CHAPTER 1

INTRODUCTION

Oh, we know you do not like Mathematics,

but we promise not to use it as an instrument of torture....

– Lillian R. Lieber

The Education of T. C. MITS

The storm starts when the drops start dropping

When the drops stop dropping then the storm starts stopping

– Dr. Seuss

Oh Say Can You Say?

1.1. The purpose of this book is to explore various problems and results on irreducible polynomials. This section discusses some basic definitions and a theorem of Gauss. We begin by clarifying the direction of this book. Throughout this book, we will make use of the following basic notation and definitions:

$\mathbb{Z}$ represents the set of integers.

$\mathbb{Q}$ represents the set of rational numbers.

$\mathbb{R}$ represents the set of real numbers.

$\mathbb{C}$ represents the set of complex numbers.

$R[x]$ represents the set of polynomials with coefficients coming from a ring R (for example, R may be $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, or the integers modulo m where m denotes an integer $m > 1$).

$\deg f(x)$ represents the degree of a non-zero polynomial $f(x)$.

As suggested in the above definition of the degree of $f(x)$, we will make an effort to avoid referring to the degree of the polynomial 0. Nevertheless, it will be convenient sometimes to do so. We will view the degree of 0 in such cases as being $-\infty$ with the understanding that $-\infty$ is less than any given integer. With the further understanding that $-\infty + a = -\infty$ for any integer a, we obtain as a consequence of this definition that the degree of the product of two polynomials is the sum of the degrees of the two polynomials.

Observe that equality, addition, and multiplication in $R[x]$, where R denotes a ring, are defined in the obvious way. More specifically, suppose that $f(x) = \sum_{j=0}^{n} a_j x^j$, $g(x) = \sum_{j=0}^{r} b_j x^j$, and $h(x) = \sum_{j=0}^{s} c_j x^j$ are all in $R[x]$. Define $a_j = 0$ for $j > n$, $b_j = 0$ for $j > r$, and $c_j = 0$ for $j > s$. Then $g(x) = h(x)$ means that $b_j = c_j$ for every $j \in \{0, 1, 2, \dots\}$, $f(x) = g(x) + h(x)$ means that $a_j = b_j + c_j$ for every $j \in \{0, 1, 2, \dots\}$, and $f(x) = g(x)h(x)$ means that $a_k = \sum_{j=0}^{k} b_j c_{k-j}$ for every $k \in \{0, 1, 2, \dots\}$.

Definition 1. *A polynomial $f(x) \in \mathbb{Z}[x]$ is irreducible (or irreducible over $\mathbb{Z}$) provided that $f(x) \not\equiv \pm 1$ and whenever $f(x) = g(x)h(x)$ with $g(x)$ and $h(x) \in \mathbb{Z}[x]$, either $g(x) \equiv \pm 1$ or $h(x) \equiv \pm 1$. If $f(x) \in \mathbb{Z}[x]$ is not irreducible and is not 0, 1, or -1, then $f(x)$ is reducible (or reducible over $\mathbb{Z}$). The polynomials 0, 1, and -1 are considered neither reducible nor irreducible.*

Definition 2. *A polynomial $f(x) \in \mathbb{Q}[x]$ is irreducible over $\mathbb{Q}$ provided that $f(x)$ is not constant and whenever $f(x) = g(x)h(x)$ with $g(x)$ and $h(x) \in \mathbb{Q}[x]$, either $\deg g(x) = 0$ or $\deg h(x) = 0$. If $f(x) \in \mathbb{Q}[x]$ has degree at least one and is not irreducible over $\mathbb{Q}$, then $f(x)$ is reducible over $\mathbb{Q}$. Constant polynomials are considered neither reducible nor irreducible over the rationals.*

Thus, for example, $x^2 + 1$ is irreducible over $\mathbb{Z}$ and over $\mathbb{Q}$, but $3x - 3$ is reducible over $\mathbb{Z}$ and irreducible over $\mathbb{Q}$. Although, at this point we will not need the more general definition, we point out that the above definitions agree with the algebraic concepts of irreducibility and reducibility. More specifically, if R is a ring, then $f \in R$ is irreducible provided that f is not a unit in R and whenever $f = gh$ with g and $h \in R$, either g or h is a unit in R; furthermore, if $f \in R$ is not irreducible, is non-zero, and is not a unit in R, then f is reducible. Note that the statements "f is irreducible" and "f is not reducible" have slightly different meanings. The reader should keep this in mind when reading the statements of theorems in this book.

We will mainly focus our attention on irreducibility over $\mathbb{Z}$ and irreducibility over $\mathbb{Q}$. Thus, we will not emphasize irreducibility over general rings. However, later, we will explore irreducibility over other fields bearing in mind that from such irreducibility results one can often deduce irreducibility results over $\mathbb{Z}$ and over $\mathbb{Q}$.

There is an important connection between irreducibility over $\mathbb{Z}$ and irreducibility over $\mathbb{Q}$. This connection was suggested in our simple irreducibility examples above. It is due to Gauss (cf. Hasse [1]).

