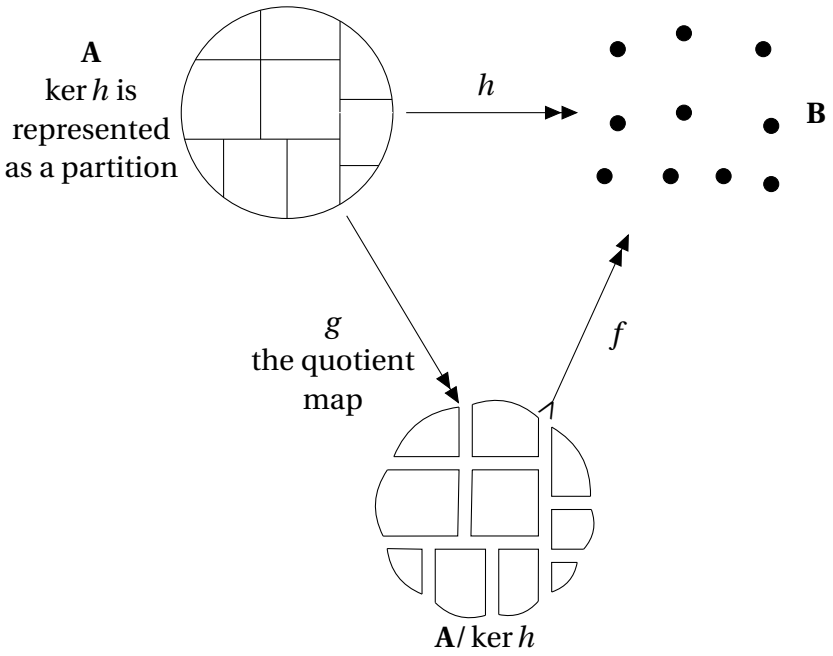
GEORGE MCNULTY

Rings and Modules

Algebra For First Year Graduate Students
Part I

DRAWINGS BY THE AUTHOR



UNIVERSITY OF SOUTH CAROLINA

2014

PREFACE

This exposition is for the use of first year graduate students pursuing advanced degrees in mathematics. In the United States, people in this position generally find themselves confronted with a battery of examinations at the beginning of their second year, so if you are among them a good part of your energy during your first year will be expended mastering the essentials of several branches of mathematics, algebra among them.

While every doctoral program in mathematics sets its own expectations, there is a fair consensus on those parts of algebra that should be part of a mathematician's repertoire. I have tried to gather here just those parts. So here you will find the basics of (commutative) rings and modules in Part I. The basics of groups and fields, constituting the content of second semester, are in Part II. The background you will need to make good use of this exposition is a good course in linear algebra and another in abstract algebra, both at the undergraduate level.

As you proceed through these pages you will find many places where the details and sometimes whole proofs of theorems will be left in your hands. The way to get the most from this presentation is to take it on with paper and pencil in hand and do this work as you go. There are also weekly problem sets. Most of the problems have appeared on Ph.D. examinations at various universities. In a real sense, the problems sets are the real heart of this presentation.

This work grew out of teaching first year graduate algebra courses. Mostly, I have done this at the University of South Carolina (but the first time I did it was at Dartmouth College and I had the delightful experience of teaching this material at the University of the Philippines). Many of the graduate students in these courses have influenced my presentation here. Before all others, I should mention Kate Scott Owens, who had the audacity to sit in to front row with her laptop, putting my classroom discussions into LaTeX on the fly. She then had the further audacity to post the results so that all the flaws and blunders I made would be available to everyone. So this effort at exposition is something in the way of self-defense....

George F. McNulty
Columbia, SC
2014

CONTENTS

LECTURE 0

THE BASICS OF ALGEBRAIC SYSTEMS

0.1 ALGEBRAIC SYSTEMS

In your undergraduate coursework you have already encountered many algebraic systems. These probably include some specific cases, like $\langle \mathbb{Z}, +, \cdot, -, 0, 1 \rangle$ which is the system of integers equipped with the usual two-place operations of addition, multiplication, the one-place operation of forming negatives, and two distinguished integers 0 and 1, which we will construe as zero-place operations (all output and no input). You have also encountered whole classes of algebraic systems such as the class of vector spaces over the real numbers, the class of rings, and the class of groups. You might even have encountered other classes of algebraic systems such are Boolean algebras and lattices.

The algebraic systems at the center of this two-semester course are rings, modules, groups, and fields. Vector spaces are special cases of modules. These kinds of algebraic systems arose in the nineteenth century and the most of the mathematics we will cover was well-known by the 1930's. This material forms the basis for a very rich and varied branch of mathematics that has flourished vigorously over the ensuing decades.

Before turning to rings, modules, groups, and fields, it pays to look at algebraic systems from a fairly general perspective. Each algebraic system consist of a nonempty set of elements, like the set $\mathbb{Z}$ of integers, equipped with a system of operations. The nonempty set of elements is called the **universe** of the algebraic system. (This is a shortening of "universe of discourse".) Each of the operations is a function that takes as inputs arbitrary r-tuples of elements of the universe and returns an output again in the universe—here, for each operation, r is some fixed natural number called the **rank** of the operation. In the familiar algebraic system $\langle \mathbb{Z}, +, \cdot, -, 0, 1 \rangle$, the operations of addition and multiplication are of rank 2 (they are two-place operations), the operation of forming negatives is of rank 1, and the two distinguished elements 0 and 1 are each regarded as operations of rank 0.

Aside. Let A be a set and r be a natural number. We use A^r to denote the set of all r-tuples of elements of A. An operation F of rank r on A is just a function from A^r into A. There is a curious case. Suppose A is the empty set and $r > 0$. Then A^r is also empty. A little reflection shows that

the empty set is also a function from A^r into A, that is the empty set is an operation of rank r. The curiosity is that this is so for any positive natural number r. This means that the rank of this operation is not uniquely determined. We also note that A^0 actually has one element, namely the empty tuple. This means that when A is empty there can be no operations on A of rank 0. On the other hand, if A is nonempty, then the rank of every operation of finite rank is uniquely determined and A has operations of every finite rank. These peculiarities explain, to some extent, why we adopt the convention that the universe of an algebraic system should be nonempty.

The notion of the signature of an algebraic system is a useful way to organize the basic notions of our subject. Consider these three familiar algebraic systems:

$$\langle \mathbb{Z}, +, \cdot, -, 0, 1 \rangle$$
$$\langle \mathbb{C}, +, \cdot, -, 0, 1 \rangle$$
$$\langle \mathbb{R}_{2\times 2}, +, \cdot, -, 0, 1 \rangle$$

The second system consists of the set of complex numbers equipped with the familiar operations, while the third system consists of the set of 2×2 matrices with real entries equipped with matrix multiplication, matrix addition, matrix negation, and distinguished elements 0 for the zero matrix, and 1 for the identity matrix. Notice that + has a different meaning on each line displayed above. This is a customary, even well-worn, ambiguity. To resolve this ambiguity let us regard + not as a two-place operation but as a symbol for a two-place operation. Then each of the three algebraic systems gives a different meaning to this symbol—a meaning that would ordinarily be understood from the context, but could be completely specified as needed. A **signature** is just a set of operation symbols, each with a uniquely determined natural number called its rank. More formally, a signature is a function with domain some set of operation symbols that assigns to each operation symbol its rank. The three algebraic systems above have the same signature.

"Algebraic system" is a mouthful. So we shorten it to "algebra". This convenient choice is in conflict another use of this word to refer to a particular kind of algebraic system obtained by adjoining a two-place operation of a certain kind to a module.

As a matter of notation, we tend to use boldface $\mathbf{A}$ to denote an algebra and the corresponding normalface A to denote its universe. For an operation symbol Q we use, when needed, $Q^{\mathbf{A}}$ to denote the operation of $\mathbf{A}$ symbolized by Q. We follow the custom of writing operations like + between its inputs (like $5+2$) but this device does not work very well if the rank of the operation is not two. So in general we write things like $Q^{\mathbf{A}}(a_0, a_1, \ldots, a_{r-1})$ where the operation symbol Q has rank r and $a_0, a_1, \ldots, a_{r-1} \in A$.

Each algebra has a signature. It is reasonable to think of each algebra as one particular way to give meaning to the symbols of the signature.

Homomorphisms and their relatives

Let $\mathbf{A}$ and $\mathbf{B}$ be algebras of the same signature. We say that a function $h : A \to B$ is a **homomorphism** provided for every operation symbol Q and all $a_0, a_1, \ldots, a_{r-1} \in A$, where r is the rank of Q, we have

$$h(Q^{\mathbf{A}}(a_0, a_1, \ldots, a_{r-1})) = Q^{\mathbf{B}}(h(a_0), h(a_1), \ldots, h(a_{r-1})).$$

That is, h preserves the basic operations. We use $h : \mathbf{A} \to \mathbf{B}$ to denote that h is a homomorphism from the algebra $\mathbf{A}$ into the algebra $\mathbf{B}$. For example, we learned in linear algebra that the determinant det is a homomorphism from $\langle \mathbb{R}_{2\times 2}, \cdot, 0, 1 \rangle$ into $\langle \mathbb{R}, \cdot, 0, 1 \rangle$. The key fact from linear algebra

is

$$\det(AB) = \det A \det B.$$

We note in passing that the multiplication on the left (that is AB) is the multiplication of matrices, while the multiplication on the right is multiplication of real numbers.

In the event that h is a homomorphism from **A** into **B** that happens to be one-to-one we call it an **embedding** and express this in symbols as

$$h : \mathbf{A} \rightarrowtail \mathbf{B}.$$

In the event that the homomorphism h is onto B we say that **B** is a **homomorphic image** of **A** and write

$$h : \mathbf{A} \twoheadrightarrow \mathbf{B}.$$

In the event that the homomorphism h is both one-to-one and onto B we say that h is an **isomorphism** and we express this in symbols as

$$h : \mathbf{A} \rightarrowtail\!\!\!\!\twoheadrightarrow \mathbf{B}$$

or as

$$\mathbf{A} \overset{h}{\cong} \mathbf{B}.$$

It is an easy exercise, done by all hard-working graduate students, that if h is an isomorphism from **A** to **B** then the inverse function h^{-1} is an isomorphism from **B** to **A**. We say that **A** and **B** are **isomorphic** and write

$$\mathbf{A} \cong \mathbf{B}$$

provided there is an isomorphism from **A** to **B**.

The algebra $\langle \mathbb{R}, +, -, 0 \rangle$ is isomorphic to $\langle \mathbb{R}^+, \cdot, ^{-1}, 1 \rangle$, where $\mathbb{R}^+$ is the set of positive real numbers. There are isomorphisms either way that are familiar to freshman in calculus. Find them.

A homomorphism from **A** into **A** is called an **endomorphism** of **A**. An isomorphism from **A** to **A** is called an **automorphism** of **A**.

Subuniverses and sublagebras

Let **A** be an algebra. A subset $B \subseteq A$ is called a **subuniverse** of **A** provided it is closed with respect to the basic operations of **A**. This means that for every operation symbol Q of the signature of **A** and for all $b_0, b_1, \ldots, b_{r-1} \in B$, where r is the rank of Q we have $Q^{\mathbf{A}}(b_0, b_1, \ldots, b_{r-1}) \in B$. Notice that if the signature of **A** has an operation symbol c of rank 0, then $c^{\mathbf{A}}$ is an element of A and this element must belong to every subuniverse of **A**. On the other hand, if the signature of **A** has no operation symbols of rank 0, then the empty set $\varnothing$ is a subuniverse of **A**.

The restriction of any operation of **A** to a subuniverse B of **A** results in an operation on B. In the event that B is a nonempty subuniverse of **A**, we arrive at the **subalgebra B** of **A**. This is the algebra of the same signature as **A** with universe B such that $Q^{\mathbf{B}}$ is the restriction to B of $Q^{\mathbf{A}}$, for each operation symbol Q of the signature. $\mathbf{B} \leq \mathbf{A}$ symbolizes that **B** is a subalgebra of **A**.

Here is a straightforward but informative exercise for hard-working graduate students. Let $\mathbb{N} = \{0, 1, 2, \ldots\}$ be the set of natural numbers. Discover all the subuniverses of the algebra $\langle \mathbb{N}, + \rangle$.

Congruence relations

Let **A** be an algebra and h be a homomorphism from **A** to some algebra. We associate with h the following set, which called here the **functional kernel** of h,

$$\theta = \{(a, a') \mid a, a' \in A \text{ and } h(a) = h(a')\}.$$

This set of ordered pairs of elements of A is evidently an equivalence relation on A. That is, θ has the following properties.

(a) It is reflexive: $(a, a) \in \theta$ for all $a \in A$.

(b) It is symmetric: for all $a, a' \in A$, if $(a, a') \in \theta$,then $(a', a) \in \theta$.

(c) It is transitive: for all $a, a', a'' \in A$, if $(a, a') \in \theta$ and $(a', a'') \in \theta$, then $(a, a'') \in \theta$.

This much would be true were h any function with domain A. Because θ is a binary (or two-place) relation on A it is useful to use the following notations interchangeably.

$$(a, a') \in \theta$$
$$a \,\theta\, a'$$
$$a \equiv a' \mod \theta$$

Here is another piece of notation which we will use often. For any set A, any $a \in A$ and any equivalence relation θ on A we put

$$a/\theta := \{a' \mid a' \in A \text{ and } a \equiv a' \mod \theta\}.$$

We also put

$$A/\theta := \{a/\theta \mid a \in A\}.$$

Because h is a homomorphism θ has one more important property, sometimes called the substitution property:
For every operation symbol Q of the signature of **A** and for all $a_0, a_0', a_1, a_1', \ldots, a_{r-1}, a_{r-1}' \in A$, where r is the rank of Q,

if

$$\begin{aligned} a_0 &\equiv a_0' \mod \theta \\ a_1 &\equiv a_1' \mod \theta \\ &\vdots \\ a_{r-1} &\equiv a_{r-1}' \mod \theta \end{aligned}$$

then

$$Q^{\mathbf{A}}(a_0, a_1, \ldots, a_{r-1}) \equiv Q^{\mathbf{A}}(a_0', a_1', \ldots, a_{r-1}') \mod \theta.$$

An equivalence relation on A with the substitution property above is called a **congruence** relation of the algebra **A**. The functional kernel of a homomorphism h from **A** into some other algebra is always a congruence of **A**. We will see below that this congruence retains almost all the information about the homomorphism h.

As an exercise to secure the comprehension of this notion, the hard-working graduate students should try to discover all the congruence relations of the familiar algebra $\langle \mathbb{Z}, +, \cdot \rangle$.

A comment of mathematical notation

The union $A \cup B$ and the intersection $A \cap B$ of sets A and B are familiar. These are special cases of more general notions. Let $\mathscr{K}$ be any collection of sets. The union $\bigcup \mathscr{K}$ is defined via

$$a \in \bigcup \mathscr{K} \Leftrightarrow a \in C \text{ for some } C \in \mathscr{K}.$$

Here is the special case $A \cup B$ rendered in this way

$$a \in A \cup B \Leftrightarrow a \in \bigcup \{A, B\} \Leftrightarrow a \in C \text{ for some } C \in \{A, B\} \Leftrightarrow a \in A \text{ or } a \in B.$$

Similarly, the intersection $\bigcap \mathscr{K}$ is defined via

$$a \in \bigcap \mathscr{K} \Leftrightarrow a \in C \text{ for all } C \in \mathscr{K}.$$

Notice that in the definition of the intersection, each set belonging to the collection $\mathscr{K}$ imposes a constraint on what elements are admitted to membership in $\bigcap \mathscr{K}$. When the collection $\mathscr{K}$ is empty there are no constraints at all on membership in $\bigcap \mathscr{K}$. This means $\bigcap \varnothing$ is the collection of all sets. However, the having the collection of all sets in hand leads to a contradiction, as discovered independently by Ernst Zermelo and Bertrand Russell in 1899. To avoid this, we must avoid forming the intersection of empty collections. This situation is analogous to division by zero. Just as when division of numbers comes up, the careful mathematician considers the possibility that the divisor is zero before proceeding, so must the careful mathematician consider the possibility that $\mathscr{K}$ might be empty before proceeding to form $\bigcap \mathscr{K}$.

We also use the notation

$$\bigcup_{i \in I} A_i \text{ and } \bigcap_{i \in I} A_i$$

to denote the union and intersection of $\mathscr{K} = \{A_i \mid i \in I\}$. The set I here is used as a set of indices. In using this notation, we impose no restrictions on I (save that in forming intersections the set I must not be empty). In particular, we make no assumption that the set I is ordered in any way.

The familiar set builder notation, for example $\{n \mid n \text{ is a prime number}\}$, has a companion in the function builder notation. Here is an example

$$f = \langle e^x \mid x \in \mathbb{R} \rangle.$$

The function f is just the exponential function on the real numbers. We take the words "function", "sequence", and "system" to have the same meaning. We also use the notation $f(c)$ and f_c interchangeably when f is a function and c is a member of its domain.

0.2 Problem Set 0

Algebra Homework, Edition 0
First Week
How Is Your Linear Algebra?

Problem 0.
Classify up to similarity all the square matrices over the complex numbers with minimal polynomial $m(x) = (x-1)^2(x-2)^2$ and characteristic polynomial $c(x) = (x-1)^6(x-2)^5$.

Problem 1.
Let $T : V \to V$ be a linear transformation of rank 1 on a finite dimensional vector space V over any field. Prove that either T is nilpotent or V has a basis of eigenvectors of T.

Problem 2.
Let V be a vector space over a field K.

(a) Prove that if U_0 and U_1 are subspaces of V such that $U_0 \nsubseteq U_1$ and $U_1 \nsubseteq U_0$, then $V \neq U_0 \cup U_1$.

(b) Prove that if U_0, U_1, and U_2 are subspaces of V such that $U_i \nsubseteq U_j$ when $i \neq j$ and K has at least 3 elements, then $V \neq U_0 \cup U_1 \cup U_2$.

(c) State and prove a generalization of (b) for n subspaces.

0.3 The Homomorphism Theorem for Algebras of the Simplest Signature

The simplest signature is, of course, empty. It provides no operation symbols. In this setting, algebras have nothing real to distinguish them from nonempty sets. Every function between two nonempty sets will be a homomorphism. Every subset will be a subuniverse. Every equivalence relation will be a congruence. Isomorphisms are just one-to-correspondences between nonempty sets and two nonempty sets will be isomorphic just in case they have the same cardinality. So doing algebra in the empty signature if a branch of combinatorics.

Nevertheless, there is an important lesson for us to learn here.

Suppose that A is a nonempty set. By a **partition** of A we mean a collection $\mathscr{P}$ of subsets of A with the following properties:

(a) Each member of **P** is a nonempty subset of A.

(b) If $X, Y \in \mathscr{P}$ and $X \neq Y$, then X and Y are disjoint.

(c) Every element of A belongs to some set in the collection $\mathscr{P}$.

There is a close connection between the notion of a function with domain A, the notion of an equivalence relation on A, and the notion of a partition of A. You may already be familiar with this connection. We present it in the following theorem:

The Homomorphism Theorem, empty version. *Let A be a nonempty set, let $f : A \twoheadrightarrow B$ be a function from A onto B, let θ be an equivalence relation on A, and let $\mathscr{P}$ be a partition of A. All of the following hold.*

(a) *The functional kernel of f is an equivalence relation on A.*

(b) *The collection $A/\theta = \{a/\theta \mid a \in A\}$ is a partition of A.*

(c) *The map η that assigns to each $a \in A$ the set in $\mathscr{P}$ to which it belongs is a function from A onto $\mathscr{P}$; moreover $\mathscr{P}$ is the collection of equivalence classes of the functional kernel of η.*

(d) *If θ is the functional kernel of f, then there is a one-to-one correspondence g from A/θ to B such that $f = g \circ \eta$.*

Figure 0.3 displays something of what this theorem asserts.

The empty version of the Homomorphism Theorem is almost too easy to prove. One merely has to check what the definitions of the various notions require. The map η is called the **quotient map**. That it is a function, i.e. that =

$$\{(a, X) \mid a \in A \text{ and } a \in X \in \mathscr{P}\}$$

is a function, follows from the disjointness of distinct members of the partition. That its domain in A follows from condition (c) in the definition of partition. The one-to-one correspondence g mentioned in assertion (d) of the Homomorphism Theorem is the following set of ordered pairs:

$$\{(a/\theta, f(a)) \mid a \in A\}.$$

The proof that this set is a one-to-one function from A onto B is straightforward, the most amusing part being the demonstration that it is actually a function.

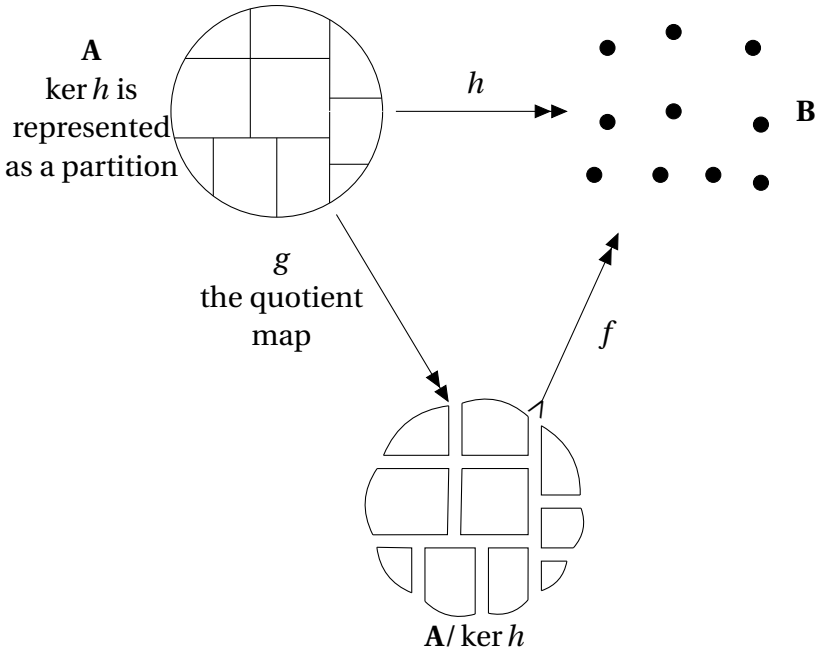


Figure 0.1: The Homomorphism Theorem

0.4 Direct Products

Just as you are familiar with $A \cup B$ and $A \cap B$, you probably already know that $A \times B$ denotes the set of all ordered pairs whose first entries are chosen from A while the second entries are chosen from B. Just as we did for unions and intersections we will extend this notion.

Let $\langle A_i \mid i \in I \rangle$ be any system of sets. We call a function $a : I \to \bigcup_{i \in I} A_i$ a **choice function** for the system $\langle A_i \mid i \in I \rangle$ provided $a_i \in A_i$ for all $i \in I$. It is perhaps most suggestive to think of a as an I-tuple (recalling that we are using function, tuple, system, and sequence interchangeably). The **direct product** of the system $\langle A_i \mid i \in I \rangle$ is just the set of all these choice functions. Here is the notation we use:

$$\prod \langle A_i \mid i \in I \rangle := \prod_{i \in I} A_i := \{a \mid a \text{ is a choice function for the system } \langle A_i \mid i \in I \rangle\}.$$

The sets A_i are called the **direct factors** of this product. If any of the sets in the system $\langle A_i \mid i \in I \rangle$ is empty, then the direct product is also empty. On the other hand, if I is empty then the direct product is $\{\varnothing\}$, since the empty set will turn out to be a choice function for the system. Notice that $\{\varnothing\}$ is itself nonempty and, indeed, has exactly one element.

Observe that $\prod \langle A, B \rangle = \{\langle a, b \rangle \mid a \in A \text{ and } b \in B\}$. This last set is, for all practical purposes, $A \times B$.

Projection functions are associated with direct products. For any $j \in I$, the j^{th} projection function p_j is defined, for all $a \in \prod \langle A_i \mid i \in I \rangle$, via

$$p_j(a) := a_j.$$

The systems of projection functions has the following useful property: it **separate points**. This means that if $a, a' \in \prod \langle A_i \mid i \in T \rangle$ and $a \neq a'$, then $p_j(a) \neq p_j(a')$ for some $j \in I$. Suppose that

$\langle A_i \mid i \in I\rangle$ is a system of sets, that B is some set, and that $\langle f_i \mid i \in I\rangle$ is a system of functions such that $f_i : B \to A_i$ for each $i \in I$. Define the map $h : B \to \prod\langle A_i \mid i_I\rangle$ via

$$h(b) := \langle f_i(b) \mid i \in I\rangle.$$

Then it is easy that $f_i = p_i \circ h$ for all $i \in I$. If the system $\langle f_i \mid i \in I\rangle$ separates points, then the function h defined just above will be one-to-one, as all hard-working graduate students will surely check.

We form direct products of systems of algebras in the following way. Let $\langle \mathbf{A}_i \mid i \in I\rangle$ be a system of algebras, all with the same signature. We take $\prod\langle \mathbf{A}_i \mid i \in I\rangle$ to be the algebra $\mathbf{P}$ with universe $P := \prod\langle A_i \mid i \in I\rangle$ and where the operations on $\mathbf{P}$ are defined coordinatewise. This means that for each operation symbol Q and all $a_0, a_1, \ldots, a_{r-1} \in P$, where r is the rank of Q we have

$$Q^{\mathbf{P}}(a_0, a_1, \ldots, a_{r-1}) = \langle Q^{\mathbf{A}_i}(a_{0,i}, a_{1,i}, \ldots, a_{r-1,i}) \mid i \in I\rangle.$$

To see more clearly what is intended here, suppose that Q has rank 3, that $I = \{0,1,2,3\}$, and that $a, b, c \in P$. Then

$$\begin{array}{rcllll} a & = & \langle a_0, & a_1, & a_2, & a_3\rangle \\ b & = & \langle b_0, & b_1, & b_2, & b_3\rangle \\ c & = & \langle c_0, & c_1, & c_2, & c_3\rangle \\ Q^{\mathbf{P}}(a,b,c) & = & \langle Q^{\mathbf{A}_0}(a_0,b_0,c_0), & Q^{\mathbf{A}_1}(a_1,b_1,c_1), & Q^{\mathbf{A}_2}(a_2,b_2,c_2), & Q^{\mathbf{A}_1}(a_3,b_3,c_3)\rangle \end{array}$$

In this way, the direct product of a system of algebras, all of the same signature, will be again an algebra of the common signature and it is evident that each projection map is a homomorphism from the direct product onto the corresponding direct factor. Even the following fact is easy to prove.

Fact. Let $\langle \mathbf{A}_i \mid i \in I\rangle$ be a system of algebras, all of the same signature. Let $\mathbf{B}$ be an algebra of the same signature as $\mathbf{A}$ and let $\langle f_i \mid i \in I\rangle$ be a system of homomorphisms so that $f_i : \mathbf{B} \to \mathbf{A}_i$ for all $i \in I$. Then there is a homomorphism $h : \mathbf{B} \to \prod_{i\in I}\mathbf{A}_i$ so that $f_i = p_i \circ h$ for all $i \in I$. Moreover, if $\langle f_i \mid i \in I\rangle$ separates points, then h is one-to-one.

LECTURE 1

THE ISOMORPHISM THEOREMS

The four theorems presented today arose over a period of perhaps forty years from the mid 1890's to the mid 1930's. They emerged from group theory and the theory of rings and modules chiefly in the work of Richard Dedekind and Emmy Noether and it was Noether who gave their first clear formulation in the context of module theory in 1927. You have probably already seen versions of these theorems for groups or rings is an undergraduate abstract algebra course.

We will frame them is the broader context of algebras in general. That way it will not be necessary to do more than add a comment or two when applying them in the context of groups, rings, and modules (these being our principal focus). In addition, you will be able to apply them in the context of lattices, Boolean algebras, or other algebraic systems.

At the center of this business is the notion of a **quotient algebra**. Let $\mathbf{A}$ be an algebra and let θ be a congruence of $\mathbf{A}$. Recall that for each $a \in A$ we use a/θ to denote the congruence class $\{a' \mid a' \in A \text{ and } a \equiv a' \mod \theta\}$. Moreover, we use A/θ to denote the partition $\{a/\theta \mid a \in A\}$ of A into congruence classes. We make the quotient algebra $\mathbf{A}/\theta$ by letting its universe be A/θ and, for each operation symbol Q of the signature of $\mathbf{A}$, and all $a_0, a_1, \ldots, a_{r-1} \in A$, where r is the rank of Q, we define

$$Q^{\mathbf{A}/\theta}(a_0/\theta, a_1/\theta, \ldots, a_{r-1}/\theta) := Q^{\mathbf{A}}(a_0, a_1, \ldots, a_{r-1})/\theta.$$

Because the elements of A/θ are congruence classes, we see that the r inputs to $Q^{\mathbf{A}/\theta}$ must be congruence classes. On the left side of the equation above the particular elements a_i have no special standing—they could be replaced by any a_i' provided only that $a_i \equiv a_i' \mod \theta$. Loosely speaking, what this definition says is that to evaluate $Q^{\mathbf{A}/\theta}$ on an r-tuple of θ-classes, reach into each class, grab an element to represent the class, evaluate $Q^{\mathbf{A}}$ at the r-tuple of selected representative to obtain say $b \in A$, and then output the class b/θ. A potential trouble is that each time such a process is executed on the same r-tuple of congruence classes, different representatives might be selected resulting in, say b' instead of b. But the substitution property, the property that distinguishes congruences from other equivalence relations, is just what is needed to see that there is really no trouble. To avoid a forest of subscripts, here is how the argument would go were

Q to have rank 3. Suppose $a, a', b, b', c, c' \in A$ with

$$a/\theta = a'/\theta$$
$$b/\theta = b'/\theta$$
$$c/\theta = c'/\theta$$

So a and a' can both represent the same congruence class—the same for b and b' and for c and c'. Another way to write this is

$$a \equiv a' \mod \theta$$
$$b \equiv b' \mod \theta$$
$$c \equiv c' \mod \theta$$

What we need is $Q^{\mathbf{A}}(a,b,c)/\theta = Q^{\mathbf{A}}(a',b',c')/\theta$. Another way to write that is

$$Q^{\mathbf{A}}(a,b,c) \equiv Q^{\mathbf{A}}(a',b',c') \mod \theta.$$

But this is exactly what the substitution property provides. Hard-working graduate students will do the work to see that what works for rank 3 works for any rank.

The theorem below, sometimes called the First Isomorphism Theorem, is obtained from its version for the empty signature replacing arbitrary functions by homomorphisms and arbitrary equivalence relations by congruence relations.

The Homomorphism Theorem. *Let* $\mathbf{A}$ *be an algebra, let* $f : \mathbf{A} \twoheadrightarrow \mathbf{B}$ *be a homomorphism from* $\mathbf{A}$ *onto* $\mathbf{B}$*, and let* θ *be a congruence relation of* $\mathbf{A}$*. All of the following hold.*

(a) *The functional kernel of* f *is an congruence relation of* A*.*

(b) $\mathbf{A}/\theta$ *is an algebra of the same signature as* $\mathbf{A}$*.*

(c) *The map* η *that assigns to each* $a \in A$ *the congruence class* a/θ *is a homomorphism from* $\mathbf{A}$ *onto* $\mathbf{A}/\theta$ *and its functional kernel is* θ*.*

(d) *If* θ *is the functional kernel of* f*, then there is an isomorphism* g *from* $\mathbf{A}/\theta$ *to* $\mathbf{B}$ *such that* $f = g \circ \eta$*.*

The proof of this theorem has been, for the most part, completed already. We just saw how to prove part (b) and part (a) was done when the notions of congruence relation and functional kernel were introduced. Even parts (c) and (d) were mostly established in the version of the theorem for algebras with empty signature. It only remains to prove that the quotient map η in part (c) and the map g in part (d) are actually homomorphisms. With the definition of how the operations in the quotient algebra work, this only requires checking that the basic oeprations are preserved by η and by g. This work is left to the diligent graduate students.

From parts (a) and (c) of the Homomorphism Theorem we draw the following corollary.

Corollary 1.0.1. *Let* $\mathbf{A}$ *be an algebra. The congruence relations of* $\mathbf{A}$ *are exactly the functional kernels of homomorphisms from* $\mathbf{A}$ *into algebras of the same signature as* $\mathbf{A}$*.*

It will be necessary, as we develop the theory of rings, modules, and groups, to determine whether certain equivalence relations at hand are in fact congruence relations. Of course, we can always check the conditions defining the concept of congruence relation. But sometimes it is simpler to show that the relation is actually the functional kernel of some homomorphism.

Now let us suppose that θ is a congruence of $\mathbf{A}$ and that $\mathbf{B}$ is a subalgebra of $\mathbf{A}$. By $\theta \restriction B$ we mean the restriction of θ to B. That is

$$\theta \restriction B := \theta \cap (B \times B).$$

Now θ partitions A into congruence classes. Some of these congruence classes may include elements of B while others may not. We can **inflate** B using θ to obtain the set θB of all elements of A related by θ to some element of B. That is

$$\theta B := \{a \mid a \equiv b \mod \theta \text{ for some } b \in B\}.$$

The diagram below illustrates the inflation of B by θ, where we have drawn lines to indicate the partition of A into θ-classes.

