theta}}}g$.

We have

$$D_{\widehat{\mathbf{r}}}(g) = \lim_{\Delta r \to 0} \frac{g(r + \Delta r, \theta) - g(r, \theta)}{\Delta r} = \frac{\partial g}{\partial r},$$

so $A(r, \theta) = \partial g / \partial r(r, \theta)$.

Now we will see why B is not simply the partial derivation of g with respect to θ.

SHERMAN: Read carefully. A combination of our latest and Woody's.

To estimate the directional derivative at P of g in the direction of $\widehat{\boldsymbol{\theta}}$ in the case where $P = (r, \theta)$ and Q is a distance Δs from P in the direction of $\widehat{\boldsymbol{\theta}}$, form the quotient

$$\frac{g(Q) - g(P)}{\Delta s}. \tag{18.8.4}$$

The distance from P to Q is not Δs. It is approximately $r\Delta\theta$ when $\Delta\theta$ is small. That tells us $\Delta\theta = \Delta s / r$ in (18.8.4) is not $\Delta\theta$ but $r\Delta\theta$. Therefore

$$D_{\theta}g = \lim_{\Delta\theta \to 0} \frac{g(r, \theta + \Delta\theta) - g(r, \theta)}{r\Delta\theta} = \frac{1}{r} \lim_{\Delta\theta \to 0} \frac{g(r, \theta + \Delta\theta) - g(r, \theta)}{\Delta\theta} = \frac{1}{r}\frac{\partial g}{\partial \theta}. \tag{18.8.5}$$

Thus $B = \frac{1}{r}\frac{\partial g}{\partial \theta}$, and we have

Gradient in Polar Coordinates

$$\mathbf{grad}\, g = \frac{\partial g}{\partial r}\widehat{\mathbf{r}} + \frac{1}{r}\frac{\partial g}{\partial \theta}\widehat{\boldsymbol{\theta}}. \tag{18.8.6}$$

The r in the denominator in the $\widehat{\boldsymbol{\theta}}$ term occurs because an angular change of $\Delta\theta$ causes a point to move approximately the distance $r\Delta\theta$, not $\Delta\theta$. SHERMAN: Could this also be explained (more efficiently) by referring to polar coordinates and magnification?

The Divergence in Polar Coordinates

The divergence of $\mathbf{F}(x, y) = P(x, y)\mathbf{i} + Q(x, y)\mathbf{j}$ is $\partial P / \partial x + \partial Q / \partial y$. But what is the divergence of a vector field described in polar coordinates, $\mathbf{G}(r, \theta) = A(r, \theta)\widehat{\mathbf{r}} + B(r, \theta)\widehat{\boldsymbol{\theta}}$? It is not the sum of $\partial A / \partial r$ and $\partial B / \partial \theta$.

To find $\nabla \cdot \mathbf{G}$, use the relation between $\nabla \cdot \mathbf{G}$ at $P = (r, \theta)$ and the flux across a small curve C that surrounds P.

$$\nabla \cdot \mathbf{G} = \lim_{\text{length of } C \to 0} \frac{\oint_C \mathbf{G} \cdot \mathbf{n}\, ds}{\text{Area within } C}. \tag{18.8.7}$$

Equation (18.8.7) provides a coordinate-free description of divergence in the plane.

We are free to choose the small closed curve C to make it easy to estimate the flux across it. A curve C corresponding to small changes Δr and $\Delta\theta$ is shown in Figure 18.8.2. We will use (18.8.7) to find the divergence at $P = (r, \theta)$. Point P is not inside C, it is on C. However, since $\mathbf{G}$ is continuous, $\mathbf{G}(P)$ is the limit of values of $\mathbf{G}$ at points inside, so we may use (18.8.7).

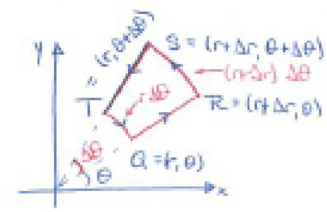


Figure 18.8.2 C is the curve $QRSTQ$

To estimate the flux across C, we estimate the flux across the four parts of the curve. Because they are short when Δr and $\Delta\theta$ are small, we may estimate the integral over each part by multiplying the value of the integrand at any point of the section, even at an end point, by the length of the section. As usual, $\widehat{\mathbf{n}}$ denotes an exterior unit vector perpendicular to C.

On QR and ST, $B\widehat{\boldsymbol{\theta}}$ contributes to the flux (on RS and TQ it does not since $\mathbf{n} \cdot \widehat{\boldsymbol{\theta}}$ is 0). On QR, $\widehat{\boldsymbol{\theta}}$ is parallel to $\mathbf{n}$, as shown in Figure 18.8.3.

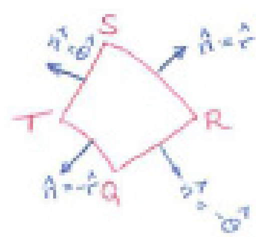


Figure 18.8.3

However, on ST it points in the opposite direction and $\widehat{\boldsymbol{\theta}} \cdot \widehat{\mathbf{n}}$ is -1. So, across ST, the flux contributed by $B\widehat{\boldsymbol{\theta}}$ is approximately

$$(B\widehat{\boldsymbol{\theta}} \cdot \widehat{\mathbf{n}})\Delta r = -B(r, \theta)\Delta r. \tag{18.8.8}$$

We would get a better estimate by using $B(r + \frac{\Delta r}{2}, \theta)$ but $B(r, \theta)$ is good enough since B is continuous.

On QR, $\widehat{\boldsymbol{\theta}}$ and $\widehat{\mathbf{n}}$ point in almost the same direction, so $\widehat{\boldsymbol{\theta}} \cdot \widehat{\mathbf{n}}$ is close to 1 when $\Delta\theta$ is small. On ST, $B\widehat{\boldsymbol{\theta}}$ contributes approximately $B(r, \theta + \Delta\theta)\Delta r$ to the flux.

The total contribution of $B\widehat{\boldsymbol{\theta}}$ to the flux across C is

$$B(r, \theta + \Delta\theta)\Delta r - B(r, \theta)\Delta r \tag{18.8.9}$$

The contribution of $A\widehat{\mathbf{r}}$ to the flux is negligible on QR and ST because $\widehat{\mathbf{r}}$ and $\widehat{\mathbf{n}}$ are perpendicular there. On TQ, $\widehat{\mathbf{r}}$ and $\widehat{\mathbf{n}}$ point in almost directly opposite directions, hence $\widehat{\mathbf{r}} \cdot \widehat{\mathbf{n}}$ is near -1. The flux of $A\widehat{\mathbf{r}}$ there is approximately

$$A(r, \theta)(\widehat{\mathbf{r}} \cdot \widehat{\mathbf{n}})r\Delta\theta = -A(r, \theta)r\Delta\theta. \tag{18.8.10}$$

On RS, which has radius $r + \Delta r$, $\widehat{\mathbf{r}}$ and $\widehat{\mathbf{n}}$ are almost identical, hence $\widehat{\mathbf{r}} \cdot \widehat{\mathbf{n}}$ is near 1. The contribution on RS, which has length $(r + \Delta r)\Delta\theta$, is approximately

$$A(r + \Delta r, \theta)\ (r + \Delta r)\ \Delta\theta. \tag{18.8.11}$$

Combining (18.8.9), (18.8.10) and (18.8.11), we see that the limit in (18.8.7) is the sum of two limits:

$$\lim_{\Delta r, \Delta\theta \to 0} \frac{A(r + \Delta r, \theta)(r + \Delta r)\Delta\theta - A(r, \theta)r\Delta\theta}{r\Delta r\Delta\theta} \tag{18.8.12}$$

The area within C is approximately, $r\Delta r\Delta\theta$.

and

$$\lim_{\Delta r,\Delta\theta\to 0} \frac{B(r,\theta+\Delta\theta)\Delta r - B(r,\theta)\Delta r}{r\Delta r\Delta\theta} \tag{18.8.13}$$

The first limit, (18.8.12), equals

$$\lim_{\Delta r,\Delta\theta\to 0} \frac{1}{r}\frac{(r+\Delta r)A(r+\Delta r,\Delta\theta) - rA(r,\theta)}{\Delta r},$$

which is

$$\frac{1}{r}\frac{\partial(rA)}{\partial r}.$$

In addition to the factor of $1/r$ the function being differentiated is rA.