Theorem 1. *Let $f(x) \in \mathbb{Z}[x]$. If $f(x)$ is not reducible over $\mathbb{Z}$, then $f(x)$ is not reducible over $\mathbb{Q}$. Furthermore, if $f(x)$ is irreducible over $\mathbb{Q}$ and the greatest common divisor of its coefficients is* 1, *then $f(x)$ is irreducible over $\mathbb{Z}$.*

To prove Theorem 1, we first define the content of $f(x) \in \mathbb{Z}[x]$, where $f(x) \not\equiv 0$, to be the greatest common divisor of the coefficients of $f(x)$. We consider the following lemma before proceeding.

Lemma. *Let $g(x)$ and $h(x) \in \mathbb{Z}[x]$ be such that the contents of $g(x)$ and $h(x)$ are both* 1. *Then the content of $g(x)h(x)$ is* 1.

Proof. Let d be the content of $g(x)h(x)$. It suffices to show that if p is a prime, then $p \nmid d$. Let p be a prime. Write $g(x)$ and $h(x)$ in the form $g(x) = \sum_{j=0}^{r} b_j x^j$ and $h(x) = \sum_{j=0}^{s} c_j x^j$. Define $b_j = 0$ for $j > r$, and define $c_j = 0$ for $j > s$. Let k be the minimal $j \in \{0, ..., r\}$ such that $p \nmid b_j$. Let ℓ be the minimal $j \in \{0, ..., s\}$ such that $p \nmid c_j$. These exist since the contents of $g(x)$ and $h(x)$ are 1. Since each of $b_0, b_1, \dots, b_{k-1}$ and each of $c_0, c_1, \dots, c_{\ell-1}$

are divisible by p, the coefficient of $x^{k+\ell}$ in $g(x)h(x)$ is

$$b_0c_{k+\ell} + b_1c_{k+\ell-1} + \cdots + b_{k-1}c_{\ell+1} + b_kc_\ell + b_{k+1}c_{\ell-1} + \cdots + b_{k+\ell}c_0 \equiv b_kc_\ell \not\equiv 0 \pmod{p}.$$

Thus, there is a coefficient in $g(x)h(x)$ which is not divisible by p, proving that $p \nmid d$. ■

Proof of Theorem 1. The theorem is easily verified in the case that $f(x)$ is a constant. We restrict then to the case that $\deg f(x) \geq 1$. We begin by showing that if $f(x) \in \mathbb{Z}[x]$ is reducible over $\mathbb{Q}$, then $f(x)$ is reducible over $\mathbb{Z}$. Suppose $f(x) = u(x)v(x)$ where $u(x)$ and $v(x) \in \mathbb{Q}[x]$, $\deg u(x) > 0$, and $\deg v(x) > 0$. Let a be the content of $f(x)$ so that $f(x) = af_0(x)$ where $f_0(x) \in \mathbb{Z}[x]$ and $f_0(x)$ has content 1. Let b and c be integers such that $bu(x) \in \mathbb{Z}[x]$ and $cv(x) \in \mathbb{Z}[x]$. Let d be the content of $bu(x)$, and let e be the content of $cv(x)$. Hence, there exist polynomials $u_0(x)$ and $v_0(x) \in \mathbb{Z}[x]$ with content 1 such that $bu(x) = du_0(x)$ and $cv(x) = ev_0(x)$. Thus,

$$abcf_0(x) = bcf(x) = (bu(x))(cv(x)) = (du_0(x))(ev_0(x)) = deu_0(x)v_0(x).$$

By the lemma, the content of $u_0(x)v_0(x)$ is 1 so that the content of $deu_0(x)v_0(x)$ is de. Therefore, since the content of $abcf_0(x)$ is abc, we get that $abc = de$. So, $f_0(x) = u_0(x)v_0(x)$. Thus, $f(x) = af_0(x) = au_0(x)v_0(x)$. Note that $au_0(x)$ and $v_0(x) \in \mathbb{Z}[x]$. Also, $\deg(au_0(x)) = \deg u(x) > 0$ and $\deg v_0(x) > 0$. Hence, $f(x)$ is reducible over $\mathbb{Z}$.

The second statement of Theorem 1 follows directly from considering its contrapositive and the definitions of reducibility over $\mathbb{Z}$ and over $\mathbb{Q}$. This completes the proof. ■

To prove that a polynomial $f(x) \in \mathbb{Z}[x]$ is irreducible over $\mathbb{Q}$, we will often use the next result which follows easily from the proof of Theorem 1 given above.

Theorem 2. *If $f(x) \in \mathbb{Z}[x]$ is a reducible polynomial over $\mathbb{Q}$, then there exist $g(x) \in \mathbb{Z}[x]$ and $h(x) \in \mathbb{Z}[x]$ with $f(x) = g(x)h(x)$, $\deg g(x) > 0$, and $\deg h(x) > 0$.*

One can conclude a little more than Theorem 2. If $f(x) \in \mathbb{Z}[x]$ and $f(x)$ factors over the rationals as $f(x) = g(x)h(x)$ with $g(x) \in \mathbb{Q}[x]$, $h(x) \in \mathbb{Q}[x]$, $\deg g(x) \geq 1$, and

$\deg h(x) \geq 1$, then there are rational numbers a and b such that $f(x) = (ag(x))(bh(x))$ (i.e., $ab = 1$) and both $ag(x)$ and $bh(x)$ have integer coefficients.