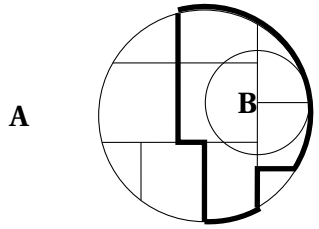


Figure 1.1: The Inflation θB of $\mathbf{B}$ by θ

The Second Isomorphism Theorem. *Let* $\mathbf{A}$ *be an algebra, let* θ *be a congruence of* $\mathbf{A}$*, and let* $\mathbf{B}$ *be a subalgebra of* $\mathbf{A}$*. Then each of the following hold.*

(a) $\theta \restriction B$ *is a congruence relation of* $\mathbf{B}$*.*

(b) θB *is a subuniverse of* $\mathbf{A}$*.*

(c) $\boldsymbol{\theta}\mathbf{B}/(\theta \restriction \theta B) \cong \mathbf{B}/\theta \restriction B$*.*

Proof. For part (a) we have to see that $\theta \restriction B$ is an equivalence relation on B and that it has the substitution property. Hard-working graduate students will check that it is indeed an equivalence relation. To see that the substitution property holds, let Q be an operation symbol. Just for simplicity, let us suppose the rank of Q is 3. Pick $a, a', b, b', c, c' \in B$ so that

$$\begin{aligned} a &\equiv a' \mod \theta \restriction B \\ b &\equiv b' \mod \theta \restriction B \\ c &\equiv c' \mod \theta \restriction B. \end{aligned}$$

We must show that $Q^{\mathbf{B}}(a,b,c) \equiv Q^{\mathbf{B}}(a',b',c') \mod \theta \restriction B$. Because all those elements come from B we see that

$$\begin{aligned} a &\equiv a' \mod \theta \\ b &\equiv b' \mod \theta \\ c &\equiv c' \mod \theta, \end{aligned}$$

and that both $Q^{\mathbf{B}}(a,b,c) = Q^{\mathbf{A}}(a,b,c)$ and $Q^{\mathbf{B}}(a',b',c') = Q^{\mathbf{A}}(a',b',c')$. It follows from the substitution property for θ that $Q^{\mathbf{A}}(a,b,c) \equiv Q^{\mathbf{A}}(a',b',c') \mod \theta$. But since both $Q^{\mathbf{A}}(a,b,c) = Q^{\mathbf{B}}(a,b,c) \in B$ and $Q^{\mathbf{A}}(a',b',c') = Q^{\mathbf{B}}(a',b',c') \in B$, we draw the desired conclusion that $Q^{\mathbf{B}}(a,b,c) \equiv Q^{\mathbf{B}}(a',b',c') \mod \theta \restriction B$.

For part (b) we have to show that θB is closed under all the basic operations of $\mathbf{A}$. So let Q be an operation symbol, which without loss of generality we assume to have rank 3. Let $a,b,c \in \theta B$. Our goal is to show that $Q^{\mathbf{A}}(a,b,c) \in \theta B$. Using the definition of θB pick $a',b',c' \in B$ so that

$$\begin{aligned} a &\equiv a' \mod \theta \\ b &\equiv b' \mod \theta \\ c &\equiv c' \mod \theta. \end{aligned}$$

Because B is a subuniverse, we see that $Q^{\mathbf{A}}(a',b',c') \in B$. Because θ is a congruence, we see that $Q^{\mathbf{A}}(a,b,b) \equiv Q^{\mathbf{A}}(a',b',c')$. Putting these together, we find that $Q^{\mathbf{A}}(a,b,c) \in \theta B$, as desired.

For part (c) we will invoke the Homomorphism Theorem. Define the map h from B to $\theta B/(\theta \restriction \theta B)$ via

$$h(b) := b/(\theta \restriction \theta B).$$

We have three contentions, namely that h is a homomorphism, that h is onto $B/(\theta \restriction \theta B)$, and that the functional kernel of h is $\theta \restriction B$. Given these, the Homomorphism Theorem provides that desired isomorphism.

To see that h is a homomorphism we have to show it respects the operations. So again take Q be to an operation symbol, of rank 3 for simplicity. Let $a,b,c \in B$. Now observe

$$\begin{aligned} h(Q^{\mathbf{B}}(a,b,c)) &= Q^{\mathbf{B}}(a,b,c)/(\theta \restriction \theta B) \\ &= Q^{\boldsymbol{\theta}\mathbf{B}}(a,b,c)/(\theta \restriction \theta B) \\ &= Q^{\boldsymbol{\theta}\mathbf{B}/(\theta\restriction\theta B)}(a/(\theta \restriction \theta B), b/(\theta \restriction \theta B), c/(\theta \restriction \theta B)) \\ &= Q^{\boldsymbol{\theta}\mathbf{B}/(\theta\restriction\theta B)}(h(a),h(b),h(c)). \end{aligned}$$

In this way we see that h respects Q. So h is a homomorphism.

To see that h is onto, let $b' \in \theta B$. Pick $b \in B$ so that $b' \equiv b \mod \theta$. We assert that $h(b) = b'/(\theta \restriction \theta B)$. So what we have to demonstrate is that

$$b/(\theta \restriction \theta B) = b'/(\theta \restriction \theta B)$$

or what is the same

$$b \equiv b' \mod \theta \restriction \theta B.$$

Now both b and b' belong to θB, so all that remains is to see that $b \equiv b' \mod \theta$. But we already know this.

Finally, we have to understand the functional kernel of h. Let $a,b \in B$ and observe

$$\begin{aligned} h(a) = h(b) &\Leftrightarrow a/(\theta \restriction \theta B) = b/(\theta \restriction \theta B) \\ &\Leftrightarrow a \equiv b \mod \theta \restriction \theta B \\ &\Leftrightarrow a \equiv b \mod \theta \restriction B. \end{aligned}$$

The last equivalence follows since a and b both belong to B. So we see that $\theta \restriction B$ is the functional kernel of h, completing the proof. □

Let $\mathbf{A}$ be an algebra and let θ and φ be congruences of $\mathbf{A}$ with $\theta \subseteq \varphi$. Let

$$\varphi/\theta := \{(a/\theta, a'/\theta) \mid a, a' \in A \text{ with } a \equiv a' \mod \varphi\}.$$

So φ/θ is a two place relation on A/θ.

The Third Isomorphism Theorem. *Let* $\mathbf{A}$ *be an algebra and let* θ *and* φ *be congruences of* $\mathbf{A}$ *with* $\theta \subseteq \varphi$. *Then*

(a) φ/θ *is a congruence of* $\mathbf{A}/\theta$, *and*

(b) $(\mathbf{A}/\theta)/(\varphi/\theta) \cong \mathbf{A}/\varphi$.

Proof. Define the function h from A/θ to A/φ so that for all $a \in A$ we have

$$h(a/\theta) := a/\varphi.$$

Here we have to worry again whether h is really a function—the definition above uses a representative element a of the congruence class a/θ to say how to get from the input to the output. What if $a/\theta = a'/\theta$? Then $(a, a') \in \theta$. Since $\theta \subseteq \varphi$, we get $(a, a') \in \varphi$. This means, of course, that $a/\varphi = a'/\varphi$. So we arrive at the same output, even using different representatives. This means our definition is sound.

Let us check that h is a homomorphism. So let Q be an operation symbol, which we suppose has rank 3 just in order to avoid a lot of indices. Pick $a, b, c \in A$. Now observe

$$\begin{aligned} h(Q^{\mathbf{A}/\theta}(a/\theta), b/\theta, c/\theta) &= h(Q^{\mathbf{A}}(a,b,c)/\theta) \\ &= Q^{\mathbf{A}}(a,b,c)/\varphi \\ &= Q^{\mathbf{A}/\varphi}(a/\varphi, b/\varphi, c/\varphi) \\ &= Q^{\mathbf{A}/\varphi}(h(a/\theta), h(b/theta), h(c/\theta)) \end{aligned}$$

In this way we see that h respects the operation symbol Q. We conclude that h is a homomorphism.

Notice that h is onto A/φ since any member of that set has the form a/φ for some $a \in A$. This means that $h(a/\theta) = a/\varphi$.

Now lets tackle the functional kernel of h. Let $a, b \in A$. Then observe

$$h(a/\theta) = h(b/\theta) \Leftrightarrow a/\varphi = b/\varphi \Leftrightarrow a \equiv b \mod \varphi.$$

So $(a/\theta, b/\theta)$ belongs to the functional kernel of h if and only if $a \equiv b \mod \varphi$. That is, the functional kernel of h is φ/θ. From the Homomorphism Theorem we see that φ/θ is a congruence of $\mathbf{A}/\theta$. Also from the Homomorphism Theorem we conclude that $(\mathbf{A}/\theta)/(\varphi/\theta) \cong \mathbf{A}/\varphi$. □

The set inclusion relation $\subseteq$ is a partial ordering of the congruence relations of an algebra $\mathbf{A}$. Some of the secrets of about $\mathbf{A}$ can be discovered by understanding how the congruence relations are ordered. The next theorem, sometimes called the Fourth Isomorphism Theorem, is a first and useful step along this road. To understand it we need the notion of isomorphism of relational structures (as opposed to algebras). Let A and B be nonempty sets and let $\sqsubseteq$ be a two-place

relation on A and $\preceq$ be a two-place relation on B. A function h from A to B is called an **isomorphism** between $\langle A, \sqsubseteq\rangle$ and $\langle B, \prec\rangle$ provided h is one-to-one, h is onto B, and for all $a, a' \in A$ we have

$$a \sqsubseteq a' \text{ if and only if } h(a) \preceq h(a').$$

As a matter of notation, let $\mathrm{Con}\mathbf{A}$ be the set of congruence relations of $\mathbf{A}$.

The Correspondence Theorem. *Let $\mathbf{A}$ be an algebra and let θ be a congruence of $\mathbf{A}$. Let $P = \{\varphi \mid \varphi \in \mathrm{Con}\mathbf{A}$ and $\theta \subseteq \varphi\}$. Then the map from P to $\mathrm{Con}\mathbf{A}/\theta$ that sends each $\varphi \in P$ to φ/θ is an isomorphism between the ordered set $\langle P, \subseteq\rangle$ and the ordered set $\langle \mathrm{Con}\mathbf{A}/\theta, \subseteq\rangle$.*

Proof. Let Ψ denote the map mentioned in the theorem. So

$$\Psi(\varphi) = \varphi/\theta$$

for all $\varphi \in \mathrm{Con}\mathbf{A}$ with $\theta \subseteq \varphi$.

To see that Ψ is one-to-one, let $\varphi, \rho \in \mathrm{Con}\mathbf{A}$ with $\theta \subseteq \varphi$ and $\theta \subseteq \rho$. Suppose that $\Psi(\varphi) = \Psi(\rho)$. This means $\varphi/\theta = \rho/\theta$. Now consider for all $a, a' \in A$

$$\begin{aligned}(a, a') \in \varphi &\Leftrightarrow (a/\theta, a'/\theta) \in \varphi/\theta \\ &\Leftrightarrow (a/\theta, a'/\theta) \in \rho/\theta \\ &\Leftrightarrow (a, a') \in \rho\end{aligned}$$

So $\varphi = \rho$. Notice that the first equivalence depends of $\theta \subseteq \varphi$ while the last depends of $\theta \subseteq \rho$. We see that Ψ is one-to-one.

To see that Ψ is onto $\mathrm{Con}\mathbf{A}/\theta$, let μ be a congruence of $\mathbf{A}/\theta$. Define

$$\varphi := \{(a, a') \mid a, a' \in A \text{ and } (a/\theta, a'/\theta) \in \mu\}.$$

This two-place relation is our candidate for a preimage of μ. First we need to see that φ is indeed a congruence of $\mathbf{A}$. The checking of reflexivity, symmetry, and transitivity is routine. To confirm the substituion property, let Q be an operation symbol (with the harmless assumption that its rank is 3). Pick $a, a', b, b', c, c' \in A$ so that

$$\begin{aligned}a &\equiv a' \mod \varphi \\ b &\equiv b' \mod \varphi \\ c &\equiv c' \mod \varphi.\end{aligned}$$

We must see that $Q^{\mathbf{A}}(a, b, c) \equiv Q^{\mathbf{A}}(a', b', c') \mod \varphi$. From the three displayed conditions we deduce

$$\begin{aligned}a/\theta &\equiv a'/\theta \mod \mu \\ b/\theta &\equiv b'/\theta \mod \mu \\ c/\theta &\equiv c'/\theta \mod \mu.\end{aligned}$$

Because μ is a congruence of $\mathbf{A}/\theta$, we obtain

$$Q^{\mathbf{A}/\theta}(a/\theta, b/\theta, c/\theta) \equiv Q^{\mathbf{A}/\theta}(a'/\theta, b'/\theta, c'/\theta) \mod \mu.$$

But given how the operations work in quotient algebras, this gives

$$Q^{\mathbf{A}}(a,b,c)/\theta \equiv Q^{\mathbf{A}}(a',b',c')/\theta \mod \mu.$$

Then the definition of φ supports the desired conclusion that $Q^{\mathbf{A}}(a,b,c) \equiv Q^{\mathbf{A}}(a',b',c') \mod \varphi$. So φ is a congruence of $\mathbf{A}$. But we also need to see that $\theta \subseteq \varphi$ to get that $\varphi \in P$. So suppose that $a, a' \in A$ with $(a,a') \in \theta$. Then $a/\theta = a'/\theta$. This entails that $(a/\theta, a'/\theta) \in \mu$ since μ is reflexive. In this way, we see that $(a,a') \in \varphi$. So $\theta \subseteq \varphi$ and $\varphi \in P$. Now consider

$$\begin{aligned}
\Psi(\varphi) &= \varphi/\theta \\
&= \{(a/\theta, a'/\theta) \mid a, a' \in A \text{ and } (a,a') \in \varphi\} \\
&= \{(a/\theta.a'/\theta) \mid a, a' \in A \text{ and } (a/\theta, a'/\theta) \in \mu\} \\
&= \mu.
\end{aligned}$$

In this way, we see that Ψ is onto $\mathrm{Con}\mathbf{A}/\theta$.

Last, we need to show that Ψ respects the ordering by set inclusion. So let $\varphi, \rho \in \mathrm{Con}\mathbf{A}$ with $\theta \subseteq \varphi$ and $\theta \subseteq \rho$. Let us first suppose that $\varphi \subseteq \rho$. To see that $\Psi(\varphi) \subseteq \Psi(\rho)$, let $a, a' \in A$ and notice

$$\begin{aligned}
(a/\theta, a'/\theta) \in \Psi(\varphi) &\implies (a/\theta, a'/\theta) \in \varphi/\theta \\
&\implies (a,a') \in \varphi \\
&\implies (a,a') \in \rho \\
&\implies (a/\theta, a'/\theta) \in \rho/\theta \\
&\implies (a/\theta, a'/\theta) \in \Psi(\rho)
\end{aligned}$$

So we find if $\varphi \subseteq \rho$, then $\Psi(\varphi) \subseteq \Psi(\rho)$. For the converse, suppose $\Psi(\varphi) \subseteq \Psi(\rho)$. Let $a, a' \in A$ and notice

$$\begin{aligned}
(a,a') \in \varphi &\implies (a/\theta, a'/\theta) \in \varphi/\theta \\
&\implies (a/\theta, a'/\theta) \in \Psi(\varphi) \\
&\implies (a/\theta, a'/\theta) \in \Psi(\rho) \\
&\implies (a/\theta, a'/\theta) \in \rho/\theta \\
&\implies (a,a') \in \rho.
\end{aligned}$$

So we find that if $\Psi(\varphi) \subseteq \Psi(\rho)$, then $\varphi \subseteq \rho$. So we have for all $\varphi, \rho \in P$

$$\varphi \subseteq \rho \text{ if and only if } \Psi(\varphi) \subseteq \Psi(\rho).$$

Finally, we can conclude that Ψ is an isomorphism between our two ordered sets of congruences. □

1.1 Problem Set 1

Algebra Homework, Edition 1
Second Week
Just Some General Notions

Problem 3.
Prove that the congruence relations of **A** are exactly those subuniverses of $\mathbf{A} \times \mathbf{A}$ which happen to be equivalence reations on A.

Problem 4.
Prove that the homomorphisms from **A** to **B** are exactly those subuniverses of $\mathbf{A} \times \mathbf{B}$ which are functions from A to B.

Problem 5.
Prove that the projection functions associated with $\mathbf{A} \times \mathbf{B}$ are homomorphisms.

LECTURE 2

COMPREHENDING PLUS AND TIMES

2.1 WHAT A RING IS

The notion of a ring arose in the nineteenth century by generalizing a collection of specific algebraic systems built around various examples of addition and multiplication. Certainly our understanding of addition and multiplication of positive integers is very old. Eudoxus of Cnidus, a contemporary of Plato, put—in modern terms—the notions of addition and multiplication of positive real numbers on a sound basis. His work can be found in Book V of Euclid's elements. Negative numbers emerged in India and China about the time of Archimedes, but met with little welcome in the Hellenistic world. This attachment of mathematical illegitimacy to negative numbers persisted in Europe into the eighteenth century. However, by the end of the eighteenth century not only negative real numbers but complex numbers in general were well in hand. Euler was a master of it all.

In the nineteenth century we had algebraic systems built around addition of multiplication of all of the following:

- integers
- rational numbers
- real numbers
- complex numbers
- algebraic numbers
- constructible numbers
- $n \times n$ matrices with entries selected from the systems listed above.
- polynomials with coefficients selected from the systems listed above.

- functions from the reals to the reals (and similarly with the reals replaced by some other systems)
- many other examples of addition and multiplication

As that century progressed, mathematicians realized that to develop the theories of each of these particular cases, one had to duplicate more or less a lot of effort. The examples had many properties is common. So it was a matter of convenience to develop the details of many of these common properties just once, before pursuing the more specialized theory of, say, the complex numbers. This led to the notion of a ring.

The signature we use to present this notion consists of a two-place operation symbol $\cdot$ to name multiplication, a two-place operation symbol $+$ to name addition, a one-place operation symbol $-$ to denote the formation of negatives, and two constant symbols 0 and 1. A **ring** is an algebraic system of this signature in which the following equations hold true.

$$\begin{aligned} x+(y+z) &= (x+y)+z & x\cdot(y\cdot z) &= (x\cdot y)\cdot z \\ x+0 &= x & x\cdot 1 &= x \\ x+y &= y+x & 1\cdot x &= x \\ -x+x &= 0 & x\cdot(y+z) &= x\cdot y+x\cdot z \\ & & (x+y)\cdot z &= x\cdot z+y\cdot x \end{aligned}$$

This collection of equations is some times called the axioms of ring theory.

You see here the familiar associative, commutative, and distributive laws, as well as equations giving the behavior of 0 and 1. It is important to realize that while the commutative law for addition is included, the commutative law for multiplication is not. The absence of the commutative law for multiplication has compelled me to include two forms of the distributive law as well as two equations to capture the behavior of 1. The ring a 2×2 matrices with real entries is an example of a ring where the commutative law for multiplication fails. A ring in which the commutative law for multiplication holds as well is called a **commutative ring**. While there is a rich theory of rings in general, in our course almost all rings will be commutative rings.

Because the axioms of ring theory are all equations it is easy to see that every subalgebra of a ring must be a ring itself, that every homomorphic image of a ring must also be a ring, and that the direct product of any system of rings is again a ring. Because the commutative law for multiplication is also an equations, the same observations apply to commutative rings.

You should also realize that in a ring the elements named by 0 and 1 might be the same. In this event, by way of a fun exercise, you can deduce from the ring axioms that such a ring can have only one element. Evidently, all one-element rings are isomorphic and, of themselves, not very interesting. They are called *trivial rings*.

According to the definition above, every ring must have an element named by the constant symbol 1 and this element must behave as described by the equations in our list. This has been the most common convention since the 1970's. However, some considerable part of the older literature and some of the contemporary literature use a different somewhat wider notion that lacks the constant symbol 1. For example, the even integers under ordinary addition and multiplication would constitute a ring in this manner, but not in the sense that I have but forward here. In that style of exposition, what we have called "rings" are referred to as "rings with unit". Nathan Jacobson, one of the great ring theorists of the twentieth century used the notion of ring I have adopted and referred to these other old-fashion algebraic systems as "rngs".

2.2 Congruences and Ideals on Rings

Let **R** be a ring and let θ be a congruence on **R**. Recall that

$$0/\theta = \{a \mid a \in R \text{ and } a \equiv 0 \mod \theta\}$$

is the θ-congruence class containing 0. Observe that the set $0/\theta$ has each of the following properties

(a) $0 \in 0/\theta$.

(b) If $a, b \in 0/\theta$, then $a + b \in 0/\theta$.

(c) if $a \in 0/\theta$ and $r \in R$, then $ra, ar \in 0/\theta$.

To obtain (b) reason as follows

$$\begin{aligned} a &\equiv 0 \mod \theta \\ b &\equiv 0 \mod \theta \\ a+b &\equiv 0+0 \mod \theta \\ a+b &\equiv 0 \mod \theta \end{aligned}$$

The third step uses the key substitution property of congruence relations, whereas the fourth step use the equation $0 + 0 = 0$, which follows easily from the ring axioms.

To obtain (c) reason as follows

$$\begin{aligned} a &\equiv 0 \mod \theta \\ r &\equiv r \mod \theta \\ ar &\equiv 0r \mod \theta \\ ar &\equiv 0 \mod \theta \end{aligned}$$

The second step uses the fact that congruence relations, being special equivalence relations, are reflexive. The last step uses the equation $0x = 0$, which can be deduced from the ring axioms. A similar line of reasoning produces the conclusion

$$ra \equiv 0 \mod \theta.$$

Any subset $I \subseteq R$ that has the three attributes listed above for $0/\theta$ is called an **ideal** of the ring **R**. This means that I is an ideal of **R** if and only if

(a) $0 \in I$.

(b) If $a, b \in I$, then $a + b \in I$.

(c) if $a \in I$ and $r \in R$, then $ra, ar \in I$.

So we have taken the definition of ideal to allow us the observe that in any ring **R**

If θ is a congruence relation of **R**, then $0/\theta$ is an ideal of **R**.

That is, every congruence relation gives rise to an ideal.

The converse is also true. Let $\mathbf{R}$ be a ring and let I be an ideal of $\mathbf{R}$. Define

$$\theta_I := \{(a,b) \mid a,b \in R \text{ and } a-b \in I\}.$$

The eager graduate students should check that θ_I is indeed a congruence relation of $\mathbf{R}$. Actually, the theorem below tells a fuller tale and its proof, which only requires pursuing all the definitions involved, is left to delight the graduate students.

Theorem on Ideals and Congruences. *Let* $\mathbf{R}$ *be any ring, let* θ *be a congruence relation of* $\mathbf{R}$ *and let* I *be any ideal of* $\mathbf{R}$. *All of the following hold.*

(a) $0/\theta$ *is an ideal of* $\mathbf{R}$.

(b) θ_I *is a congruence relation of* $\mathbf{R}$.

(c) $I = 0/(\theta_I)$.

(d) $\theta = \theta_{0/\theta}$.

(e) *The collection of all ideals of* $\mathbf{R}$ *is ordered by* $\subseteq$ *and the map* $I \mapsto \theta_I$ *is an isomorphism of the ordered set of all ideals of* $\mathbf{R}$ *with the ordered set of all congruence relations of* $\mathbf{R}$.

The significance of this theorem is that when dealing with rings we can replace the study of congruence relations with the study of ideals. After all, each congruence θ is a set of ordered pairs, that is $\theta \subseteq R \times R$.; whereas each ideal I is merely a set of elements of R, that is $I \subseteq R$. Of course, there are places, in ring theory, where congruence relations are more convenient than ideals, so we need to remember both.

Here is some notation for using ideals in place of congruences. Let $\mathbf{R}$ be any ring and let θ and I be a congruence relation and and ideal that correspond to each other, let $a,b \in R$ and let J be an ideal of $\mathbf{R}$ so that $I \subseteq J$.

$$\begin{aligned} \mathbf{R}/I &:= \mathbf{R}/\theta \\ a+I &:= a/\theta = \{a+b \mid b \in I\} \\ J/I &:= \{b+I \mid b \in J\} = 0/(\theta_J/\theta_I) \\ a \equiv b \mod I \text{ means } & a \equiv b \mod \theta \end{aligned}$$

The graduate students should work out the details to see that these conventions really do the job. Incidentally, the notation $a+I$ is a special case of $U+V := \{u+v \mid u \in U \text{ and } v \in V\}$, where $U < V \subseteq R$.

Suppose that $\mathbf{R}$ is a ring and $h : \mathbf{R} \to \mathbf{S}$ is a homomorphism. The **kernel** of h is the following set

$$\ker h := \{a \mid a \in R \text{ and } h(a) = 0\}.$$

The graduate students should check that if θ denotes that functional kernel of h, then

$$\ker h = 0/\theta.$$

So $\ker h$ is an ideal of $\mathbf{R}$ and the congruence corresponding to this ideal is the functional kernel of h.

2.3 The Isomorphism Theorems for Rings

With this sort of lexicon in hand, all the isomorphism theorems can be rendered into ring theoretic versions, with no need for further proofs. Here they are.

The Homomorphism Theorem, Ring Version. *Let* $\mathbf{R}$ *be a ring, let* $f : \mathbf{R} \twoheadrightarrow \mathbf{S}$ *be a homomorphism from* $\mathbf{R}$ *onto* $\mathbf{S}$*, and let* I *be an ideal of* $\mathbf{R}$*. All of the following hold.*

(a) *The kernel of* f *is an ideal of* $\mathbf{R}$*.*

(b) $\mathbf{R}/I$ *is a ring.*

(c) *The map* η *that assigns to each* $a \in R$ *the congruence class* $a + I$ *is a homomorphism from* $\mathbf{R}$ *onto* $\mathbf{A}/I$ *and its kernel is* I*.*

(d) *If* I *is thel kernel of* f*, then there is an isomorphism* g *from* $\mathbf{R}/I$ *to* $\mathbf{S}$ *such that* $f = g \circ \eta$*.*

The Second Isomorphism Theorem, Ring Version. *Let* $\mathbf{R}$ *be a ring, let* I *be an ideal of* $\mathbf{R}$*, and let* $\mathbf{S}$ *be a subring of* $\mathbf{R}$*. Then each of the following hold.*

(a) $I \cap S$ *is an ideal of* $\mathbf{S}$*.*

(b) $I + S$ *is a subuniverse of* $\mathbf{R}$*.*

(c) $\mathbf{I} + \mathbf{S}/I \cong \mathbf{S}/I \cap S$*.*

The Third Isomorphism Theorem, Ring Version. *Let* $\mathbf{R}$ *be a ring and let* I *and* J *be ideals of* $\mathbf{R}$ *with* $I \subseteq J$*. Then*

(a) J/I *is an ideal of* $\mathbf{R}/I$*, and*

(b) $(\mathbf{R}/I)/(J/I) \cong \mathbf{R}/J$*.*

The Correspondence Theorem, Ring Version. *Let* $\mathbf{R}$ *be a ring and let* I *be an ideal of* $\mathbf{R}$*. Let* $P = \{J \mid J \text{ is an ideal of } \mathbf{R} \text{ and } I \subseteq J\}$*. Then the map from* P *to the ordered set of ideals of* $\mathbf{R}/I$ *that sends each* $J \in P$ *to* J/I *is an isomorphism between the ordered set* $\langle P, \subseteq \rangle$ *and the ordered set of ideals of* $\mathbf{R}/I$*.*

2.4 Dealing with Ideals

Let $\mathbf{R}$ be a ring. Then R and $\{0\}$ will be ideals of $\mathbf{R}$. (They might be the same ideal, but only if $\mathbf{R}$ is a one-element ring. By a **proper ideal** of $\mathbf{R}$ we mean one that is different from R. By a **nontrivial ideal** we mean one that is different from $\{0\}$. The collection of all ideals of $\mathbf{R}$ is ordered by $\subseteq$. Under this ordering, $\{0\}$ is the unique least ideal and R is the unique largest ideal.

Let $\mathbf{R}$ be a ring and let $\mathcal{K}$ be any nonempty collection of ideals of $\mathbf{R}$. It is a routine exercise (why not put pen to paper?) that $\bigcap \mathcal{K}$ is also an ideal of $\mathbf{R}$ and this ideal is the greatest (in the sense of $\subseteq$) ideal included in every ideal belonging to $\mathcal{K}$. So every nonempty collection of ideals has a greatest lower bound in the ordered set of ideals. Let $W \subseteq R$ and take $\mathcal{K} = \{I \mid I \text{ is an ideal of } \mathbf{R} \text{ and } W \subseteq I\}$. Then $\bigcap \mathcal{K}$ is the smallest ideal of $\mathbf{R}$ that includes W. This ideal is denoted by (W) and is called the **ideal generated by** W.

Unlike the situation with intersection, when $\mathscr{K}$ is a nonempty collection of ideals of the ring $\mathbf{R}$ it is usually not the case that the union $\bigcup\mathscr{K}$ will turn out to be an ideal. However, $(\bigcup\mathscr{K})$ will be an ideal—indeed, it is the least ideal in the ordered set of ideals that includes every ideal in $\mathscr{K}$.

So the collection of all ideals of any ring is an ordered set with a least member, a great member, and every nonempty collection of ideals has both a greatest lower bound and least upper bound. Such ordered sets are called **complete lattice-ordered sets**.

While in general the union of a collection of ideals is unlikely to be an ideal, there are collections for which the union is an ideal. A collection $\mathscr{K}$ of ideals is said to be **updirected** provided if $I, J \in \mathscr{K}$, then there is $K \in \mathscr{K}$ so that $I \subseteq K$ and $J \subseteq K$.

Theorem 2.4.1. *Let* $\mathbf{R}$ *be a ring and let* $\mathscr{K}$ *be a nonempty updirected collection of ideals of* $\mathbf{R}$. *Then* $\bigcup\mathscr{K}$ *is an ideal of* $\mathbf{R}$.

Proof. First observe that $0 \in \bigcup\mathscr{K}$, since $\mathscr{K}$ is nonempty and every ideal must contain 0.

Now suppose that $a, b \in \bigcup\mathscr{K}$. Pick $I, J \in \mathscr{K}$ so that $a \in I$ and $b \in J$. Because $\mathscr{K}$ is updirected, pick $K \in \mathscr{K}$ so that $I \cup J \subseteq K$. So $a, b \in K$. Because K is an ideal, we see $a + b \in K \subseteq \bigcup\mathscr{K}$.

Finally, suppose $a \in \bigcup\mathscr{K}$ and $r \in R$. Pick $I \in \mathscr{K}$ so that $a \in I$. Then $ar, ra \in I$ since I is an ideal. Hence $ar, ra \in \bigcup\mathscr{K}$

In this way, we see that $\bigcup\mathscr{K}$ is an ideal. □

One kind of updirected set is a chain. The collection $\mathscr{C}$ is a chain of ideals provided for all $I, J \in \mathscr{C}$ either $I \subseteq J$ or $J \subseteq I$. As a consequence, we see that the union of any nonempty chain of ideals is again an ideal.

A little reflection shows that this result is not particularly ring theoretic. In fact, for algebras generally the union of any updirected collection of congruence relations is again a congruence relation.

Now let $\mathbf{R}$ and a ring and $W \subseteq R$. The ideal (W) that is generated by W was defined in what might be called a shrink wrapped manner as the intersection of all the ideals containing W. It is also possible to describe this ideal by building it up from W is stages using the following recursion.

$$\begin{aligned} W_0 &:= W \cup \{0\} \\ W_{n+1} &:= W_n \cup \{ra \mid r \in R \text{ and } a \in W_n\} \cup \{ar \mid r \in R \text{ and } a \in W_n\} \cup \{a + b \mid a, b \in W_n\} \\ &\qquad \text{for all natural numbers } n. \end{aligned}$$

Notice $W \subseteq W_0 \subseteq W_1 \subseteq W_2 \subseteq \ldots$ and each set along this chain repairs potential failures of the earlier sets along the chain to be ideals. It does this by adding new elements. Unfortunately, these new elements, while they repair earlier failures may introduce failures of their own. For this reason the construction continues through infinitely many stages. Now let $W_\omega := \bigcup_{n\in\omega} W_n$ be the union of this chain of sets. Our expectation is that all the failures have been fixed and that W_ω is an ideal. The eager graduate students are invited to write out a proof of this. But more is true. Actually, $W_\omega = (W)$. Here are some suggestions for how to prove this. To establish $W_\omega \subseteq (W)$ prove by induction on n that $W_n \subseteq I$ for every ideal I that includes W. Observe that $(W) \subseteq W_\omega$ once we know that W_ω is an ideal that includes W.

This process that shows that shrink wrapping and building up from the inside works not only here in the context of ideals, but in several other contexts as well.

A more transparent version of the building up from the inside is available in our particular context. By a *combination* of W over $\mathbf{R}$ we mean an element of the form

$$r_0 w_0 s_0 + r_1 w_1 s_1 + \cdots + r_{n-1} w_{n-1} s_{n-1}$$

where n is a natural number, $r_0, s_0, r_1, s_1, \ldots, r_{n-1}, s_{n-1} \in R$, and $w_0, w_1, \ldots, w_{n-1} \in W$. In case $n = 0$, we take the element represented to be the zero of the ring. It is straightforward, with the help of the distributive laws, to see that the set of all combinations of W over $\mathbf{R}$ is an ideal that includes the subset W. An induction on the length of combinations shows that all these combinations belong to (W). So the set of all combinations of W over $\mathbf{R}$ must be the ideal (W) generated be W. In commutative rings it is only necessary to consider combinations of the form

$$r_0 w_0 + r_1 w_1 + \cdots + r_{n-1} w_{n-1}.$$

In particular, if $\mathbf{R}$ is commuative , $w \in R$, and I is an ideal of $\mathbf{R}$, then the ideal $(\{w\} \cup I)$ generated by the element w and the ideal I consists of all elements of the form

$$r w + u \text{ where } r \in R \text{ and } u \in I.$$

2.5 Problem Set 2

ALGEBRA HOMEWORK, EDITION 2
THIRD WEEK
PRIME IDEALS

PROBLEM 6.