The second limit, (18.8.13), equals

$$\lim_{\Delta r,\Delta\theta\to 0} \frac{1}{r}\frac{B(r,\theta+\Delta\theta) - B(r,\theta)}{\Delta\theta} = \frac{1}{r}\frac{\partial B}{\partial\theta}.$$

Here there is only the factor of $1/r$.

Thus we have the divergence formula

We reserve the use of $\nabla\cdot$ for rectangular coordinates, and use div in other coordinate systems.

Divergence in Polar Coordinates

$$\operatorname{div}(A\widehat{\mathbf{r}} + B\widehat{\boldsymbol{\theta}}) = \frac{1}{r}\frac{\partial(rA)}{\partial r} + \frac{1}{r}\frac{\partial B}{\partial\theta}. \tag{18.8.14}$$

The Curl in the Plane

The curl of $\mathbf{F}(x,y) = P(x,y)\mathbf{i} + Q(x,y)\mathbf{j} + 0\mathbf{k}$, a vector field in the plane, is given by

$$\mathbf{curl}\,\mathbf{F} = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathbf{k}.$$

What is the formula for the curl when the field is described in polar coordinates: $\mathbf{G}(r,\theta) = A(r,\theta)\widehat{\mathbf{r}} + B(r,\theta)\widehat{\mathbf{n}}$? To find out we will reason as we did with divergence. This time we use

$$(\mathbf{curl}\,\mathbf{G})\cdot\widehat{\mathbf{n}} = \lim_{\text{length of } C\to 0} \frac{\oint_C \mathbf{G}\cdot\mathbf{k}\,ds}{\text{Area bounded by } C}.$$

See (18.6.8) on page 1658.

where C is a closed curve around a point in the (r,θ) plane, and the limit is taken as the length of C approaches 0. The curl is evaluated at a point, which is on or within C.

We compute the circulation of $\mathbf{G} = A\widehat{\mathbf{r}} + B\widehat{\boldsymbol{\theta}}$ around the same curve used in the derivation of divergence in polar coordinates.

On TQ and RS, $A\widehat{\mathbf{r}}$, being perpendicular to the curve, contributes nothing to the circulation of $\mathbf{G}$ around C. On QR it contributes approximately

$$A(r,\theta)(\widehat{\mathbf{r}} \cdot \mathbf{T})\Delta r = A(r,\theta)\Delta r.$$

On ST, since there $\widehat{\mathbf{r}} \cdot \mathbf{T} = -1$, it contributes approximately

$$A(r,\theta+\Delta\theta)(\mathbf{r} \cdot \mathbf{T})\Delta r = -A(r,\theta+\Delta\theta)\Delta r.$$

A similar computation shows that $B\widehat{\boldsymbol{\theta}}$ contributes to the total circulation approximately

$$B(r+\Delta r,\theta)(r+\Delta r)\Delta\theta - B(r,\theta)r\Delta\theta. \tag{18.8.15}$$

Therefore $(\mathbf{curl}\,\mathbf{G}) \cdot \mathbf{k}$ is the sum of two limits:

$$\lim_{\Delta r,\Delta\theta\to 0} \frac{A(r,\theta)\Delta r - A(r,\theta+\Delta\theta)\Delta r}{r\Delta r\Delta\theta} = -\frac{1}{r}\frac{\partial A}{\partial \theta}$$

and

$$\lim_{\Delta r,\Delta\theta\to 0} \frac{B(r+\Delta r,\theta)(r+\Delta r)\Delta\theta - B(r,\theta)r\Delta\theta}{r\Delta r\Delta\theta} = \frac{1}{r}\frac{\partial (rB)}{\partial r}.$$

All told, we have

We reserve the use of $\nabla\times$ for rectangular coordinates, and use **curl** in other coordinate systems.

Curl in Polar Coordinates

$$\mathbf{curl}(A\widehat{\mathbf{r}} + B\widehat{\boldsymbol{\theta}}) = \left(-\frac{1}{r}\frac{\partial A}{\partial \theta} + \frac{1}{r}\frac{\partial (rB)}{\partial r}\right)\mathbf{k}. \tag{18.8.16}$$

EXAMPLE 1 Find the divergence and curl of $\mathbf{F} = r\theta^2\widehat{\mathbf{r}} + r^3\tan(\theta)\widehat{\boldsymbol{\theta}}$.

SOLUTION The calculations are applications of (18.8.14) and (18.8.16). First, the divergence:

$$\begin{aligned} \operatorname{div}\mathbf{F} &= \frac{1}{r}\frac{\partial}{\partial r}\left(r \cdot r\theta^2\right) + \frac{1}{r}\frac{\partial}{\partial \theta}\left(r^3\tan(\theta)\right) \\ &= \frac{1}{r}\left(2r\theta^2\right) + \frac{1}{r}\left(r^3\sec^2(\theta)\right) = 2\theta^2 + r^2\sec^2(\theta). \end{aligned}$$

Then the curl:

$$\begin{aligned} \mathbf{curl}\,\mathbf{F} &= \left(-\frac{1}{r}\frac{\partial}{\partial \theta}\left(r\theta^2\right) + \frac{1}{r}\frac{\partial}{\partial r}\left(r \cdot r^3\tan(\theta)\right)\right)\mathbf{k} \\ &= \left(-\frac{1}{r}(2r\theta) + \frac{1}{r}\left(4r^3\tan(\theta)\right)\right)\mathbf{k} = \left(-2\theta + 4r^2\tan(\theta)\right)\mathbf{k}. \end{aligned}$$

◇

Cylindrical Coordinates

In cylindrical coordinates the gradient of $g(r,\theta,z)$ is

Gradient in Cylindrical Coordinates

$$\textbf{grad}\, g = \frac{\partial g}{\partial r}\widehat{\mathbf{r}} + \frac{1}{r}\frac{\partial g}{\partial \theta}\widehat{\boldsymbol{\theta}} + \frac{\partial g}{\partial z}\widehat{\mathbf{z}} \tag{18.8.17}$$

Here $\widehat{\mathbf{z}}$ is the unit vector in the positive z direction, denoted $\mathbf{k}$ in Chapter 14. Formula (18.8.17) differs from (18.8) only by the extra term $(\partial g/\partial z)\widehat{\mathbf{z}}$. It can be obtained by computing directional derivatives of g along $\widehat{\mathbf{r}}$, $\widehat{\boldsymbol{\theta}}$, and $\widehat{\mathbf{z}}$. The derivation is similar to the one that gave us the formula for the gradient of $g(r,\theta)$.

The divergence of $\mathbf{G}(r,\theta,z) = A\widehat{\mathbf{r}} + B\widehat{\boldsymbol{\theta}} + C\widehat{\mathbf{z}}$ is given by

Divergence in Cylindrical Coordinates

$$\operatorname{div}\mathbf{G} = \frac{1}{r}\frac{\partial(rA)}{\partial r} + \frac{1}{r}\frac{\partial B}{\partial \theta} + \frac{\partial(C)}{\partial z}. \tag{18.8.18}$$

The partial derivatives with respect to r and z are similar in that the factor r is present in both $\partial(rA)/\partial r$ and $\partial(rC)/\partial r$. The proof of (18.8.18) uses the relation between $\nabla\cdot\mathbf{G}$ and the flux across the small surface determined by small changes Δr, $\Delta\theta$, and Δz.

The curl of $\mathbf{G} = A\widehat{\mathbf{r}} + B\widehat{\boldsymbol{\theta}} + C\widehat{\mathbf{z}}$ is given by a determinant:

Curl in Cylindrical Coordinates

$$\textbf{curl}\,\mathbf{G} = \frac{1}{r}\det\begin{pmatrix} \widehat{\mathbf{r}} & r\widehat{\boldsymbol{\theta}} & \mathbf{k} \\ \frac{\partial}{\partial r} & \frac{\partial}{\partial \theta} & \frac{\partial}{\partial z} \\ A & rB & C \end{pmatrix}$$

This formula is obtained from the circulation around three small closed curves lying in planes perpendicular to $\widehat{\mathbf{r}}$, $\widehat{\boldsymbol{\theta}}$, and $\mathbf{k}$.