1.2. We will need the notion of fields at various points in this book, and so we briefly discuss them now. A field F is a set of at least 2 elements together with 2 operations, addition $(+)$ and multiplication $(\times)$, having the properties: (i) the sum or product of 2 elements of F is in F, (ii) sums and products each commute (i.e., $a + b = b + a$ and $a \times b = b \times a$ for every a and b in F), (iii) the associative laws for addition and multiplication hold ($(a + b) + c = a + (b + c)$ and $(a \times b) \times c = a \times (b \times c)$), (iv) there is an additive identity element (denoted 0) and a multiplicative identity element (denoted 1) in F with $0 \neq 1$ ($a + 0 = a$ and $a \times 1 = a$ for every $a \in F$), (v) there is an additive inverse for every $a \in F$ (denoted $-a$ so that $-a + a = 0$) and a multiplicative inverse for every non-zero $a \in F$ (denoted a^{-1} so that $a^{-1} \times a = 1$), and (vi) the distributive law for multiplication over addition holds ($a \times (b + c) = a \times b + a \times c$ for every $a, b, c \in F$). In other words, F is a ring in which the non-zero elements form an abelian group under multiplication. As usual, we will often use $a \cdot b$ or simply ab to denote $a \times b$. The rational numbers, the real numbers, the complex numbers, and integers modulo a prime p with their usual operations of addition and multiplication are easily seen to be fields.

Let F be a field. Corresponding to the definition of irreducibility over $\mathbb{Q}$, we define a non-constant polynomial $f(x) \in F[x]$ to be irreducible over F if $f(x) = g(x)h(x)$ with $g(x)$ and $h(x)$ in $F[x]$ implies that either $g(x)$ or $h(x)$ is a constant. If a polynomial $f(x) \in F[x]$ is not constant and is not irreducible over F, then we say that $f(x)$ is reducible over F. Certainly of some relevance to the subject of irreducibility is the fact that there is unique factorization in $F[x]$.

Theorem 3. *Every non-constant monic polynomial in $F[x]$ can be uniquely written as a product of monic irreducible polynomials in $F[x]$ except for the order in which the factors occur.*

There is really no loss in generality in considering only monic polynomials in the statement of Theorem 3. For example, if $f(x) = 2x^3 - x^2 + 2x - 1$ and one wishes to apply the theorem with $F = \mathbb{Q}$ (or any field with characteristic different from 2), then one simply considers the theorem with $(1/2)f(x) = x^3 - (1/2)x^2 + x - (1/2)$. To establish the theorem, we first establish a lemma (the Division Algorithm for Polynomials). We note, though, that the proof of the theorem will only depend on the existence (and not the uniqueness) of the polynomials $q(x)$ and $r(x)$ in the lemma.

Lemma. *Let $f(x)$ and $g(x)$ be in $F[x]$ with $g(x) \not\equiv 0$. Then there exist unique polynomials $q(x)$ and $r(x)$ in $F[x]$, with either $r(x) \equiv 0$ or $\deg r(x) < \deg g(x)$, such that*

$$f(x) = g(x)q(x) + r(x). \tag{1.1}$$

Proof. We first show the existence of such $q(x)$ and $r(x)$. If (1.1) is possible with $r(x) \equiv 0$, then we are through. Suppose then that (1.1) is not possible with $r(x) \equiv 0$. In particular, $f(x) \not\equiv 0$. Observe that (1.1) (without any degree restrictions on $r(x)$) is satisfied with $q(x) \equiv 0$ and $r(x) = f(x)$. Consider a $q(x)$ and $r(x) \in F[x]$ satisfying (1.1) with $\deg r(x)$ as small as possible. We write $g(x) = \sum_{j=0}^{r} b_j x^j$ and $r(x) = \sum_{j=0}^{s} c_j x^j$, where b_r and c_s are non-zero. To show the existence of $q(x)$ and $r(x)$ as in the theorem, it suffices to show that $s < r$. If $s \ge r$, then observe that $f(x) = g(x)q_0(x) + r_0(x)$ where $q_0(x) = q(x) + b_r^{-1}c_s x^{s-r}$ and $r_0(x) = r(x) - b_r^{-1}c_s x^{s-r} g(x)$. Since either $r_0(x) \equiv 0$ or $\deg r_0(x) < \deg r(x)$, this situation is impossible (since (1.1) is impossible with $r(x) \equiv 0$ and since we chose $r(x)$ so that $\deg r(x)$ is minimal). Therefore, $s < r$.