(a) Let I and J be ideals of a commutative ring $\mathbf{R}$ with $I + J = R$. Prove that $IJ = I \cap J$.

(b) Let I, J, and K be ideals of a principal ideal domain. Prove that $I \cap (J + K) = I \cap J + I \cap K$.

PROBLEM 7.
Let $\mathbf{R}$ be a commutative ring and I be a proper prime ideal of $\mathbf{R}$ such that $\mathbf{R}/I$ satisfies the descending chain condition on ideals. Prove that $\mathbf{R}/I$ is a field.

PROBLEM 8.
Let $\mathbf{R}$ be a commutative ring and I be an ideal which is contained in a prime ideal P. Prove that the collection of prime ideals contained in P and containing I has a minimal member.

PROBLEM 9.
Let X be a finite set and let $\mathbf{R}$ be the ring of functions from X into the field $\mathbb{R}$ of real numbers. Prove that an ideal M of $\mathbf{R}$ is maximal if and only if there is an element $a \in X$ such that

$$M = \{f \mid f \in R \text{ and } f(a) = 0\}.$$

LECTURE 3

RINGS LIKE THE INTEGERS

3.1 INTEGRAL DOMAINS

The ring $\langle \mathbb{Z}, +, \cdot, -, 0, 1 \rangle$ of integers is one of the most familiar mathematical objects. Its investigation lies at the heart of number theory that, together with geometry, is among the oldest parts of mathematics. This ring is commutative and has a host of other very nice properties. Among these is that the product of any two nonzero integers must itself be nonzero. This property may fail, even in rings closely connected to the ring of integers. For example, let **R** be the direct square of the ring of integers. The elements of this ring will be ordered pairs of integers which the ring operations defined coordinatewise. That is

$$\begin{aligned} (a,b)+(c,d) &= (a+c, b+d) \\ (a,b)\cdot(c,d) &= (ac, bd) \\ -(a,b) &= (-a,-b) \end{aligned}$$

The zero of **R** is the pair $(0,0)$ while the unit (the one) is $(1,1)$. But observe that the product of $(1,0)$ with $(0,1)$ is $(1\cdot 0, 0\cdot 1) = (0,0)$.

A ring **D** is called an **integral domain** provided

(a) **D** is a commutative ring,

(b) 0 and 1 name different elements of D, and

(c) If $a, b \in D$ and $a \neq 0 \neq b$, then $ab \neq 0$.

Integral domains used to be called by a more charming name: domains of integrity. Condition (b) above is equivalent to the stipulation that integral domains must have at least two elements. Condition (c) can be replaced by either of the following conditions.

(c′) If $a, b \in D$ and $ab = 0$, then either $a = 0$ or $b = 0$.

(c″) If $a, b, c \in D$ with $a \neq 0$ and $ab = ac$, then $b = c$

Condition (c′) is just a contrapositive form of Condition (c). Condition c″) is the familiar cancellation law. The graduates student can find amusement by showing the equivalence of this condition.

While, as observed above, the direct product of a system of integral domains need not be an integral domain (is it ever?), every subring of an integral domain will be again an integral domain. What about homomorphic images of integral domains? Well, the trivial one-element ring is a homomorphic image of every ring, including every integral domain, and the trivial ring is not an integral domain. But suppose **D** is an integral domain and h is a homomorphism mapping **D** onto the nontrivial ring **S**. Must **S** be an integral domain? Certainly, conditions (a) and (b) hold for **S**. Consider a concrete example. Let I be the set of integers that are multiples of 4. It is easy to check that I is an ideal of the ring of integers. The quotient ring $\mathbb{Z}/I$ has just four elements:

$$0+I \qquad 1+I \qquad 2+I \quad \text{and} \quad 3+I.$$

In the quotient ring we have the product $(2+I)\cdot(2+I) = 2\cdot 2+I = 4+I = 0+I$. This violates condition (c) in the definition of integral domain. So while some homomorphic images of some integral domain will be integral domains, it is not true generally. Perhaps some property of the ideal I would ensure that the quotient ring is an integral domain.

Let **R** be a commutative ring and let I be an ideal of **R**. I is said to be a **prime ideal** provided

- I is a proper ideal of **R** [that is, $I \neq R$], and
- if $a, b \in R$ with $ab \in I$, then either $a \in I$ or $b \in I$.

The graduate students can prove the following theorem by chasing definitions.

Theorem 3.1.1. *Let* **R** *be a commutative ring and let* I *be an ideal of* **R**. **R**/I *is an integral domain if and only if* I *is a prime ideal of* **R**.

Suppose **R** is a ring. Consider the list of elements of R below:

$$1, 1+1, 1+1+1, 1+1+1+1, \ldots.$$

This looks like a list of the positive integers, but we mean something different. The element 1 is the unit of multiplication in **R** and + names the addition operation in **R**. The ring **R** may not contain any integers at all. The list above might even be finite, depending on the ring **R**. If the list is infinite we say that **R** has **characteristic** 0. If the list is finite, then (as pigeons know) two distinct members of this list must actually be the same element. That is

$$\underbrace{1+\cdots+1}_{n \text{ times}} = \underbrace{1+\cdots+1}_{n \text{ times}} + \underbrace{1+\cdots+1}_{k \text{ times}}$$

for some positive natural numbers n and k. This entails that

$$0 = \underbrace{1+\cdots+1}_{k \text{ times}}$$

for some positive natural number k. In this case, we say that the **characteristic** of **R** is the smallest such positive natural number. On reflection, it might have been better to say that rings of

characteristic 0 had infinite characteristic. However, the use of characteristic 0 for this notion is so well entrenched that we are stuck with it.

The characteristic of a ring **R** is a useful invariant of **R**. It will play a prominent role in the spring semester during our development of the theory of fields. Observe that every finite ring must have a characteristic that is not 0. Because 1 must belong to every subring of **R**, we see that all the subrings of **R** have the same characteristic as **R**. On the other hand, the homomorphic images of **R** may have characteristic differing from the characteristic of **R**. To begin with, trivial rings have characteristic 1 (these are the only rings of characteristic 1) and trivial rings are homomorphic images of every ring. The ring of integers has characteristic 0, but $\mathbb{Z}/(6)$ evidently has characteristic 6. On the other hand, it is easy to verify (do it, why not?) that the characteristic of a homomorphic image of **R** can be no larger than the characteristic of **R** (well, taking 0 to be larger than all the positive natural numbers...). We leave it to the eager graduate students to figure out the characteristic of $\mathbf{R} \times \mathbf{S}$ when the characteristic of **R** is r and the characteristic of **S** is s.

Here is a useful fact.

Fact. Let **D** be an integral domain. The characteristic of **D** is either 0 or it is a prime number.

We won't prove this, but here is a hint as to why an integral domain cannot have characteristic 6.

$$0 = 1+1+1+1+1+1 = 1\cdot(1+1+1)+1\cdot(1+1+1) = (1+1)\cdot(1+1+1).$$

3.2 Principal Ideal Domains

A route to a deeper understanding of the ring of integers is to investigate the congruence relations of this ring. This is the route chosen by Gauss in his 1801 masterpiece *Disquistiones Arithmeticæ*. Of course, we see that the investigation of congruences of a ring amounts to the investigation of its ideals. The notion of an ideal of a ring arose in the work of Kummer, Kronecker, and Dedekind in the second half of the nineteenth century to be refined still later by Hilbert and by Emmy Noether. Still, the discoveries of Gauss needed changes of only the most modest kind to fit with the later theoretical apparatus.

We begin with an important observation that surely must have been known to Euclid.

A Key Fact About the Integers. *Let d be any nonzero integer and let n be any integer. There are unique integers q and r satisfying the following constraints:*

(a) $n = qd + r$, *and*

(b) *Either* $r = 0$ *or* $0 < r < |d|$.

Graduate students with itchy fingers who turn their hands to this are advised that there are two things to show: the *existence* of integers q and r and the *uniqueness* of these integers. Here is a hint. Consider the set $\{|n - xd| \mid x \in \mathbb{Z}\}$. This is a set of natural numbers. It is nonempty (why?). Every nonempty set of natural numbers has a least element.

The uniquely determined integers q and r mentioned in this Key Fact are called the *quotient* of n upon division by d and the *remainder* of n upon division by d, respectively. We will also call r the *residue* of n upon division by d.

Let I be any nontrivial ideal of the ring of integers. Since I is not trivial, it must have a member other than 0 and, because I is an ideal, there must be a positive integer in I. Hence there must be

a least positive integer d in I. Now let $n \in I$ be chosen arbitrarily. Using the Key Fact, pick integers q and r so that

(a) $n = qd + r$, and

(b) Either $r = 0$ or $0 < r < |d|$.

Then $r = n - qd$. Notice that $n, d \in I$ because that's the way we chose them. So $r = n - qd \in I$ because I is an ideal. But $0 < r < |d| = d$ is impossible, by the minimality of the choice of d. So we conclude that $r = 0$ and therefore that n is a multiple of d. Thus

$$I = \{qd \mid q \in \mathbb{Z}\} = (d).$$

So we have the conclusion that every ideal of the ring of integers is generated by some one of its members (and, in fact, by the smallest postive integer belonging to the ideal if the ideal in not trivial).

A **principal ideal domain** is an integral domain for which every ideal is generated by some one of its members. In an arbitrary ring, we will say an ideal is **principal** provided it is generated by some one of its members. So a principal ideal domain is an integral domain for which every ideal is principal.

The ring of integers is a principal ideal domain. Many interesting properties of the ring of integers also hold for principal ideal domains in general. This includes the powerful Fundamental Theorem of Arithmetic:

> *Every nonzero integer, other than* 1 *and* -1*, can be written in a unique way as a product of primes.*

In order to formulate this result for rings more generally, we need to introduce some further notions.

A **unit** in a commutative ring is an element u such that there is an element v is the ring so that $uv = 1 = vu$. So a unit is just an element with a multiplicative inverse. The units of the ring of integers are just 1 and -1. (Notice the appearance of these numbers in the statement above.) Two elements a and b of a commutative ring are said to be **associates** provided $au = bu$ for some unit u. It is routine (and you know the routine when the word routine comes up in these notes...) to show that relation "is an associate of" is an equivalence relation on any commutative ring. We will use $a \sim b$ to denote that a and b are associates. Do you think $\sim$ is a congruence relation on the ring?

An element a of a commutative ring is said to be **irreducible** provided it is neither 0 nor a unit and if $a = bc$ for some elements b and c in the ring, then either b is a unit or c is a unit. So irreducible elements of a ring are the ones that cannot be factored, except in some trivial manner. (Observe that $2 = (-1) \cdot (-1) \cdot 1$ is a factorization of the integer 2 is such a trivial manner.).

An integral domain **D** is said to be a **unique factorization domain** provided

(a) Every nonzero nonunit in D can be expressed as a (finite) product of irreducibles.

(b) If m and n are natural numbers and $a_0, a_1, \ldots, a_{m-1} \in D$ and $b_0, b_1, \ldots, b_{n-1} \in D$ are irreducibles such that

$$a_0 a_1 \ldots a_{m-1} \sim b_0 b_1 \ldots b_{n-1},$$

then $m = n$ and there is a permutation σ of $\{0, 1, \ldots, m-1\}$ so that

$$a_i \sim b_{\sigma(i)} \text{ for all } i \text{ with } 0 \le i < m.$$

The point of the permutation σ is that we don't really want to consider $2 \cdot 3$ and $3 \cdot 2$ as distinct factorizations of 6. Observe that stipulation (a) asserts the existence of a factorization into irreducibles, while stipulation (b) asserts the uniqueness of such factorization.

The Fundamental Theorem of Arithmetic asserts that the ring of integers is a unique factorization domain. So is every principal ideal domain and that is what we tackle below.

You might wonder that we have used the word "irreducible" instead of "prime" in formulating these notions. (You might also be wondering now if prime ideals have anything to do with primes....) Euclid realized long ago that an irreducible (positive) integer p had the property

$$\text{If } p \mid ab, \text{ then either } p \mid a \text{ or } p \mid b.$$

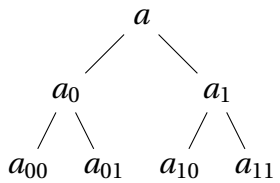
Here $p \mid a$ means that p divides a—-that is, $a = pc$ for some integer c. The divisibility relation, denoted by $\mid$, makes sense in any ring and we use it without further comment.

We will say that an element p of a commutative ring is **prime** provided p is neither 0 nor a unit and for all a and b in the ring

$$\text{If } p \mid ab, \text{ then either } p \mid a \text{ or } p \mid b.$$

The **primeness condition** is just that every irreducible element is prime. Incidentally, the converse is always true in any integral domain: a prime element is always irreducible. Indeed, if a is prime and $a = bc$, then we see that either $a \mid b$ or $a \mid c$. Consider, for instance, the first alternative. Pick d so that $b = ad$. Then $a \cdot 1 = bc = adc$. Now cancel a (we are in a integral domain) to obtain $1 = dc$. This means that c is a unit. The second alternative is similar.

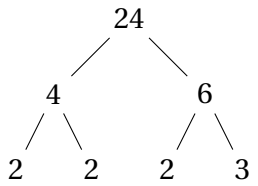
An attempt to factor a nonzero nonunit a into irreducibles might look like this:



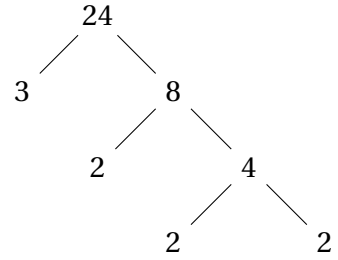
This tree represents two steps in an attempt to factor a.

$$a = a_0 a_1$$
$$a_0 = a_{00} a_{01}$$
$$a_1 = a_{10} a_{11}$$

So we have the factorization $a = a_{00} a_{01} a_{10} a_{11}$. The diagram displayed is a tree (look at it while standing on your head) with three levels. Each node branches into two succeeding nodes (except the nodes on the bottom level). This tree has four branches that start at the root (a) and extend down to the bottom level. Now our intention is that all the nodes should themselves be nonzero nonunits. So if we run into an irreducible then we will not attempt to factor it. Here is a tree showing a factorization of the integer 24.



Suppose we try again. Here is another way to factor 24.



These trees and their labellings reflect the actual processes of the factorizations. We see that they are not unique. But the irreducibles (counting how often they appear but not the order of their appearance) is unique. In each of these trees, every node has either 0 or 2 suceeding nodes, since multiplication is a two-place operation. In any case, each node has only finitely many nodes as immediate successors. We say the tree is *finitely branching*. There is a useful combinatorial fact about trees that comes into play.

König's Infinity Lemma. *Any finitely branching tree with infinitely many nodes must have an infinite branch.*

Proof. We can build the desired infinite branch by the following recursion.

Let a_0 be the root of the tree. There are only finitely many nodes immediately below a_0. Every node, apart from a_0 lies somewhere below a_0. Since the union of finitely many finite sets is always finite, there must be a node immediately below a_0 which itself has infinitely many nodes immediately below it. Let a_1 be such a node. Now apply the same reasoning to a_1 to obtain a node immediately below a_1 that is above infinitely many nodes. Continuing in this way, we obtain a branch $a_0, a_1, a_2, \ldots$ that is infinite. □

The graduate students should be a bit unhappy with the informality of this proof. For one thing, it describes an infinite process. For another it is not terribly specific about how to pick any of the nodes along the infinite branch, apart from a_0. Producing the infinite branch requires making infinitely many choices. These issues might be addressed in two stages. The first stage would secure the validity of definition by recursion. To see what is at issue consider the following familiar definition of the factorial function.

$$0! = 1$$
$$(n+1)! = n!(n+1) \text{ for all natural numbers } n$$

The issue is two-fold: first, is there any function, here indicated by !, that fulfills the two conditions laid out above? Second, is there exactly one such function? After all definitions should be, well, definite. Here is a slightly more general situation. Suppose that a is a member of some set U

and h is a function from $U \times \mathbb{N}$ into U. Is there exactly one function f from the natural numbers to U satisfying the following constraints?

$$f(0) = a$$
$$f(n+1) = h(f(n), n+1) \text{ for all natural numbers } n.$$

The answer to this question is YES. It is among the simplest cases of a theorem known as the Recursion Theorem. You might try to prove this—remember there is an existence part and a uniqueness part. Induction may help in your proof.

After securing some version of the Recursion Theorem in the first stage, the second stage of cleaning up König's Infinity Lemma is to remove the ambiguity about how to pick the "next element of the infinite branch". This amounts to producing a suitable function to play the role of h in your definition by recursion. Here is what you need h to accomplish. Call a node in the tree *good* provide there are infinitely many nodes beneath it. Given a good node c we see that the set of good nodes immediately beneath it is always a nonempty set. We want $h(c, n+1)$ to pick some element of this nonempty set. (In our case, h turns out not to depend on its second input.) Functions like h always exist. They are called choice functions.

A commutative ring has the **divisor chain condition** provided whenever $a_0, a_1, a_2, \ldots$ are elements of the ring so that $a_{k+1} \mid a_k$ for all natural numbers k, then there is a natural number n so that $a_n \sim a_{n+k}$ for all natural numbers k. This means that, ignoring the distinction between associates, every descending divisor chain is finite.

Theorem Characterizing Unique Factorization Domains. *Let* **D** *be an integral domain.* **D** *is a unique factorization domain if and only if* **D** *has both the primeness condition and the divisor chain condition.*

Proof. First, suppose that **D** has the divisor chain condition and the primeness condition. Let $a \in D$ be any nonzero nonunit. Consider any factorization tree with root a. This tree is finitely branching (in fact, the branching is bounded by 2) and it cannot have any infinite branch, according the the divisor chain condition. By König the factorization tree is finite. So we see that a can be written as a product of irreducibles.

Now let $a_0 \ldots a_{m-1} \sim b_0 b_1 \cdots b_{n-1}$ be products of irreducibles. We assume, without loss of generality, that $n \leq m$. We will deduce the required uniqueness by induction on m. Leaving in the hands of the capable graduate students the base step ($m = 0$) of the inductive argument, we turn to the inductive step. Let $m = k+1$. Now since a_k is irreducible, the primeness condition ensures that it is also prime. Evidently, $a_k \mid b_0 \ldots b_{n-1}$. A little (inductive) thought shows us that since a_k is prime there must be $j < n$ so that $a_k \mid b_j$. Since b_j is irreducible, we find that $a_k \sim b_j$. Using the cancellation law (we are in an integral domain!) we see that

$$a_0 a_1 \ldots a_{k-1} \sim b_0 \ldots b_{j-1} b_{j+1} \cdots b_{n-1}$$

or something easier if $j = n-1$. The left side as $k = m-1$ factors in the product whereas the right side has $n-1$ factors. Applying the induction hypothesis, we find that $m-1 = n-1$ (and hence that $m = n$) and we can pick a one-to-one map σ' from $\{0, 1, \ldots, m-2\}$ onto $\{0, 1, \ldots, j-1\} \cup \{j+1, \ldots, n-2\}$ so that

$$a_i \sim b_{\sigma'(i)} \text{ for all } i < m-1.$$

Now extend σ' to the set $\{0, 1, 2, \ldots, m-1\}$ by putting $\sigma(m-1) = j$. Then σ is a permutation of $\{0, 1, 2, \ldots, m-1\}$ that fulfills the uniqueness requirement. So **D** is a unqie factorization domain.

Second, suppose for the converse, that **D** is a unique factorization domain. Let us check the divisor chain condition. Let $\cdots \mid a_2 \mid a_1 \mid a_0 = a$ be a divisor chain that is proper in the sense that no entry in the chain is an associate of any other entry. We must show that this chain is finite. For i less than the length of our chain, pick b_{i+1} so that $a_i = b_{i+1}a_{i+1}$. (This will be a proper factorization with neither a_{i+1} nor b_{i+1} being units.) Let $a = c_0 \dots c_{n-1}$ be a factorization of a into irreducibles. Suppose, for contradiction, that our divisor chain has more than n entries. Notice

$$c_0 c_1 \dots c_{n-1} = a = b_0 b_1 \dots b_{n-1} b_n a_n.$$

Each of $b_0, \dots, b_n$ as well as a_n can be written as a product of irreducibles. Clearly the right side of the equation above has more factors that the left side. This violates the unique factorization property, providing the contradiction we seek. So we find that every unique factorization domain has the divisor chain condition.

To see that primeness condition, suppose $a, b, c \in D$ where a is irreducible and $a \mid bc$. Pick $d \in D$ so that $bc = ad$. Factor $b = b_0 \dots b_{m-1}, c = c_0 \dots c_{n-1}$ and $d = d_0 \dots d_{\ell-1}$ into irreducibles. This gives

$$b_0 \dots b_{m-1} c_0 \dots c_{n-1} = a d_0 \dots d_{\ell-1}$$

By the uniqueness of factorizations, there must be j so that either $a \sim b_j$ (and $j < m$) or $a \sim c_j$ (and $j < n$). In the first alternative, we get $a \mid b$ while in the second we get $a \mid c$. □

Example. The ring $\mathbb{Z}[\sqrt{-5}]$ is an integral domain that is not a unique factorization domain.

Proof. The ring $\mathbb{Z}[\sqrt{-5}]$ is, by definition, the smallest subring of the field $\mathbb{C}$ of complex numbers that includes $\mathbb{Z} \cup \{\sqrt{-5}\}$. Since it is a subring of a field it must be an integral domain. You probably see easily that

$$\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\}.$$

To see that $\mathbb{Z}[\sqrt{-5}]$ is not a unique factorization domain consider the following factorizations of 9.

$$9 = 3 \cdot 3 = (2 + \sqrt{-5})(2 - \sqrt{-5}).$$

What we need is to show that each of $3, 2 + \sqrt{-5}$, and $2 - \sqrt{-5}$ are irreducible and the none of these is an associate of any other of them. We do this with the help of a function $N : \mathbb{Z}[\sqrt{-5}] \to \mathbb{N}$ defined by

$$N(a + b\sqrt{-5}) = a^2 + 5b^2 \text{ for all integers } a \text{ and } b.$$

This function has the following nice properties.

- $N(0) = 0$.
- $N(1) = 1$.
- $N(rt) = N(r)N(t)$ for all $r, t \in \mathbb{Z}[\sqrt{-5}]$.

Functions with these nice properties are sometimes called *norms*.

First, let's determine the units of $\mathbb{Z}[\sqrt{-5}]$. Suppose that u is unit and pick v so that $uv = 1$. Then

$$1 = N(1) = N(uv) = N(u)N(v).$$

Since N outputs natural numbers, we see that $N(u) = 1$. Pick integers a and b so that $u = a + b\sqrt{-5}$. Then

$$1 = N(u) = N(a + b\sqrt{-5}) = a^2 + 5b^2.$$

Notice that $5b^2$ cannot be 1. It follows that $b = 0$ and $a = 1$ or $a = -1$. This means that our unit u is either 1 or -1. So we find that the units of $\mathbb{Z}[\sqrt{-5}]$ are just 1 and -1. It follows at once that none of $3, 2 + \sqrt{-5}$, and $2 - \sqrt{-5}$ is an associate of any other of them.

It remains to see that our three members listed of $\mathbb{Z}[\sqrt{-5}]$ are irreducible. Below is an argument for $2 + \sqrt{-5}$. I leave the other two listed elements in the capable hands of the graduate students. Pick $r, t \in \mathbb{Z}[\sqrt{-5}]$ so that $2 + \sqrt{-5} = rt$. We need to see that one of r and t is a unit. So consider

$$9 = 4 + 5 = N(2 + \sqrt{-5}) = N(rt) = N(r)N(t).$$

The only possibilities for $N(r)$ are $1, 3$, and 9. If $N(r) = 1$, then, as we saw above, r must be a unit. Likewise, if $N(r) = 9$, then $N(t) = 1$ and t is a unit. So it only remains to consider the case that $N(r) = 3$. Pick integers a and b so that $r = a + b\sqrt{-5}$. Then $3 = N(r) = N(a + b\sqrt{-5}) = a^2 + 5b^2$. The only possibility for b is 0, since otherwise $a^2 + 5b^2$ must be at least 5. But then $3 = a^2$. Since there is no integer a whose square in 3, we reject the alternative that $N(r) = 3$.

In this way, we see that 9 has two quite distinct factorizations into irreducibles in $\mathbb{Z}[\sqrt{-5}]$. So $\mathbb{Z}[\sqrt{-5}]$ is not a unique factorization domain. □

The Fundemental Factorization Theorem for Principal Ideal Domains. *Every principal ideal domain is a unique factorization domain.*

Proof. We just need to demonstrate that every principal ideal domain has both the primeness condition and the divisor chain condition.

Let **D** be a principal ideal domain and suppose that $a \in D$ is irreducible and that $a \mid bc$ where $b, c \in D$. We must argue that either $a \mid b$ or $a \mid c$. So let us reject the first alternative: we assume $a \nmid b$. Let $M = (a)$. My contention is that M is maximal among proper ideals. Certainly, $M \neq D$ since a is not a unit. So M is a proper ideal. Suppose that I is an ideal that includes M. Since I is a principal ideal pick d to be a generator of of I. Now $a \in M \subseteq I = (d)$. So a is a multiple of d. That is, $a = dw$ for some w. Since a is irreducible, either d is a unit, in which case $I = D$, or w is a unit, in which case $M = I$. In this way, we see that M is maximal. Since we have $a \nmid b$ we see that $b \notin M$. So the ideal (a, b) generated by a and b must be all of D. This means $1 \in (a, b)$. So pick $x, y \in D$ so that

$$1 = xa + yb.$$

This yields $c = xac + ybc$. But $bc \in M = (a)$ since $a \mid bc$ and $xac \in (a)$ as well. So $c \in (a)$, since (a) is an ideal. This means that $a \mid c$ and so the primeness condition holds.

Now consider the divisor chain condition. Suppose that $\cdots \mid a_2 \mid a_1 \mid a_0 = a$ is a proper divisor chain in **D**. The

$$(a_0) \subsetneqq (a_1) \subsetneqq (a_2) \subsetneqq \ldots$$

is a properly increasing chain of ideals in **D**. Let I be the union of this chain. We know that the union of any chain of ideals is again an ideal. So I is an ideal. Let d be a generator I. Pick a natural number k so that $d \in (a_k)$. Then $I = (d) \subseteq (a_k) \subseteq I$. Thus, $I = (a_k)$ and the chain of ideals displayed above must be finite. This means our original divisor chain must also be finite, proving the divisor chain condition. □

So we have an immediate corollary.

The Fundamental Theorem of Arithmetic. *The ring of integers is a unique factorization domain.*

Actually, the line of reasoning we have just described is a kind of reorganization of the reasoning in Gauss's Disquisitiones.

We can extract from our proof of the Fundamental Factorization Theorem for Principal Ideal Domains the following result:

In a principal ideal domain every prime ideal is maximal among all the proper ideals.

To see it, let P be a prime ideal of the principal ideal domain **D** and pick a so that $P = (a)$. Observe that $b \in P$ if and only if $a \mid b$ for all $b \in D$. This allows us to conclude that a is prime. By a contention mentioned in the proof of the Fundamental Factorization Theorem, we see that P is a maximal ideal.

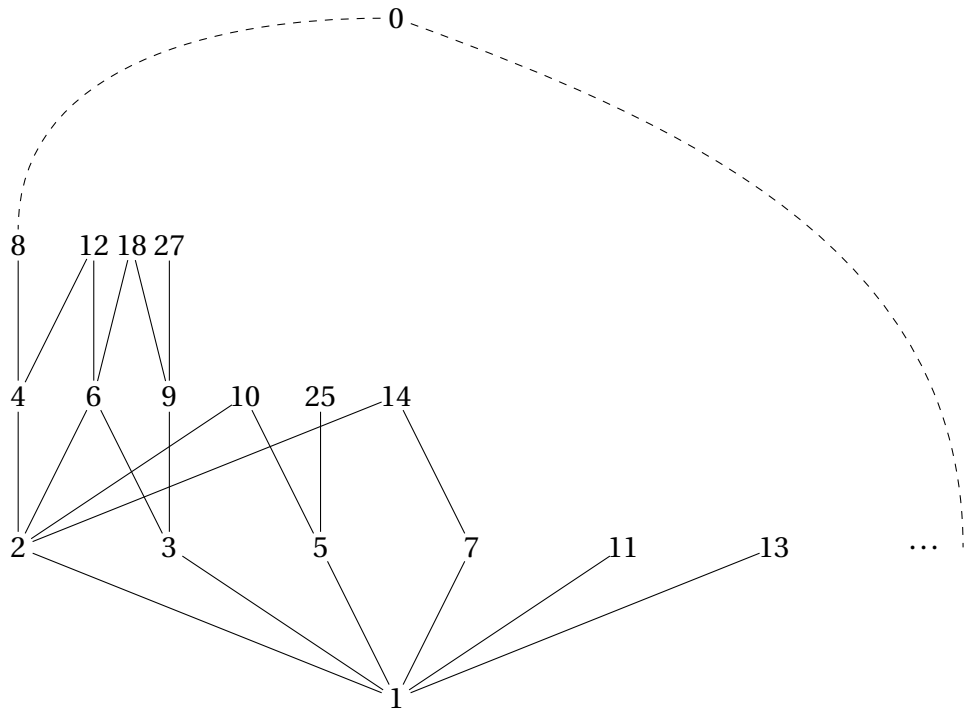
The converse, that every maximal proper ideal is prime, holds in every commutative ring **R**. For suppose M is a maximal proper ideal and $ab \in M$. In case $b \notin M$ we have $(b, M) = R$. So we can pick $u \in R$ and $w \in M$ so that $1 = ub + w$. It follows that $a = uab + aw$. Since both $ab \in M$ and $w \in M$ we conclude that $a \in M$, as desired.

Of course, this more general business s is only interesting if there are significant examples of principal ideal domains other than the ring of integers. There are and we will meet some others soon.

3.3 Divisibility

We have already used divisibility above. Given a commutative ring **R** we say that an element a **divides** an element b provided there is an element c so that $ac = b$. We denote this relation by $a \mid b$. Observe that $\mid$ is a two-place relation on R. Moreover, all the graduate students will see easily that this relation is both reflex and transitive. It just misses being an order relation because it fails the antisymmetry property—that is, $a \mid b$ and $b \mid a$ may hold even though $a \neq b$. For example, $1 \mid -1$ and $-1 \mid 1$ in the ring of integers, but $-1 \neq 1$. Suppose **R** is an integral domain and $a \mid b$ and $b \mid a$. Pick elements u and v so that $au = b$ and $bv = a$. Then we have $a(uv) = a$. This means that either $a = 0$ or $uv = 1$. In the first alternative we find that $b = 0$ as well, so that $a = 0 = b$, while in the second alternative we see that $a \sim b$. So in either case a and b are associates. The relation of association is an equivalence relation on R and up to this equivalence relation, the divisibility relation is an ordering.

Let us suppose that **R** is an integral domain and consider the divisibility ordering $\mid$ on (the $\sim$-classes of) R. The element 0 is the largest element in this ordering since $a \mid 0$ for all a (because $a \cdot 0 = 0$). Likewise the element 1 is the least element of this ordering (well, actually we are ordering the $\sim$-classes and we really mean the set of units is the least thing...). The figure below sketches part of the divisibility ordering on the natural numbers (these are the representatives of the $\sim$-classes of the integers by taking the nonnegative member of each class).



A Finite Fragment of the Divisibility Relation on the Natural Numbers

The set of natural numbers order by divisibility has some properties that may be discerned from this diagram (or perhaps more easily if more of the diagram were to be filled in...). As noted, it has a least element and a greatest element. Also the elements are evidently organized into levels, depending on the number of factors occurring in their decomposition into primes. So $8, 12, 18$, and 27 belong on the same level since they each have 3 factors in their decompositions:

$$\begin{aligned} 8 &= 2 \cdot 2 \cdot 2 \\ 12 &= 2 \cdot 2 \cdot 3 \\ 18 &= 2 \cdot 3 \cdot 3 \\ 27 &= 3 \cdot 3 \cdot 3 \end{aligned}$$

In this way, 1 is the only element at level 0, which suggests we might think it has 0 factors in its decomposition into primes. The primes themselves occupy level 1, and so on. There will be countably many levels—one level for each natural number—and each level is itself countably infinite (an extension of a famous result of Euclid: the graduate students are invited to prove this extension).

Another thing to notice is that any two elements, for example 6 and 9, have a greatest lower bound (in this case 3) and a least upper bound (in this case 18). Ordered sets in which every pair of distinct elements as both a least upper bound and a greatest lower bound are called **lattice-ordered sets**. It is important to realize that the words "greatest" and "least" refer to divisibility and *not* to that other familiar order $\leq$. In rings, we issue special names. Given to elements a and b of a commutative ring we say that an element d is a **greatest common divisor** of a and b provided

- $d \mid a$ and $d \mid b$, and

- if $e \mid a$ and $e \mid b$, then $e \mid d$.

You should notice that greatest common divisors are not unique—both 3 and -3 are greatest common divisors of 6 and 9. However, in any integral domain, any two greatest common divisors of a and b must be associates. Likewise, we say an element ℓ is a **least common multiple** of a and b provided

- $a \mid \ell$ and $b \mid \ell$, and
- if $a \mid m$ and $b \mid m$,then $\ell \mid m$.

Like greatest common divisors, least common multiples need not be unique. In integral domains, they are unique up to association. We say that the elements a and b are **relatively prime** provided 1 is a greatest common divisor of a and b.

While in the ring of integers, it is easy to see that greatest common divisors and least common multiple always exist, this is less obvious for other rings. After some reflection, you can convince yourselves that the existence of greatest common divisors and least common multiples can be established with the help of the Fundamental Theorem of Arithmetic. Only a bit more reflection leads us to be conclusion that greatest common divisors and least common multiples always exist in unique factorization domains.