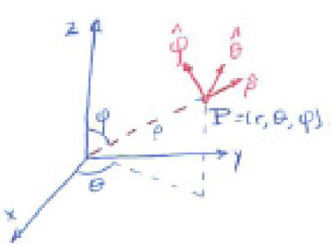


Figure 18.8.4

Spherical Coordinates

In mathematics texts, spherical coordinates are denoted ρ, ϕ, θ. In physics and engineering a different notation is standard. There ρ is replaced by r, θ

is the angle with z-axis, and ϕ plays the role of θ; the roles of ϕ and θ are interchanged. The formulas we state are in the mathematicians' notation.

The three basic unit vectors for spherical coordinates are denoted $\widehat{\boldsymbol{\rho}}$, $\widehat{\boldsymbol{\phi}}$, $\widehat{\boldsymbol{\theta}}$. For instance, $\widehat{\boldsymbol{\rho}}$ points in the direction of increasing ρ. See Figure 18.8.4. Note that, at the point P, $\widehat{\boldsymbol{\phi}}$ and $\widehat{\boldsymbol{\theta}}$ are tangent to the sphere through P and center at the origin, while $\widehat{\boldsymbol{\rho}}$ is perpendicular to that sphere. Also, any two of $\widehat{\boldsymbol{\rho}}$, $\widehat{\boldsymbol{\phi}}$, $\widehat{\boldsymbol{\theta}}$ are perpendicular.

To obtain the formulas for $\nabla \cdot \mathbf{G}$ and $\nabla \times \mathbf{G}$, we would use the region corresponding to small changes $\Delta\rho$, $\Delta\phi$, and $\Delta\theta$, shown in Figure 18.8.5. That computation yields the following formulas:

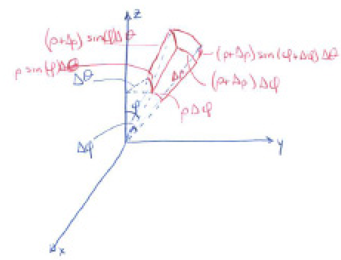


Figure 18.8.5

Gradient, Divergence, and Curl in Spherical Coordinates

If $g(\rho, \phi, \theta)$ is a scalar function,

$$\mathbf{grad}\, g = \frac{\partial g}{\partial \rho}\widehat{\boldsymbol{\rho}} + \frac{1}{\rho}\frac{\partial g}{\partial \phi}\widehat{\boldsymbol{\phi}} + \frac{1}{\sin(\phi)}\frac{\partial g}{\partial \theta}\widehat{\boldsymbol{\theta}}.$$

If $\mathbf{G}(\rho, \phi, \theta) = A\widehat{\boldsymbol{\rho}} + B\widehat{\boldsymbol{\phi}} + C\widehat{\boldsymbol{\theta}}$ then

$$\operatorname{div} \mathbf{G} = \frac{1}{\rho^2}\frac{\partial(\rho^2 A)}{\partial r}) + \frac{1}{\rho\sin(\phi)}\frac{\partial(\sin(\phi)B)}{\partial \phi} + \frac{1}{\sin(\phi)}\frac{\partial C}{\partial \theta}$$

and

$$\begin{aligned}\mathbf{curl}\,\mathbf{G} \;&=\; \frac{1}{\rho}\left(\frac{1}{\sin(\phi)}\frac{\partial(\sin(\phi))C)}{\partial \phi} - \frac{1}{\sin(\phi)}\frac{\partial B}{\partial \theta}\right)\widehat{\boldsymbol{\rho}} \\ &\quad + \frac{1}{\rho}\left(\frac{1}{\sin(\phi)}\frac{\partial A}{\partial \theta} - \frac{\partial(\rho C)}{\partial \rho}\right)\widehat{\boldsymbol{\phi}} + \frac{1}{\rho}\left(\frac{\partial(\rho B)}{\partial \rho} - \frac{\partial A}{\partial \phi}\right)\widehat{\boldsymbol{\theta}}.\end{aligned}$$

Each can be obtained by the method we used for polar coordinates. The change in ϕ or θ is not the same as the distance the point moves. However, a change in ρ is the same as the distance the point moves. The distance between (ρ, ϕ, θ) and $(\rho, \phi + \Delta\phi, \Delta\theta)$ is approximately $\rho\Delta\phi$ and the distance between (ρ, ϕ, θ) and $(\rho, \phi, \theta + \Delta\theta)$ is approximately $\rho\sin(\phi)\Delta\theta$.

An Application to Rotating Fluids

If a fluid is rotating in a cylinder, for instance, in a centrifuge, it rotates as a rigid body and its velocity at a distance r from the axis of rotation is

$$\mathbf{G}(r, \theta) = cr\widehat{\boldsymbol{\theta}},$$

where c is a positive constant.

Then

$$\mathbf{curl}\,\mathbf{G} = \frac{1}{r}\frac{\partial(cr^2)}{\partial r}\mathbf{k} = 2c\mathbf{k}.$$

The curl is independent of r. That means that an imaginary paddle wheel held with its axis in a fixed position would rotate at the same rate no matter where it is placed. The paddle wheel would complete one rotation around its axis when the fluid completes one rotation around its axis.

In the more general case with

$$\mathbf{G}(r,\theta) = cr^n\widehat{\boldsymbol{\theta}},$$

and n is an integer, we have

$$\mathbf{curl}\,\mathbf{G} = \frac{1}{r}\frac{\partial(cr^{n+1})}{\partial r}\mathbf{k} = c(n+1)r^{n-1}\mathbf{k}.$$

We just considered $n = 1$. For $n > 1$ the curl increases as r increases. The paddle wheel rotates faster if placed farther from the axis of rotation. The direction of rotation is the same as that of the fluid, counterclockwise.

When $n = -2$ the speed of the fluid decreases as r increases and

$$\mathbf{curl}\,\mathbf{G} = c(-2+1)r^{-2-1}\mathbf{k} = -cr^{-3}\mathbf{k}.$$

The minus sign in the curl means the paddle wheel spins clockwise even though the fluid rotates counterclockwise. The farther the paddle wheel is from the axis, the slower it rotates.

Summary

We expressed gradient, divergence, and curl in several coordinate systems. Though the basic unit vectors in each system may change direction from point to point, they remain perpendicular to each other. That simplified the computation of flux and circulation. The formulas are more complicated than those in rectangular coordinates because the amount a parameter changes is not the same as the distance the corresponding point moves.

EXERCISES for Section 18.8

In Exercises 1 through 4 find and draw the gradient of the function of (r, θ) at $(2, \pi/4)$.

1. r

2. $r^2\theta$

3. $e^{-r}\theta$

4. $r^3\theta^2$

In Exercises 5 through 8 find the divergence of the function.

5. $5\widehat{\mathbf{r}} + r^2\theta\widehat{\boldsymbol{\theta}}$

6. $r^3\theta\widehat{\mathbf{r}} + 3r\theta\widehat{\boldsymbol{\theta}}$

7. $r\widehat{\mathbf{r}} + r^3\widehat{\boldsymbol{\theta}}$

8. $r\sin(\theta)\widehat{\mathbf{r}} + r^2\cos(\theta)\widehat{\boldsymbol{\theta}}$

In Exercises 9 through 12 compute the curl of the function.

9. $r\widehat{\boldsymbol{\theta}}$

10. $r^3\theta\widehat{\mathbf{r}} + e^r\widehat{\boldsymbol{\theta}}$

11. $r\cos(\theta)\widehat{\mathbf{r}} + r\theta\widehat{\boldsymbol{\theta}}$

12. $1/r^3\widehat{\boldsymbol{\theta}}$

13. Find the directional derivative of $r^2\theta^3$ in the direction

(a) $\widehat{\mathbf{r}}$

(b) $\widehat{\boldsymbol{\theta}}$

(c) $\mathbf{i}$

(d) $\mathbf{j}$

14. What property of rectangular coordinates makes the formulas for gradient, divergence, and curl in those coordinates relatively simple?

15. Estimate the flux of $\mathbf{F}(r, \theta) = r^2\theta^3\widehat{\boldsymbol{\theta}}$ around the circle of radius 0.01 with center at $(r, \theta) = (2, \pi/6)$.