To prove uniqueness, for $j = 1$ and 2, consider $q_j(x)$ and $r_j(x)$ in $F[x]$ such that either $r_j(x) \equiv 0$ or $\deg r_j(x) < \deg g(x)$ and such that $f(x) = g(x)q_j(x) + r_j(x)$. Then

$$g(x)\left(q_2(x) - q_1(x)\right) = r_1(x) - r_2(x).$$

Hence, $r_1(x) \equiv r_2(x)$ since otherwise $g(x)$ divides a non-zero polynomial of degree $<$ $\deg g(x)$ (namely, $r_1(x) - r_2(x)$). It follows that $q_1(x) \equiv q_2(x)$, completing the proof. ■

Proof of Theorem 3. If $f(x) \in F[x]$ with $\deg f(x) > 0$, then either $f(x)$ can be written as the product of smaller degree polynomials or $f(x)$ is irreducible. The existence of a factorization for non-constant monic polynomials $f(x) \in F[x]$ into a product of monic irreducible polynomials easily follows by induction on the degree of $f(x)$. To establish the uniqueness of such a factorization, we show that whenever

$$g_1(x)g_2(x)\cdots g_\ell(x) = h_1(x)h_2(x)\cdots h_k(x), \tag{1.2}$$

where each $g_j(x)$ and each $h_j(x)$ is a monic irreducible polynomial in $F[x]$, the polynomials $g_1(x), \dots, g_\ell(x)$ are some rearrangement of the polynomials $h_1(x), \dots, h_k(x)$. We assume that there exist monic irreducible $g_j(x)$ and $h_j(x)$ in $F[x]$ as in (1.2) with $g_1(x), \dots, g_\ell(x)$ not some rearrangement of the polynomials $h_1(x), \dots, h_k(x)$. Among all such examples, we consider one with $\deg g_1(x)$ minimal. Set $g(x) = g_1(x)$. For each $j \in \{1, \dots, k\}$, we consider $q_j(x)$ and $r_j(x)$ in $F[x]$ with $\deg r_j(x) < \deg g(x)$ and with $h_j(x) = g(x)q_j(x) + r_j(x)$ (such $q_j(x)$ and $r_j(x)$ exist by the lemma). From (1.2), we get that

$$r_1(x)\cdots r_k(x) = g(x)w(x) \tag{1.3}$$

for some $w(x) \in F[x]$. In (1.3), we factor out the leading coefficients from $w(x)$ and each $r_j(x)$ and cancel to obtain a factorization of the form (1.3) where $w(x)$ and each $r_j(x)$ is monic. We express these polynomials as possibly empty products of monic irreducible polynomials from $F[x]$. Since $\deg r_j(x) < \deg g(x)$ for each $j \in \{1, \dots, k\}$, each of the resulting monic irreducible factors on the left-hand side of (1.3) will be different from $g(x)$. We divide both sides of the resulting equation by whatever monic irreducible factors both sides of the equation have in common. Since $g(x)$ will remain as a factor on the right-hand side, there will still be at least one monic irreducible factor on the left-hand side as well. We are left with an equation of the form (1.2) where each polynomial on the left-hand side is a monic irreducible polynomial in $F[x]$ of degree $< \deg g(x) = \deg g_1(x)$ and where $g(x)$ is a factor occurring on the right-hand side. This contradicts the minimality of $\deg g_1(x)$, completing the proof. ■

Before closing this section, we make a remark on the last lemma that will be of use to us later. Suppose that $f(x)$ and $g(x)$ are in $\mathbb{Z}[x]$. Since $\mathbb{Z}[x] \subseteq \mathbb{Q}[x]$, the lemma implies that there exist unique polynomials $q(x)$ and $r(x)$ in $\mathbb{Q}[x]$ such that $f(x) = g(x)q(x)+r(x)$ and either $r(x) \equiv 0$ or $\deg r(x) < \deg g(x)$. It is often the case that the polynomials $q(x)$ and $r(x)$ are not in $\mathbb{Z}[x]$ (for example, consider $f(x) = x^2 + 1$ and $g(x) = 2x + 1$). However, whenever $g(x)$ is monic, $q(x)$ and $r(x)$ are in $\mathbb{Z}[x]$. To see this, observe that it suffices to show that $q(x) \in \mathbb{Z}[x]$. Assume $q(x) \notin \mathbb{Z}[x]$, and let k be the maximal non-negative integer for which the coefficient of x^k in $q(x)$, say c, is in $\mathbb{Q}$ but not in $\mathbb{Z}$. If $m = \deg g(x)$, then it follows that the coefficient of x^{m+k} on the right side of $f(x) = g(x)q(x) + r(x)$ is of the form c plus an integer and, hence, is not an integer. This contradicts that the coefficients on the left side are integral. Thus, $q(x) \in \mathbb{Z}[x]$.