It takes a bit more work (but what else should the graduate students be doing?) to establish the following fact.

Fact. Let **D** be an integral domain. If any two elements of D have a greatest common divisor, then **D** has the primeness condition.

This means that in the Theorem Characterizing Unique Factorization Domains we can replace the primeness condition with the condition that any pair of elements have a greatest common divisor.

3.4 The Chinese Remainder Theorem

The Chinese Remainder Theorem, the focus of this section, appeared in its earliest known form in China in the 3rd century C.E. and after various refinements has taken its place among the most widely known theorems of number theory. It actually holds in all commutative rings and even in some much broader contexts. In its most familiar form, it deals with the simultaneous solution of certain congruences with respect to pairwise relatively prime moduli. To frame this for commutative rings in general we will replace the integer moduli by ideals. Suppose that a and b are relatively prime integers. Observe that the ideal $(a) + (b)$ must have a generator d since the ring of integers in a principal ideal domain. Thus $(a) + (b) = (d)$. Because $(a) \subseteq (d)$ we see that $d \mid a$. Likewise, $d \mid b$. So d is a common divisor of a and b. It must be a greatest common divisor of a and b since (d) is the least ideal that contains both (a) and (b). But recall that a and b are relatively prime. Hence $(d) = (1)$. So we can draw two conclusions:

$$1 = au + bv \quad \text{for some integers } u \text{ and } v$$

and that $(a) + (b) = \mathbb{Z}$. Actually, either of these conclusions imply that a and b are relatively prime.

The Chinese Remainder Theorem. *Let* **R** *be a commutative ring and let* $I_0, I_1, \ldots, I_{n-1}$ *be finitely many ideals of* **R** *such that*

$$I_j + I_k = R \quad \text{for all } j,k < n \text{ with } j \neq k.$$

Let $a_0, a_1, \ldots, a_{n-1} \in R$. *There is* $b \in R$ *such that*

$$\begin{aligned} b &\equiv a_0 \mod I_0 \\ b &\equiv a_1 \mod I_1 \\ &\vdots \\ b &\equiv a_{n-1} \mod I_{n-1}. \end{aligned}$$

Proof. The first interesting case happens when $n = 2$. Let us examine it. Since $I_0 + I_1 = R$ we can pick $r_0 \in I_0$ and $r_1 \in I_1$ so that $1 = r_0 + r_1$. Let us that $b = r_0 a_1 + r_1 a_0$. Then observe

$$\begin{aligned} b = r_0 a_1 + r_1 a_0 &\equiv 0 \cdot a_1 + 1 \cdot a_0 \mod I_0 \\ &\equiv a_0 \mod I_0 \\ b = r_0 a_1 + r_1 a_0 &\equiv 1 \cdot a_1 + 0 \cdot a_0 \mod I_1 \\ &\equiv a_1 \mod I_1 \end{aligned}$$

So the stipulations of the theorem are strong enough to assert that each *pair* of the listed congruences can be satisfied by some appropriately chosen element b.

Now for each j with $0 < j < n$ we have that $I_0 + I_j = R$. So pick $s_j \in I_0$ and $t_j \in I_j$ so that $1 = s_j + t_j$. Then we obtain

$$1 = (s_1 + t_1)(s_2 + t_2)\ldots(s_{n-1} + t_{n-1}).$$

Using the laws of commutative rings and the properties of ideals we can expand this to obtain

$$1 = s + t_1 t_2 \ldots t_{n-1}$$

where $s \in I_0$. Notice that $t_1 t_2 \ldots t_{n-1} \in \bigcap_{0<j<n} I_j$. This means that

$$I_0 + \bigcap_{0<j<n} I_j = R.$$

As we observed above, there is an element $d_0 \in R$ so that

$$\begin{aligned} d_0 &\equiv 1 \mod I_0 \\ d_0 &\equiv 0 \mod \bigcap_{0<j<n} I_j. \end{aligned}$$

Since $\bigcap_{0<j<n} I_j \subseteq I_k$ for all k with $0 < k < n$, we find that

$$\begin{aligned} d_0 &\equiv 1 \mod I_0 \\ d_0 &\equiv 0 \mod I_j \text{ for all } j \text{ with } 0 \neq j < n. \end{aligned}$$

We can apply this reasoning that worked for the index 0 to any of the indices. In this way, for each $k < n$ we can have $d_k \in R$ such that

$$\begin{aligned} d_k &\equiv 1 \mod I_k \\ d_k &\equiv 0 \mod I_j \text{ for all } j \text{ with } k \neq j < n. \end{aligned}$$

Now put $b = \sum_{k<n} d_k a_k$. Then for all $j < n$ we obtain

$$\begin{aligned} b &= d_0 a_0 + \cdots + d_{j-1} a_{j-1} + d_j a_j + d_{j+1} a_{j+1} + \cdots + d_{n-1} a_{n-1} \\ b &\equiv 0 \cdot a_0 + \cdots + 0 \cdot a_{j-1} + 1 \cdot a_j + 0 \cdot a_{j+1} + \cdots + 0 \cdot a_{n-1} \mod I_j \\ b &\equiv a_j \mod I_j \end{aligned}$$

as desired. □

We can cast the Chinese Remainder Theorem as a structure theorem for commutative rings. Recall that in § 1.3 we discussed direct products of algebraic systems in general. For commutative rings we can enhance the fact at the end of that section.

The Chinese Remainder Theorem: Structural Version. *Let* $\mathbf{R}$ *be a commutative ring and* $I_0, I_1, \ldots, I_{n-1}$ *be a finite list of ideals of* $\mathbf{R}$. *Then*

$$\mathbf{R}/\bigcap_{j<n} I_j \text{ is embeddable into } \mathbf{R}/I_0 \times \cdots \times \mathbf{R}/I_{n-1}.$$

Moreover, if $I_j + I_k = R$ *for all* j *and* k *with* $j, k < n$ *and* $j \neq k$, *then the embedding is an isomorphism.*

Proof. We need a map h from $\mathbf{R}$ into the direct product whose kernel is $\bigcap_{j<n} I_j$. Then we can invoke the Homomorphism Theorem to obtain the desired embedding. The map is the one that comes most easily to hand:

$$h(a) := (a + I_0, a + I_1, \ldots, a + I_{n-1}) \text{ for all } a \in R.$$

This map is assembled from the quotient maps. It is routine to demonstrate that it is a homomorphism and that its kernel is $\bigcap_{j<n} I_j$. An appeal to the Homomorphism Theorem gives us the desired embedding. So the first part of this theorem just rests on general considerations. The power resides in the "moreover" part of the statement. For that, what is needed is to see that h maps R onto the direct product.

Consider any element of the direct product. It has the form

$$(a_0 + I_0, a_1 + I_1, \ldots, a_{n-1} + I_{n-1}).$$

We must see that there is a $b \in R$ so that

$$h(b) = (a_0 + I_0, a_1 + I_1, \ldots, a_{n-1} + I_{n-1}).$$

Given the definition of h, we see that this is the same as finding a $b \in R$ so that

$$b + I_j = a_j + I_j \text{ for all } j < n.$$

In other words, that

$$b \equiv a_j \mod I_j \text{ for all } j < n.$$

Of course, this is precisely what the Chinese Remainder Theorem does for us. □

3.5 Problem Set 3

ALGEBRA HOMEWORK, EDITION 3
FOURTH WEEK
MORE IDEALS

PROBLEM 10.
Let **R** be a commutative ring and let n be a positive integer. Let $J, I_0, I_1, \ldots, I_{n-1}$ be ideals of **R** so that I_k is a prime ideal for every $k < n$ and so that $J \subseteq I_0 \cup \cdots \cup I_{n-1}$. Prove that $J \subseteq I_k$ for some $k < n$.

PROBLEM 11.
Let **R** be a nontrivial commutative ring and let J be the intersection of all the maximal proper ideals of **R**. Prove that $1 + a$ is a unit of **R** for all $a \in J$.

PROBLEM 12.
Let **R** be a commutative ring. Define

$$N := \{a \mid a \in R \text{ and } a^n = 0 \text{ for some positive integer } n\}.$$

a. Prove that N is an ideal of **R**.

b. Prove that $N \subseteq P$ for every prime ideal P of **R**.

PROBLEM 13.
Let **R** be a commutative ring and I be an ideal of **R**. Prove each of the following:

a. Suppose P_0 and P_1 are ideals of **R**. If $I \subseteq P_0 \cup P_1$, then either $I \subseteq P_0$ or $I \subseteq P_1$.

b. Suppose P_0, P_1, and P_2 are prime ideals of **R**. if $I \subseteq P_0 \cup P_1 \cup P_2$, then either $I \subseteq P_0$ or $I \subseteq P_1$ or $I \subseteq P_2$.

LECTURE 4

ZORN'S LEMMA

Zorn's Lemma is a transfinite existence principle which has found a number of useful and informative applications in algebra and analysis. While the Lemma bears the name of Max Zorn, equal credit should be extended to Felix Hausdorff and Kazimierz Kuratowski who found closely related results decades before Zorn.

A **chain** or **linearly ordered set** is just a partially ordered set in which any two elements are comparable. We also refer to any subset of a partially ordered set as a **chain** when it is linearly ordered by the ordering inhereted from the larger ordered set. This means that, where a and b are elements of the chain and $\leq$ denotes the order relation, we have either $a \leq b$ or $b \leq a$. Let C be a subset of a partially ordered set P and $b \in P$. We say that b is an **upper bound** of C provided $a \leq b$ for all $a \in C$. We say b is a **strict upper bound** provided $a < b$ for all $a \in C$. An element d is **maximal** in C if $d \in C$ and whenever $d \leq a \in C$ it follows that $d = a$.

Zorn's Lemma. *Let P be a partially ordered set and suppose that every chain in P has an upper bound in P. Then P has a maximal member.*

Proof. Let g be a function which chooses an element from each nonempty subset of P. That is the domain of g is the collection of nonempty subsets of P and $g(D) \in D$ for each nonempty subset $D \subseteq P$. The function g, which is called a *choice function*, exists according to the Axiom of Choice.

Denote the ordering on P by $\leq$. For each set $C \subseteq P$ let $\widehat{C}$ denote the set of all strict upper bounds of C. Notice that the empty set $\varnothing$ is a chain in P. According to our hypothesis it must have an upper bound in P. Since $\varnothing$ is empty this upper bound must be a proper upper bound. This means $\widehat{\varnothing}$ is nonempty. (Hence, P is nonempty.)

We will say that $K \subseteq P$ is a g**-chain** provided

- K is not empty.
- K is a chain.
- if $C \subseteq K$ and C has a strict upper bound in K, then $g(\widehat{C})$ is a minimal member of $\widehat{C} \cap K$.

Here is a useful fact about how elements in g-chains compare.

Fact. *Let K and J be g-chains so that $a \in K - J$ and $b \in J$. Then $b < a$.*

Proof. Let $C = \{d \mid d \in K \cap J \text{ and } d < a\}$. So C has a strict upper bound in K. Since K is a g-chain, we have $g(\widehat{C})$ is a minimal member of $\widehat{C} \cap K$. Also, $g(\widehat{C}) \leq a$. Now $\widehat{C} \cap J$ must be empty, since otherwise $g(\widehat{C}) \in K \cap J$, putting $g(\widehat{C}) \in C$, which is impossible. So if $b \in J$ then there is $d \in C$ with $b \leq d < a$. Hence, $b < a$. □

Claim. *The union of any nonempty collection of g-chains is a g-chain.*

Proof. Let L be a union of some nonempty family $\mathscr{F}$ of g-chains. We first have to check that L is linearly ordered. So suppose $a, b \in L$. Pick $K, J \in \mathscr{F}$ so that $a \in K$ and $b \in J$. We need to show that a and b are comparable. We might as well consider that $a \notin J$, since if $a \in J$ we see, J being a chain, that a and b are comparable. But the Fact above then tells us that $b < a$, so a and b are comparable. This means that L is a chain.

Of course, L is not empty since it is union of a nonempty collection of nonempty sets. So it remains to verify the last condition in the definition of g-chain. To this end, let $C \subset L$ such that C has a strict upper bound $b \in L$. Pick $J \in \mathscr{F}$ so that $b \in J$. To see that $C \subseteq J$, pick $a \in C$ and, for contradiction, suppose $a \notin J$. Pick $K \in \mathscr{F}$ so that $a \in K$. Now the Fact yields $b < a$. But here we also have $a < b$. So we find $C \subseteq J$. Since J is a g-chain, we have $g(\widehat{C})$ is a minimal member of $\widehat{C} \cap J$. But we need to see that $g(\widehat{C})$ is a minimal member of $\widehat{C} \cap L$. Suppose not. Pick $a' \in \widehat{C} \cap L$ so that $a' < g(\widehat{C})$. To simplify notation, let $g(\widehat{C}) = b'$. So $a' < b'$. Now pick $K' \in \mathscr{F}$ so that $a' \in K'$. Now the Fact above again yields $b' < a'$, which is contrary to $a' < b'$. This verifies for L the last condition in the definition of g-chain. So L is a g-chain, as claimed. □

Now let M be the union of the collection of all g-chains. Were $\widehat{M}$ nonempty we could form $M \cup \{g(\widehat{M})\}$, which would be a chain properly extending M. A routine check of the definition shows that $M \cup \{g(\widehat{M})\}$ would again be a g-chain. This produces $g(\widehat{M}) \in M \cup \{g(\widehat{M})\} \subseteq M$. But we know $g(\widehat{M}) \notin M$. So $\widehat{M}$ must be empty. So M has no strict upper bounds. But by the hypothesis, every chain has an upper bound. So M must have a largest element m. That is $a \leq m$ for all $a \in M$. As there is no strict upper bound of M, there can be no element which is strictly above m. That is m is the maximal element we seek.

Zorn's Lemma is proven.

□

LECTURE 5

RINGS LIKE THE RATIONALS

5.1 FIELDS

In the commutative ring of rational numbers every nonzero element has a multiplicative inverse—every nonzero element is a unit. Other rings you are acquainted with have this property as well. A **field** is a nontrivial commutative in which every nonzero element has a multiplicative inverse. Fields evidently satisfy the cancellation law. So every field is an integral domain. Moreover, since every nonzero element is a unit we see that every nontrivial ideal of a field must actually be the whole field. In other words, every field has exactly two ideals: the trivial ideal and the whole field. Both of these ideals are principal ideals, so every field is a principal ideal domain. So every field is also a unique factorization domain, but in itself, this is not too interesting since fields have no nonzero nonunits to factor and there are no irreducible elements.

Theorem 5.1.1. *Let* $\mathbf{R}$ *be a commutative ring and* I *be an ideal of* $\mathbf{R}$*. Then*

(a) $\mathbf{R}$ *is a field if and only if* $\mathbf{R}$ *has exactly two ideals.*

(b) $\mathbf{R}/I$ *is a field if and only if* I *is a maximal among the proper ideals of* $\mathbf{R}$*.*

Proof. For part (a) we have already observed the implication from left to right. For the converse, suppose $\mathbf{R}$ has exactly two ideals and let $a \in R$ be nonzero. We have to show that a is invertible. The ideal (a) must be the whole of R, so in particular $1 \in (a)$. This means we can (and do) pick $b \in R$ so that $1 = ab$. So b is the desired inverse of a.

Part (b) is an immediate consequence of part (a) and the Correspondence Theorem. □

To simplify the language we will say that I is a **maximal ideal** of the ring $\mathbf{R}$ provided

(a) I is a proper ideal of $\mathbf{R}$, and

(b) Either $I = J$ or $J = R$ whenever J is an ideal of $\mathbf{R}$ with $I \subseteq J$.

So the theorem of above asserts, in part, that, for a commutative ring $\mathbf{R}$, we have that $\mathbf{R}/I$ is a field if and only if I is a maximal ideal of $\mathbf{R}$.

The Maximal Ideal Theorem.

(a) *Every proper ideal of a ring is included in a maximal ideal of the ring.*

(b) *Every nontrivial commutative ring has a homomorphic image that is a field.*

Proof. For part (a) let I be a proper ideal of the ring $\mathbf{R}$. Let

$$\mathscr{F} = \{J \mid I \subseteq J \text{ and } J \text{ is a proper ideal of } \mathbf{R}\}.$$

Any maximal element of $\mathscr{F}$ will be a maximal ideal that includes I. We invoke Zorn's Lemma to see that $\mathscr{F}$ has a maximal member. Indeed, suppose $\mathscr{C}$ is a chain included in $\mathscr{F}$. If $\mathscr{C}$ is empty, then I will be an upper bound of $\mathscr{C}$. So we suppose that $\mathscr{C}$ is not empty. Observe that $\bigcup\mathscr{C}$ is an ideal of $\mathbf{R}$ since it is a union of a chain of ideals. Plainly, $I \subseteq \bigcup\mathscr{C}$. Finally, were $\bigcup\mathscr{C}$ not proper we would have $1 \in \bigcup\mathscr{C}$. But that would mean that $1 \in J$ for some $J \in \mathscr{C}$. However, the members of $\mathscr{C}$ belong to $\mathscr{F}$ so they are proper ideals. So we find that $\bigcup\mathscr{C}$ is a proper ideal of $\mathbf{R}$ that includes I. This means that $\bigcup\mathscr{C}$ belongs to $\mathscr{F}$. That is, every nonempty chain in $\mathscr{F}$ has an upper bound in $\mathscr{F}$. According to Zorn, $\mathscr{F}$ must have maximal members. This establishes part (a).

Part (b) is an immediate consequence of part (a) and the first theorem in this section. □

5.2 Fields of Fractions

In addition to the field of rational numbers, you are also acquainted with the field of real numbers, as well as the field of complex numbers. We also have in hand finite fields like $\mathbb{Z}/(p)$, where p is a prime number. This is because we know that (p) is a prime ideal of $\mathbb{Z}$ and we know that in a principal ideal domain prime ideals are maximal. Some of these fields don't seem much like the field of rational numbers. We know there is a close connection between the integers and the rationals. We can build the field of rationals from the ring of integers. An interesting thing is that the same procedure can be applied to any integral domain to produce a closely associated field. Here is how.

Fix an integral domain $\mathbf{D}$ throughout this section. The idea is to enhance D by adjoining all the multiples of the multiplicative inverses of the nonzero elements of D. There is a little wrinkle in this process. When we do this for the integers we have to throw in $\frac{1}{4}$ to ensure that 4 will have a multiplicative inverse and then we have to throw in $\frac{2}{4} = 2 \cdot \frac{1}{4}$. Of course, we have to identify $\frac{2}{4}$ and $\frac{1}{2}$. There is a two-step process to smooth out this wrinkle.

Let $E = \{(a,b) \mid a, b \in D \text{ with } b \neq 0\}$. On E define the binary relation $\asymp$ be

$$(a,b) \asymp (c,d) \text{ if and only if } ad = bc$$

for all $(a,b),(c,d) \in E$. The eager graduate students will write out a proof that $\asymp$ is an equivalence relation on E. As a second step, we name the equivalence classes in a convenient manner. For $a, b \in D$ with $b \neq 0$ we put

$$\frac{a}{b} := \{(c,d) \mid (c,d) \in E \text{ and } (a,b) \asymp (c,d)\}.$$

So we have

$$\frac{a}{b} = \frac{c}{d} \text{ if and only if } ad = bc,$$

for all $a, b, c, d \in D$ with $b \neq 0 \neq d$.

Let $F' = \{\frac{a}{b} \mid a, b \in D \text{ with } b \neq 0\}$. Our plan is to make F' into a field by defining the ring operations in some appropriate manner. Here is how. For all $\frac{a}{b}, \frac{c}{d} \in F$ let

$$\begin{aligned} \frac{a}{b} + \frac{c}{d} &:= \frac{ad + cb}{bd} \\ \frac{a}{b} \cdot \frac{c}{d} &:= \frac{ac}{bd} \\ -\frac{a}{b} &:= \frac{-a}{b} \\ 0^* &:= \frac{0}{1} \\ 1^* &:= \frac{1}{1} \end{aligned}$$

The last two equations define the one and the zero of the ring of fractions. Of course, these definitions are very familiar from the days in school when we learned how to deal with fractions. It is worth noting that the soundness to these definitions depends on the fact that **D** is an integral domain—to ensure that $bd \neq 0$ when $b \neq 0 \neq d$. Here is the question:

> Is the algebra $\langle F', +, \cdot, -, 0^*, 1^* \rangle$ really a field?

Unfortunately, we seem to be forced to check all the equations defining commutative rings as well as checking that every nonzero element has a multiplicative inverse. This checking is tedious but must be done (by the graduate students!). The most strenuous case is checking the associative law for addition. Here is a verification of a distributive law to show how it is done.

$$\begin{aligned} \frac{a}{b}\left(\frac{c}{d} + \frac{e}{f}\right) &= \frac{a}{b}\frac{cf + ed}{df} \\ &= \frac{a(cf + ed)}{b(df)} \\ &= \frac{a(cf) + a(ed)}{b(df)} \\ &= \frac{((ac)f + (ae)d) \cdot 1}{((bd)f) \cdot 1} \\ &= \frac{(ac)f + (ae)d}{(bd)f}\frac{1}{1} \\ &= \frac{(ac)f + (ae)d}{(bd)f}\frac{b}{b} \\ &= \frac{(ac)(bf) + (ae)(bd)}{(bd)(bf)} \\ &= \frac{ac}{bd} + \frac{ae}{bf} \\ &= \frac{a}{b}\frac{c}{d} + \frac{a}{b}\frac{e}{f} \end{aligned}$$

In the reasoning above, we used $\frac{1}{1} = \frac{b}{b}$ where we know $b \neq 0$. This is a little lemma that is helpful in the other parts of the proof.

Let $D' = \{\frac{a}{1} \mid a \in D\}$. It is easy to check that D' is a subuniverse of the field $\langle F', +, \cdot, -, 0^*, 1^*\rangle$ and that the map sending $a \mapsto \frac{a}{1}$ for $a \in D$ is an embedding of **D** into the field. But we would rather regard **D** as a subring of its field of fractions, just as we regard $\mathbb{Z}$ as a subring of $\mathbb{Q}$. We accomplish this by letting

$$F := D \cup (F' \setminus \{\frac{a}{1} \mod a \in D\}).$$

We have to define the operations on F. Here is how to define addition for $u, v \in F$.

$$u + v := \begin{cases} u + v & \text{if } u, v \in D \\ u + v & \text{if } u, v \in F' \\ \frac{ub+a}{b} & \text{if } u \in D \text{ and } v = \frac{a}{b} \in F' \\ \frac{a+vb}{b} & \text{if } v \in D \text{ and } u = \frac{a}{b} \in F' \end{cases}$$

The first two lines of this may look a bit strange. The + in the first case refers to the addition in **D**, whereas on the second line the + refers to the addition defined above over F'. The other operations can be defined in a similar fashion. In effect, what we have done is a bit of transplant surgery. We have sliced out D' and put D in its place making sure to stitch things up so the operations work right. The result is a field **F** that has **D** has a subring. This field **F** is called the **field of fractions** of the integral domain **D**.

We have provided one construction that starts with an integral domain **D** and ends up with an extension **F** that can be rightful called a "field of fractions". However, it should be clear that this construction is not really unique—it is possible to make small changes that will produce other fields that could also be called fields of fractions but that are technically different from the one we have just constructed. There is, however, a strong uniqueness result for fields of fractions. field!field of fractions

Theorem on the Existence and Uniqueness of Fields of Fractions. *Let* **D** *be any integral domain. There is a field* **F** *such that* **D** *is a subring of* **F***, and moreover, if* **S** *is any ring and* **K** *is any field so that* **S** *is a subring of* **K** *and if* $h : \mathbf{D} \to \mathbf{S}$ *is any isomorphism from* **D** *onto* **S***, then* h *has a unique extension to an embedding of* **F** *into* **K**.

Proof. We already established the existence of a field **F** of fractions. Suppose the field **K** and the embedding h are given to us. We define the extension $\hat{h}$ from F into K as follows. For any $u \in F$ let

$$\hat{h}(u) = \begin{cases} h(u) & \text{if } u \in D \\ h(a)(h(b))^{-1} & \text{if } u = \frac{a}{b} \notin D \end{cases}$$

In the second alternative, $h(b)$ will be a nonzero element of K and it will have a multiplicative inverse in K, which we have denoted by $(h(b))^{-1}$. It is a routine work for the delight of the graduate students to demonstrate that $\hat{h}$ is actually an embedding. Staring hard at the definition of $\hat{h}$ should suggest a proof that this is the only way to get such an extension. □

So the field of fraction of an integral domain **D** is, in the sense described above, the smallest field extending **D**.

5.3 Problem Set 4

Algebra Homework, Edition 4
Fifth Week
Fields

Problem 14.
Let **F** be a field and let $p(x) \in \mathbf{F}[x]$ be a polynomial of degree n. Prove that $p(x)$ has at most n distinct roots in **F**.

Problem 15.
Let **R** be a commutative ring and let $a \in R$ with $a^n \neq 0$ for every natural number n. Prove that **R** has an ideal P such that each of the following properties holds:

i. $a^n \notin P$ for every natural number n, and

ii. for all ideals I of **R**, if $P \subseteq I$ and $P \neq I$, then $a^n \in I$ for some natural number n.

Problem 16.
Let **F** be a field and let $\mathbf{F}^*$ be its (multiplicative) group of nonzero elements. Let **G** be any finite subgroup of $\mathbf{F}^*$. Prove that **G** must be cyclic.

Problem 17.
Suppose that **D** is a commutative ring such that $\mathbf{D}[x]$ is a principal ideal domain. Prove that **D** is a field.

LECTURE 6

RINGS OF POLYNOMIALS

6.1 POLYNOMIALS OVER A RING

$5x^3+3x^2-7x+1$ is a polynomial with integer coefficients. Our experience in school and even through calculus leads us to think of polynomials as functions, but here in algebra we take a different view. We consider that polynomials are formal expressions that describe functions. We regard polynomials as certain kinds of strings of symbols. We could also regard the polynomial at the start of this paragraph has a polynomial over the ring $\frac{\mathbb{Z}}{(8)}$. That ring has just 8 elements and there are only 8^8 one-place operations on the underlying set $\{0,1,2,3,4,5,6,7\}$. However, there is a countable infinity of polynomials, some of each degree, with coefficients in that ring. This means that some (actually many) polynomials will name the same function.

The interesting thing about treating polynomials as strings of symbols is that we can define an addition and a multiplication, as well as the formation of negatives and in this way produce a ring. We know well how to add and multiply polynomials in a formal manner—having had lots of drill in Algebra I. To help in formalizing addition and multiplication, it is convenient to write polynomials backwards from how most of us were taught. In fact, it is reasonable to imagine each polynomial as an infinitely long expression where after some point all the coefficients are 0 (and so have been neglected...).

Here is how addition works, of course.

$$\begin{array}{ccccccc}
a_0 & + & a_1x & + & a_2x^2 & +\cdots+ & a_nx^n \\
b_0 & + & b_1x & + & b_2x^2 & +\cdots+ & b_nx^n \\
\hline
(a_0+b_0) & + & (a_1+b_1)x & + & (a_2+b_2)x^2 & +\cdots+ & (a_n+b_n)x^n
\end{array}$$

Notice that while this looks like we have assumed that the polynomials are both of degree n, we have not made such an assumption. Some (or all) of the coefficients above can be 0. So this description of addition works for all polynomials. It is important to realize that the +'s occurring in the parentheses on the last line actually refer to the addition in the ring of coefficients. So the idea is that, unlike the other +'s, which are formal symbols, those in the parentheses should

actually be executed to produce elements of the ring of coefficients to get the coefficients of the sum of the polynomials.

Multiplication is more complicated.

$$\begin{array}{ccccccc} a_0 & + & a_1x & + & a_2x^2 & +\cdots+ & a_nx^n \\ b_0 & + & b_1x & + & b_2x^2 & +\cdots+ & b_nx^n \\ \hline (a_0b_0) & + & (a_0b_1+a_1b_0)x & + & (a_0b_2+a_1b_1+a_2b_0)x^2 & +\cdots+ & (\sum_{i+j=n} a_ib_j)x^n \end{array}$$

In general, the k^{th} coefficient is

$$\sum_{i+j=k} a_ib_j.$$

Here is a smaller example

$$\begin{aligned}(a_0+a_1x)(b_0+b_1x+b_2x^2) &= a_0(b_0+b_1x+b_2x^2)+a_1x(b_0+b_1x+b_2x^2)\\ &= a_0b_0+a_0b_1x+a_0b_2x^2+a_1b_0x+a_1b_1x^2+a_1b_3x^3\\ &= a_0b_0+(a_0b_1+a_1b_0)x+(a_0b_2+a_1b_1+a_2b_0)x^2+(a_0b_3+a_1b_2+a_0b_3)x^3\end{aligned}$$

This looks like a deduction, but isn't really. Rather, it is the basis for the given definition of multiplication. But this does show that while we didn't allow the commutative law to sneak into the calculation of the coefficients, we have somehow assumed here that the variable x commutes with everything.

Given a ring $\mathbf{R}$ we make the **ring $\mathbf{R}[x]$ of polynomials with coefficients from $\mathbf{R}$** by imposing the addition and multiplication described above on the set of polynomials. The zero of the ring of polynomials is the polynomials where all the coefficients are 0. The one of this ring is the polynomial with constant coefficient 1 and all other coefficients 0. Form negatives of polynomials we leave to the imagination of the graduate students.

Well, is $\mathbf{R}[x]$ really a ring? We need to check the equations that we used to define the notion of a ring. The equations only involving $=, -$ and 0 are easy. The associative law for multiplication and the distributive laws are messy and best not displayed in public (but the disciplined graduate students will not flinch from checking this stuff). Notice that $\mathbf{R}$ is a subring of $\mathbf{R}[x]$.

The **zero polynomial** is the one whose coefficients are all 0. Every nonzero polynomial

$$a_0+a_1+\cdots+a_nx^n$$

has a rightmost coefficient that is not 0. This coefficient is the **leading coefficient** of the polynomial and the exponent of the associated x is called the **degree** of the polynomial. It is convenient to assign no degree to the zero polynomial.

If the sum of two polynomials is not the zero polynomial then the degree of the sum can be no larger than the maximum of the degree of the summands. Likewise, if the product of two polynomials is not the zero polynomial, then the degree of the product is no larger than the sum of the degrees of two factors. If $\mathbf{R}$ is an integral domain, then the degree of the product of nonzero polynomials in $\mathbf{R}[z]$ is the sum of the degrees of the factors.

Once we are convinced that $\mathbf{R}[x]$ is a ring we can repeat the construction to from the ring $\mathbf{R}[x][y]$. Here is a member of $\mathbb{Z}[x][y]$.

$$(1+2x+3x^3)+(2-x)y+(5+x^3)y^2$$

Observe that the coefficients of this polynomial (namely, the parts in parentheses) are members of $\mathbf{R}[x]$. We identify this polynomial with

$$1+2x+2y-xy+5y^2+3x^3+x^3y^2$$

Now notice that the polynomial below

$$(1+2y+5y^2)+(2-y)x+(3+y^2)x^3$$

is a member of $\mathbb{Z}[y][x]$ that we also identify with

$$1+2x+2y-xy+5y^2+3x^3+x^3y^2$$

By similar reasoning we identify $\mathbb{Z}[x][y]$ with $\mathbb{Z}[y][x]$. We use the notation $\mathbb{Z}[x,y]$ to denote this ring. More generally, we arrive at the polynomial ring $\mathbf{R}[x_0,x_1,\ldots,x_n]$ in any finite number of variables. It is even possible to consider rings of polynomials over infinite sets of variables, although we will not pursue this.

Here are some easily deduced facts.

Fact. Let $\mathbf{R}$ be a ring. $\mathbf{R}[x]$ is a commutative ring if and only if $\mathbf{R}$ is a commutative ring.

Fact. Let $\mathbf{R}$ be a ring. $\mathbf{R}[x]$ is an integral domain if and only if $\mathbf{R}$ is an integral domain.

A very useful result about rings of polynomial is next.

The Homomorphism Extension Property for $\mathbf{R}[x]$. *Let $\mathbf{R}$, $\mathbf{S}$, and $\mathbf{T}$ be rings so that $\mathbf{S}$ is a subring of $\mathbf{T}$ and let h be a homomorphism from $\mathbf{R}$ onto $\mathbf{S}$. For any $t \in T$ there is exactly one homomorphism $\hat{h}$ extending h that maps $\mathbf{R}[x]$ into $\mathbf{T}$ such that $\hat{h}(x) = t$.*

This theorem is illustrated in Figure 6.1.