16. Estimate the circulation of the field in the preceding exercise around the same circle.

When translating between rectangular and polar coordinates, it may be necessary to express $\widehat{\mathbf{r}}$ and $\widehat{\boldsymbol{\theta}}$ in terms of $\mathbf{i}$ and $\mathbf{j}$ and also $\mathbf{i}$ and $\mathbf{j}$ in terms of $\widehat{\mathbf{r}}$ and $\widehat{\boldsymbol{\theta}}$. Exercises 17 and 18 concern this matter.

17. Let (r, θ) be a point that has rectangular coordinates (x, y).

(a) Show that $\widehat{\mathbf{r}} = \cos(\theta)\mathbf{i} + \sin(\theta)\mathbf{j}$, which equals $\frac{x\mathbf{i}+y\mathbf{j}}{\sqrt{x^2+y^2}}$

(b) Show that $\widehat{\boldsymbol{\theta}} = -\sin(\theta)\mathbf{i} + \cos(\theta)\mathbf{j}$, which equals $\frac{-y\mathbf{i}+x\mathbf{j}}{\sqrt{x^2+y^2}}$

(c) Draw a picture to accompany the calculations done in (a) and (b).

So we have $\widehat{\mathbf{r}}$ and $\boldsymbol{\theta}$ in terms of $\mathbf{i}$ and $\mathbf{j}$:

$$\begin{aligned} \widehat{\mathbf{r}} &= \frac{x\mathbf{i}+y\mathbf{j}}{\sqrt{x^2+y^2}} \\ \widehat{\boldsymbol{\theta}} &= \frac{-y\mathbf{i}+x\mathbf{j}}{\sqrt{x^2+y^2}} \end{aligned} \tag{18.8.19}$$

18. Show that if (x, y) has polar coordinates (r, θ), then

$$\begin{cases} \mathbf{i} = \cos(\theta)\theta\ \widehat{\mathbf{r}} - \sin(\theta)\ \widehat{\boldsymbol{\theta}} \\ \mathbf{j} = \sin(\theta)\theta\ \widehat{\mathbf{r}} + \cos(\theta)\ \widehat{\boldsymbol{\theta}} \end{cases}$$

by solving the equations (18.8.19) for $\mathbf{i}$ and $\mathbf{j}$.

In Exercises 19 through 22

(a) Find the gradient of the function, using the formula for gradient in rectangular coordinates.

(b) Find it by first expressing the function and its gradient in polar coordinates.

(c) Show that the gradients in (a) and (b) agree.

show that the two results agree.

19. $x^2 + y^2$

20. $\sqrt{x^2 + y^2}$

21. $3x + 2y$

22. $x/\sqrt{x^2 + y^2}$

In Exercises 23 through 26

(a) Find the gradient of the function, using its formula in polar coordinates, that is (18.8).

(b) Find it by expressing the function in rectangular coordinates.

(c) Show that the gradients in (a) and (b) agree.

23. r^2

24. $r^2 \cos(\theta)$

25. $r \sin(\theta)$

26. e^r

In Exercises 27 and 28

(a) Find the divergence of the vector field in rectangular coordinates.

(b) Find it by expressing the function in polar coordinates and using (18.8.14).

(c) Show that the divergences in (a) and (b) agree.

27. $x^2\mathbf{i} + y^2\mathbf{j}$

28. $xy\mathbf{i}$

In Exercises 29 and 30

(a) Find the curl of the vector field in rectangular coordinates,

(b) Find it by expressing the function in polar coordinates and using (18.8.16),

(c) Show that the curls found in (a) and (b) agree.

29. $xy\mathbf{i} + x^2y^2\mathbf{j}$

30. $\frac{x}{\sqrt{x^2+y^2}}\mathbf{i}$

Exercises 31 and 32 are useful in developing the formula for the gradient in cylindrical and spherical coordinates.

31. Find the distance from (r, θ, z) to each of the points

(a) $(r + \Delta r, \theta, z)$,

(b) $(r, \theta + \Delta\theta, z)$,

(c) $(r, \theta, z + \Delta z)$.

32. Approximate the distance from the point (ρ, ϕ, θ) to

(a) $(\rho + \Delta\rho, \phi, \theta)$?

(b) $(\rho, \phi + \Delta\phi, \theta)$?

(c) $(\rho, \phi, \theta + \Delta\theta)$?

33. Using the formulas for the gradient of $g(r, \phi, \theta)$, find the directional derivative of g in the direction

(a) $\widehat{\boldsymbol{\rho}}$,

(b) $\widehat{\boldsymbol{\phi}}$,

(c) $\widehat{\boldsymbol{\theta}}$.

34. Using the formulas for the gradient of $g(r, \theta, z)$, find the directional derivative of g in the direction

(a) $\widehat{\mathbf{r}}$,

(b) $\widehat{\boldsymbol{\theta}}$,

(c) $\mathbf{k}$.

35. Using as few mathematical symbols as you can, state the formula for the divergence of a vector field given relative to $\widehat{\mathbf{r}}$ and $\widehat{\boldsymbol{\theta}}$.

36. Using as few mathematical symbols as you can, state the formula for the curl of a vector field given relative to $\widehat{\mathbf{r}}$ and $\widehat{\boldsymbol{\theta}}$.

37. In the formula for the divergence of $A\widehat{\mathbf{r}} + B\widehat{\boldsymbol{\theta}}$, why do the terms rA and $1/r$ appear in $\frac{1}{r}\frac{\partial}{\partial r}(rA)$?

38. Obtain the formula for the gradient in cylindrical coordinates.

39. Obtain the formula for curl in cylindrical coordinates.

40. Obtain the formula for divergence in cylindrical coordinates.

41. Obtain the formula for the gradient in spherical coordinates.

42. Where did we use the fact that $\widehat{\mathbf{r}}$ and $\widehat{\boldsymbol{\theta}}$ are perpendicular when developing the expression for divergence in polar coordinates?

43. Obtain the formula for the gradient of $g(r,\theta)$ in polar coordinates by starting with the formula for the gradient of $f(x,y)$ in rectangular coordinates. During the calculations some as complicated expressions cancel and $\cos^2(\theta)+\sin^2(\theta)=1$ simplifies expressions. (See Exercise 18.8.19.)
Assume $g(r,\theta)=f(x,y)$, where $x=r\cos(\theta)$ and $y=r\sin(\theta)$. To express $\nabla f = \partial f/\partial x\mathbf{i} + \partial f/\partial y\mathbf{j}$ in terms of polar coordinates, it is necessary to express $\partial f/\partial x$, $\partial f/\partial y$, $\mathbf{i}$, and $\mathbf{j}$ in terms of partial derivative of $g(r,\theta)$, $\widehat{\mathbf{r}}$, and $\widehat{\boldsymbol{\theta}}$.

(a) Show that $\partial r/\partial x = \cos(\theta)$, $\partial r/\partial y = \sin(\theta)$, $\partial\theta/\partial x = -(\sin(\theta))/r$, $\partial\theta/\partial y = (\cos\theta)/r$.

(b) Use the chain rule to express $\partial f/\partial x$ and $\partial f/\partial y$ in terms of partial derivatives of $g(r,\theta)$.

(c) Recalling the expression of $\mathbf{i}$ and $\mathbf{j}$ in terms of $\widehat{\mathbf{r}}$ and $\widehat{\boldsymbol{\theta}}$ in Exercise 18 obtain the gradient of $g(r,\theta)$ in polar coordinates.

44. In Exercise 26 of Section 18.3 we found the divergence of $\mathbf{F} = r^n\widehat{\mathbf{r}}$ using rectangular coordinates. Find the divergence using polar coordinates. The second way is easier.

45. In Exercise 31 of Section 18.6 we used rectangular coordinates to show that an irrotational planar central field is conservative. Use the formula for curl in polar coordinates to obtain the same result. This way is easier.

46. In Exercise 21 in Section 18.4 we used rectangular coordinates to show that an incompressible central field in the plane must have the form $\mathbf{F}(\mathbf{r}) = (k/r)\widehat{\mathbf{r}}$. Obtain this result using the formula for divergence in polar coordinates.

47.

(a) Using a diagram, show that $\widehat{\mathbf{r}}$, $\widehat{\boldsymbol{\theta}}$, and $\widehat{\mathbf{z}}$ are mutually perpendicular.

(b) Using a diagram, show that $\widehat{\boldsymbol{\rho}}$, $\widehat{\boldsymbol{\phi}}$, and $\widehat{\boldsymbol{\theta}}$ are mutually perpendicular.