1.3. As in the previous section, we let F denote a field. Fix $g(x) = \sum_{j=0}^{r} b_j x^j$ and $h(x) = \sum_{j=0}^{s} c_j x^j$ in $F[x]$ with $b_r \neq 0$. Consider the set of all polynomials of the form

$$u(x)g(x) + v(x)h(x), \tag{1.4}$$

where $u(x)$ and $v(x)$ denote elements in $F[x]$. Observe that if $u(x) = b_r^{-1}$ and $v(x) = 0$, then the polynomial in (1.4) is monic. Let $w(x)$ be a monic polynomial of the form given in (1.4) with smallest possible degree. In particular, $\deg w(x) \le \deg g(x)$. Observe that if there is a polynomial $d(x) \in F[x]$ that divides both $g(x)$ and $h(x)$, then since $w(x)$ is of the form given in (1.4), $d(x)$ divides $w(x)$. Thus, every common divisor of $g(x)$ and $h(x)$ divides $w(x)$. We show that $w(x)$ is itself a common divisor of $g(x)$ and $h(x)$. By the lemma to Theorem 3, we know that there is a $q(x)$ and an $r(x)$ in $F[x]$ such that $g(x) = w(x)q(x)+r(x)$ and either $r(x) \equiv 0$ or $\deg r(x) < \deg w(x)$. If $r(x) \not\equiv 0$, then there is a $b \in F$ such that $br(x)$ is monic and $\deg(br(x)) < \deg w(x)$. If $u(x)$ and $v(x)$ are fixed so that the expression in (1.4) is equal to $w(x)$, we then obtain that

$$(b - bu(x)q(x))\, g(x) + (-bv(x)q(x))\, h(x) = br(x),$$

contradicting the minimality of $\deg w(x)$. Thus, we get that $r(x) \equiv 0$ so that $w(x)$ divides $g(x)$. Similarly, $w(x)$ divides $h(x)$. Observe that, although we fixed $g(x)$ above to be a non-zero polynomial, the same conclusions would follow if instead we only knew that $h(x)$ was non-zero.

Definition. *The greatest common divisor of two polynomials $g(x)$ and $h(x)$ in $F[x]$, with at least one of $g(x)$ or $h(x)$ non-zero, is the monic polynomial in $F[x]$ of largest degree which divides both $g(x)$ and $h(x)$ and is denoted by* $\gcd(g(x), h(x))$.

Let $g(x)$ and $h(x)$ be as in the definition. Clearly, there is a monic polynomial of largest degree as in the definition. We justify that this definition uniquely defines this polynomial (as suggested by the wording in the definition). As before, we set $w(x) \in F[x]$ to be a non-zero monic polynomial of smallest possible degree that can be written in the form (1.4). Suppose $\widetilde{w}(x)$ satisfies the definition of the greatest common divisor of $g(x)$ and $h(x)$ above. Then, as we have just seen, $\widetilde{w}(x)$ divides $w(x)$ so that $\deg \widetilde{w}(x) \le \deg w(x)$. On the other hand, $w(x)$ divides $g(x)$ and $h(x)$, so by the definition of the greatest common divisor of $g(x)$ and $h(x)$, we obtain $\deg \widetilde{w}(x) \ge \deg w(x)$. Therefore, $\deg \widetilde{w}(x) = \deg w(x)$ and $w(x)$ divides $\widetilde{w}(x)$. We deduce that $\widetilde{w}(x) = cw(x)$ for some $c \in F$. Since $\widetilde{w}(x)$ and $w(x)$ are both monic, we obtain $c = 1$ and $\widetilde{w}(x) = w(x)$. This is true for any $\widetilde{w}(x)$ satisfying the definition above; hence, the greatest common divisor of $g(x)$ and $h(x)$ is uniquely determined by this definition. Furthermore, we obtain that there must be exactly one non-zero monic polynomial in $F[x]$ of smallest degree which can be written in the form (1.4) and that it is the same as the greatest common divisor of $g(x)$ and $h(x)$.

We consider the trivial example of $g(x) = 2x$ and $h(x) = x + 1$ and $F = \mathbb{Q}$. Here, it is easy to see from the definition that the greatest common divisor of $g(x)$ and $h(x)$ is 1. Also, the expression in (1.4) is 1 with $u(x) = -1/2$ and $v(x) = 1$. Observe that for every choice of $u(x)$ and $v(x)$ in $\mathbb{Z}[x]$, we have that

$$u(x)(2x) + v(x)(x + 1) \ne 1$$

(since otherwise we could let $x = 1$ to get that $1 = 2u(1) + 2v(1)$, an even number). It is important to keep in mind that if (1.4) is the greatest common divisor of $g(x)$ and $h(x)$ in $F[x]$, then the coefficients of $u(x)$ and $v(x)$ belong to the *field* F.

For computational purposes, it is convenient to know Euclid's algorithm for computing the greatest common divisor of two polynomials. Euclid's algorithm can be described as follows. Let $g(x)$ and $h(x)$ be in $F[x]$ with $g(x) \not\equiv 0$. Define $r_{-1}(x) = g(x)$ and $r_0(x) = h(x)$. For k a positive integer, define $r_k(x)$ recursively by taking $q_k(x)$ and $r_k(x)$ to be the unique polynomials in $F[x]$ (see the lemma to Theorem 3) such that