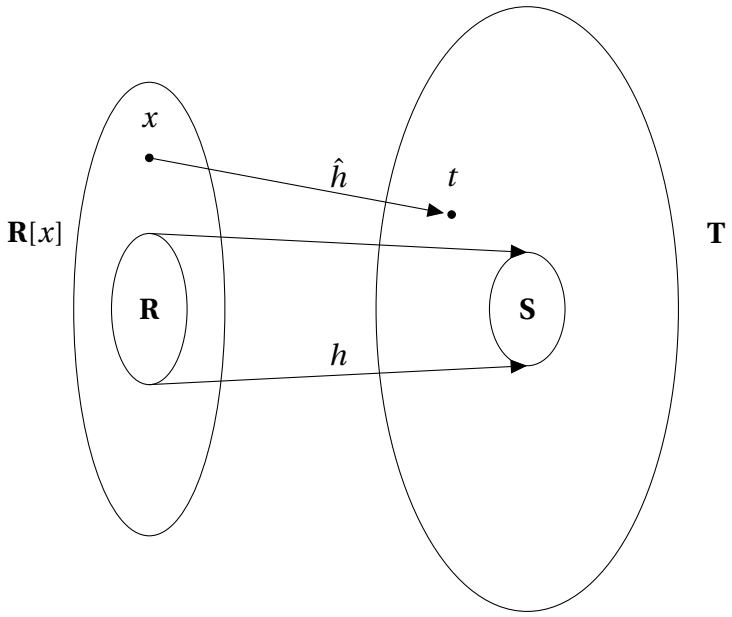


Figure 6.1: The Homomorphism Extension Property

Proof. Consider an arbitrary polynomial

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n.$$

Were there to be any extension $\hat{h}$ of h as desired, then we would have to have

$$\begin{aligned}\hat{h}(p(x)) &= \hat{h}(a_0) + \hat{h}(a_1\hat{h}(x) + \cdots + \hat{h}(a_n)(\hat{h}(x))^n \\ &= h(a_0) + h(a_1)t + \cdots + h(a_n)t^n.\end{aligned}$$

In this way we see that there can be at most one possibility for $\hat{h}$. Moreover, we can define the desired extension by

$$\hat{h}(p(x)) := h(a_0) + h(a_1)t + \cdots + h(a_n)t^n.$$

The only issue is whether this function is actually a homomorphism. This we leave in the capable hands of the graduate students. □

An interesting special case of this theorem is when $\mathbf{R} = \mathbf{S}$ and h is just be identity map. In that case the extension $\hat{h}$ gives us

$$\hat{h}(p(x)) = \hat{h}(a_0 + a_1 x + \cdots + a_n x^n) = a_0 + a_1 t + \cdots + a_n t^n = p(t).$$

Notice that the $p(x)$ on this line is a polynomial whereas the $p(t)$ is an element of T. If we construe the polynomial $p(x)$ as a name for a function from T to T, then what $\hat{h}$ does is *evaluate* the named function at the input t. For this reason, $\hat{h}$, which depends on t, is called an *evaluation map*. In this context, we say that t is a **root** of $p(x)$ provided $\hat{h}(p(x)) = 0$; that is provided $p(t) = 0$ in $\mathbf{T}$.

We saw a key fact about the integers that had to do with quotients and remainders. This very useful fact led us to the conclusion that the ring of integers is a principal ideal domain. Something like this fact holds for polynomial rings.

Theorem on Quotients and Remainders for Polynomials. *Let* $\mathbf{R}$ *be a commutative ring, let*$d(x) \in \mathbf{R}[x]$ *be a nonzero polynomial, and let* b *be the leading coefficient of* $d(x)$*. Let* $f(x) \in \mathbf{R}[x]$ *be any polynomial. There is a natural number* k *and polynomials* $q(x)$ *and* $r(x)$ *such that*

(a) $b^k f(x) = q(x)d(x) + r(x)$ *and*

(b) *Either* $r(x)$ *is the zero polynomial or* $\deg r(x) < \deg d(x)$.

Moreover, given such a k *the polynomials* $q(x)$ *and* $r(x)$ *are unique, provided* $\mathbf{R}$ *is an integral domain.*

Proof. Observe that if the degree of $d(x)$ is larger than the degree of $f(x)$ then we can take $r(x) = f(x)$ and $q(x) = 0$ and we can put $k = 0$. So the existence part of this theorem only needs a proof when $\deg f(x) \geq \deg d(x)$. We prove the existence part of the theorem by induction on $\deg f(x)$.

Base Step: $\deg f(x) = \deg d(x)$
Let a be the leading coefficient of $f(x)$. Put $k = 1, q(x) = a$, and $r(x) = bf(x) - ad(x)$. This works.

Inductive Step
We suppose that $\deg f(x) = n + 1 > \deg d(x)$. Let m be the degree of $d(x)$. Once more let a be the leading coefficient of $f(x)$. Observe

$$\hat{f}(x) := bf(x) - ax^{n1-m}xd(x)$$

is a polynomial of degree no more than n. We can apply the induction hypothesis to obtain a natural number ℓ and polynomials $\hat{q}(x)$ and $r(x)$ so that

(a) $b^\ell \hat{f}(x) = \hat{q}(x)d(x) + \hat{r}(x)$ and

(b) $r(x)$ is the zero polynomial or $\deg r(x) < \deg d(x)$.

But this entails

$$\begin{aligned} b^{\ell+1} f(x) &= \hat{q}(x)d(x) + ax^{n+1-m}d(x) + r(x) \\ &= (\hat{q}(x) + ax^{n+1-m})d(x) + r(x). \end{aligned}$$

Taking $q(x) := \hat{q}(x) = ax^{n+1-m}$ establishes the inductive step.

So the existence part of the theorem is finished. For the uniqueness part, we suppose that $\mathbf{R}$ is an integral domain and we take k to be a fixed natural number and $f(x), q_0(x), q_1(x), r_0(x)$, and $r_1(x)$ to be polynomials such that

(a) $b^k f(x) = q_0(x)d(x) + r_0(x)$,

(b) $r_0(x)$ is the zero polynomial or $\deg r_0(x) < \deg d(x)$,

(c) $b^k f(x) = q_1(x) + r_1(x)$, and

(d) $r_1(x)$ is the zero polynomial or $\deg r_1(x) < \deg d(x)$.

It follows that

$$(q_0(x) - q_1(x))d(x) = r_1(x) - r_0(x).$$

Now the polynomial on the right is either the zero polynomial or it had degree less than the degree of $d(x)$. The polynomial on the left is either the zero polynomial or it has degree at least the degree of $d(x)$. It follows that both sides of this equation are the zero polynomial. In particular, $r_0(x) = r_1(x)$. (At this point we have yet to invoke the fact that $\mathbf{R}$ is an integral domain.) So we have

$$(q_0(x) - q_1(x))d(x) = 0.$$

We know that $d(x)$ is not the zero polynomial. Since $\mathbf{R}[x]$ is an integral domain, we find that $q_0(x) = q_1(x)$, as desired. □

Here are three important immediate corollaries of this theorem.

Corollary 6.1.1. *Let* $\mathbf{R}$ *be a commutative ring, let* $d(x) \in \mathbf{R}[x]$ *be a nonzero polynomial whose leading coefficient is a unit. Let* $f(x) \in \mathbf{R}[x]$ *be any polynomial. There are polynomials* $q(x)$ *and* $r(x)$ *such that*

(a) $f(x) = q(x)d(x) + r(x)$ *and*

(b) *Either* $r(x)$ *is the zero polynomial or* $\deg r(x) < \deg d(x)$.

Moreover, the polynomials $q(x)$ *and* $r(x)$ *are unique, provided* $\mathbf{R}$ *is an integral domain.*

Corollary 6.1.2. $\mathbf{F}[x]$ *is a principal ideal domain provided* $\mathbf{F}$ *is a field. Hence* $\mathbf{F}[x]$ *is a unique factorization domain, provided* $\mathbf{F}$ *is a field.*

Corollary 6.1.3. *Let* $\mathbf{R}$ *be a commutative ring, let* $f(x) \in \mathbf{R}[x]$ *be a polynomial with coefficients in* R *and let* $r \in R$. *Then* r *is a root of* $f(x)$ *if and only if* $(x - r) \mid f(x)$.

The second of the corollaries displayed above can be deduced in the same manner that we used to establish that $\mathbb{Z}$ is a principal ideal domain.

There is one more general observation to make here.

The Binomial Theorem holds in every commutative ring.

This means that in any commutative ring we have

$$(x+y)^n = \sum_{k\le n} \binom{n}{k} x^k y^{n-k} \text{ for all elements } x \text{ and } y \text{ of the ring.}$$

This must be understood carefully. The binomial coefficient $\binom{n}{k}$ that appear here are positive natural numbers, not elements of the ring at hand. We must understand them as indicating repeated additions within the ring. That is we take $\binom{n}{k}$ to be

$$\underbrace{1+\cdots+1}_{\binom{n}{k} \text{ times}}.$$

With this in mind, it is routine to see that only the laws of commutative rings are needed to establish the Binomial Theorem. Now notice that $n \mid \binom{n}{k}$ for all k such that $0 < k < n$, while $\binom{n}{0} = 1 = \binom{n}{n}$. This observation yields

Fact. Let **R** be a commutative ring of characteristic n. Then for all $x, y \in R$

$$(x+y)^n = x^n + y^n.$$

Moreover, the map sending $a \mapsto a^n$ for all $a \in R$ is a homomorphism.

This map, used in later parts of algebra, is called the **Frobenius map**.

Finally, we know that $\mathbb{Z}$ is an integral domain and so $\mathbb{Z}[x]$ is also an integral domain. However, even though $\mathbb{Z}$ is a principal ideal domain, it turns out that $\mathbb{Z}[x]$ is *not* a principal ideal domain. Establishing this fact is a task left to the graduate students in one of the Problem Sets. Even though $\mathbb{Z}[x]$ is not a principal ideal domain, it turns out that it is still a unique factorization domain and that while some of the ideals of $\mathbb{Z}[x]$ cannot be generated by some single element, it is nevertheless true that all the ideals of $\mathbf{Z}[x]$ can be generated by some *finite set* of elements. These are consequences of more general theorems (of Gauss and Hilbert) that are the primary objectives of this sequence of lectures.

6.2 Polynomials over a Unique Factorization Domain

We know that both $\mathbb{Z}$ and $\mathbf{F}[x]$, where **F** is a field, are principal ideal domains, but that neither $\mathbb{Z}[x]$ nor $\mathbf{F}[x, y]$ are principal ideal domains. It is a theorem of Gauss that $\mathbb{Z}[x]$ is nevertheless a unique factorization domain. With little change to the proof of Gauss, we have the following theorem.

Unique Factorization Theorem for Polynomials over a Unique Factorization Domain. *Let* **D** *be a unique factorization domain. Then* $\mathbf{D}[x]$ *is also a unique factorization domain.*

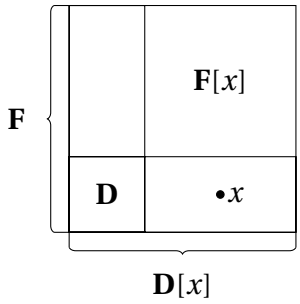


Figure 6.2: Linking **D** and **F**[x]

The proof of this theorem depends on three lemmas. Let **F** be the field of fractions of the unique factorization domain **D**. The diagram below may help to understand how the proof will work.

The information we start with is that **D** is a unique factorization domain, the **F** is the field of fractions of **D** (and therefore closely linked to **D**), and that **F**[x] is also a unique factorization domain. Observe that

$$\mathbf{D} \subseteq \mathbf{D}[x] \subseteq \mathbf{F}[x].$$

When Gauss tackled this problem, he had $\mathbb{Z}$ in place of **D** and $\mathbb{Q}$ in place of **F**.

We will say that a polynomial in **D**[x] is **primitive** provided there is no irreducible of **D** that divides all of the coefficients of the polynomial.

Lemma A.
Let **D** *be a unique factorization domain and let* **F** *be its field of fractions. Let $p(x)$ be a nonzero polynomial with coefficients in* **F**. *There is an element $c \in F$ and a primitive polynomial $q(x) \in$* **D**[x] *such that $p(x) = cq(x)$. Moreover, up to multiplication by units of* **D**, *the element c and the coefficients of $q(x)$ are unique.*

Proof. Pick $a_0, b_0, a_1, b_1, \ldots, a_n, b_n \in D$ so that

$$p(x) = \frac{a_0}{b_0} + \frac{a_1}{b_1}x + \cdots + \frac{a_n}{b_n}x^n.$$

Let $b = b_0 b_1 \cdots b_n$. Then

$$bp(x) = c_0 + c_1 x + \cdots + c_n x^n$$

for certain elements $c_0, c_1, \ldots, c_n \in D$. Let d be a greatest common divisor of $c_0, c_1, \ldots, c_n$. Factoring d out of $c_0 + c_1 x + \cdots + c_n x^n$ leaves a primitive polynomial $q(x)$ such that

$$bp(x) = dp_1(x).$$

So that $c = \frac{d}{b}$. Then $p(x) = cp_1(x)$, establishing the existence part of Lemma A. This argument was just clearing the denominators and factoring the remaining coefficients are must as possible.

Now consider the uniqueness assertion. Suppose $c, c^* \in F$ and $q(x), q^*(x) \in D[x]$ with both $q(x)$ and $q^*(x)$ primitive such that

$$cq(x) = p(x) = c^* q^*(x).$$

Pick $r, s, r^*, s^* \in D$ so that $c = \frac{r}{s}$ and $c^* = \frac{r^*}{s^*}$ and so that r and s are relatively prime as are r^* and s^*. So we have

$$s^* r q(x) = s r^* q^*(x).$$

Now let t be any prime in $\mathbf{D}$ so that $t \mid s^*$. We know that t cannot divide r^* and it cannot divide each of the coefficients of $q^*(x)$ since that polynomial is primitive. Therefore it must divide s. So, factoring s^* into primes, we see that $s^* \mid s$. In a like manner, we can conclude that $s \mid s^*$. This means that s and s^* are associates. Pick a unit u of $\mathbf{D}$ so that $s^* = su$. So we find after cancellation

$$rq(x) = ur^*q^*(x).$$

Applying the same reasoning to r^* and r, we can find a unit v of $\mathbf{D}$ so that $r^* = vr$. This gives

$$q(x) = uvq^*(x) \text{ and } c = \frac{r}{s} = \frac{ru}{su} = \frac{ru}{s^*} = \frac{rvu}{s^*v} = \frac{r^*u}{s^*v} = \frac{r^*}{s^*}\frac{u}{v} = c^*\frac{u}{v}.$$

But both uv and $\frac{u}{v}$ are units of $\mathbf{D}$. This finishes the proof of uniqueness up to multiplication by units of $\mathbf{D}$. □

An immediate consequence of Lemma A is that if $p(x)$ and $q(x)$ are primitive polynomials in $\mathbf{D}[x]$ that are associates in $\mathbf{F}[x]$, then they are already associates in $\mathbf{D}[x]$.

Gauss's Lemma. *Let* $\mathbf{D}$ *be a unique factorization domain. The product of two primitive polynomials in* $\mathbf{D}[x]$ *is again primitive.*

Proof. Let $f(x)$ and $g(x)$ be primitive and put $h(x) = f(x)g(x)$. Suppose, for the sake of contradiction, that t is an irreducible of $\mathbf{D}$ that divides all the coefficients of $h(x)$. So t is prime and the ideal (t) is a prime ideal. This means that $\mathbf{D}/(t)$ is an integral domain. Let η denote the quotient map from $\mathbf{D}$ to $\mathbf{D}/(t)$. Using the Homomorphism Extension Theorem for Polynomials, we know there is a unique homomorphism $\hat{\eta} : \mathbf{D}[x] \to \mathbf{D}/(t)[x]$ so that $\hat{\eta}(x) = x$. What $\hat{\eta}$ does is simply apply η to each of the coefficients of the polynomial given as input.

Now observe in $\mathbf{D}/(t)[x]$ we have

$$\begin{aligned} 0 &= \hat{\eta}(h(x)) \text{ since each coefficient of } h(x) \text{ is divisible by } t. \\ &= \hat{\eta}(f(x)g(x)) \\ &= \hat{\eta}(f(x)\hat{\eta}(g(x)) \end{aligned}$$

But t cannot divide all the coefficients of $f(x)$ nor all the coefficients of $g(x)$, since these polynomials are primitive. So $\hat{\eta}(f(x)) \neq 0 \neq \hat{\eta}(g(x))$ in $\mathbf{D}/(t)[x]$. Since $\mathbf{D}/(t)[x]$ is an integral domain, we have uncovered a contradiction. So $h(x)$ must be primitive. □

The proof just given is certainly in the fashion of the 20$^{\text{th}}$ century. Here is a proof that appeals directly to basic principles. It is much more like the reasoning of Gauss.

A more basic proof of Gauss's Lemma. Let

$$\begin{aligned} f(x) &= a_0 + a_1x + \cdots + a_nx^n \\ g(x) &= b_0 + b_1x + \cdots + b_mx^m \end{aligned}$$

be primitive polynomials. Put $h(x) = f(x)g(x) = c_0 + c_1x + \cdots + c_{n+m}x^{n+m}$. We know that

$$c_k = \sum_{i+j=k} a_ib_j.$$

Let t be a prime of $\mathbf{D}$. Pick ℓ as small as possible so that $t \nmid a_\ell$ and pick r as small as possible so that $t \nmid b_r$. We can do this since $f(x)$ and $g(x)$ are primitive. Then

$$c_{\ell+r} = (a_0 b_{\ell+r} + a_1 b_{\ell+r-1} + \cdots + a_{\ell-1} b_{r+1}) + a_\ell b_r + (a_{\ell+1} b_{r-1} + \cdots + a_{\ell+r} b_0).$$

(Be generous in understanding this equation. Depending on the values of ℓ and r some terms in the first and last pieces of the sum may be missing.) Then t divides the first and last pieces of this sum, but not the middle term $a_\ell b_r$. This means that t cannot divide $c_{\ell+r}$. Hence, no prime can divide all the coefficients of $h(x)$. So $h(x)$ must be primitive. □

Lemma B.
Let $\mathbf{D}$ *be a unique factorization domain and let* $\mathbf{F}$ *be its field of fractions. If* $f(x) \in \mathbf{D}[x]$ *is irreducible and of positive degree, then* $f(x)$ *is also irreducible in* $\mathbf{F}[x]$

Proof. Observe that $f(x)$ must be primitive since it is of positive degree and irreducible in $\mathbf{D}[x]$. Now suppose that $f(x) = g(x)h(x)$ for some polynomials $g(x), h(x) \in F[x]$. According to Lemma A, we can pick $c \in F$ and primitive polynomials $g^*(x)$ and $h^*(x)$ so that $f(x) = cg^*(x)h^*(x)$. Pick $a, b \in D$ so that a and b are relatively prime and $c = \frac{a}{b}$. Then

$$bf(x) = c\left(g^*(x)h^*(x)\right).$$

Gauss's Lemma tells us that $g^*(x)h^*(x)$ is primitive. So the uniqueness assertion of Lemma A gives us two unit u and v of $\mathbf{D}$ such that

$$ub = c \text{ and } f(x) = vg^*(x)h^*(x).$$

Since $f(x)$ is irreducible in $\mathbf{D}[x]$ it must be that one of $g^*(x)$ and $h^*(x)$ is a unit and thus has degree 0. But then one of $g(x)$ and $h(x)$ must also have degree 0 and be, therefore, a unit of $\mathbf{F}$. This means that $f(x)$ is irreducible in $\mathbf{F}[x]$. □

Here is a proof of the Unique Factorization Theorem for Polynomials over a Unique Factorization Domain.

Proof. Let $f(x) \in D[x]$ be a nonzero polynomial. We begin by letting c be a greatest common divisor of the coefficients of $f(x)$ we obtain a primitive polynomial $g(x)$ so that

$$f(x) = cg(x).$$

Either c is a unit of $\mathbf{D}$ or else we can factor it into irreducibles over $\mathbf{D}$. Observe that apart from units $g(x)$ has no factors of degree 0 in $\mathbf{D}[x]$ since $g(x)$ is primitive. Thus any proper factorization of $g(x)$ over $\mathbf{D}[x]$ must produce factors of properly smaller degree. In this way we see that $f(x)$ can be factored into irreducibles over $\mathbf{D}[x]$.

To see that the factorization of $f(x)$ is unique suppose

$$f(x) = c_0 c_1 \cdots c_m g_0(x) g_1(x) \cdots g_n(x)$$
$$f(x) = d_0 d_1 \cdots d_k h_0(x) h_1(x) \cdots h_\ell(x)$$

are factorization of $f(x)$ into irreducibles over $\mathbf{D}[x]$ so that $c_0, c_1, \ldots, c_m, d_0, d_1, \ldots d_k$ are irreducibles of degree 0 while the remaining irreducible factors have positive degree. Irreducibles in $\mathbf{D}[x]$ of positive degree are primitive. Using Gauss's Lemma and Lemma A, we see that

(a) $c_0c_1\cdots c_m$ and $d_0d_1\cdots d_k$ are associates over $\mathbf{D}$. Since $\mathbf{D}$ is a unique factorization domain, we find that $m = k$ and perhaps after some reindexing c_i and d_i are associates for all $i \le m$.

(b) $g_0(x)g_1(x)\cdots g_n(x)$ and $h_0(x)h_1(x)\cdots h_\ell(x)$ are associates over $\mathbf{D}[x]$ and hence over $\mathbf{F}[x]$. By Lemma B, these polynomials are irreducible over $\mathbf{F}[x]$. Because $\mathbf{F}[x]$ is a unique factorization domain, we find that $n = \ell$ and, after a suitable reindexing, that $g_j(x)$ and $h_j(x)$ are associates over $\mathbf{F}[x]$, for all $j \le n$. But the $g_j(x)$'s and the $h_j(x)$'s are primitive, so by Gauss's Lemma and Lemma A they must also be associates over $\mathbf{D}[x]$.

This establishes the uniqueness of the factorization. □

An easy induction shows that

if $\mathbf{D}$ is a unique factorization domain, then so is $\mathbf{D}[x_0, x_1, \ldots, x_{n-1}]$.

Eisenstein's Criteria. *Let* $\mathbf{D}$ *be a unique factorization domain and let* $\mathbf{F}$ *be its field of fractions. Let* $f(x) = a_0 + a_1x + \cdots + a_nx^n \in D[x]$ *where* $a_n \neq 0$ *and* n *is positive. If there is an irreducible* $p \in D$ *such that*

1. $p \mid a_i$ *for all* $i < n$,
2. $p \nmid a_n$, *and*
3. $p^2 \nmid a_0$,

then $f(x)$ *is irreducible in* $\mathbf{F}[x]$. *If, in addition,* $f(x)$ *is primitive, then* $f(x)$ *is irreducible in* $\mathbf{D}[x]$.

Proof. First suppose that $f(x)$ is primitive and that it satisfies the given criteria. Suppose $f(x) = g(x)h(x)$ is a factorization of $f(x)$ over $\mathbf{D}[x]$. Let $g(x) = b_0 + b_1x + \ldots$ and $h(x) = c_0 + c_1x + dots$. Then $a_0 = b_0c_0$. Now $p \mid a_0 = b_0c_0$ but $p^2 \nmid b_0c_0$. So p divides exactly one of b_0 and c_0. It is harmless to suppose that $p \mid b_0$ but $p \nmid c_0$. Now p cannot divide all the coefficients of $g(x)$ since then it would divide all the coefficients of $f(x)$, even a_n. Pick k as small as possible so that $p \nmid b_k$. Observe that

$$a_k = b_0c_k + b_1c_{k-1} + \cdots + b_{k-1}c_1 + b_kc_0.$$

Now $p \mid b_0c_k + b_1c_{k-1} + \cdots + b_{k-1}c_1$ but $p \nmid b_k$ and $p \nmid c_0$. Since p is prime we get $p \nmid b_kc_0$. But this implies that $p \nmid a_k$. We conclude that $k = n$. But this means that $\deg f(x) = \deg g(x)$ and that $\deg h(x) = 0$. So $h(x) \in D$. Since $f(x) = g(x)h(x)$ and $f(x)$ is primitive, we find that $h(x)$ must actually be a unit of $\mathbf{D}$. So $f(x)$ is irreducible in $\mathbf{D}[x]$. By Lemma B it is also irreducible in $\mathbf{F}[x]$.

Now consider the case when $f(x)$ is not primitve. Let c be the greatest common divisor of the coefficients of $f(x)$. So $f(x) = cf^*(x)$ where $f^*(x)$ primitive. Now observe that $p \nmid c$ since $c \mid a_n$. By the primeness of p, it follows that $f^*(x)$ satisfies Eisenstein's Criteria for the prime p. Hence $f^*(x)$ is irreducible in $\mathbf{D}[x]$ and hence in $\mathbf{F}[x]$ by Lemma B. But c is a unit of $\mathbf{F}[x]$ so $f(x)$ is an associate over $\mathbf{F}[x]$ of an irreducible. This makes $f(x)$ irreducible over $\mathbf{F}[x]$, as desired. □

Here is an example of what is at stake. The polynomial $6 + 3x$ has integer coefficients and it satisfies Eisenstein's Criteria with $p = 2$. So it is irreducible over $\mathbb{Q}[x]$ by Eisenstein (but really, every polynomial of degree 1 is irreducible over $\mathbb{Q}[x]$). However, $6 + 3x = 3(2 + x)$ is a proper factorization over $\mathbb{Z}[x]$ since 3 is not a unit for $\mathbb{Z}[x]$. Of course, $6 + 3x$ is also not primitive.

Eisenstein's Criteria is one of a large assortment of techniques for show that polynomials are irreducible—especially polynomial is rings like $\mathbb{Z}[x], \mathbb{Z}[x, y], \ldots$.

6.3 Hilbert's Basis Theorem

Now we know that if **D** is a principal ideal domain, then **D**$[x]$ is a unique factorization domain, even though it might not be a principal ideal domain. Here we will see that, in some measure, **D**$[x]$ retains some features of a principal ideal domain.

We will call a ring **R Noetherian** provided every ideal of **R** is finitely generated. So every principal ideal domain is Noetherian.

Theorem Characterizing Noetherian Rings. *Let* **R** *be a ring. The following are logically equivalent.*

(a) **R** *is a Noetherian ring.*

(b) *Every ascending chain of ideals of* **R** *is finite.*

(c) *Every nonempty collection of ideals of* **R** *has a maximal member with respect to the ordering by inclusion.*

Proof.

(a)⇒(b)
Suppose $I_0 \subseteq I_1 \subseteq I_2 \subseteq \ldots$ is an ascending chain of ideals of **R**. Then $\bigcup_{i\in\mathbb{N}} I_i$ is also an ideal of **R**. Because **R** is Noetherian there is a finite set X so that $(X) = \bigcup_{i\in\mathbb{N}} I_i$. Because X is finite and the ideals form a chain, there must be a natural number k so that $X \subseteq I_k$. But then

$$I_k \subseteq \bigcup_{i\in\mathbb{N}} I_i = (X) \subseteq I_k.$$

It follows that $I_k = I_{k+1} = I_{k+2} = \ldots$. So the ascending chain of ideals is finite.

(b)⇒(c)
Let $\mathscr{F}$ be a nonempty family of ideals of **R**. Since every ascending chain of ideals of **R** is finite, it follows that every chain of ideals in $\mathscr{F}$ has an upper bound in $\mathscr{F}$. By Zorn's Lemma, $\mathscr{F}$ must have maximal members.

(c)⇒(a)
Let I be an ideal of **R**. Let $\mathscr{F} = \{J \mid J \subseteq I \text{ and } J \text{ is a finitely generated ideal}\}$. Let M be a maximal member of $\mathscr{F}$. Then $M \subseteq I$. Were $M \neq I$ we could pick $a \in I \setminus M$. But then $M \subsetneq (M \cup \{a\}) \subseteq I$. Since $(M \cup \{a\})$ is finitely generated, this violate the maximality of M. So $I = M$, which is finitely generated. □

Hilbert's Basis Theorem. *If* **R** *is a commutative Noetherian ring, then so is* **R**$[x]$.

Proof. Let I be any ideal of **R**$[x]$ and let m be any natural number. Define

$$I(m) := \{a \mid a \text{ is the leading coefficient of a polynomial of degree } m \text{ that belongs to } I\} \cup \{0\}$$

The graduate students should routinely check that $I(m)$ is always an ideal of **R**. It should also be clear that $I(m) \subseteq I(m+1)$.

Fact. Suppose I and J are ideals of **R**$[x]$ with $I \subseteq J$. If $I(m) = J(m)$ for all natural numbers m, then $I = J$.

To establish this fact one should consider $f(x) \in J$ with the object of proving that $f(x) \in I$. This can be done by induction on the degree of $f(x)$. This induction is left for the pleasure of the graduate students.

Now consider an ascending chain $I_0 \subseteq I_1 \subseteq I_2 \subseteq \dots$ of ideals of $\mathbf{R}[x]$. There is an associated grid of ideals on $\mathbf{R}$.

$$\begin{array}{ccccccc}
\vdots & & \vdots & & \vdots & & \\
\cup & & \cup & & \cup & & \\
I_0(2) & \subseteq & I_1(2) & \subseteq & I_2(2) & \subseteq & \cdots \\
\cup & & \cup & & \cup & & \\
I_0(1) & \subseteq & I_1(1) & \subseteq & I_2(1) & \subseteq & \cdots \\
\cup & & \cup & & \cup & & \\
I_0(0) & \subseteq & I_1(0) & \subseteq & I_2(0) & \subseteq & \cdots
\end{array}$$

The family $\mathscr{F} = \{I_i(j) \mid i\, j \in \mathbb{N}\}$ displayed on this grid is a nonempty family of ideals of $\mathbf{R}$. It must have a maximal member, say $I_n(m)$. Each of the finitely many rows associated an argument j with $j \leq m$ is an ascending chain and can only extend to the right finitely far. Let ℓ be a natural number large enough so that none of these finitely many rows extends beyond ℓ steps. Notice that $n \leq \ell$. Then $I_\ell(i) = I_{\ell+k}(i)$ for all $i \leq m$ for all natural numbers k, while $I_\ell\, i = I_n(m) = I_{\ell+k}(i)$ whenever $i > m$. Now the Fact asserted above tells us that $I_\ell = I_{\ell+k}$ for all natural numbers k. So the ascending chain $I_0 \subseteq I_1 \subseteq I_2 \subseteq \cdots$ is finite, as desired. □

It follows that if $\mathbf{R}$ is any commutative Noetherian ring, then $\mathbf{R}[x_0, x_1, \dots, x_{n-1}]$ is also a commutative Neotherian ring. This theorem has a fundamental role to play in commutative algebra and algebraic geometry. The proof I gave above has the charm of an illuminating diagram, but it doesn't allow us to lay our hands directly on a finite generating set for an ideal I of $\mathbf{R}[x]$. Coupled with the proof of the Fact embedded in our proof, some headway could be made in this direction.

6.4 Problem Set 5

Algebra Homework, Edition 5
Sixth Week
Rings of Polynomials

Problem 18.
Is the polynomial $y^3 - x^2y^2 + x^3y + x + x^4$ irreducible in $\mathbb{Z}[x, y]$?

Problem 19.
Let **R** be a principal ideal domain, and let I and J be ideals of **R**. IJ denotes the ideal of **R** generated by the set of all elements of the form ab where $a \in I$ and $b \in J$. Prove that if $I + J = R$, then $I \cap J = IJ$.

Problem 20.
Let **D** be a unique factorization domain and let I be a nonzero prime ideal of $\mathbf{D}[x]$ which is minimal among all the nonzero prime ideals of $\mathbf{D}[x]$. Prove that I is a principal ideal.

Problem 21.
Let **D** be a subring of the field **F**. An element $r \in F$ is said to be **integral over D** provided there is a monic polynomial $f(x) \in \mathbf{D}[x]$ such that r is a root of $f(x)$. For example, the real number $\sqrt{2}$ is integral over the ring of integers since it is a root of $x^2 - 2$.

Now suppose **D** is a unique factorization domain and **F** is its field of fractions. Prove that the set of elements of **F** that are integral over **D** coincides with D itself.

Problem 22.
Let **R** be a commutative ring and let **S** be a subring of **R** so that **S** is Noetherian. Let $a \in R$ and let $\mathbf{S}'$ be the subring of **R** generated by $S \cup \{a\}$. Prove that $\mathbf{S}'$ is Noetherian.

LECTURE 7

MODULES, A GENERALIZATION OF VECTOR SPACES

7.1 MODULES OVER A RING

A vector space over a field **F** is a set of vectors, including a zero vector, that has a two-place operation for vector addition, a one-place operation for form the negative of a vector, a one-place operation for each element of F that can be used to scale vectors. Most of us were brought up to consider a kind of two-place operation from multiplying a vector by a scalar. For example, in the standard 2-dimensional vector space over the reals, the when the vector $(2,6)$ is multiplied by the scalar 0.5 the resulting vector is $(1,3)$, a vector pointing in the same direction as $(2,6)$ but it is scaled down—it is only half as long. The one-place operation that sends each real pair (a,b) to $(0.5a, 0.5b)$ precisely captures the effect of multiplication by the scalar 0.5. Of course, the advantage to us of construing scalar multiplication as a system of one-place operations is that then vector spaces fit into our overall view of algebraic systems in general.

Let **F** be a field. We say that $\langle V, +, , -, 0, r\cdot\rangle_{r\in F}$ is a **vector space over F** provided all of the equation below hold.

$$\begin{aligned} x+(y+z) &= (x+y)+z & (r+s)\cdot x &= rx+sx \\ x+y &= y+x & r(x+y) &= rx+ry \\ -x+x &= 0 & (rs)x &= r(sx) \\ x+0 &= x & 1x &= x \end{aligned}$$

for all $x, y, z \in V$ and $r, s \in F$.

We have followed the customary practice of using the same symbol + to denote both the addition in the field of scalars and the addition in the vector space. Really, they are different in all but a few cases. The same might be said for using juxtaposition to denote the multiplication in the ring a the (one-place functional) action of a scalar on a vector. In the equations above rs is the product in the ring whereas $r(sx)$ means the action, consecutively, of two scalings.

We obtain the notion of a module over a ring by replacing the field **F** with an arbitrary ring **R**. So let **R** be a ring. We say that $\langle V, +, , -, 0, r\cdot\rangle_{r\in R}$ is a **module over R** provided all of the equation below hold.

$$\begin{aligned} x+(y+z) &= (x+y)+z & (r+s)\cdot x &= rx+sx \\ x+y &= y+x & r(x+y) &= rx+ry \\ -x+x &= 0 & (rs)x &= r(sx) \\ x+0 &= x & 1x &= x \end{aligned}$$

for all $x, y, z \in V$ and $r, s \in R$. Some people would say *left unitary* **R**-*module* for this notion. The "left" comes from writing the scalars on the left—there is a companion notion of right modules. The "unitary" comes from the stipulation $1x = x$. Many of the most striking properties of vector spaces, rely on the fact that every nonzero element of a field in a unit. Still, modules in general retain some of the nice features of vector spaces.