Coordinate systems whose basic unit vectors are mutually perpendicular are called **orthogonal**. Orthogonality is useful in developing the formulas for ∇, $\nabla\cdot$, and $\nabla\times$ in such systems.

18.9 Maxwell's Equations

At any point in space there is an electric field $\mathbf{E}$ and a magnetic field $\mathbf{B}$. The electric field is due to charges (electrons and protons), whether stationary or moving. The magnetic field is due to moving charges.

To assure yourself that the magnetic field $\mathbf{B}$ is everywhere, hold up a pocket compass. The magnetic field, produced within the Earth, makes the needle point north in the northern hemisphere.

All electrical phenomena and their applications can be explained by four equations called **Maxwell's equations**. They allow $\mathbf{B}$ and $\mathbf{E}$ to vary in time. We state them for the simpler case when $\mathbf{B}$ and $\mathbf{E}$ are constant: $\partial \mathbf{B}/\partial t = \mathbf{0}$ and $\partial \mathbf{E}/\partial t = \mathbf{0}$. We met the first equation in the previous section. Here are the four equations

I. $\int_S \mathbf{E} \cdot \mathbf{n}\, dS = Q/\epsilon_0$, where S is a surface bounding a spatial region and Q is the charge in that region. (Gauss's Law for Electricity)

II. $\oint_C \mathbf{E} \cdot d\mathbf{r} = 0$ for a closed curve C. (Faraday's Law of Induction)

III. $\int_S \mathbf{B} \cdot \mathbf{n}\, dS = 0$ for a surface S that bounds a spatial region. (Gauss's Law for Magnetism)

IV. $\oint_C \mathbf{B} \cdot d\mathbf{r} = \mu_0 \int_S \mathbf{J} \cdot \mathbf{n}\, dS$, where C bounds the surface $\mathcal{S}$ and $\mathbf{J}$ is the density of the electric current flowing through $\mathcal{S}$. (Ampere's Law)

The constants ϵ_0 and μ_0 ("myoo naught") depend on the units used. They will be important in the CIE on Maxwell's Equations at the end of this chapter.

The statements about integrals can be translated into information about the behavior of $\mathbf{E}$ or $\mathbf{B}$ at each point.

In derivative or local form they are

I′. $\operatorname{div} \mathbf{E} = \rho/\epsilon_0$, where ρ is the charge density (Coulomb's Law)

II′. $\mathbf{curl}\, \mathbf{E} = \mathbf{0}$

III′. $\operatorname{div} \mathbf{B} = 0$

IV′. $\mathbf{curl}\, \mathbf{B} = \mu_0 \mathbf{J}$

It turns out that $1/(\mu_0\epsilon_0)$ equals the square of the speed of light. Why that is justified is an astonishing story told in CIE 26.

Going Back and Forth Between Local and Global.

Examples 1 and 2 show that Gauss's law is equivalent to Coulomb's law.

EXAMPLE 1 Obtain Gauss's law for electricity (I) from Coulomb's law (I′).
SOLUTION Let $\mathcal{V}$ be the solid region whose boundary is $\mathcal{S}$. Then

$$\begin{aligned} \int_{\mathcal{S}} \mathbf{E}\cdot\mathbf{n}\,dS &= \int_{\mathcal{V}} \nabla\cdot\mathbf{E}\,dV && \text{(Divergence theorem)}\\ &= \int_{\mathcal{V}} \tfrac{\rho}{\epsilon_0}\,dV && \text{(Coulomb's law)}\\ &= \tfrac{1}{\epsilon_0}\int_{\mathcal{V}} \rho\,dV = \tfrac{Q}{\epsilon_0}. \end{aligned}$$

The total charge in $\mathcal{V}$ is $Q = \int_{\mathcal{V}} \rho\,dV$. ◇

Does Gauss's law imply Coulomb's law? Example 2 shows that the answer is yes.

EXAMPLE 2 Deduce Coulomb's law (I′) from Gauss's law for electricity (I).
SOLUTION Let $\mathcal{V}$ be any spatial region and let $\mathcal{S}$ be its surface. Let Q be the total charge in $\mathcal{V}$. Then

$$\begin{aligned} \frac{Q}{\epsilon_0} &= \int\limits_{\mathcal{S}} \mathbf{E}\cdot\mathbf{n}\,dS && \text{(Gauss's law)}\\ &= \int\limits_{\mathcal{V}} \nabla\cdot\mathbf{E}\,dV && \text{(Divergence theorem).} \end{aligned}$$

On the other hand,

$$Q = \int\limits_{\mathcal{V}} q\,dV,$$

where q is the charge density. Thus

$$\int\limits_{\mathcal{V}} \frac{q}{\epsilon_0}\,dV = \int\limits_{\mathcal{V}} \nabla\cdot\mathbf{E}\,dV \qquad \text{or} \qquad \int\limits_{\mathcal{V}} \left(\frac{q}{\epsilon_0} - \nabla\cdot\mathbf{E}\right)\,dV = 0,$$

for all spatial regions. Since the integrand is assumed to be continuous, the zero-integral principle tells us that it is identically 0. That is,

$$\frac{q}{\epsilon_0} - \nabla\cdot\mathbf{E} = 0,$$

which gives us Coulomb's law. ◇

EXAMPLE 3 Show that II implies II′. That is, $\oint_C \mathbf{E}\cdot\,d\mathbf{r} = 0$ for closed curves implies $\mathbf{curl}\,\mathbf{E} = \mathbf{0}$.

SOLUTION By Stokes' theorem, for any orientable surface $\mathcal{S}$ bound by a closed curve,

$$\int_{\mathcal{S}} (\mathbf{curl}\,\mathbf{E}) \cdot \mathbf{n}\; dS = 0$$

The zero-integral principle implies that $(\mathbf{curl}\,\mathbf{E}) \cdot \mathbf{n} = 0$ at each point on the surface. Choosing $\mathcal{S}$ such that $\mathbf{n}$ is parallel to $\mathbf{curl}\,\mathbf{E}$ (if $\mathbf{curl}\,\mathbf{E}$ is not $\mathbf{0}$), implies that the magnitude of $\mathbf{curl}\,\mathbf{E}$ is 0, hence $\mathbf{curl}\,\mathbf{E}$ is $\mathbf{0}$. $\diamond$

Maxwell, by studying I′, II′, III′, IV′, deduced that electromagnetic waves travel at the speed of light, and therefore light is an electromagnetic phenomenon.

The exercises present the analogy of the equations in integral form for the general case where $\mathbf{B}$ and $\mathbf{E}$ vary with time. In this generality they are known as **Maxwell's equations**, in honor of James Clerk Maxwell (1831-1879), who put them in their final form in 1865.

Mathematics and Electricity

Benjamin Franklin, in his book *Experiments and Observations Made in Philadelphia*, published in 1751, made electricity into a science. For his accomplishments, he was elected a Foreign Associate of the French Academy of Sciences, an honor bestowed on no other American for over a century. In 1865, Maxwell completed the theory that Franklin had begun.

At the time that Newton published his *Principia* on the gravitational field (1687), electricity and magnetism were the subjects of little scientific study. The experiments of Franklin, Oersted, Henry, Ampère, Faraday, and others in the late eighteenth and early nineteenth centuries gradually built up a mass of information subject to mathematical analysis. All the phenomena could be summarized in four equations, which in their final form appeared in Maxwell's *Treatise on Electricity and Magnetism*, published in 1873.

Summary

We stated the four equations that describe electrostatic and magnetic fields that do not vary with time. Then we showed how to use the divergence theorem or Stokes' theorem to translate between their global and local forms. The exercises include the four equations in their general form, where $\mathbf{E}$ and $\mathbf{B}$ vary with time.

EXERCISES for Section 18.9

1. Obtain II from II$'$.

2. Obtain III$'$ from III.

3. Obtain III from III$'$.

4. Obtain IV$'$ from IV.

5. Obtain IV from IV'.

In Exercises 6 to 9 use terms such as circulation flux, current, and charge densityto express the equation in words.