$$r_{k-2}(x) = r_{k-1}(x)q_k(x) + r_k(x) \tag{1.5}$$

with

$$r_k(x) \equiv 0 \quad \text{or} \quad \deg r_k(x) < \deg r_{k-1}(x).$$

Since $r_k(x) \equiv 0$ or $r_k(x)$ has smaller degree than $r_{k-1}(x)$, we deduce that there exists a smallest integer $t \geq 0$ such that $r_t(x) \equiv 0$. If $t = 0$, then it is easy to see that there is an $a \in F$ such that $ar_{t-1}(x) = \gcd(g(x), h(x))$ (a in this case being the element in F which is an inverse of the leading coefficient of $g(x)$). We show in general that a multiple of $r_{t-1}(x)$ is the greatest common divisor of $g(x)$ and $h(x)$. Beginning with $k = 1$ in (1.5) and by incrementing k, we obtain inductively that any common divisor of $g(x) = r_{-1}(x)$ and $h(x) = r_0(x)$ divides $r_k(x)$. In particular, $\gcd(g(x), h(x))$ divides $r_{t-1}(x)$. Beginning with $k = t$ in (1.5) and by decrementing k, we obtain that $r_{t-1}(x)$ is a divisor of $r_{k-2}(x)$ for $k \in \{1, 2, \ldots, t\}$. Hence, $r_{t-1}(x)$ divides both $g(x)$ and $h(x)$. The desired result that a multiple of $r_{t-1}(x)$ is $\gcd(g(x), h(x))$ easily follows.

The above outlines a method for determining the greatest common divisor of two given polynomials in $F[x]$. The method is called the Euclidean algorithm (for polynomials). Observe that if $h(x) \not\equiv 0$, then $t \leq 1 + \deg h(x)$ so that at most $1 + \deg h(x)$ "steps" are needed above to determine $\gcd(g(x), h(x))$. If $g(x)h(x) \not\equiv 0$, then we can replace the roles of $g(x)$ and $h(x)$ if necessary to compute the greatest common divisor in $\leq$ $\min\{\deg g(x), \deg h(x)\} + 1$ steps.

As an example, consider

$$g(x) = x^5 - x^4 + 3x^3 + 2 \quad \text{and} \quad h(x) = x^4 - 2x^3 + 4x^2 - 3x + 2$$

as polynomials in $\mathbb{Q}[x]$. Then we get

$$r_{-1}(x) = x^5 - x^4 + 3x^3 + 2, \quad r_0(x) = x^4 - 2x^3 + 4x^2 - 3x + 2,$$

$$r_1(x) = x^3 - x^2 + x, \quad r_2(x) = 2x^2 - 2x + 2, \quad r_3(x) = 0,$$

$$q_1(x) = x + 1, \quad q_2(x) = x - 1, \quad \text{and} \quad q_3(x) = x/2.$$

To obtain the greatest common divisor of $g(x)$ and $h(x)$, we multiply $r_2(x)$ by an appropriate constant (namely, $1/2$) to obtain a monic polynomial. Thus,

$$\gcd(g(x), h(x)) = x^2 - x + 1.$$

In this example, we have used the Euclidean algorithm directly. When computing greatest common divisors by hand, there are many short cuts one can make along the way. Using the notation before the example, we know that $\gcd(g(x), h(x))$ divides $r_k(x)$ for every k. Thus, the monic factor of $r_k(x)$ of largest degree which divides both $g(x)$ and $h(x)$ will be the greatest common divisor of $g(x)$ and $h(x)$. In the example, we could have noticed that $r_1(x) = x^3 - x^2 + x = x(x^2 - x + 1)$ and checked whether each of x and $x^2 - x + 1$ divides $g(x)$ and $h(x)$. The latter does and the former doesn't, so we could have concluded that $\gcd(g(x), h(x)) = x^2 - x + 1$ without computing $r_2(x)$ or $r_3(x)$. More can be said in this direction; for every $k \le t$, $\gcd(r_{k-1}(x), r_k(x)) = \gcd(g(x), h(x))$. Hence, after factoring $r_1(x)$, we could have simply noticed that x is not a factor of $h(x)$, checked that $x^2 - x + 1$ is a factor of $h(x)$, and concluded that $\gcd(g(x), h(x)) = x^2 - x + 1$. We were also fortunate that there were not a great deal of non-integral coefficients that came into play in this example. When working in $\mathbb{Q}[x]$, one can in general multiply $q_k(x)$ and $r_k(x)$ by non-zero constants as one computes them if one wishes to avoid dealing with non-integral coefficients.

Let $d(x) = \gcd(g(x), h(x))$. Recall that we have already seen that there must exist $u(x)$ and $v(x)$ in $F[x]$ such that

$$u(x)g(x) + v(x)h(x) = d(x).$$

One can calculate such a $u(x)$ and $v(x)$ as one does the Euclidean algorithm. Define

$$u_{-1}(x) = 1, \; u_0(x) = 0, \; v_{-1}(x) = 0, \; \text{and} \; v_0(x) = 1.$$

For $k \geq 1$, define

$$u_k(x) = q_k(x)u_{k-1}(x) + u_{k-2}(x) \quad \text{and} \quad v_k(x) = q_k(x)v_{k-1}(x) + v_{k-2}(x).$$