There is another source of modules. Let I be an ideal of **R**. Then $\langle I, +, -, 0, r\cdot\rangle_{r\in R}$ is clearly an **R**-module. This would even be true if I were a "left" ideal of **R**. Indeed, the left ideals of **R** are essentially the same as the sub **R**-modules of **R**. Below we will only be concerned with **R**-modules when **R** is a commutative ring. In this case, the ideals of **R** and the submodules of **R** coincide.

In fact, we will be almost exclusively concerned with modules whose underlying ring is a principal ideal domain. The familiar ring $\mathbb{Z}$ of integers and rings of the form $\mathbf{F}[x]$, where **F** is a field, are examples of principal ideal domains. Reflect a moment on the $\mathbb{Z}$-modules. Did you notice that the investigation of the $\mathbb{Z}$-modules differs in no important way from the investigation of Abelian groups?

Let **V** be a finite dimensional vector space over a field **F**. The linear operators (alias endomorphisms) of **V** can be acted on in an obvious way by the polynomials in $\mathbf{F}[x]$. Under this action, the linear operators of **V** form a module over $\mathbf{F}[x]$. Investigation of the structure of such modules leads to some of the deeper results in linear algebra.

7.2 Free Modules

A module **F** over a nontrivial ring **R** is said to be **free on** a set $B \subseteq F$ provided for every **R**-module **M** and every function $\varphi : B \to M$ there is a unique homomorphism $\psi : \mathbf{F} \to \mathbf{M}$ that extends φ. We will say that **F** is a **free R-module** provided it is free on some set. In the context of vector spaces, we know that every vector space has a basis B and that any map from B into another vector space over the same field extends uniquely to a linear transformation. This means that every vector space is free on any of its bases. This fails for modules in general. The free modules are much more like vector spaces.

The Uniqueness Theorem for Free Modules. *Let* **F** *be a module over a nontrivial ring* **R** *that is free on* B *and let* $\mathbf{F}*$ *be a module over* **R** *that is free on* B^*. *If* $|B| = |B^*|$, *then* **F** *and* $\mathbf{F}^*$ *are isomorphic.*

Proof. Let φ be a one-to-one map from B onto B^*. Let ψ extend φ to a homomorphism from **F** into $\mathbf{F}^*$. Likewise let ψ^* extend φ^{-1} to a homomorphism from $\mathbf{F}^*$ into **F**. Then $\psi \circ \psi^*$ is an endomorphism of **F** extending the identity map on B. The identity on F is also such an endomorphism. By the uniqueness of such extensions, we find $\psi \circ \psi^*$ is that identity map on F. Likewise, $\psi^* \circ \psi$ is the identity map on F^*. So ψ is an isomorphism from **F** onto $\mathbf{F}^*$ and ψ^* is its inverse. □

Observe that $\mathbf{R}$ is an $\mathbf{R}$-module that is free on $\{1\}$ and that the trivial $\mathbf{R}$-module is free on $\varnothing$. We will see that free $\mathbf{R}$-modules have a simple form. For this we employ the notion of direct sum of modules. Let $\mathbf{M}_i$ be an $\mathbf{R}$-module for each $i \in I$, where I is any set. We define the **direct sum**

$$\bigoplus_{i\in I} M_i := \{\langle v_i | i \in I\rangle | v_i \in M_i \text{ for all } i \in I \text{ and all but finitely many of } v_i\text{'s are } 0\}$$

It is routine to check that this set is a subuniverse of $\prod_{i\in I}\mathbf{M}_i$. So $\bigoplus_{i\in I}\mathbf{M}_i$ is an $\mathbf{R}$-module.

The Structure Theorem for Free Modules. *Suppose* $\mathbf{F}$ *is a module over a nontrivial ring that is free on* B*. For each* $b \in B$ *let* $\mathbf{R}_b = \mathbf{R}$*. Then* $\mathbf{F}$ *is isomorphic to* $\bigoplus_{b\in B}\mathbf{R}_b$*.*

Proof. All be need to do is prove that $\bigoplus_{b\in B} R_b$ is free on a set of cardinality $|B|$. The set we are after is the set of all B-tuples that have 1 in exactly one position and 0 is all other positions. This is the "standard" basis familiar from linear algebra. The graduate students should enjoy filing in the rest of this proof. □

As in vector spaces, in modules generally we will say that a set X is **linearly independent** provided that for any finitely many distinct $v_0, v_1, \ldots, v_{n-1} \in X$ if $a_0 v_0 + a_1 v_1 + \cdots + a_{n-1} v_{n-1} = 0$, then $a_0 = a_1 = \cdots = a_{n-1} = 0$. In any module $\mathbf{M}$, a linearly independent subset that generates $\mathbf{M}$ is said to be a **basis** for $\mathbf{M}$.

Theorem Characterizing Free Modules. *Let* $\mathbf{R}$ *be a nontrivial ring and* $\mathbf{F}$ *be an* $\mathbf{R}$*-module.* $\mathbf{F}$ *is a free* $\mathbf{R}$*-module if and only if* $\mathbf{F}$ *has a basis.*

Proof. Suppose that $\mathbf{F}$ is a module over $\mathbf{R}$ that is free on B. Let $\mathbf{M}$ be the submodule of $\mathbf{F}$ generated by B. I leave it to the graduate students to check that $\mathbf{M}$ is also free on B. So there is an isomorphism from $\mathbf{M}$ onto $\mathbf{F}$ that extends the identity map on B. But any such extension must fix each element of M since $\mathbf{M}$ is generated by B. This means that $M = F$, and so we see that B generates $\mathbf{F}$. Next, observe that the subset of $\bigoplus_{b\in B} R_b$ consisting of those B-tuples with exactly one entry 1 and the rest 0 is evidently linearly independent. But $\bigoplus_{b\in B}\mathbf{R}_b$ and $\mathbf{F}$ are isomorphic via an isomorphism that sends our linear independent subset of the direct sum to B. As the image of a linearly independent set under an isomorphism is again linearly independent, we find the B is linearly independent. Therefore, B is a basis for $\mathbf{F}$.

Now suppose that B is a basis for $\mathbf{F}$. Just as in linear algebra, we can show that every element of F can be expressed uniquely as a linear combination of elements of B. Suppose that $\mathbf{M}$ is an $\mathbf{R}$-module and let $\varphi : B \to M$. Define $\psi : F \to M$ via

$$\psi(w) := a_0\varphi(v_0) + \cdots + a_n\varphi(v_n)$$

for all $w \in F$, where $a_0 v_0 + \cdots + a_n v_n$ is the unique linear combination of distinct elements $v_0, \ldots, v_n \in B$ that represents w. It is routine to prove that ψ is a homomorphism. So $\mathbf{F}$ is free on B. □

One of the most useful features of vector spaces is that any two bases of the same space have the same cardinality. This gives us a notion of dimension in vector spaces. This property is lost in some free modules. On the other hand, it is often true.

The Dimension Theorem for Free Modules. *Let* $\mathbf{R}$ *be a ring that has a division ring has a homomorphic image. Any two bases of a free* $\mathbf{R}$*-module have the same cardinality.*

Proof. Let I be an ideal of $\mathbf{R}$ so that $\mathbf{R}/I$ is a division ring. Let $\mathbf{F}$ be a free $\mathbf{R}$ with basis B. Let E be the collection of all elements that can be written as linear combinations of elements of F using only coefficients from I. You should check that E is closed under the module operations, so $\mathbf{E}$ is a submodule of $\mathbf{F}$.

Observe that $\mathbf{F}/E$ can be construed as an $\mathbf{R}/I$-module in a natural way. (Hint: define $(a+I)(v+E)$ to be $av+E$. Be sure to check that this definition is sound.) Now let $B^* = \{v+E \mid v \in B\}$.

We want to demonstrate that B^* is linearly independent for the $\mathbf{R}/I$-module $\mathbf{F}/E$. Suppose $v_0 + E, v_1 + E, \ldots, v_n + E$ are distinct members of $B^* = B/E$. Take $a_0, \ldots, a_n \in R$. Observe the following sequence of steps.

$$\begin{aligned} 0 + E &= \sum_{i \le n} (a_i + I)(v_i + E) \\ &= \sum_{i \le n} (a_i v_i + E) \\ &= \left(\sum_{i \le n} a_i v_i \right) + E \end{aligned}$$

This means that if $0 + E = \sum_{i \le n} (a_i + I)(v_i + E)$, then $\sum_{i \le n} a_i v_i \in E$. By the definition of E, there are $w_0, \ldots, w_m \in F$ and $c_0, \ldots, c_m \in I$ so that

$$\sum_{i \le n} a_i v_i = \sum_{j \le m} c_j w_j.$$

Now because B generates $\mathbf{F}$ we see that each of the w_j's can be written as a linear combination of elements of B. This entails that $\sum_{j \le m} c_j w_j$ can be rewritten as a linear combination of elements of B with the coefficients all belonging to I. Let $\sum_{k \le \ell} d_k u_k$ be such a linear combination. But now

$$0 = \sum_{i \le n} a_i v_i - \sum_{k \le \ell} d_k u_k.$$

The expression on the right can be rewritten as a linear combination of distinct elements of B. The coefficients of the linear combination can be of three forms:

$$a_i - d_k \qquad \text{or} \qquad a_i \qquad \text{or} \qquad -d_k$$

depending on whether $v_i = u_k$. All of these coefficients must be 0. Notice that in the first alternative that we get $a_i \in I$ and in the second $a_i = 0 \in I$. So we find that $a_i \in I$ for all i. This means that $a_i + I = 0 + I$ for all i and concludes the proof that B^* is linearly independent.

That B^* generates $\mathbf{F}/E$ follows easily from the fact that B generates $\mathbf{F}$. So B^* is a basis of $\mathbf{F}/E$.

Contention. $|B| = |B^*|$.

In fact, the quotient map that send $v \mapsto v + E$ for $v \in B$ is one-to-one. To see this, suppose $v, v' \in B$ and $v + b = v' + E$. Then $v - v' \in E$. By the same device we used above, we can write $v - v'$ as a linear combination of distinct elements of B with coefficients drawn from I. Let $\sum_{k \le \ell} d_k u_k$ be such a linear combination. This gives

$$0 = v' - v + \sum_{k \le \ell} d_k u_k.$$

Notice that v' and v might well appear among the u_k's, but none of the d_k's is a unit of **R** since I must be a proper ideal. Nevertheless, rewriting the right side as a linear combination of distinct elements of B must result in all the coefficients being 0. This can only happen if $v' = v$, establishing our contention.

At this point we know that every basis of **F** is the same cardinality has some basis of $\mathbf{F}/E$. So the last thing we need is that any two bases of a free module over a division ring have the same cardinality. Proving this only requires a careful examination of any standard proof that any two bases of a vector space have the same cardinality. One must see that the commutative property of multiplication in the field plays no role in such a proof. It also pays to notice the role division has to play in such a proof. So commit due diligence on some linear algebra book to complete this proof. □

The unique dimension guaranteed by the theorem above is called the **rank** of the free modules.

By the Maximal Ideal Theorem, we know that any nontrivial commutative ring **R** has a maximal ideal I and so $\mathbf{R}/I$ is actually a field. So we have the following corollary.

Corollary 7.2.1. *Let* **R** *be a nontrivial commutative ring. Any two bases of the same free* **R***-module must have the same cardinality.*

Suppose **R** is a nontrivial commutative ring and **F** is a free **R**-module. By the **rank** of **F** we mean the cardinality of any base of **F**.

7.3 Problem Set 6

Algebra Homework, Edition 6
Seventh Week
Ideals Yet Again

Problem 23.

(a) Prove that $(2, x)$ is not a principal ideal of $\mathbb{Z}[x]$.

(b) Prove that (3) is a prime ideal of $\mathbb{Z}[x]$ that is not a maximal ideal of $\mathbb{Z}[x]$.

Problem 24.
Show that any integral domain satisfying the descending chain condition on ideals is a field.

Problem 25.
Prove the following form of the Chinese Remainder Theorem: Let $\mathbf{R}$ be a commutative ring with unit 1 and suppose that I and J are ideals of R such that $I + J = R$. Then

$$\frac{\mathbf{R}}{I \cap J} \cong \frac{\mathbf{R}}{I} \times \frac{\mathbf{R}}{J}.$$

Problem 26.
Prove that there is a polynomial $f(x) \in \mathbb{R}[x]$ such that

(a) $f(x) - x$ belongs to the ideal $(x^2 + 2x + 1)$;

(b) $f(x) - x^2$ belongs to the ideal $(x - 1)$, and

(c) $f(x) - x^3$ belongs to the ideal $(x^2 - 4)$.

LECTURE 8

SUBMODULES OF FREE MODULES OVER A PID

The objective here is to prove that, over a principal ideal domain, every submodule of a free is also a free module and that the rank of a free module is always at least as large of the ranks of its submodules.

So let **R** be a (nontrivial) principal ideal domain. We know that **R** is a free **R**-module of rank 1. What about the submodules of **R**? Suppose **E** is such a submodule. It is clear that E is an ideal and, in fact, that the ideals of **R** coincide with the submodules of **R**. In case **E** is trivial (that is the sole element of E is 0) we see that **E** is the free **R**-module of rank 0. So consider the case that **E** is nontrivial. Since **R** is a principal ideal domain we pick $w \neq 0$ so that **E** is generated by w. That is $E = \{r\,w \mid r \in R\}$. Since we know that **R** has $\{1\}$ as a basis, we see that the map that sends 1 to w extends to a unique module homomorphism from **R** onto **E**. Indeed, notice $h(r \cdot 1) = r \cdot h(1) = r\,w$ for all $r \in R$. But the homomorphism h is also one-to-one since

$$\begin{aligned} h(r) &= h(s) \\ r\,h(1) &= s\,h(1) \\ r\,w &= s\,w \\ r &= s \end{aligned}$$

where the last step follows because integral domains satisfy the cancellation law and $w \neq 0$. In this way we see that **E** is isomorphic to the free **R**-module of rank 1. We also see that $\{w\}$ is a basis for **E**.

So we find that at least all the submodules of the free **R**-module of rank 1 are themselves free and have either rank 0 or rank 1. We can also see where the fact that **R** is a principal ideal domain came into play.

The Freedom Theorem for Modules over a PID.
Let **R** *be a principal ideal domain, let* **F** *be a free* **R***-module and let* **E** *be a submodule of* **F**. *Then* **E** *is a free* **R***-module and the rank of* **E** *is no greater than the rank of* **F**.

Proof. Since trivial modules (those whose only element is 0) are free modules of rank 0, we suppose below that **E** is a nontrivial module. This entails that **F** is also nontrivial.

Let B be a basis for $\mathbf{F}$ and $C \subseteq B$. Because $\mathbf{F}$ is not the trivial module, we see that B is not empty. Let $\mathbf{F}_C$ be the submodule of $\mathbf{F}$ generated by C. Let $\mathbf{E}_C = \mathbf{E} \cap \mathbf{F}_C$. Evidently, C is a basis for $\mathbf{F}_C$. To see that $\mathbf{E}_C$ is free we will have to find a basis for it.

Suppose, for a moment, that C has been chosen so that $\mathbf{E}_C$ is known to be free and that $w \in B$ with $w \notin C$. Put $D := C \cup \{w\}$. Consider the map defined on D into R that sends all the elements of C to 0 and that sends w to 1. This map extends uniquely to a homomorphism φ from $\mathbf{F}_D$ onto $\mathbf{R}$ and it is easy to check (as hardworking graduate student will) that the kernel of φ is just F_C. By the Homomorphism Theorem, we draw the conclusion that $\mathbf{F}_D/F_C$ is isomorphic to $\mathbf{R}$ and that it is free of rank 1. What about $\mathbf{E}_D/E_C$? Observe that $E_C = E \cap F_C = E \cap F_D \cap F_C = E_D \cap F_C$. So we can apply the Second Isomorphism Theorem:

$$\mathbf{E}_D/E_C = \mathbf{E}_D/E_D \cap F_C \cong \mathbf{E}_D + \mathbf{F}_C/F_C.$$

But $\mathbf{E}_D + \mathbf{F}_C/F_C$ is a submodule of $\mathbf{F}_D/F_C$. This last is a free $\mathbf{R}$-module of rank 1. We saw above that every submodule of a free $\mathbf{R}$-module of rank 1 must be itself a free $\mathbf{R}$-module and have rank either 0 or 1. In this way, we find that either $\mathbf{E}_D = \mathbf{E}_C$ (in the rank 0 case) or else $\mathbf{E}_D/E_C$ is a free $\mathbf{R}$-module of rank 1. Let us take up this latter case. Let X be a basis for $\mathbf{E}_C$, which we assumed, for the moment, was free. Pick $u \in E_D$ so that $\{u/E_C\}$ is a basis for $\mathbf{E}_D/E_C$.

We contend that $X \cup \{u\}$ is a basis for $\mathbf{E}_D$. To establish linear independence, suppose $x_0, \ldots, x_{n-1}$ are distinct element of X, that $r_0, \ldots, r_n \in R$ and that

$$0 = r_0 x_0 + \cdots + r_{n-1} x_{n-1} + r_n u.$$

First notice that

$$r_n(u/E_C) = r_n u/E_C = (r_0 x_0 + \cdots + r_{n-1} x_{n-1} + r_n u)/E_C = 0/E_C.$$

Since $\{u/E_C\}$ is a basis for $\mathbf{E}_D/E_C$, we must have $r_n = 0$. This leads to

$$0 = r_0 x_0 + \cdots + r_{n-1} x_{n-1}.$$

But now since X is a basis for $\mathbf{E}_C$ we see that $0 = r_0 = \cdots = r_{n-1}$. So we find that $X \cup \{u\}$ is linearly independent.

To see that $X \cup \{u\}$ generates E_D, pick $z \in E_D$. Since $\{u/E_C\}$ is a basis for $\mathbf{E}_D/E_C$, pick $r \in R$ so that

$$z/E_C = r\,u/E_C.$$

This means that $z - r\,u \in E_C$. But X is a basis of $\mathbf{E}_C$. So pick $x_0, \ldots, x_{n-1} \in X$ and $r_0, \ldots, r_{n-1} \in R$ so that

$$z - r\,u = r_0 x_0 + \cdots + r_{n-1} x_{n-1}.$$

Surely this is enough to see that z is in the submodule generated by $X \cup \{u\}$. So this set generates $\mathbf{E}_D$ and we conclude that it must be a basis of $\mathbf{E}_D$.

In this way we see that for $C \subseteq D \subseteq B$ where D arises from adding an element to C, if E_C is free, then so is $\mathbf{E}_D$ and that either $E_D = E_C$ or a basis for $\mathbf{E}_D$ can be produced by adding just one element to a basis for $\mathbf{E}_C$.

With this in mind, we can envision a procedure for showing that $\mathbf{E}$ is free and its rank cannot be larger than that of $\mathbf{F}$. Notice that $E = E \cap F = E \cap F_B$. So $\mathbf{E} = \mathbf{E}_B$. The idea is simple. We will start with $\varnothing \subseteq B$. We observe that $\mathbf{F}_\varnothing = \mathbf{E}_\varnothing$ is the module whose sole element is 0. It is free of rank 0.

Next we select an element $w \in B$ and form $\varnothing \cup \{w\} = \{w\}$. We find that $\mathbf{E}_{\{w\}}$ is free of rank 0 or rank 1. We select another element and another and another... until finally all the elements of B have been selected. At this point we would have E_B is free and its rank can be no more than the total number of elements we selected, namely $|B|$ which is the rank of $\mathbf{F}$.

To carry out this program, in case B were finite or even countable, we could mount a proof by induction. You can probably see how it might be done. But we want to prove this for arbitary sets B. We could still pursue this inductive strategy openly by well-ordering B and using transfinite induction. By using the well-ordering we would always know what was meant by "pick the next element of B."

Instead, we will invoke Zorn's Lemma to short-circuit this rather long induction.

Let $\mathscr{F} = \{f \mid f \text{ is a function with } \operatorname{dom} f \subseteq B \text{ and } \operatorname{range} f \text{ a basis for } \mathbf{E}_{\operatorname{dom} f}\}$. Recalling that functions are certain kinds of sets of order pairs, we see that $\mathscr{F}$ is paritally ordered by set inclusion. Maybe it helps to realize that to assert $f \subseteq g$ is the same as asserting that g extends f. We note that $\mathscr{F}$ is not empty since the empty function (the function with empty domain) is a member of $\mathscr{F}$. To invoke Zorn's Lemma, let $\mathscr{C}$ be any chain included in $\mathscr{F}$. Let $h = \bigcup \mathscr{C}$. Evidently $f \subseteq h$ for all $f \in \mathscr{C}$. So h is an upper bound of $\mathscr{C}$. We contend that $h \in \mathscr{F}$. We ask the hard-working graduate students to check that the union of any chain of functions is itself a function. Once you do that bit of work, it should be evident that $\operatorname{dom} h = \bigcup\{\operatorname{dom} f \mid f \in \mathscr{C}\}$ and that $\operatorname{range} h = \bigcup\{\operatorname{range} f \mid f \in \mathscr{C}\}$. So it remains to show that $\operatorname{range} h$ is a basis for $E_{\operatorname{dom} h}$. To see that $\operatorname{range} h$ is a generating set, let z be an arbitrary element of $E_{\operatorname{dom} h} = E \cap F_{\operatorname{dom} h}$. Hence z must be generated by some finitely many elements belong in $\operatorname{dom} h$. This means there are finitely many functions $f_0, \ldots, f_{n-1} \in \mathscr{C}$ so that z is generated by finitely many elements of $\operatorname{dom} f_0 \cup \cdots \cup \operatorname{dom} f_{n-1}$. But $\operatorname{dom} f_0, \ldots, \operatorname{dom} f_{n-1}$, rearranged in some order, forms a chain under inclusion. So $z \in F_{\operatorname{dom} f_\ell}$ for some $\ell < n$. Hence $z \in E_{\operatorname{dom} f_\ell}$. But $\operatorname{range} f_\ell$ is a basis for $\mathbf{E}_{\operatorname{dom} f_\ell}$. Because $\operatorname{range} f_\ell \subseteq \operatorname{range} h$ we find that $\operatorname{range} h$ has enough elements to generate z. Since z was an arbitrary element of $E_{\operatorname{dom} h}$ we conclude that $\operatorname{range} h$ generates $E_{\operatorname{dom} h}$. It remains to show that $\operatorname{range} h$ is linearly independent. But $\operatorname{range} h$ is the union of the chain $\{\operatorname{range} f \mid f \in \mathscr{C}\}$. I ask the hard-working graduate students to prove that the union of any chain of linearly independent sets must also be linearly independent. Once you have done this you will be certain that h belongs to $\mathscr{F}$. By Zorn, let g be a maximal element of $\mathscr{F}$.

We would be done if $\operatorname{dom} g = B$, since then $E = E \cap F = E \cap F_B = E_B = E_{\operatorname{dom} g}$. In which case, $\operatorname{range} g$ would be a basis for $\mathbf{E}$ and $\operatorname{rank} \mathbf{E} = |\operatorname{range} g| \leq |\operatorname{dom} g| = |B| = \operatorname{rank} \mathbf{F}$.

Consider the possibility that $\operatorname{dom} g$ is a proper subset of B. Put $C = \operatorname{dom} g$ and put $X = \operatorname{range} g$. Let $w \in B$ with $w \notin \operatorname{dom} g$. Put $D = C \cup \{w\}$. As we have seen above, either $E_D = E_C$ or $X \cup \{u\}$ is a basis for $\mathbf{E}_D$, for some appropriately chosen u. We can now extend g to a function g' by letting $g'(w)$ be any element of $\operatorname{range} g$ in the case when $E_D = E_C$ and by letting $g'(w) = u$ is the alternative case. In this way, $g' \in \mathscr{F}$, contradicting the maximality of g. So we reject this possibility.

This completes the proof. □

Corollary 8.0.1. *Let* $\mathbf{R}$ *be a principal ideal domain. Every submodule of a finitely generated* $\mathbf{R}$*-module must itself be finitely generated.*

Proof. Suppose $\mathbf{M}$ is an $\mathbf{R}$-module generated by n elements. Let $\mathbf{N}$ by a submodule of $\mathbf{M}$.

Now let $\mathbf{F}$ be the free $\mathbf{R}$-module with a basis of n elements. There is a function that matches this basis with the generating set of $\mathbf{M}$. So, appealing to freeness, there is a homomorphism h from $\mathbf{F}$ onto $\mathbf{M}$. Let $E = \{v \mid v \in F \text{ and } h(v) \in N\}$. It is straightforward to check (will you do it?) that E is

closed under the module operations. So we get a submodule **E** of **F**. Moreover, the restriction of h to E is a homomorphism from **E** onto **N**. But by our theorem **E** is generated by a set with no more than n elements. Since the image, under a homomorphism, of any generating set for **E** must be a generating set of **N** (can you prove this?), we find that **N** is finitely generated. □

8.1 Problem Set 7

Algebra Homework, Edition 7
Eighth Week
More on Polynomials and then Some

Problem 27.
Let **D** be an integral domain and let $c_0, \ldots, c_{n-1}$ be n distinct elements of D. Further let $d_0, \ldots, d_{n-1}$ be arbitrary elements of D. Prove there is at most one polynomial $f(x) \in D[x]$ of degree $n-1$ such that $f(c_i) = d_i$ for all $i < n$.

Problem 28.
Let **F** be a field and let $c_0, \ldots, c_{n-1}$ be n distinct elements of F. Further let $d_0, \ldots, d_{n-1}$ be arbitrary elements of F. Prove there is at least one polynomial $f(x) \in F[x]$ of degree n such that $f(c_i) = d_i$ for all $i < n$.

Problem 29.
Let **R** be the following subring of the field of rational functions in 3 variables with complex coefficients:

$$R = \left\{ \frac{f}{g} : f, g \in \mathbb{C}[x,y,z] \text{ and } g(1,2,3) \neq 0 \right\}$$

Find 3 prime ideals P_1, P_2, and P_3 in R with

$$0 \subsetneqq P_1 \subsetneqq P_2 \subsetneqq P_3 \subsetneqq R.$$

Problem 30.
Let **R** be a commutative ring. An **R**-module **P** is said to be **projective** provided for all **R**-modules **M** and **N** and all homomorphisms f from **M** onto **N**, if g is a homomorphism from **P** into **N**, then there is a homomorphism h from **P** into **M** so that $f \circ h = g$.

Prove that every free **R**-module is projective.

LECTURE 9

DIRECT DECOMPOSITION OF FINITELY GENERATED MODULES OVER A PID

9.1 THE FIRST STEP

The objective here is to show how to obtain a direct decomposition of a finitely generated module over a principal ideal domain. We would like that the direct factors admit no further nontrivial direct decomposition.

The operations of modules work in a direct product of modules in a coordinatewise manner. So knowing how to perform the operations in each direct factor leads immediately to a knowledge of how the operations work in the direct product. One point of inconvenience with direct products is that very few modules actually arise as direct products—simply because the elements of your favorite module are not tuples of any kind. So our direct decompositions make use of isomorphisms.

For the task at hand, our direct decompositions turn out to have only finitely many direct factors. In this situation, it is easy to replace the direct product with the notion of a direct sum. Suppose that we have the following direct decomposition of the $\mathbf{R}$-module $\mathbf{M}$:

$$\mathbf{M} \cong \mathbf{N} \times \mathbf{L}.$$

Then composing the isomorphism with the projection functions on the direct product, we find two homomorphisms $f : \mathbf{M} \twoheadrightarrow \mathbf{N}$ and $g : \mathbf{M} \twoheadrightarrow \mathbf{L}$ and these homomorphisms have the following properties:

(a) For every $v \in N$ and $w \in L$ there is $u \in M$ so that $f(u) = v$ and $g(u) = w$.

(b) For every $u \in M$, if $f(u) = 0$ and $g(u) = 0$, then $u = 0$.

Another way to frame these two properties is in terms of the kernels of these homomorphism. Let $\mathbf{N}'$ be the submodule that is the kernel of g and let $\mathbf{L}'$ be the submodule that is the kernel of f.

(a') For every $u \in M$ there are $v \in N'$ and $w \in L'$ so that $u = v + w$.

(b') The intersection of N' and L' is trivial.

Here is how to derive (a') from (a) and (b). Use (a) to get $w \in M$ such that $f(w) = 0$ and $g(w) = g(u)$. The $g(u-w) = g(u) - g(w) = 0$. Now observe that $u = (u-w) + w$ and $u - w \in \ker g = N'$ and $w \in \ker f = L'$ as desired. I leave it to the hard-working graduate students to show that these two views (one from homomorphisms and one from kernels) are logically equivalent. The Homomorphism Theorem, after just a bit of work, yields that $\mathbf{N} \cong \mathbf{N}'$ and $\mathbf{L} \cong \mathbf{L}'$. This leads to the following definition. Let $\mathbf{N}'$ and $\mathbf{L}'$ be submodules of $\mathbf{M}$ that satisfy (a') and (b'). We say that $\mathbf{M}$ is a **direct sum** of $\mathbf{N}'$ and $\mathbf{L}'$ and we write $\mathbf{M} = \mathbf{N}' \oplus \mathbf{L}'$. Evidently, $\mathbf{N}' \oplus \mathbf{L}' \cong \mathbf{N}' \times \mathbf{L}'$.

We can extend this notion to three submodules $\mathbf{N}_0, \mathbf{N}_1$, and $\mathbf{N}_2$. Here is what works.

(a') For every $u \in M$ there are $v_0 \in N_0$, $v_1 \in N_1$, and $v_2 \in N_2$ so that $u = v_0 + v_1 + v_2$.

(b') The intersection $N_0 \cap (N_1 + N - 2)$, $N_1 \cap (N_0 + N_2)$, and $N_2 \cap (N_0 + N_1)$ are all trivial.

The hard-working graduate students should verify that this works and also that the obvious extension to any finite number of direct summands also succeeds.

Now let us turn to our task of decomposing modules. Here is a first step.

Fact. Let $\mathbf{R}$ be a nontrivial integral domain. As an $\mathbf{R}$-module $\mathbf{R}$ is directly indecomposable.

Proof. We know that $\mathbf{R}$ can be itself construed as an $\mathbf{R}$-module and as such it is a free $\mathbf{R}$-module of rank 1. To see that this module is directly indecomposable, suppose that $\mathbf{M}$ and $\mathbf{N}$ are $\mathbf{R}$-modules and that φ is an isomorphism from $\mathbf{R}$ onto $\mathbf{M} \times \mathbf{N}$. Let $M' = \{r \mid \varphi(r) = (u, 0) \text{ for some } u \in M\}$. Likewise, let $N' = \{r \mid \varphi(r) = (0, v) \text{ for some } v \in N\}$. Plainly, M' and N' are ideals in $\mathbf{R}$ and $M' \cap N' = \{0\}$ since φ is one-to-one. Since $\mathbf{R}$ is nontrivial, we see that $\mathbf{M}$ and $\mathbf{N}$ cannot both be trivial. Suppose, without loss of generality, that $\mathbf{M}$ is nontrivial. So M' is nontrivial. Pick $r \in M'$ with $r \neq 0$. We want to see that $\mathbf{N}$ must be a trivial module, or, what is the same, the N' is a trivial ideal. Let s be an arbitrary element of N'. Then $rs \in M' \cap N' = \{0\}$. That is, $rs = 0$. Since $r \neq 0$ and $\mathbf{R}$ is an integral domain, we conclude that $s = 0$. Since s was an arbitrary element of N', we have that N', and hence $\mathbf{N}$, is trivial. This means that $\mathbf{R}$ is a directly indecomposable $\mathbf{R}$-module, since $\mathbf{R}$ is itself nontrivial but in any direct decomposition we find that one of the direct factors must be trivial. □

This means that over any nontrivial integral domain any free module of finite rank directly decomposes into the direct product of finitely many copies of the ring and this direct decomposition is into directly indecomposable modules.

Here is another step we can take in directly decomposing a module.

Fact. Let $\mathbf{R}$ be a commutative ring. Suppose $\mathbf{M}$ is an $\mathbf{R}$-module, that $\mathbf{F}$ is a free $\mathbf{R}$-module, and the f is a homomorphism from $\mathbf{M}$ onto $\mathbf{F}$ with kernel $\mathbf{N}$. Then there is a free $\mathbf{R}$-module $\mathbf{E}$ so that $\mathbf{M} \cong \mathbf{N} \times \mathbf{E}$.

Proof. Let B be a basis for $\mathbf{F}$. For each $u \in B$ pick $v_u \in M$ so that $f(v_u) = u$. The set $C = \{v_u \mid u \in B\}$ is a linearly independent subset of M. Here is how to see it:

Let $w_0, \ldots, w_{n-1}$ be finitely many distinct elements of C and let $r_0, \ldots, r_{n-1} \in R$ with

$$r_0 w_0 + \cdots + r_{n-1} w_{n-1} = 0.$$

Applying f to both sides we obtain

$$r_0 f(w_0) + \cdots + r_{n-1} f(w_{n-1}) = 0.$$

But $f(w_0), \ldots, f(w_{n-1})$ are distinct elements of B, which is linearly independent. So $r_0 = \cdots = r_{n-1} = 0$, as desired.