6. I

7. II

8. III

9. IV

10. Which of the laws tell us that an electric current produces a magnetic field?

11. Which of the laws tells us that a magnetic field produces an electric current?

In this section we assumed that the fields **E** and **B** do not vary in time, that is, $\partial \mathbf{E}/\partial t = \mathbf{0}$ and $\partial \mathbf{B}/\partial t = \mathbf{0}$. The general case, in empty space, where **E** and **B** depend on time, is also described by four equations, which we call 1, 2, 3, 4. Numbers 1 and 3, do not involve time; they are similar to I$'$ and III$'$.

i. $\nabla \cdot \mathbf{E} = q/\epsilon_0$

ii. $\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$

iii. $\nabla \cdot \mathbf{B} = 0$

iv. $\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}$

(**J** is the current density.)

12. Which equation implies that a changing magnetic field creates an electric field?

13. Which equation implies that a changing electrostatic field creates a magnetic field?

14. Show that ii. is equivalent to

$$\oint_C \mathbf{E} \cdot \ dt = -\frac{\partial}{\partial t} \int_{\mathcal{S}} \mathbf{B} \cdot \mathbf{n} \, dS$$

Here, C bounds S. (You may assume that $\frac{\partial}{\partial t} \int_{\mathcal{S}} \mathbf{B} \cdot \mathbf{n} \, dS$ equals $\int_{\mathcal{S}} (\partial B/\partial t) \cdot \mathbf{n} \ dS$.)

15. Show that iv. is equivalent to

$$\oint_C \mathbf{B} \cdot \ d\mathbf{r} = \mu_0 \int_{\mathcal{S}} \mathbf{J} \cdot \mathbf{n} \, dS + \mu_0 \epsilon_0 \frac{\partial}{\partial t} \int_{\mathcal{S}} \mathbf{E} \cdot \mathbf{n} \, dS$$

(The circulation of $\mathbf{B}$ is related to the total current through the surface $\mathcal{S}$ that C bounds and to the rate at which the flux of $\mathbf{E}$ through $\mathcal{S}$ changes.)

18.S Chapter Summary

The first six sections developed three theorems: Green's theorem, Gauss' law (also called the divergence theorem), and Stokes' theorem. The final four sections applied them to geometry, to physics, and to non-rectangular coordinate systems.

Name	Mathematical Expression	Physical Description
Green's theorem	$\oint_C \mathbf{F} \cdot \mathbf{n}\, ds = \int_{\mathcal{R}} \nabla \cdot \mathbf{F}\, dA$ $\oint_C (-Qdx + Pdy) = \int_{\mathcal{R}} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right)\, dA$ $\oint_C \mathbf{F} \cdot \mathbf{T}\, ds = \oint_C \mathbf{F} \cdot d\mathbf{r} = \int_{\mathcal{R}} (\nabla \times \mathbf{F}) \cdot \mathbf{k}\, dA$ $\oint_C (Pdx + Qdy) = \int_{\mathcal{R}} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\, dA$	flux of $\mathbf{F}$ across C differential form circulation of $\mathbf{F}$ around C
Gauss' law (divergence theorem)	$\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{N}\, dS = \int_R \nabla \cdot \mathbf{F}\, dV$	
Stokes' theorem	$\oint_C \mathbf{F} \cdot \mathbf{T}\, ds = \int_{\mathcal{S}} (\nabla \times \mathbf{F}) \cdot \mathbf{n}\, dS$ ($\mathcal{S}$ is a surface bounded by C with $\mathbf{n}$ compatible with orientation of C)	

Green's theorem can be viewed as the planar version of either the divergence theorem or Stokes' theorem.

Though div $\mathbf{F}$ and **curl** $\mathbf{F}$ were defined in rectangular coordinates, they have a meaning that is independent of coordinates. For instance, if $\mathbf{F}$ is a vector field in space, the divergence of $\mathbf{F}$ at a point multiplied by the volume of a small region containing the point approximates the flux of $\mathbf{F}$ across the surface of the small region. More precisely,

$$\operatorname{div} \mathbf{F} \text{ at } P \text{ equals the limit of } \frac{\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n}\, ds}{\text{volume of } \mathcal{R}} \text{ as the diameter of } \mathcal{R} \text{ approaches } 0$$

The **curl** of $\mathbf{F}$ at P is a vector, so it is a bit harder to describe physically. Let $\mathbf{n}$ be a unit vector and C a small curve that lies in a plane through P, is perpendicular to $\mathbf{n}$, and surrounds P. Then the product of the scalar component of **curl** $\mathbf{F}$ at P in the direction $\mathbf{n}$ and the area of the region bounded by C gives the circulation of $\mathbf{F}$ along C:

$$((\mathbf{curl}\, \mathbf{F} \text{ at } P) \cdot \mathbf{n}) \;\; (\text{area enclosed by } C) \;\; \text{equals circulation of } \mathbf{F} \text{ along } C\,.$$

A field whose curl is $\mathbf{0}$ is called irrotational. A field whose divergence is 0 is called incompressible (or divergence-free).

Of particular interest are conservative fields. A field $\mathbf{F}$ is conservative if its circulation on a curve depends only on its endpoints. If the domain of $\mathbf{F}$ is simply connected, $\mathbf{F}$ is conservative if and only if its curl is $\mathbf{0}$. A conservative field is expressible as the gradient of a scalar function.

Among the conservative fields are the central fields. If, in addition, they are divergence-free, they take a special form that depends on the dimension.

In the table $\mathbf{R}$ stands for the set of real numbers, $\mathbf{R}^2$ for the set of pairs (x, y), $\mathbf{R}^3$ for the set of triplets (x, y, z), and $\mathbf{R}^n$ for the set of n-tuplets $(x_1, x_2, \dots, x_n)$.

Geometry	General Form of Divergence-Free Symmetric Central Fields	Description
$\mathbf{R}^2$ (plane)	$c\frac{\widehat{\mathbf{r}}}{r}$	inverse power
$\mathbf{R}^3$ (space)	$c\frac{\widehat{\mathbf{r}}}{r^2}$	inverse square power
$\mathbf{R}^n$	$c\frac{\widehat{\mathbf{r}}}{r^{n-1}}$	inverse $n-1$ power

When $\mathbf{curl}\,\mathbf{F} = \mathbf{0}$ we can replace an integral $\int_A^B \mathbf{F} \cdot d\mathbf{r}$ by an integral over another curve joining A and B, which can be helpful if the new line integral is easier to evaluate than the original one. Similarly, in a region where $\nabla \cdot \mathbf{F} = 0$ we can replace an integral $\int_{\mathcal{S}} \mathbf{F} \cdot \mathbf{n}\, dS$ over the surface $\mathcal{S}$ with a more convenient integral over a different surface.

In applications in space the most important field is the inverse square central field, $\mathbf{F} = \frac{\widehat{\mathbf{r}}}{r^2}$. Its flux over a closed surface that does not enclose the origin is 0, but its flux over a surface that encloses the origin is 4π. If one thinks in terms of steradians, it is clear why the second integral is 4π: the flux of $\widehat{\mathbf{r}}/r^2$ also measures the solid angle subtended by a surface. Also, the first case becomes clear in terms of solid angle when the parts of the surface where $\mathbf{n} \cdot \mathbf{r}$ is positive and where it is negative are treated separately.

EXERCISES for 18.S

SHERMAN/DOUG: Check the wording for proper use of all terms.

1. Match the vector fields (a.-e.) with their descriptions (1.-5.)

a.	$\mathbf{F}(\mathbf{r})$	1.	an inverse cube field
b.	$f(\mathbf{r})\widehat{\mathbf{r}}$	2.	a central field (center at origin)
c.	$f(r)\widehat{\mathbf{r}}$	3.	an arbitrary vector field
d.	$\widehat{\mathbf{r}}/r^2$	4.	an inverse square field
e.	$\mathbf{r}/r^4$	5.	a radially symmetric field (center at origin)

Is this note true? Answer: a-3, b-2, c-4, d-5, e-1

There is not a one-to-one relation between the columns.

2. Use Green's theorem to evaluate $\oint_C (xy\ dx + e^x\ dy)$, where C is the curve that goes from $(0, 0)$ to $(2, 0)$ on the x-axis and returns from $(2, 0)$ to $(0, 0)$ on the parabola $y = 2x - x^2$.

3. A curve C in the xy-plane bounds a region $\mathcal{R}$ of area A.

(a) If $\oint_C \mathbf{F} \cdot d\mathbf{r} = -2$, estimate $\nabla \times \mathbf{F}$ at points in R.

(b) Would you use $\odot$ or $\oplus$ to indicate the curl?