By observing that

$$\begin{aligned} u_{k-1}&(x)v_k(x) - u_k(x)v_{k-1}(x) \\ &= u_{k-1}(x)\left(q_k(x)v_{k-1}(x) + v_{k-2}(x)\right) - \left(q_k(x)u_{k-1}(x) + u_{k-2}(x)\right)v_{k-1}(x) \\ &= -\left(u_{k-2}(x)v_{k-1}(x) - u_{k-1}(x)v_{k-2}(x)\right), \end{aligned}$$

one gets by using induction that

$$u_{k-1}(x)v_k(x) - u_k(x)v_{k-1}(x) = (-1)^k \quad \text{for every } k \geq 0.$$

Similar induction arguments give

$$u_{k-1}(x)r_k(x) + u_k(x)r_{k-1}(x) = h(x) \quad \text{for every } k \geq 0$$

and

$$v_{k-1}(x)r_k(x) + v_k(x)r_{k-1}(x) = g(x) \quad \text{for every } k \geq 0.$$

In particular, by letting $k = t$, we get that

$$u_t(x)r_{t-1}(x) = h(x), \quad v_t(x)r_{t-1}(x) = g(x),$$

and

$$\begin{aligned} (-1)^t r_{t-1}(x) &= \left(u_{t-1}(x)v_t(x) - u_t(x)v_{t-1}(x)\right) r_{t-1}(x) \\ &= u_{t-1}(x)\left(v_t(x)r_{t-1}(x)\right) - v_{t-1}(x)\left(u_t(x)r_{t-1}(x)\right) \\ &= u_{t-1}(x)g(x) - v_{t-1}(x)h(x). \end{aligned}$$

Recalling that $r_{t-1}(x)$ is a multiple of $\gcd(g(x), h(x))$, we can multiply through by an appropriate constant above to get the representation of $\gcd(g(x), h(x))$ in the form (1.4).

Next, we turn to roots of polynomials and prove

Theorem 4. *If $f(x)$ is a non-zero polynomial in $F[x]$ of degree n, then there are at most n distinct $a \in F$ such that $f(a) = 0$.*

Proof. Assume that the theorem is not true, and let $f(x)$ be a non-zero polynomial in $F[x]$ of minimal degree n with more than n distinct roots in F. There exist distinct elements $a_1, \dots, a_{n+1}$ in F with $f(a_j) = 0$ for each $j \in \{1, \dots, n+1\}$. By the lemma to Theorem 3, there exist $q(x)$ and $r(x)$ in $F[x]$ such that

$$(x - a_1)(x - a_2) \cdots (x - a_n) = f(x)q(x) + r(x) \tag{1.6}$$

and such that either $r(x) \equiv 0$ or $\deg r(x) < \deg f(x)$. By substituting each of $a_1, \dots, a_n$ for x into (1.6), we get that

$$r(a_1) = r(a_2) = \cdots = r(a_n) = 0.$$

By the minimality condition on $\deg f(x)$ above, we get that $r(x) \equiv 0$. Since $\deg f(x) = n$, we get that $\deg q(x) = 0$. Thus, there is an $a \in F$ such that

$$f(x) = a(x - a_1)(x - a_2) \cdots (x - a_n).$$

Since $f(x) \not\equiv 0$, $a \neq 0$. Hence,

$$f(a_{n+1}) = a(a_{n+1} - a_1)(a_{n+1} - a_2) \cdots (a_{n+1} - a_n) \neq 0,$$

giving a contradiction and completing the proof. ∎

Finally, we mention that throughout this book, we will make use of the Fundamental Theorem of Algebra that a polynomial $f(x)$ with complex coefficients of degree ≥ 1 has at least one root in the set of complex numbers. From the Fundamental Theorem of Algebra, it is not difficult to conclude that if $f(x) \in \mathbb{C}[x]$ and $\deg f(x) = n$, then there exist n complex numbers $\alpha_1, \dots, \alpha_n$ (not necessarily distinct) and a complex number a such that

$$f(x) = a(x - \alpha_1) \cdots (x - \alpha_n).$$

We do not prove the Fundamental Theorem of Algebra here but simply comment that its proof can be found in (almost) any book on complex analysis and in many books on algebra.

Problems

(1.1) Let a be an integer, and let $f(x) \in \mathbb{Z}[x]$. Suppose that $f(x+a)$ is irreducible over $\mathbb{Z}$ (or over $\mathbb{Q}$). Prove that $f(x)$ is irreducible over $\mathbb{Z}$ (or over $\mathbb{Q}$, respectively). (We will usually use this result in this form, but observe that the role of $x+a$ in this problem can be replaced by any $g(x) \in \mathbb{Z}[x]$ with $\deg g(x) \geq 1$.)

(1.2) Let $f(x) \in \mathbb{Z}[x]$ with $f(0) \neq 0$, and let $n = \deg f(x)$. Suppose that $x^n f(1/x)$ is irreducible over $\mathbb{Z}$ (or over $\mathbb{Q}$). Prove that $f(x)$ is irreducible over $\mathbb{Z}$ (or over $\mathbb{Q}$, respectively).