Now let $\mathbf{E}$ be the submodule of $\mathbf{M}$ generated by C. So $\mathbf{E}$ is free since C is a basis. I contend that $\mathbf{M} \cong \mathbf{N} \times \mathbf{E}$. Here is how to define the isomorphism φ:

Let w be an arbitrary element of M. Let $f(w) = r_0 u_0 + \cdots + r_{n-1} u_{n-1}$ where $u_0, \ldots, u_{n-1}$ are distinct elements of B and all the r_i's are nonzero. Let $v_0, \ldots, v_{n-1}$ be elements of C so that $f(v_i) = u_i$ for all $i < n$. Put $x = r_0 v_0 + \cdots + r_{n-1} v_{n-1}$. Then $x \in E$ and $f(x) = f(w)$. This means $w - x \in N$ since N is the kernel of f. So define $\varphi(w) = (w - x, x)$.

It is a straightforward piece of work (done by all hard working graduate students) to see that φ is an isomorphism. □

To invoke this last fact for a particular module $\mathbf{M}$ we essentially have to find a free submodule of $\mathbf{M}$. Such a submodule would have a basis C. For $w \in C$ we would have to have the implication

$$rw = 0 \implies r = 0, \text{ for all } r \in R,$$

since this is just part of the definition of linear independence. Indeed, when $\mathbf{R}$ is an integral domain, all the elements of E, not just those in C would have to have this property. This suggests that N should consist of those elements that fail this property. That is elements $x \in M$ such that $rx = 0$ for some $r \neq 0$. Such elements are called **torsion** elements. The 0 of a module is always a torsion element, provided the ring in nontrivial. The module $\mathbf{M}$ is said to be **torsion free** provided 0 is its only torsion element. The step along the way is the following fact.

Fact. Let $\mathbf{R}$ be a nontrivial integral domain and let $\mathbf{M}$ be an $\mathbf{R}$-module. Then the set T of torsion elements is a submodule of $\mathbf{M}$ and $\mathbf{M}/T$ is torsion free.

Proof. We have already noted that $0 \in T$. To see that T is closed under addition, let $u, v \in T$. Pick nonzero elements $r, s \in R$ so that $ru = 0 = sv$. Then $rs \neq 0$ since $\mathbf{R}$ is an integral domain. Now observe $(rs)(u+v) = (rs)u + (rs)v = (sr)u + (rs)v = s(ru) + r(sv) = 0 + 0 = 0$ is $\mathbf{R}$ is commutative. So $u + v \in T$. Finally, suppose that $t \in R$. Then $r(tu) = (rt)u = (tr)U = t(ru) = 0$, so $tu \in T$. In this way, we see that T is closed under the module operations and we can form the submodule $\mathbf{T}$.

To see that $\mathbf{M}/T$ is torsion free, pick a nonzero element u/T of M/T and a scalar $r \in R$ so that $r(u/T) = 0/T$. Since u/T is nonzero we know that $u \notin T$. On the other hand $r(u/T) = (ru)/T = 0/T$ means that $ru \in T$. So pick a nonzero $s \in R$ so that $s(ru) = 0$. This means that $(sr)u = 0$. But, since $u \notin T$ we know that u is not a torsion element. So $sr = 0$. Since $s \neq 0$ and $\mathbf{R}$ is an integral domain, we see that $r = 0$. This means that u/T is not a torsion element. So $\mathbf{M}/T$ is a torsion free module. □

So when is a torsion free module actually free?

Fact. Let $\mathbf{R}$ be a nontrivial integral domain. Every finitely generated torsion free $\mathbf{R}$-module is free of finite rank.

Proof. Let $\mathbf{M}$ be a torsion free $\mathbf{R}$-module generated by the finite set X. Let Y be a maximal linearly independent subset of X. Let $\mathbf{F}$ be the submodule of $\mathbf{M}$ generated by Y. Of course, $\mathbf{F}$ is free of finite rank. For each $x \in X$ pick $s_x \in R$ so that $s_x x \in F$. This is possible since if $x \in Y$ we can let $s_x = 1$, while if $x \notin Y$, then $Y \cup \{x\}$ is linearly dependent. This means that for some distinct $y_0, \ldots, y_{n-1} \in Y$ there are $s_x, r_0, \ldots, r_{n-1} \in R \setminus \{0\}$ so that

$$s_x x + r_0 y_0 + \cdots + r_{n-1} y_{n-1} = 0.$$

In this way, $s_x x \in F$. Now let s be the product of all the s_x's as x runs through X. Then $sx \in F$ for all $x \in X$. Since X generates $\mathbf{M}$, we see that $sv \in F$ for all $v \in M$. Now let φ get the map from M into F defined via

$$\varphi(v) := sv \text{ for all } v \in M.$$

It is routine to check that φ is a homomorphism. The kernel of φ must be trivial since $\mathbf{M}$ is torsion free. So φ is one-to-one. This means that $\mathbf{M}$ is isomorphic with a submodule of the free module $\mathbf{F}$. Since $\mathbf{R}$ is a principal ideal domain, by the Freedom Theorem we conclude that $\mathbf{M}$ is free. Moreover, since $\mathbf{F}$ has finite rank, so must $\mathbf{M}$. □

The First Decomposition Theorem for Modules over an Integral Domain.
Let $\mathbf{R}$ *be a nontrivial integral domain, let* $\mathbf{M}$ *be a finitely generated* $\mathbf{R}$*-module, and let* $\mathbf{T}$ *be the torsion submodule of* $\mathbf{M}$. *There is a free module* $\mathbf{F}$ *of finite rank such that*

$$\mathbf{M} \cong \mathbf{T} \times \mathbf{F}.$$

Moreover, the rank of $\mathbf{F}$ *is determined by* $\mathbf{M}$.

Proof. According to the Facts established above, we can take $\mathbf{F}$ to be $\mathbf{M}/T$. So only the "moreover" part of the theorem remains to be established. To this end, suppose that $\mathbf{F}'$ is some free module so that

$$\mathbf{M} \cong \mathbf{T} \times \mathbf{F}'.$$

The conclusion we want is $\mathbf{F} \cong \mathbf{F}'$.

What are the torsion elements of $\mathbf{T} \times \mathbf{F}'$? Suppose (u, v) is torsion. Pick $r \neq 0$ so that $r(u, v) = (0, 0)$. So $rv = 0$. But $v \in F'$, which being free is also torsion free. So $v = 0$. This means that the torsion elements of $\mathbf{T} \times \mathbf{F}'$ are exactly the elements of $T' := \{(u, 0) \mid u \in T\}$. In this way we see

$$\mathbf{F} = \mathbf{M}/T \cong (\mathbf{T} \times \mathbf{F}')/T' \cong \mathbf{F}'.$$

The rightmost isomorphism above comes from the Homomorphism Theorem since T' is the kernel of the project of the direct product onto its rightmost direct factor. □

Both the torsion module $\mathbf{T}$ and the free module $\mathbf{F}$ may admit further direct decomposition. As regards the free module, we know it can be decomposed as the direct product of n copies of the $\mathbf{R}$-module $\mathbf{R}$, which we have seen is directly indecomposable.

9.2 Problem Set 8

ALGEBRA HOMEWORK, EDITION 8
NINTH WEEK
POLYNOMIALS AGAIN

PROBLEM 31.
Prove that the polynomial $x^3y + x^2y - xy^2 + x^3 + y$ is irreducible in $\mathbb{Z}[x, y]$.

PROBLEM 32.
Let **F** and **M** be modules over the same ring and let **F** be a free module. Let $h : \mathbf{M} \twoheadrightarrow \mathbf{F}$ be a homomorphism from **M** onto **F**. Prove each of the following.

(a) There is an embedding $g : \mathbf{F} \rightarrowtail \mathbf{M}$ of **F** into **M** such that $h \circ g = \mathrm{id}_F$. (Here id_F denotes the identity map of the set F.)

(b) $\mathbf{M} = \ker h \oplus \mathbf{F}'$, where $\mathbf{F}'$ is the image of **F** with respect to g.

PROBLEM 33.
Prove that there is a polynomial $f(x) \in \mathbb{R}[x]$ such that

(a) $f(x) - 1$ belongs to the ideal $(x^2 - 2x + 1)$;

(b) $f(x) - 2$ belongs to the ideal $(x + 1)$, and

(c) $f(x) - 3$ belongs to the ideal $(x^2 - 9)$.

9.3 The Second Step

Let **M** be any **R**-module and X be any subset of M. Define

$$\operatorname{ann} X := \{r \mid r \in R \text{ and } rx = 0 \text{ for all } x \in X\}.$$

This set is called the **annihilator** of X. It is routine to check that $\operatorname{ann} X$ is always an ideal of **R**, provided **R** is commutative. Running this game in the other direction, let $S \subseteq R$ and define

$$M[S] := \{u \mid u \in M \text{ and } ru = 0 \text{ for all } r \in S\}.$$

Again, it is routine to check that $M[S]$ is closed under the module operations, provided that **R** is commutative. So we obtain a submodule $\mathbf{M}[S]$ of the module **M**. For a single element $r \in R$ we write $\mathbf{M}[r]$ for $\mathbf{M}[\{r\}]$.

Let **T** be a torsion module of a principal ideal domain **R**. Suppose that the finite set X generates **T**. As we did in one of the proofs in the preceding lecture, for each $x \in X$ we can pick a nonzero $s_x \in R$ so that $s_x x = 0$. Let s be the product of these finitely many s_x's as x runs through X. Since **R** is an integral domain we see that $s \neq 0$ and since **R** is commutative and X generates **T** we see that $su = 0$ for all $u \in T$. This means that $\operatorname{ann} T$ is a nontrivial ideal. Because **R** is a principal ideal we can pick $r \in R$ so that $(r) = \operatorname{ann} T$. This nonzero element r, which is unique up to associates, is called the **exponent of T**.

More generally, if u is a torsion element of an **R**-module, where **R** is a principal ideal domain then there will be a nonzero element r so that $(r) = \operatorname{ann}\{u\}$. We call r the **order** of u and sometimes refer to $\operatorname{ann}\{u\}$ as the **order ideal** of u. If v is also a torsion element and s is the order of v, where r and s are relatively prime, then rs will be the order of $u + v$. (This could be proven by a hard working graduate student.)

We are ready to begin decomposing our torsion module.

Fact. Let **T** be a torsion **R**-module with exponent r, where **R** is a principal ideal domain. Suppose that $r = sq$ where s and q are relatively prime. Then $\mathbf{T} \cong \mathbf{T}[s] \times \mathbf{T}[q]$.

Proof. Using the relative primeness of s and q select elements $a, b \in R$ so that $1 = as + bq$. So for any $u \in T$ we have

$$u = 1 \cdot u = (as + bq)u = q(bu) + s(au).$$

Observe that $q(bu) \in T[s]$ and $s(au) \in T[q]$. So every element of T can be expressed as a sum of an element of $T[s]$ and an element of $T[q]$. The expression is unique since if $u = v + w$ where $v \in T[s]$ and $w \in T[q]$, then $v - qbu = sau - w \in T[s] \cap T[q]$. But the order of any element of this intersection must divide both s and q, which are relatively prime. So the intersection is just $\{0\}$ and it follows that $v = qbu$ and $w = sav$. The map that sends u to (v, w) where $v \in T[s]$, $w \in T[q]$, and $u = v + w$ is easily seen to be an isomorphism. □

Suppose that r is the exponent of our torsion **R**-module **T**, where **R** is a principal ideal domain. Let $p_0, \ldots, p_{n-1} \in R$ be distinct primes and let $e_0, \ldots, e_{n-1}$ be positive integers so that

$$r = p_0^{e_0} \ldots p_{n-1}^{e_{n-1}}.$$

Then applying the Fact above over and over again we find

$$\mathbf{T} \cong \mathbf{T}[p_0^{e_0}] \times \cdots \times \mathbf{T}[p_{n-1}^{e_{n-1}}].$$

Modules of the form $\mathbf{T}[p^e]$ where $p \in R$ is prime and e is a positive integer are said to be **primary** or sometimes more specifically p-**primary**. These are just the torsion modules of exponent a power of p. So now we know that every finitely generated torsion module over a principal ideal domain can be decomposed as a product of primary modules. We state this as a theorem.

The Primary Decomposition Theorem.
Every finitely generated module over a principal ideal domain is isomorphic to a direct products of finitely many primary modules.

Still, these primary modules may admit further direct decomposition.
Which p-primary modules are directly indecomposable?

Fact. Let $\mathbf{R}$ be a principal ideal domain and $p \in R$ be prime. Every cyclic p-primary $\mathbf{R}$-module is directly indecomposable.

Proof. Let $\mathbf{T}$ be an $\mathbf{R}$-module of exponent p^e that is generated by w. Suppose that $\mathbf{T} \cong \mathbf{M} \times \mathbf{N}$. We need to argue that one of $\mathbf{M}$ and $\mathbf{N}$ is trivial. We know that $p^e w = 0$ but $p^{e-1} w \neq 0$. Pick $u \in M$ and $v \in N$ so that (u, v) generates $\mathbf{M} \times \mathbf{N}$. Then we have $p^e u = 0$ and $p^e v = 0$ and either $p^{e-1} u \neq 0$ or $p^{e-1} v \neq 0$. Without loss of generality, suppose $p^{e-1} u \neq 0$. This makes $\mathbf{M}$ nontrivial, so our ambition is to show that $\mathbf{N}$ is trivial. Now since the element (u, v) generates all elements of $\mathbf{M} \times \mathbf{N}$, we see that every element of $M \times N$ can be obtained by multiplying (u, v) by an appropriate scalar. So pick $r \in R$ so that $r(u, v) = (0, v)$. It follows that $ru = 0$ and $rv = v$. Since the order of u is p^e, we have that $p^e \mid r$. So $r = sp^e$ for some s. But then $v = rv = sp^e v = 0$. But this entails that $\mathbf{N}$ is trivial. □

Of course, we can get a cyclic submodule easily enough just by selecting a single element and using it to generate a submodule. Something more clever is possible.

Fact. Let $\mathbf{R}$ be a principal ideal domain and let $\mathbf{T}$ be a torsion $\mathbf{R}$-module of exponent r. Then $\mathbf{T}$ has an element of order r.

Proof. Pick distinct primes $p_0, \ldots, p_{n-1} \in R$ and positive integers $e_0, \ldots, e_{n-1}$ so that

$$r = p_0^{e_0} \ldots p_{n-1}^{e_{n-1}}.$$

For each $j < n$, let $r_j = \frac{r}{p_j} = p^{e_0} \ldots p_{j-1}^{e_{j-1}} p_j^{e_j - 1} p_{j+1}^{e_{j+1}} \ldots p_{n-1}^{e_{n-1}}$. Notice that $r \nmid r_j$ for all $j < n$. This allows us, for each $j < n$, to pick $u_j \in T$ so that $r_j u_j \neq 0$. Now, for each $j < n$, put

$$v_j = \frac{r}{p_j^{e_j}} u_j.$$

Then, for each $j < n$, we have $p_j^{e_j} v_j = 0$ but $p_j^{e_j - 1} v_j \neq 0$. So v_j is an element of order $p_j^{e_j}$. It now follows that $v_0 + \cdots + v_{n-1}$ is an element of order r, as desired. □

There is one additional fact that proves useful.

Fact. Let $\mathbf{T}$ be a torsion $\mathbf{R}$-module of exponent r, where $\mathbf{R}$ is a principal ideal domain. Let $\mathbf{M}$ and $\mathbf{N}$ be submodules of $\mathbf{T}$, where $\mathbf{M}$ is generated by an element of order r. Let f be a homomorphism from $\mathbf{N}$ into $\mathbf{M}$. Finally, let $v \in T$ with $v \notin N$. Then f can be extended to a homomorphism from the submodule generated by $N \cup \{v\}$ into $\mathbf{M}$.

Proof. Let u be the element of order r that generates $\mathbf{M}$. Let $\mathbf{N}'$ be the submodule generated by $N \cup \{v\}$. Evidently, $rv = 0 \in N$, so v/N has some nonzero order s in $\mathbf{N}'/N$ and $s \mid r$. So $sv \in N$. Pick $p \in R$ so that

$$r = ps.$$

Now $f(sv) \in M$ and $pf(sv) = f(psv) = f(rv) = f(0) = 0$. So the order of $f(sv)$ divides p.

Since $f(sv) \in M$ and $\mathbf{M}$ is generated by u pick $q \in R$ so that $f(sv) = qu$. Then we see $0 = pf(sv) = pqu$. This means

$$r \mid pq.$$

Hence, $ps \mid pq$. Since $\mathbf{R}$ is an integral domain, we have

$$s \mid q.$$

So pick $t \in R$ such that

$$q = st.$$

This entails $f(sv) = qu = stu$. Let $w = tu \in M$. So $f(sv) = sw$.

Now every element of N' can be written in the form $y + av$ for some choices of $y \in N$ and $a \in R$. There may be several ways to make these choices.

Here is how we define our extension f' of f:

$$f'(y + av) = f(y) + aw.$$

It is not clear that this definition is sound. Let us verify that. Suppose that $y + av = y' + a'v$ where $y, y' \in N$ and $a, a' \in R$. We need to see that

$$f(y) + aw = f(y') + a'w \text{ or written another way } f(y) - f(y') = (a' - a)w.$$

But notice that $y - y' = (a' - a)v$. So $(a' - a)v \in N$. This means $s \mid a' - a$. Pick $m \in R$ so that $a' - a = ms$. But this leads to

$$f(y) - f(y') = f(y - y') = f((a' - a)v) = f(msv) = mf(sv) = msw = (a' - a)w,$$

just as we desire. So our definition is sound. Since $f'(y) = f'(y + 0v) = f(y) + 0w = f(y)$, for all $y \in N$, we see that f' extends f. We must also check that f' is a homomorphism, a task we leave to hard working graduate students. □

So our scheme is to grab an element whose order is the exponent of $\mathbf{T}$, let $\mathbf{M}$ be the submodule generated by that element, and hope to find another submodule $\mathbf{N}$ so that $\mathbf{T} \cong \mathbf{M} \times \mathbf{N}$. If we are lucky maybe the exponent of $\mathbf{N}$ will be smaller. Here is what we need.

Fact. Let $\mathbf{R}$ be a principal ideal domain and let $\mathbf{T}$ be a torsion $\mathbf{R}$-module of exponent r. Then $\mathbf{T}$ has a cyclic submodule $\mathbf{M}$ of exponent r and a submodule $\mathbf{N}$ of exponent s so that $s \mid r$ and $\mathbf{T} \cong \mathbf{M} \times \mathbf{N}$.

Proof. Let $u \in T$ have order r and let $\mathbf{M}$ be the submodule of $\mathbf{T}$ generated by u. Let f be the identity map on $\mathbf{M}$. Let

$$\mathscr{F} = \{g \mid g : \mathbf{N} \to \mathbf{M} \text{ is a homomorphism extending } f \text{ for some submodule } \mathbf{N} \text{ with } M \subseteq N \subseteq T\}.$$

We will apply Zorn's Lemma to $\mathscr{F}$, which is partially ordered by set-incluion. Notice that $f \in \mathscr{F}$, so $\mathscr{F}$ is not empty. Let $\mathscr{C}$ be a nonementy chain in $\mathscr{F}$. Certainly $\bigcup \mathscr{C}$ is an upper bound on $\mathscr{C}$. We

must show it is in $\mathscr{F}$. We have noted before that the union of a chain of functions is again a function. The hard-working graduate students will see that the union of a chain of homomorphism is itself a homomorphism. It is routine to see that the union of a chain of submodules (the domains of those homomorphisms) is again a submodule. So we see indeed that $\bigcup\mathscr{C}$ belongs to $\mathscr{F}$. Let g be a maximal element of $\mathscr{F}$. Since the fact just above would otherwise allow the extension of g to larger member of $\mathscr{F}$, a thing impossible by the maximality of g, we see that g is a homomorphism from $\mathbf{T}$ into $\mathbf{M}$ which extends the identity map on $\mathbf{M}$. Let $\mathbf{N}$ be the kernel of g. Let s be the exponent of $\mathbf{N}$. Evidently, $s \mid r$.

For any $w \in T$ we have $g(w) \in M$. But then $g(g(w)) = f(g(w))$ since g extends f. But f is the identity map. So we see that $g(g(w)) = g(w)$ for all $w \in T$. But notice

$$g(w - g(w)) = g(w) - g(g(w)) = g(w) - g(w) = 0.$$

This means that $w - g(w) \in N$ for all $w \in T$, since N is the kernel of g. So we find that

$$w = g(w) + (w - g(w)) \in M + N, \text{ for all } w \in T.$$

Another way to write this is

$$T = M + N.$$

Now suppose $w \in M \cap N$. Then, on the one hand, $g(w) = f(w) = w$ since $w \in M$ and g extends f (which is the identity function), while on the other hand $g(w) = 0$ since $w \in N = \ker g$. Taken together we find that $w = 0$. This means

$$M \cap N = \{0\}.$$

So we see that $\mathbf{T} = \mathbf{M} \oplus \mathbf{N}$. This yields our desired conclusion. □

The Invariant Factor Theorem.
Let $\mathbf{T}$ *be a nontrivial finitely generated torsion* $\mathbf{R}$*-module, where* $\mathbf{R}$ *is a principal ideal domain. Then for some natural number* n *there are* $r_0, r_1, \ldots, r_n \in R$ *with*

$$r_n \mid r_{n-1} \mid \cdots \mid r_1 \mid r_0$$

and cyclic submodules $\mathbf{M}_0$ *of exponent* r_0*,*$\ldots$*,* $\mathbf{M}_n$ *of exponent* r_n *so that*

$$\mathbf{T} \cong \mathbf{M}_0 \times \cdots \times \mathbf{M}_n.$$

Proof. Let the order of $\mathbf{T}$ be r_0. (Recall that we have already proven that a nontrivial finitely generated torsion module has an exponent.) Let $u_0 \in T$ have order r_0 and let $\mathbf{M}_0$ be the submodule generated by u_0. By the preceding Fact, there is a submodule $\mathbf{N}_0$ of order r_1 with $r_1 \mid r_0$ such that $\mathbf{T} \cong \mathbf{M}_0 \times \mathbf{N}_0$. If $\mathbf{N}_0$ is trivial, we can stop since $\mathbf{T} \cong \mathbf{M}_0$ in that case. Otherwise, pick $u_1 \in N_0$ with order r_1. Take $\mathbf{M}_1$ to be the submodule of generated by u_1 and invoke the immediately preceding fact to get a proper submodule $\mathbf{N}_1$ of $\mathbf{N}_0$ of exponent r_2 so that $r_2 \mid r_1$ and $\mathbf{N}_0 \cong \mathbf{M}_1 \times \mathbf{N}_1$. At this stage we have

$$\mathbf{T} \cong \mathbf{M}_0 \times \mathbf{M}_1 \times \mathbf{N}_1 \text{ and } r_2 \mid r_1 \mid r_0 \text{ and } N_1 \subsetneq N_0,$$

where the exponent of $\mathbf{M}_0$ is r_0, the exponent of $\mathbf{M}_1$ is r_1, and the exponent of $\mathbf{N}_1$ is r_2. Again our process terminates in the event $\mathbf{N}_1$ is trivial, but otherwise the process can be continued. In this

process to chains of submodules of **T** are constructed. One is the descending chain consisting of the submodules $\mathbf{N}_j$. The other is the ascending chain

$$\mathbf{M}_0 \subsetneqq \mathbf{M}_0 \oplus \mathbf{M}_1 \subsetneqq \ldots.$$

Now we know, as a corollary of the Freedom Theorem that every submodule of **T** is finitely generated. We saw for rings that if every ideal was finitely generated then there could be no infinite ascending chains of ideals. The same reasoning applies here (as the hard working graduate student will establish) to see that there can be no infinite ascending chair of submodules of **T**. This must mean the process described above terminates at some finite stage. This completes our proof. □

The $r_0, r_1, \ldots, r_n$ mentioned in the theorem are called **invariant factors**.

Now we are ready for the chief existence theorem for direct decomposition of finitely generated modules.

The Elementary Divisor Theorem.
Let **T** *be a nontrivial finitely generated torsion* **R***-module, where* **R** *is a principal ideal domain. Then for some natural number n, there are cyclic primary submodules* $\mathbf{M}_0 \ldots \mathbf{M}_n$ *of* **T** *so that*

$$\mathbf{T} \cong \mathbf{M}_0 \times \cdots \times \mathbf{M}_n.$$

The exponents of the various cyclic primary submodules are referred to as the **elementary divisors** of **T**. The proof of the Elementary Divisor Theorem is obtained by applying the Invariant Factor Theorem to each of the direct factors arising from an application of the Primary Decompostion Theorem.

9.4 Problem Set 9

ALGEBRA HOMEWORK, EDITION 9
TENTH WEEK
A GRAB BAG

PROBLEM 34.
Let A be the 4×4 real matrix

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ -2 & -2 & 2 & 1 \\ 1 & 1 & -1 & 0 \end{pmatrix}$$

(a) Determine the rational canonical form of A.

(b) Determine the Jordan canonical form of A.

PROBLEM 35.
Suppose that N is a 4×4 nilpotent matrix over a field $\mathbf{F}$ with minimal polynomial x^2. What are the possible rational canonical forms for N?

PROBLEM 36.
Let $\mathbf{F}$ be the subring of the field of complex numbers consisting of those numbers of the form $a+ib$ where a and b are rational. Let $\mathbb{G}$ be the subring of the field of complex numbers consisting of those numbers of the form $m+ni$ where m and n are integers.

(a) Describe all the units of $\mathbb{G}$.

(b) Prove that $\mathbf{F}$ is (isomorphic to) the field of fractions of $\mathbb{G}$.

(c) Prove that $\mathbb{G}$ is a principal ideal domain.

[Hint: In this problem it is helpful to consider the function that sends each complex number z to $z\bar{z} = |z|^2$.]

LECTURE 10

THE STRUCTURE OF FINITELY GENERATED MODULES OVER A PID

Here is one of the key results in our course.

The Structure Theorem for Finitely Generated Modules over a Principal Ideal Domain. *Let* $\mathbf{M}$ *be a finitely generated* $\mathbf{R}$*-module, where* $\mathbf{R}$ *is a principal ideal domain. There is a natural number* n *such that:*

(a) *there are* n *finitely generated directly indecomposable submodules* $\mathbf{M}_0, \ldots, \mathbf{M}_{n-1}$ *of* $\mathbf{M}$ *such that*

$$\mathbf{M} \cong \mathbf{M}_0 \times \cdots \times \mathbf{M}_{n-1}, \text{ and}$$

(b) *for any natural number* m*, if* $\mathbf{N}_0, \ldots, \mathbf{N}_{m-1}$ *are directly indecomposable* $\mathbf{R}$*-modules such that* $\mathbf{M} \cong \mathbf{N}_0 \times \cdots \times \mathbf{N}_{m-1}$*, then* $n = m$ *and there is permutation* σ *of* $\{0, \ldots, n-1\}$ *so that* $\mathbf{M}_i \cong \mathbf{N}_{\sigma(i)}$ *for all* $i < n$*.*

Moreover, the finitely generated directly indecomposable $\mathbf{R}$*-modules are, up to isomorphism, the* $\mathbf{R}$*-module* $\mathbf{R}$ *(that is the free* $\mathbf{R}$*-module of rank* 1*), and the* $\mathbf{R}$*-modules of the form* $\mathbf{R}/(r)$ *where* r *is a positive power of some prime element of the ring* $\mathbf{R}$ *(these are the cyclic primary* $\mathbf{R}$*-modules). Finally, the free* $\mathbf{R}$*-module of rank* 1 *is not primary and if* $r, s \in R$ *are prime powers and* $\mathbf{R}/(r) \cong \mathbf{R}/(s)$*, then* r *and* s *are associates in the ring* $\mathbf{R}$*.*

Before turning to the proof a few remarks are in order.

First, we have allowed $n = 0$. This results in the direct product of an empty system of $\mathbf{R}$-modules. A careful, but easy, examination of the definition of direct products reveals that such a direct product produces the trivial $\mathbf{R}$-module—that is the module whose only element is 0. Evidently, the trivial $\mathbf{R}$-module is the direct product of exactly one system of directly indecomposable $\mathbf{R}$-modules, namely of the empty system.

This theorem has three parts:

- the assertion of the existence of a decomposition into indecomposables,
- the assertion that such a decomposition is unique, and
- a description of the indecomposables.

These are the hallmarks of a good structure theorem. There are other theorems of this kind in mathematics. Perhaps the most familiar is the Fundamental Theorem of Arithmetic. To make the connection plain, consider the least complicated algebraic systems, namely just nonempty sets equipped with no operations. Then the finitely generated algebraic systems are just the finite nonempty sets and isomorphisms are just one-to-one correspondences. So two of these algebraic systems will be isomorphic if and only if they have the same number of elements. The Fundamental Theorem of Arithmetic says that every finite set is isomorphic to a direct product of directly indecomposable finite sets in a way that is unique up to isomorphism and rearranging the factorization and that a set is directly indecomposable if and only if it has a prime number of elements.

This theorem also resonates with the notion of a unique factorization domain. We could reformulate our structure theorem to make this more apparent. Each finitely generated **R**-module is isomorphic to a lot of other **R**-modules (in fact, to a proper class of **R**-modules). Pick a representative from each of these isomorphism classes, taking care to include among these representatves the **R**-modules **R** and $\mathbf{R}/(r)$ where $r \in R$ is a positive power of some prime element of **R**. Let $\mathscr{M}$ be the set of all these representative and let **1** be the representative trivial **R**-module. Then $\langle \mathscr{M}, \times, \mathbf{1} \rangle$ is an algebraic system (actually a monoid) with the unique factorization property.

Finally, the structure theorem above has a number of far-reaching consequences. Taking **R** to be $\mathbb{Z}$ we obtain a structure theorem for finitely generated Abelian group. Taking **R** to be $\mathbf{F}[x]$, where **F** is field leads to the canonical form theorems of linear algebra.

Proof. Let us first dispose of the descriptions of the directly indecomposable **R**-modules that could arise in any factorization of **M**. These must be finitely generated because they will be isomorphic to submodules of **M** and, according to the Corollary of the Freedom Theorem every submodule of **M** must be finitely generated. We have already seen that the free **R**-module of rank 1 (namely the module **R**) is directly indecomposable and every other free **R**-module of finite $n > 1$ is a direct product of n-copies of **R**. We have also seen that the cyclic primary **R**-modules are the only finitely generated directly indecomposable torsion modules. Can there be any others finitely generated directly indecomposable **R**-module? By the First Decomposition Theorem every finitely generated **R**-module is isomorphic to a direct product of the form $\mathbf{T} \times \mathbf{F}$, where **T** is the torsion submodule and **F** is a submodule that is free. For a directly indecomposable module we must have either **F** trivial (and the our module would be torsion) or else **T** trivial (and the our module would be free). So the only finitely generated directly indecomposable **R**-modules are the ones already in hand, the **R**-module **R** and the cyclic primary **R**-modules.

We can say more about the cyclic primary **R**-modules. Let $r \in R$ be a positive power of a prime element of R. Then the ideal (r) is a submodule of the **R**-module **R**. The element $1/(r)$ of the quotient module $\mathbf{R}/(r)$ generates the quotient module and has order r. So the quotient module is cyclic and of exponent r. In this way we know cyclic **R**-modules of exponent r exist. Suppose that **N** is an **R**-module of exponent r which is generated by the single element u. Since $\{1\}$ is a basis for **R**, we know there is a homomorphism h from **R** onto **N** that takes 1 to u. Now for all $s \in R$ we

have $h(s) = h(s \cdot 1) = sh(1) = su$. From this we see that

$$s \in \ker h \Leftrightarrow h(s) = 0 \Leftrightarrow su = 0 \Leftrightarrow r \mid s \Leftrightarrow s \in (r).$$

That is, $\ker h = (r)$. So by the Homomorphism Theorem $\mathbf{N} \cong \mathbf{R}/(r)$. So, up to isomorphism, the only cyclic $\mathbf{R}$-module of exponent r (where r is a positive power of some prime) is $\mathbf{R}/(r)$.

Now observe that the free $\mathbf{R}$-module $\mathbf{R}$ of rank 1 is not a torsion module since $s \cdot 1 = 0 \implies s = 0$. So the $\mathbf{R}$-module $\mathbf{R}$ cannot be isomorphic with any of the modules $\mathbf{R}/(r)$ where r is a positive power of some prime. (One needs to observe here, as the hard-working graduate students will verify, that r cannot be a unit.) Now suppose that r and s are both positive powers of primes (we don't assume the primes are tne same) and that $\mathbf{R}/(r) \cong \mathbf{R}/(s)$. Then r is the order of a generator of this cyclic module and so is s. This means that $r \mid s$ and $s \mid r$. Consequently, $(r) = (s)$ and r and s are associates.

Now consider part (a) of the theorem. This is an immediate consequence of the First Decomposition Theorem and the Elementary Divisor Theorem.

Finally, consider part (b). Some of the $\mathbf{N}_i$'s can be cyclic primary modules and others can be free of rank 1, according to our description of the directly indecomposable finitely generated modules. Without loss of generality, we assume that the primary modules come first. So pick $k \leq m$ so that $\mathbf{N}_0, \ldots, \mathbf{N}_{k-1}$ are cyclic primary modules and $\mathbf{N}_k, \ldots, \mathbf{N}_{m-1}$ are free of rank 1. Let $\mathbf{T}$ be the direct product of the first group and $\mathbf{F}$ be the direct product of the second. So we find $\mathbf{M} \cong \mathbf{T} \times \mathbf{F}$. It is routine (according to hard-working graduate students) that $\mathbf{T}$ is a torsion module and also that $\mathbf{F}$ is free of $m-k$. Let $(v,u) \in T \times F$ be a torsion element of order r. Then $(0,0) = r(v,u) = (rv, ru)$. In particular, $ru = 0$. The element u can be written as a linear combination of the basis elements of $\mathbf{F}$. By distributing r through the linear combination and invoking both linear independence and the fact the r is a nonzero element of an integral domain, we see that $u = 0$. What we conclude is that the torsion elements of $\mathbf{T} \times \mathbf{F}$ are exactly those of the form $(v,0)$ where $v \in T$ is nonzero. Thus under the isomorphism $\mathbf{M} \cong \mathbf{T} \times \mathbf{F}$, the module $\mathbf{T}$ corresponds to the torsion submodule of $\mathbf{M}$. Then according to the First Decomposition Theorem the rank of $\mathbf{F}$ is determined by $\mathbf{M}$.