4. A curve C in the xy-plane bounds a region $\mathcal{R}$ of area A.

(a) If $\oint_C \mathbf{F} \cdot \mathbf{n}\, ds = -2$, estimate $\nabla \cdot \mathbf{F}$ at points in $\mathcal{R}$.

(b) How did you decide whether $\nabla \cdot \mathbf{F}$ is positive or negative?

5. A field $\mathbf{F}$ is called **uniform** if all its vectors are the same. Let $\mathbf{F}(x, y, z) = 3\mathbf{i}$.

(a) Find the flux of $\mathbf{F}$ across each of the faces of the cube in Figure 18.S.1 of side 3.

(b) Find the total flux of $\mathbf{F}$ across the surface of the cube.

(c) Verify the divergence theorem for $\mathbf{F}$.

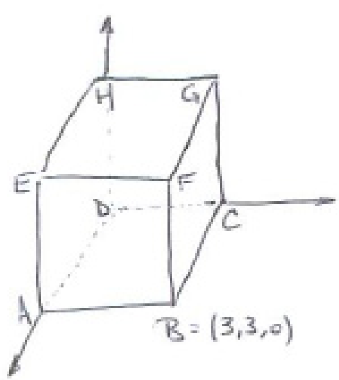


Figure 18.S.1

6. Let $\mathbf{F}$ be the uniform field $\mathbf{F}(x, y, z) = 2\mathbf{i} + 3\mathbf{j} + 0\mathbf{k}$. Carry out the three parts of Exercise 5 for this field.

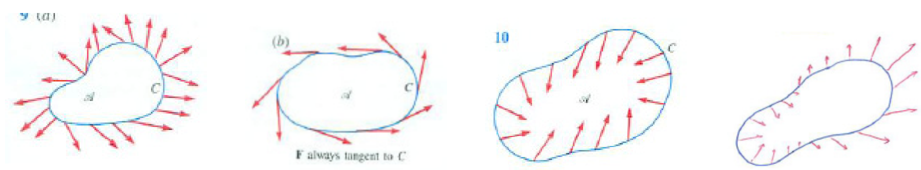
7. See Exercise 7 in Section 18.7. Suppose you placed the point at which $\mathbf{E}$ is evaluated at $(a, 0, 0)$ instead of at $(0, 0, a)$.

(a) What integral in spherical coordinates arises?

(b) Would you like to evaluate it?

SHERMAN: You have a note that (a) and (d) are similar. In (d) half of the boundary has an outflow and half has an inflow; in (a) the entire boundary is an outflow. Aren't these different enough?

In Exercises 8 to 11, **F** is defined on the whole plane but indicated only at points on a curve C bounding a region $\mathcal{R}$. What can be said about $\int_{\mathcal{R}} \nabla \cdot \mathbf{F}\, dA$?



(a) (b) (c) (d)

Figure 18.S.2

8. See Figure 18.S.2(a).

9. See Figure 18.S.2(b).

10. See Figure 18.S.2(c).

11. See Figure 18.S.2(d).

Exercises 12 to 15, involve the same **F** as for Exercises 8 to 11. What can be said about $\int_{\mathcal{S}} \nabla \times \mathbf{F} \cdot \mathbf{k}\, dA$ in each case?

12. See Figure 18.S.2(a).

13. See Figure 18.S.2(b).

14. See Figure 18.S.2(c).

15. See Figure 18.S.2(d).

SHERMAN: Your note here says "they all have $+k$. I believe you are suggesting that one should have all normals and one should have all backward pointing tangents. Are you suggesting new figures in Figure 18.S.2? The current figures (a) - (c) are from V, I believe; only (d) is hand-drawn.

16. Let C be the circle of radius 1 with center $(0, 0)$.

(a) What does Green's theorem say about the line integral

$$\oint_C \left((x^2 - y^3)\, dx + (y^2 + x^3)\, dy\right)?$$

(b) Use Green's theorem to evaluate the integral.

(c) Evaluate it directly.

17. In Example 5, Section 18.1, we computed $\oint_C \mathbf{F} \cdot d\mathbf{r}$ where $\mathbf{F} = \frac{-y\mathbf{i} + x\mathbf{j}}{x^2+y^2}$ and C is the unit circle with center at the origin.

(a) *Without doing any new computations*, evaluate $\oint_C \mathbf{F} \cdot d\mathbf{r}$ where C is the square path with corners $(1, 0)$, $(2, 0)$, $(2, 1)$, and $(1, 1)$.

(b) Evaluate the integral in (a) by a direct computation, breaking the integral into four integrals, one over each edge.

18. Let $\mathbf{F}(x, y) = (x + y)\mathbf{i} + x^2\mathbf{j}$ and let C be the counterclockwise path around the triangle whose vertices are $(0, 0)$, $(1, 1)$, and $(-1, 1)$.

(a) Use the planar divergence theorem to evaluate $\int_C \mathbf{F} \cdot \mathbf{n}\ ds$, where $\mathbf{n}$ is the outward unit normal.

(b) Evaluate the line integral in (a) directly.

19. Let b and c be positive and $\mathcal{S}$ the infinite rectangle parallel to the xy-plane, consisting of the points (x, y, c) such that $0 \leq x \leq b$ and $y \geq 0$.

(a) If b were replaced by ∞, what is the solid angle that $\mathcal{S}$ subtends at the origin? (No integration is needed.)

(b) Find the solid angle subtended by $\mathcal{S}$ when b is finite. (See Exercise 91 in Section 8.6.)

(c) Is the limit of your answer in (b) as $b \to \infty$ the same as your answer in (a)?

20. Look back at the fundamental theorem of calculus (Section 6.4), Green's theorem (Section 18.2), the divergence theorem (Section 18.6), and Stokes' theorem (Section 18.4). What single theme runs through all of them?

Calculus is Everywhere # 25
Heating and Cooling

Engineers who design a car radiator or a home air conditioner are interested in the distribution of temperature of a fin. We present one of the mathematical tools they use in an example that shows how Green's theorem is applied.

A plane region $\mathcal{A}$ with boundary curve C is occupied by a sheet of metal. By heating and cooling devices, the temperature along the border is held constant, independent of time. Assume that the temperature in $\mathcal{A}$ eventually stabilizes. The steady-state temperature at point P in $\mathcal{A}$ is denoted $T(P)$. What does that imply about the function $T(x, y)$?

Heat tends to flow from high to low temperatures, that is, in the direction of $-\nabla T$. According to Fourier's law, flow is proportional to the conductivity of the material k (a positive constant) and the magnitude of the gradient $|\nabla T|$. Thus

$$\oint_C (-k\nabla T) \cdot \mathbf{n}\; ds$$

measures the rate of heat loss across C.

Since the temperature in the metal is at a steady state, the heat in the region bounded by C remains constant. Thus

$$\oint_C (-k\nabla T) \cdot \mathbf{n}\; ds = 0.$$

Green's theorem then tells us that

$$\int_{\mathcal{A}} \nabla \cdot (-k\nabla T)\; dA = 0$$

for any region $\mathcal{A}$ in the metal plate. Since $\nabla \cdot \nabla T$ is the Laplacian of T and k is not 0, we conclude that

$$\int_{\mathcal{A}} \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right) dA = 0. \tag{C.25.1}$$

By the zero-integrals theorem, the integrand must be 0 throughout $\mathcal{A}$,

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0.$$

This reduces the study of the temperature distribution to solving a partial differential equation.

Calculus is Everywhere # 26

How Maxwell Did It

In a letter to his cousin Charles Cay, dated January 5, 1965, Maxwell wrote:

> I have also a paper afloat containing an electromagnetic theory of light, which, till I am convinced to the contrary, I hold to be great guns. [Everitt, F., James Clerk Maxwell: a force for physics, *Physics World*, December 2006, `http://physicsworld.com/cws/article/print/26527`]

It indeed was great guns, for out of his theory has come countless inventions, such as television, cell phones, and remote garage door openers. In a dazzling feat of imagination, Maxwell predicted that electrical phenomena create waves, that light is one such phenomenon, and that the waves travel at the speed of light in a vacuum.

In this section we will see how those predictions came out of the equations (I′), (II′), (III′), and (IV′) in Section 18.9.