(1.3) Find a necessary and sufficient condition for a quadratic $f(x) = ax^2+bx+c \in \mathbb{Z}[x]$ to be irreducible. Your condition should be written in terms of a, b, and c and should make use of Theorem 1.

(1.4) If $f(x) \in \mathbb{Z}[x]$ is monic, of degree ≥ 1, and reducible over the rationals, then show that there are monic polynomials $g(x)$ and $h(x)$ in $\mathbb{Z}[x]$ each of degree ≥ 1 and such that $f(x) = g(x)h(x)$.

(1.5) Let $f(x)$ denote a non-zero polynomial of degree n with arbitrary coefficients. Prove that

$$f(m) = \sum_{j=0}^{n} (-1)^{n-j} \binom{m}{j} \binom{m-j-1}{n-j} f(j)$$

holds for all integers $m > n$. Deduce that if $f(x) \in \mathbb{Z}[x]$, then $\gcd(f(0), f(1), \dots, f(n))$ is a divisor of $f(m)$ for every $m \in \mathbb{Z}$.

(1.6) If $f(x) \in \mathbb{Z}[x]$ and p is a prime, then we can view $f(x)$ as a polynomial in $\mathbb{Z}_p[x]$ where $\mathbb{Z}_p$ denotes the finite field with p elements (i.e., arithmetic modulo p). Thus, $f(x)$ is reducible modulo p if and only if there exist $g(x)$ and $h(x)$ in $\mathbb{Z}_p[x]$ of degree > 0 such that $f(x) \equiv g(x)h(x) \pmod{p}$.

(a) Give an example of a polynomial $f(x) \in \mathbb{Z}[x]$ and a prime p such that $f(x)$ is irreducible modulo p but $f(x)$ is reducible over $\mathbb{Q}$.

(b) Let $f(x) \in \mathbb{Z}[x]$, and let p be a prime. Suppose that the leading coefficient of $f(x)$ is not divisible by p. Prove that if $f(x)$ is irreducible modulo p, then $f(x)$ is irreducible over $\mathbb{Q}$.

(1.7) Let $f(x)$ and $g(x)$ be irreducible polynomials in $\mathbb{Z}[x]$. Prove that if there is a number α which is a root of both $f(x)$ and $g(x)$, then $f(x) \equiv \pm g(x)$. Also, explain what the analogous result is if $f(x)$ and $g(x)$ are irreducible polynomials in $\mathbb{Q}[x]$.

(1.8) Let $f(x)$ be an irreducible polynomial in $\mathbb{Z}[x]$ or $\mathbb{Q}[x]$. Prove that $f(x)$ has no multiple roots (i.e., prove that there does not exist an $\alpha \in \mathbb{C}$ such that $(x-\alpha)^2$ is a factor of $f(x)$ in $\mathbb{C}[x]$).

(1.9) Let $f(x) = \sum_{j=0}^{n} a_j x^j \in \mathbb{Z}[x]$. Let a and b be relatively prime integers with $b \neq 0$. Prove that if $f(a/b) = 0$, then $a|a_0$ and $b|a_n$. (This is the Rational Root Test and provides a way to find the linear factors for a given $f(x) \in \mathbb{Z}[x]$.)

(1.10) Let $f(x)$ and $g(x)$ be in $\mathbb{Z}[x]$ with $g(x) \not\equiv 0$ and the content of $g(x)$ being 1. Suppose that $f(x) = g(x)h(x)$ with $h(x) \in \mathbb{Q}[x]$. Prove that $h(x) \in \mathbb{Z}[x]$.

(1.11) (a) Let $f(x)$ and $g(x)$ be in $\mathbb{Z}[x]$, and suppose that there are infinitely many integers m such that $g(m)|f(m)$. Prove that there is a polynomial $h(x) \in \mathbb{Q}[x]$ such that $f(x) = g(x)h(x)$. (Hint: Use the lemma to Theorem 3.)

(b) Find an example of an $f(x)$ and $g(x)$ as in part (a) for which no $h(x) \in \mathbb{Z}[x]$ exists such that $f(x) = g(x)h(x)$.

(1.12) Prove the assertions

$$u_{k-1}(x)r_k(x) + u_k(x)r_{k-1}(x) = h(x) \quad \text{for every } k \geq 0$$

and

$$v_{k-1}(x)r_k(x) + v_k(x)r_{k-1}(x) = g(x) \quad \text{for every } k \geq 0$$

given in Section 1.3.

(1.13) Let $f(x) = 2x^7 + 3x^6 + 2x^5 + 8x^4 + 4x^3 + 3x^2 + 4x + 1$ and $g(x) = 2x^6 + x^5 + x^4 + 5x^3 + x + 2$. Calculate the greatest common divisor of $f(x)$ and $g(x)$ in $\mathbb{Q}[x]$.

(1.14) Recall the example with $g(x) = x^5 - x^4 + 3x^3 + 2$ and $h(x) = x^4 - 2x^3 + 4x^2 - 3x + 2$ in Section 1.3. Find $u(x)$ and $v(x)$ such that $g(x)u(x) + h(x)v(x) = x^2 - x + 1$.