First, we look at the dimensions of the constants ε_0 and μ_0 that appear in (IV′),

$$\frac{1}{\mu_0\varepsilon_0}\nabla \times \mathbf{B} = \frac{\mathcal{J}}{\varepsilon_0}.$$

The constant ε_0 appears in

$$\text{Force} = F = \frac{1}{4\pi\varepsilon_0}\frac{qq_0}{r^2}. \tag{C.26.1}$$

Since the force F is mass times acceleration its dimensions are

$$\text{mass}\cdot\frac{\text{length}}{\text{time}^2},$$

or, in symbols

$$m\frac{L}{T^2}.$$

The number 4π is a pure number, without any physical dimension.

The quantity qq_0 has the dimension of charge squared, q^2, and R^2 has dimension L^2, where L denotes length.

Solving (C.26.1) for ε_0, we find the dimension of ε_0. Since

$$\varepsilon_0 = \frac{q^2}{4\pi F r^2},$$

its dimensions are

$$\left(\frac{T^2}{mL}\right)\left(\frac{q^2}{L^2}\right) = \frac{T^2q^2}{mL^3}.$$

To find the dimension of μ_0, we will use its appearance in calculating the force between two wires of length L each carrying a current I in the same direction and separated by a distance R. Each generates a magnetic field that draws the other towards it. The equation for the force is

$$\mu_0 = \frac{2\pi RF}{I^2L}.$$

Since R has the dimension of length L and F has dimensions mL/T^2, the numerator has dimension mL^2/T^2. The current I is "charge q per second," so I^2 has dimension q^2/T^2. The dimension of the denominator is, therefore,

$$\frac{q^2L}{T^2}.$$

Hence μ_0 has the dimension

$$\frac{mL^2}{T^2}\cdot\frac{T^2}{q^2L} = \frac{mL}{q^2}.$$

The dimension of the product $\mu_0\varepsilon_0$ is therefore

$$\frac{mL}{q^2}\cdot\frac{T^2q^2}{mL^3} = \frac{T^2}{L^2}.$$

The dimension of $1/\mu_0\varepsilon_0$, the same as the square of speed. In short, $1/\sqrt{\mu_0\varepsilon_0}$ has the dimension of speed, length divided by time.

Now we are ready to do the calculations leading to the prediction of waves traveling at the speed of light. We will use the equations (I′), (II′), (III′), and (IV′), as stated on page 1696, where the fields $\mathbf{B}$ and $\mathbf{E}$ vary with time. However, we are looking for a solution in free-space where there is no matter, no charge, and no currents, so $q = 0$ and $f = 0$.

Differentiating (IV′)

$$\nabla \times \mathbf{B} = \mu_0\varepsilon_0\frac{\partial \mathbf{E}}{\partial t}.$$

Differentiating this equation with respect to time t we obtain

$$\frac{\partial}{\partial t}(\nabla \times \mathbf{B}) = \mu_0\varepsilon_0\frac{\partial^2 \mathbf{E}}{\partial t^2}. \tag{C.26.2}$$

The operator $\frac{\partial}{\partial t}$ can be moved past the $\nabla\times$ to operate directly on $\mathbf{B}$. Thus (C.26.2) becomes

$$\nabla \times \frac{\partial \mathbf{B}}{\partial t} = \mu_0\varepsilon_0\frac{\partial^2 \mathbf{E}}{\partial t^2}. \tag{C.26.3}$$

Taking the curl of both sides of (II′)

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}.$$

we get

$$\nabla\left(-\nabla \times \mathbf{E}\right) = \nabla \times \frac{\partial \mathbf{B}}{\partial t}. \tag{C.26.4}$$

Combining (C.26.3) and (C.26.4) gives us an equation that involves $\mathbf{E}$ alone:

$$\nabla \times \left(-\nabla \times \mathbf{E}\right) = \mu_0 \varepsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2}. \tag{C.26.5}$$

The vector identity

$$\nabla \times \left(\nabla \times \mathbf{E}\right) = \nabla\left(\nabla \cdot \mathbf{E}\right) - \left(\nabla \cdot \nabla\right)\mathbf{E}. \tag{C.26.6}$$

combined with $\nabla \cdot \mathbf{E} = 0$ is one of the four assumptions, namely (I), on electromagnetic fields, (C.26.5), and (C.26.6) yields we arrive at

$$\left(\nabla \cdot \nabla\right)\mathbf{E} = \mu_0 \varepsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \qquad \text{or} \qquad \frac{\partial^2 \mathbf{E}}{\partial t^2} - \frac{1}{\mu_0 \varepsilon_0}\nabla^2 \mathbf{E} = \mathbf{0}. \tag{C.26.7}$$

The expression ∇^2 in (C.26.7) is short for

$$\begin{aligned} \nabla \cdot \nabla &= \left(\frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}\right) \cdot \left(\frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}\right) \\ &= \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \end{aligned} \tag{C.26.8}$$

In $(\nabla \cdot \nabla)\mathbf{E}$ we apply (C.26.8) to the components of $\mathbf{E}$. Thus $\nabla^2\mathbf{E}$ is a vector. So is $\partial^2\mathbf{E}/\partial t^2$ and (C.26.8) makes sense.

For simplicity, suppose E has only a y-component, which depends on x and t. THus $\mathbf{E}(x, y, z, t) = E(x, t)\mathbf{j}$, where E is a scalar function. We have

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) E(x, y) = \frac{\partial^2 E}{\partial x^2} + -0 + 0 = \frac{\partial^2 E}{\partial x^2}.$$

Then (C.26.8) becomes

$$\frac{\partial^2}{\partial t^2} E(x, t)\mathbf{j} - \frac{1}{\mu_0 \varepsilon_0}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) E(x, t)\mathbf{j} = \mathbf{0},$$

from which it follows that

$$\frac{\partial^2}{\partial t^2} E(x, t) - \frac{1}{\mu_0 \varepsilon_0}\frac{\partial^2 E}{\partial x^2} = 0. \tag{C.26.9}$$

Multiply (C.26.9) by $-\mu_0\varepsilon_0$ to obtain

$$\frac{\partial^2 E}{\partial x^2} - \mu_0\varepsilon_0 \frac{\partial^2 E}{\partial t^2} = 0.$$

This is an instance of the wave equation ((16.3.11) in Section 16.3). See the Wave in a Rope CIE at the end of Section 16.3 amd in Chapter 16. The solutions are waves traveling with speed $1/\sqrt{\mu_0\varepsilon_0}$.

Maxwell then compares $\sqrt{\mu_0\varepsilon_0}$ with the velocity of light:

> "In the following table, the principal results of direct observation of the velocity of light, are compared with the principal results of the comparison of electrical units ($1/\sqrt{\mu_0 v_0}$).

Velocity of light (meters per second)		Ratio of electrical units	
Fizeau	314,000,000	Weber	310,740,000
Sun's Parallax	308,000,000	Maxwell	288,000,000
Foucault	298,360,000	Thomson	282,000,000

Table C.26.1

> It is significant that the velocity of light and the ratio of the units are quantities of the same order of magnitude. Neither of them can be said to be determined as yet with such a degree of accuracy as to enable us to assert that the one is greater or less than the other. It is to be hoped that, by further experiment, the relation between the magnitude of the two quantities may be more accurately determined.
>
> In the meantime our theory, which asserts that these two quantities are equal, and assigns a physical reason for this equality, is certainly not contradicted by the comparison of these results such as they are. [James Clerk Maxwell, Treatise on Electricity and Magnetism, Vol. 2, third edition, Oxford University Press, (1904), first edition 1873, p. 436]

On this basis Maxwell concluded that light is an electromagnetic disturbance, uniting the fields of optics and electromagnetism. and predicted the existence of other electromagnetic waves. Earlier, in 1848, Gustav Kirchanoff hasn't noticed that the ratio of the electrical uints was clear to the velocity of light, but varies it as a coincedence. In 1887, eight years after Maxwell's death, Heinrich Hertz produced the predicted waves, whose frequency placed them outside what the eye can see.

By 1890 experiments had confirmed Maxwell's conjecture. They gave the velocity of light as 299,766,000 meters per second and $\sqrt{1/\mu_0\varepsilon_0}$ as 299,550,000 meters per second.

Newton, in his *Principia* of 1687, related gravity on Earth with gravity in the heavens. Benjamin Franklin with his kite experiments showed that lightning was simply an electric phenomenon. From then through the early nineteenth century, experimenters showed that electricity and magnetism were inseparable.