It remains to show that if $\mathbf{T} \cong \mathbf{M}_0 \times \cdots \times \mathbf{M}_{\ell-1} \cong \mathbf{N}_0 \times \cdots \times \mathbf{N}_{k-1}$, where all the $\mathbf{M}_i$'s and $\mathbf{N}_j$'s are cyclic primary $\mathbf{R}$-modules, then $\ell = k$ and, after a suitable reindexing $\mathbf{M}_i \cong \mathbf{N}_i$ for all $i < \ell$.

Let $p \in R$ be prime. For any $\mathbf{R}$-module $\mathbf{Q}$ let

$$Q(p) = \{v \mid v \in Q \text{ and } p^e v = 0 \text{ for some positive integer } e\}.$$

It is routine to check that this set is closed under the module operations, show we have the submodule $\mathbf{Q}(p)$. It is also not hard to see (as hard-working graduate students will check) that

$$\mathbf{T}(p) \cong \mathbf{M}_0(p) \times \cdots \times \mathbf{M}_{\ell-1}(p) \cong \mathbf{N}_0(p) \times \cdots \times \mathbf{N}_{k-1}(p).$$

In this decomposition, if $\mathbf{M}_i$ (or $\mathbf{N}_j$) were primary with respect to a prime not associate to p, then the module $\mathbf{M}_i(p)$ (respectively $\mathbf{N}_j(p)$) would be trivial. On the other hand, if they were primary with respect to an associate of p, then $\mathbf{M}_i(p) = \mathbf{M}_i$ and $\mathbf{N}_j(p) = \mathbf{N}_j$. Since this holds for arbitrary primes p, we do not lose any generality by assuming the the primes underlying all the $\mathbf{M}_i$'s and $\mathbf{N}_j$'s are the same prime p.

Now suppose $\mathbf{Q}$ is a cyclic primary $\mathbf{R}$-module, where p^e is the exponent and u is a generator. Then $\mathbf{Q}[p]$ is generated by $p^{e-1}u$. So $\mathbf{Q}[p]$ is cyclic of exponent p. In this case, we know that $\mathbf{Q}[p] \cong \mathbf{R}/(p)$. Now (p) is a prime ideal of the ring $\mathbf{R}$. In a principal ideal domain, the maximal

ideals and the prime ideals coincide. So the ring $\mathbf{R}/(p)$ is a field. This allows us to construe the $\mathbf{R}$-module $\mathbf{Q}[p]$ as a one-dimensional vector space over the field $\mathbf{R}/(p)$. In doing this, we are changing the scalar multiplication, but leaving the addition and the zero the same. Now we have

$$\mathbf{T}[p] \cong \mathbf{M}_0[p] \times \cdots \times \mathbf{M}_{\ell-1}[p] \cong \mathbf{N}_0[p] \times \cdots \times \mathbf{N}_{k-1}[p]$$

construed as vector spaces over the field $\mathbf{R}/(p)$, with each of the direct factors being a copy of the one-dimensional vector space. This means

$$\ell = \dim \mathbf{T}[p] = k.$$

So we have discovered that $\ell = k$, one of our desired conclusions.

So we are reduced to considering the following situation:

$$\begin{aligned}\mathbf{T} &\cong \mathbf{M}_0 \times \cdots \times \mathbf{M}_{\ell-1} \\ &\cong \mathbf{N}_0 \times \cdots \times \mathbf{N}_{\ell-1}\end{aligned}$$

where $\mathbf{M}_i$ is a cyclic module of exponent p^{e_i} and $\mathbf{N}_i$ is a cyclic module of exponent p^{f_i} for all $i < \ell$ and

$$\begin{aligned}e_0 \geq e_1 \geq \cdots \geq e_{\ell-1} \\ f_0 \geq f_1 \geq \cdots \geq f_{\ell-1}.\end{aligned}$$

It remains only to show that $e_i = f_i$ for all $i < \ell$. Suppose, for the sake of contradiction, that this were not so. Let i be as small as possible so that $e_i \neq f_i$. It is harmless to also suppose that $e_i > f_i$. Let $r = p^{f_i}$. Now multiplication by r is homomorphism and it is easy to also see that

$$\begin{aligned}r\mathbf{T} &\cong r\mathbf{M}_0 \times \cdots \times r\mathbf{M}_{i-1} \times r\mathbf{M}_i \times \cdots \times r\mathbf{M}_{\ell-1} \\ &\cong r\mathbf{N}_0 \times \cdots \times r\mathbf{N}_{i-1} \times r\mathbf{N}_i \times \cdots \times r\mathbf{N}_{\ell-1}.\end{aligned}$$

Being homomorphic images of cyclic modules, each of the direct factors above is also cyclic. Because r is a positive power of the prime p, we see that the factor modules above are either primary (with prime p) or trivial. But exponents of all the $\mathbf{N}_j$'s where $i \leq j$ are factors of r, we see that these modules are all trivial. On the other hand, the exponent of $\mathbf{M}_i$ is p^{e_i} whereas $r = p^{f_i}$ with $e_i > f_i$. So $r\mathbf{M}_i$ is not trivial. This would mean

$$\begin{aligned}r\mathbf{T} &\cong r\mathbf{M}_0 \times \cdots \times r\mathbf{M}_{i-1} \times r\mathbf{M}_i \times \cdots \times r\mathbf{M}_{\ell-1} \\ &\cong r\mathbf{N}_0 \times \cdots \times r\mathbf{N}_{i-1},\end{aligned}$$

where the top direct factorization has at least $i+1$ nontrivial cyclic primary factors but the bottom has only i. But we have just proven that the number of such factors must be the same no matter how the direct factorization is accomplished. This contradiction means our supposition must be rejected. So $e_i = f_i$ for all $i < \ell$. This establishes the uniqueness of our direct factorization into directly indecomposable modules. The proof of the last remaining part of our theorem, namely part (b), is complete. □

The Structure Theorem above is an extension of the Elementary Divisor Theorem formulated in the previous lecture. We can also extend the Invariant Factor Theorem.

The Extended Invariant Factor Theorem.
Let $\mathbf{T}$ *be a nontrivial finitely generated torsion* $\mathbf{R}$*-module, where* $\mathbf{R}$ *is a principal ideal domain. Then for some natural number* n *there are* $r_0, r_1, \ldots, r_n \in R$ *with*

$$r_n \mid r_{n-1} \mid \cdots \mid r_1 \mid r_0$$

and cyclic submodules $\mathbf{M}_0$ *of exponent* $r_0, \ldots,$ $\mathbf{M}_n$ *of exponent* r_n *so that*

$$\mathbf{T} \cong \mathbf{M}_0 \times \cdots \times \mathbf{M}_n.$$

Moreover, the natural number n *is uniquely determined by* $\mathbf{T}$*, the sequence* $r_n \mid r_{n-1} \mid \cdots \mid r_1 \mid r_0$ *is uniquely determined up to associates, and cyclic submodules* $\mathbf{M}_0, \ldots, \mathbf{M}_n$ *are determined but to isomorphism.*

Only the various aspects of uniqueness require proof at this point. However, these proofs follow the lines of the uniqueness portion of the proof and the Structure Theorem. We leave the details in the hands of the hard working graduate students. It is useful to note that the cyclic modules which are the factors in this direct decomposition may not themselves be directly indecomposable.

Using the Structure Theorem, for each principal ideal domain $\mathbf{R}$ we can define a function d such that $d(p^e, \mathbf{M})$ to be the number of direct factors isomorphic to the module $\mathbf{R}/(p^e)$ in any direct factorization of $\mathbf{M}$ into directly indecomposable modules, where $p \in R$ is prime, e is a positive natural number, and $\mathbf{M}$ is a finitely generated $\mathbf{R}$-module. In addition, we take $d(0, \mathbf{M})$ to be the number of direct factors isomorphic to the $\mathbf{R}$-module $\mathbf{R}$ (that is, the directly indecomposable free module).

Then we have the useful

Corollary. *Let* $\mathbf{R}$ *be and principal ideal domain and* $\mathbf{M}$ *and* $\mathbf{N}$ *be finitely generated* $\mathbf{R}$*-modules. Then* $\mathbf{M} \cong \mathbf{N}$ *if and only if* $d(q, \mathbf{M}) = d(q, \mathbf{N})$ *for all* q *such that either* $q = 0$ *or* q *is a positive power of a prime in* $\mathbf{R}$*.*

What this corollary asserts is that the system natural numbers

$$\langle d(q, \mathbf{M}) \mid q = 0 \text{ or } q \text{ is the positive power of a prime of } \mathbf{R} \rangle$$

is a complete system of invariants of $\mathbf{M}$—that is this system of natural numbers determines $\mathbf{M}$ up to isomorphism.

As noted earlier, modules over the ring $\mathbb{Z}$ of integers are essentially the same as Abelian groups since, for instance, $3u = (1+1+1)u = u+u+u$ and $-7v = -(v+v+v+v+v+v+v)$. In this way, we see that for a $\mathbb{Z}$-module $\mathbf{M} = \langle M, +, -, 0, a\cdot \rangle_{a \in \mathbb{Z}}$ the scalar multiplication is expressible by means of the additive structure $+, -$, and 0. In particular, any map between $\mathbb{Z}$-modules that respects $+, -$, and 0 must also respect the scalar multiplication, any subset of a $\mathbb{Z}$-module that is closed under $+, 1$, and contains 0 will also be closed under all the scalar multiplications, and a similar remark holds for direct products—in any of these constructions one may ignore the scalar multiplications along the way, but impose then on the result (the homomorphic image, the subalgebra, or the direct product) by means of repeated addition.

With this in mind, noting that $\mathbb{Z}$ is a principal ideal domain, we obtain

The Fundamental Theorem for Finitely Generated Abelian Groups.
Let $\mathbf{A}$ *be a finitely generated Abelian group. There is a natural number* n *such that:*

(a) *there are n finitely generated directly indecomposable subgroups $\mathbf{A}_0, \ldots, \mathbf{A}_{n-1}$ of $\mathbf{A}$ such that*

$$\mathbf{A} \cong \mathbf{A}_0 \times \cdots \times \mathbf{A}_{n-1}, \text{ and}$$

(b) *for any natural number m, if $\mathbf{B}_0, \ldots, \mathbf{B}_{m-1}$ are directly indecomposable Abelian groups such that $\mathbf{A} \cong \mathbf{B}_0 \times \cdots \times \mathbf{B}_{m-1}$, then $n = m$ and there is permutation σ of $\{0, \ldots, n-1\}$ so that $\mathbf{A}_i \cong \mathbf{B}_{\sigma(i)}$ for all $i < n$.*

Moreover, the finitely generated directly indecomposable Abelian groups are, up to isomorphism, the group $\langle \mathbb{Z}, +, -, 0 \rangle$ of integers with respect to addition (that is the free Abelian group of rank 1), and the cyclic groups of prime power order (these are the groups $\mathbb{Z}_q$ where q is a positive power of a prime number, the set of elements is $\{0, 1, \ldots, q-1\}$, and addition works modulo q). Finally, the free Abelian group of rank 1 is not of prime power order and if $r, s \in R$ are prime powers and $\mathbb{Z}_r \cong \mathbb{Z}_s$, then $r = s$.

This could be regarded as the elementary divisor version of the structure theorem for finitely generated Abelian groups. One could as easily formulate a structure theorem from the invariant factor point of view. To see how these two points of view compare consider a description, up to isomorphism, of all the Abelian groups of order 100. The prime factorization gives $100 = 2^2 5^2$. Using the elementary divisor perspective we see that the list, representative up to isomorphism and also pairwise nonisomorphic, of Abelian groups of order 100 is

$$\mathbb{Z}_4 \times \mathbb{Z}_{25} \qquad \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_{25} \qquad \mathbb{Z}_4 \times \mathbb{Z}_5 \times \mathbb{Z}_5 \qquad \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_5 \times \mathbb{Z}_5$$

while from the invariant factor perspective the list is

$$\mathbb{Z}_{100} \qquad \mathbb{Z}_2 \times \mathbb{Z}_{50} \qquad \mathbb{Z}_5 \times \mathbb{Z}_{20} \qquad \mathbb{Z}_{10} \times \mathbb{Z}_{10}.$$

At work here are the following direct decompositions:

$$\mathbb{Z}_{100} \cong \mathbb{Z}_4 \times \mathbb{Z}_{25} \qquad \mathbb{Z}_{50} \cong \mathbb{Z}_2 \times \mathbb{Z}_{25} \qquad \mathbb{Z}_{20} \cong \mathbb{Z}_4 \times \mathbb{Z}_5 \qquad \mathbb{Z}_{10} \cong \mathbb{Z}_2 \times \mathbb{Z}_5.$$

10.1 Problem Set 10

ALGEBRA HOMEWORK, EDITION 10
ELEVENTH WEEK
DECOMPOSING MODULES

PROBLEM 37.
Let **R** be a nontrivial integral domain and **M** be an **R**-module. Prove the set T of torsion elements is a submodule of **M** and $\mathbf{M}/T$ is torsion free.

PROBLEM 38.
Let **R** be a principal ideal domain and let **T** be a torsion **R**-module of exponent r. Prove that **T** has an element of order r.

PROBLEM 39.
Prove that the sequence of invariant factors (i.e. the sequence $r_0, r_1, \ldots, r_n$) mentioned in the Invariant Factor Theorem is uniquely determined by the module.

PROBLEM 40.
Let **M** be a finitely generated **R**-module, where **R** is a principal ideal domain. Prove each of the following.

(a) The direct decomposition using the Invariant Factor Theorem is the one using the smallest number of factors that are all cyclic.

(b) The direct decomposition using the Elementary Divisor Theorem is the one using the largest number of factors that are all cyclic.

LECTURE 11

DECOMPOSITION OF VECTOR SPACE WITH A DESIGNATED LINEAR OPERATOR

Let **V** be a finite dimensional vector space over a field **F** and let T be a linear operator on **V**—that is, T is an endomorphism of the vector space **V**. Our objective here is to decompose **V** as a direct product of subspaces that are invariant with respect to T. The most straightforward way to proceed with such a project is to adjoin T to the vector space as a new one place operation. This new algebraic system would have a binary operation + (the old vector addition), a designated element 0 (the zero vector), a one-place operation − of forming negations, a one-place operation aI for each $a \in F$ (the scalar muliplications), and the new one-place operation T. The program would then become the direct decomposition of this new algebraic system into directly indecomposable factors. It is possible to carry out this program, to prove the corresponding structure theorem (which would prove the existence and uniqueness of such decompositions and describe the directly indecomposable algebras, much as in the last section).

However, there is an alternate route to the same result that allows us to take advantage of the work we have done with modules. The idea is to regard **V** as a module over the principal ideal domain $\mathbf{F}[x]$ instead of over the field **F**. This means we have to define what $f(x) \cdot v$ means for every vector $v \in V$ and every polynomial $f(x) \in \mathbf{F}[x]$. Here is the definition:

$$f(x) \cdot v := f(T)(v).$$

Here $f(T) = a_0 I + a_1 T + a_2 T^2 + \cdots + a_n T^n$ where $f(x) = a_0 + a_1 x + \cdots + a_n x^n$ and $T^2 = T \circ T$, $T^3 = T \circ T \circ T$, and so on. It is easy to see that each $f(T)$ is a linear operator (that is, an endomorphism of the vector space **V**). The polynomials of degree 0 provide the ordinary scalar multiplications of the vector space. So construing **V** as a module over $\mathbf{F}[x]$ in effect adjoins many more one-place operations than our first approach, but they are all built up from the $a \cdot I$ and T by addition and composition. This is why the two approaches are equivalent.

Recall from linear algebra that the linear operators on a finite-dimensional vector space **V** constitute a finite dimensional vector space themselves. So for any linear operator T the set $\{I, T, T^2, T^3, \ldots\}$ is linearly dependent. This means that for some natural number m there are $a_0, a_1, \ldots, a_m \in F$

with $a_m \neq 0$ so that $a_0 I + a_1 T + \cdots + a_m T^m = 0$. In other words, there is a nonzero polynomial $f(x) \in \mathbf{F}[x]$ so that $f(T)$ is the zero linear operator (the map taking all vectors to the zero vector). Evidently, $\{f(x) \mid f(T) \text{ is the zero operator}\}$ is an ideal of $\mathbf{F}[x]$. Since $\mathbf{F}[x]$ is a principal ideal domain first ideal is generated by a single polynomial. In fact we can take this polynomial to be the monic polynomial $m_T(x)$ of least degree in this ideal. This polynomial is called the **minimal polynomial** of T.

Now fix a linear operator T on the finite dimensional vector space $\mathbf{V}$. We use $\mathbf{V}_T$ to denote the module over $\mathbf{F}[x]$ described above. We know that this module can be decomposed into a direct product of cyclic submodules. What do these cyclic submodules look like? Well, suppose that $v \in V$ is a generator. Then the submodule consists of all the vectors of the form $f(T)(v)$ as $f(x)$ runs through the ring of polynomials. Hence the linear span of the set $\{v, Tv, T^2 v, \ldots\}$ is the whole submodule. Since this submodule is, among other things, a subspace of $\mathbf{V}$ (with additional operations), we know that some finite subset must span the submodule. Let m be as small as possible so that $\{v, Tv, \ldots, T^m v\}$ spans the submodule. Then there are $a_0, \ldots, a_m \in F$ so that

$$T^{m+1} v = a_0 v + \cdots + a_m T^m v.$$

This leads to

$$T^{m+2} v = a_0 Tv + \cdots + a_m T^{m+1} v = a_0 Tv + \cdots + (a_0 v + \cdots + a_m T^m v).$$

In this way we see that m is also the smallest natural number so that $T^{m+1} v$ is a linear combination of $\{v, Tv, \ldots, T^m v\}$. I contend that this set is linearly independent. Suppose

$$b_0 v + b_1 Tv + \cdots + b_m T^m v = 0.$$

Now b_m must be 0, otherwise $T^m v = -\frac{b_0}{b_m} v - \cdots - \frac{b_{m-1}}{b_m} T^{m-1} v$. Once the term $b_m T^m v$ has been eliminated (because it is 0), we can apply the same reasoning the see that $b_{m-1} = 0$, and then that $b_{m-2} = 0$, and so on. In this way we establish the linear independence of $\{v, Tv, \ldots, T^m v\}$. Thus we see that our cyclic submodule, construed as an ordinary vector space over the field $\mathbf{F}$ has a very nice basis. We call this kind of basis a T**-cyclic** basis.

Here is what happens if we represent T with respect to this basis. Put

$$v_0 = v, v_1 = Tv, \ldots, v_m = T^m v.$$

Then

$$\begin{aligned}
Tv_0 &= v_1 = 0v_0 + 1v_1 + 0v_2 + \cdots + 0v_m \\
Tv_1 &= v_2 = 0v_0 + 0v_1 + 1v_2 + \cdots + 0v_m \\
&\vdots \\
Tv_{m-1} &= v_m = 0v_0 + 0v_1 + 0v_2 + \cdots + 1v_m \\
Tv_m &= a_0 v_0 + a_1 v_1 + \cdots + a_m v_m
\end{aligned}$$

This produces the matrix

$$\begin{pmatrix}
0 & 0 & 0 & \ldots & 0 & a_0 \\
1 & 0 & 0 & \ldots & 0 & a_1 \\
0 & 1 & 0 & \ldots & 0 & a_2 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & 0 & a_{m-1} \\
0 & 0 & 0 & \ldots & 1 & a_m
\end{pmatrix}$$

This is a very pleasant matrix with lots of entries 0 and the nonzero entries located in rather restrained positions. Let us rewrite that last equation:

$$\begin{aligned} Tv_m &= a_0v_0 + a_1v_1 + \cdots + a_mv_m \\ T^{m+1}v &= a_0v + a_1Tv + \cdots + a_mt^mv \\ 0 &= -T^{m+1}v + a_mT^mv + \cdots + a_1Tv + a_0v \\ 0 &= T^{m+1}v - a_mT^mv - \cdots - a_1Tv - a_0v \\ 0 &= (T^{m+1} - a_mT^m - \cdots - a_1T - a_0)v \\ 0 &= m_T(T)v \end{aligned}$$

where $m_T(x) = x^{m+1} - a_mx^m - \cdots - a_0 \in \mathbf{F}[x]$. Notice that this is a monic polynomial of least degree which belongs to the annihilator of $\mathbf{V}_T$. So it is an exponent of the module $\mathbf{V}_T$.

We could start with any monic polynomial $f(x) = b_0 + b_1x + \cdots + b_mx^m + x^{m+1}$ of positive degree. If this polynomial happened to be the minimal polynomial of some linear operator T so that $\mathbf{V}_T$ was a cyclic module, then the associated matrix, as above, would be

$$C_f = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & -b_0 \\ 1 & 0 & 0 & \dots & 0 & -b_1 \\ 0 & 1 & 0 & \dots & 0 & -b_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & -b_{m-1} \\ 0 & 0 & 0 & \dots & 1 & -b_m \end{pmatrix}.$$

This matrix is called the **companion matrix** of the polynomial $f(x)$. Observe that this is a $m+1 \times m+1$ matrix, where $m+1$ is the degree of $f(x)$. It easy to write down the companion matrix given the monic polynomial and, vice versa, given the companion matrix to write down the monic polynomial. A routine calculation also reveals that for any monic polynomial $f(x)$ we have

$$f(x) = \det(xI - C_f).$$

This means that $f(x)$ is the **characteristic polynomial** of its companion matrix.

We summarize these findings in the following Fact.

Fact. Let $\mathbf{V}$ be a finite dimensional vector space over a field $\mathbf{F}$ and T be a linear operator on $\mathbf{V}$ such that $\mathbf{V}_T$ is a cyclic module over $\mathbf{F}[x]$ with generator v and with minimal polynomial $m_T(x)$ of degree $n+1$. Then $\{v, Tv, T^2v, \dots, T^nv\}$ is a basis for $\mathbf{V}_T$ and, with respect to this basis, the matrix of T is the companion matrix of $m_T(x)$ and the characteristic polynomial of T is the same as the minimal polynomial of T.

Now let's apply the Invariant Factor Theorem:

The Rational Canonical Form Theorem.
Let $\mathbf{V}$ be a finite dimensional vector space over the field $\mathbf{F}$. Let T be a linear operator of $\mathbf{V}$. Then for some natural number n there are monic polynomials $f_0(x), f_1(x), \dots, f_n(x) \in \mathbf{F}[x]$ with $f_0(x) = m_T(x)$, the minimal polynomial of T, such that

$$f_n(x) \mid f_{n-1}(x) \mid \cdots \mid f_1(x) \mid f_0(x)$$

and cyclic submodules of $\mathbf{V}_T$ *(sometimes called* T*-cyclic subspaces of* $\mathbf{V}$*)* $\mathbf{V}_0$ *of exponent* $f_0(x)$*,...,* $\mathbf{V}_n$ *of exponent* $f_n(x)$ *so that*

$$\mathbf{V}_T \cong \mathbf{V}_0 \times \cdots \times \mathbf{V}_n.$$

Moreover, the natural number n *is uniquely determined by* T*, the sequence* $f_n(x) \mid f_{n-1}(x) \mid \cdots \mid f_1(x) \mid f_0(x)$ *of monic polynomials is uniquely determined, and cyclic submodules* $\mathbf{V}_0,\ldots,\mathbf{V}_n$ *are determined up to isomorphism. Furthermore, each of the submodules* $\mathbf{V}_k$ *for* $k \leq n$ *has a* T*-cyclic basis* B_k*, and* $B_n \cup \cdots \cup B_0$ *is a basis for* $\mathbf{V}$*. The linear operator* T *is represented with respect to this basis by*

$$\begin{pmatrix} C_{f_n} & 0 & 0 & \ldots & 0 \\ 0 & C_{f_{n-1}} & 0 & \ldots & 0 \\ 0 & 0 & C_{f_{n-2}} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & C_{f_0} \end{pmatrix}$$

where the companion matrices of the $f_k(x)$*'s are placed in order as square blocks along the diagonal, with all remaining entries of the matrix* 0*.*

The matrix representing T in this theorem is said to be in **rational canonical form**. The uniqueness assertions of the Extended Invariant Factor Theorem ensure that a linear operator has exactly one rational canonical form. The following important theorem is now a corollary.

The Cayley-Hamilton Theorem.
Let T *be a linear operator on a finite dimensional vector space over a field. The minimal polynomial of* T *divides the characteristic polynomial of* T*. Hence,* $f(T)$ *is the constantly zero linear operator, when* $f(x)$ *is the characteristic polynomial of* T*.*

Actually, it is easy to see that the characteristic polynomial is just the product of the invariant factors.

Recall from linear algebra that if A and B are $m \times m$ matrices with entries in the field $\mathbf{F}$, then we say that A and B are **similar** provided there is a linear operator T on the m-dimensional vector space over $\mathbf{F}$ such that T can be represented by both A and B (using appropriated bases). So we find

Rational Canonical Form Theorem: Matrix Version.
Let $\mathbf{F}$ *be a field and* m *be a positive natural number. Every* $m \times m$ *matrix with entries is* $\mathbf{F}$ *is similar to exactly one matrix in rational canonical form.*

Now let's turn to the elementary divisor perspective. Consider the case when $\mathbf{V}_T$ is a cyclic primary $\mathbf{F}[x]$ module. In this case, there is a vector $v \in V$, an irreducible monic polynomial $f(x) \in \mathbf{F}[x]$, and a positive natural number e so that $(f(x))^e$ is the order of v. As long as the field $\mathbf{F}$ is arbitrary, the polynomial $f(x)$ could be quite complicated—for instance it might have arbitrarily large degree. In such a situation, it would be difficult to improve on the process we used above the obtain the rational canonical form. However, two fields immediately come to mind wherre the situation is much more restrained. The field $\mathbb{C}$ of complex numbers has the property that all irreducible polynomials in $\mathbb{C}[x]$ have degree 1, while over the field $\mathbb{R}$ of real numbers there can also be irreducible polynomials of degree 2 but of no higher degrees. Both of these facts willl be

proved in the next semester. So let us consider that $f(x) = x - a$ for some $a \in F$. Then put

$$\begin{aligned} v_0 &= v \\ v_1 &= (T - aI)v_0 = Tv_0 - av_0 \\ v_2 &= (T - aI)v_1 = Tv_1 - av_1 \\ &\vdots \\ v_{e_1} &= (T - aI)v_{e-2} = Tv_{e-2} - av_{e-2} \\ 0 &= (T - aI)v_{e-1} = Tv_{e-1} - av_{e-1} \end{aligned}$$

Rearranging this just a bit we get

$$\begin{aligned} Tv_0 &= av_0 + v_1 \\ Tv_1 &= av_1 + v_2 \\ &\vdots \\ Tv_{e_2} &= av_{e-2} + v_{e-1} \\ Tv_{e-1} &= av_{e-1} \end{aligned}$$

Now by an argument similar (hard working graduate students will provide the variations needed) to the ones used above, we can see that the e distinct vectors $v_0, v_1, \ldots, v_{e-1}$ form a basis for the vector space $\mathbf{V}$. With respect to this basis, the linear operator T has the following matrix

$$\begin{pmatrix} a & 0 & 0 & 0 & \ldots & 0 & 0 \\ 1 & a & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & a & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & a & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & a & 0 \\ 0 & 0 & 0 & 0 & \ldots & 1 & a \end{pmatrix}.$$

A matrix of this form is called a **Jordan block** and the basis underlying is called a **Jordan basis**.

Observe that it is easy given the polynomial $(x - a)^e$ to write down its Jordan block and, vice versa, given the Jordan block, the polynomial can be recovered at once. Moreover, $(x - a)^e$ is the characteristic polynomial $\det(xI - J)$ of the Jordan block J.

This time, appealing the the Structure Theorem we get

The Jordan Canonical Form Theorem.
Let $\mathbf{V}$ *be a finite dimensional vector space over the field* $\mathbf{F}$. *Let* T *be a linear operator of* $\mathbf{V}$ *such that the irreducible factors of the minimal polynomial of* T *are all of degree* 1. *Then for some natural number* n *there are polynomials* $(x - a_0)^{e_0}, (x - a_1)^{e_1}, \ldots, (x - a_{n-1})^{e_{n-1}} \in \mathbf{F}[x]$, *namely the elementary divisors of* $\mathbf{V}_T$, *and and cyclic primary submodules of* $\mathbf{V}_T$ $\mathbf{V}_0$ *of exponent* $(x - a_0)^{e_0}, \ldots,$ $\mathbf{V}_n$ *of exponent* $(x - a_{n-1})^{e_{n-1}}$ *so that*

$$\mathbf{V}_T \cong \mathbf{V}_0 \times \cdots \times \mathbf{V}_{n-1}.$$

Moreover, the natural number n *is uniquely determined by* T, *the polynomials* $(x - a_k)^{e_k}$ *for* $k <$ n *are uniquely determined, and cyclic submodules* $\mathbf{V}_0, \ldots, \mathbf{V}_n$ *are determined up to isomorphism.*

Furthermore, each of the submodules $\mathbf{V}_k$ *for* $k \leq n$ *has a Jordan basis* B_k*, and* $B_n \cup \cdots \cup B_0$ *is a basis for* $\mathbf{V}$*. The linear operator* T *is represented with respect to this basis by*

$$\begin{pmatrix} J_0 & 0 & 0 & \dots & 0 \\ 0 & J_1 & 0 & \dots & 0 \\ 0 & 0 & J_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & J_{n-1} \end{pmatrix}$$

where the Jordan blocks of the $(x-a_k)^{e_k}$*'s are placed in some order as square blocks along the diagonal, with all remaining entries of the matrix* 0.

The matrix mentioned at the conclusion of this theorem is said to be in **Jordan canonical form**. It should be noted here, that there may be several matrices in Jordan form associated with the linear operator T according to this theorem. This happens because the order in which the Jordan blocks appear along the diagonal is arbitrary. The uniqueness part of the Structure Theorem The permutation mentioned in the statement of the Structure Theorem reflects the same point. It is clear that if A and B are two matrices in Jordan form that can be obtained from each other by rearranging the Jordan blocks along the diagonal, the A and B are similar. So the Structure Theorem gives us

Jordan Canonical Form Theorem: Matrix Version.
Let $\mathbf{F}$ *be a field and* m *be a positive natural number. Let* A $m \times m$ *matrix with entries is* $\mathbf{F}$ *with the additional property that the irreducible factors of the minimal polynomial of the matrix are all of degree* 1*. Then* A *is similar to a matrix in Jordan canonical form and any matrices in Jordan canonical form that are similar to* A *can be obtained from each other by rearranging the Jordan blocks.*

11.1 Problem Set 11

Algebra Homework, Edition 11
Twelfth Week
Modules

Problem 41.
Let **R** and **S** be commutative Noetherian rings. Prove that $\mathbf{R} \times \mathbf{S}$ is also Noetherian.

Problem 42.
Let **R** be a commutative ring such that every submodule of a free **R**-module is also a free **R**-module. Prove that **R** is a principal ideal domain.

Problem 43.
Let **R** by a principal ideal domain and let **M** and **N** be finitely generated **R**-modules such that $\mathbf{M} \times \mathbf{M} \cong \mathbf{N} \times \mathbf{N}$. Prove $\mathbf{M} \cong \mathbf{N}$.

Problem 44.
Give an example of two 4×4 matrices with real entries that have the same minimal polynomial and the same characteristic polynomial but are not similar.

BIBLIOGRAPHY

Artin, Emil

(1998) *Galois theory*. second. Edited and with a supplemental chapter by Arthur N. Milgram. Mineola, NY: Dover Publications Inc., pp. iv+82. ISBN: 0-486-62342-4.

(2007) *Algebra with Galois theory*. Vol. 15. Courant Lecture Notes in Mathematics. Notes by Albert A. Blank, Reprint of the 1947 original [ıt Modern higher algebra. Galois theory, Courant Inst. Math. Sci., New York]. Courant Institute of Mathematical Sciences, New York, pp. viii+126. ISBN: 978-0-8218-4129-7.

Artin, Michael

(1991) *Algebra*. Englewood Cliffs, NJ: Prentice Hall Inc., pp. xviii+618. ISBN: 0-13-004763-5.

Birkhoff, Garrett and Saunders Mac Lane

(1965) *A survey of modern algebra*. Third edition. New York: The Macmillan Co., pp. x+437.

Dummit, David S. and Richard M. Foote

(2004) *Abstract algebra*. Third. Hoboken, NJ: John Wiley & Sons Inc., pp. xii+932. ISBN: 0-471-43334-9.

Grillet, Pierre Antoine

(2007) *Abstract algebra*. Second. Vol. 242. Graduate Texts in Mathematics. New York: Springer, pp. xii+669. ISBN: 978-0-387-71567-4.

Grove, Larry C.

(2004) *Algebra*. Reprint of the 1983 original, with an errata list on pp. xv–xvi. Mineola, NY: Dover Publications Inc., pp. xviii+299. ISBN: 0-486-43947-X.

Herstein, I. N.

(1975) *Topics in algebra*. Second. Lexington, Mass.: Xerox College Publishing, pp. xi+388.

Hungerford, Thomas W.

(1980) *Algebra*. Vol. 73. Graduate Texts in Mathematics. Reprint of the 1974 original. New York: Springer-Verlag, pp. xxiii+502. ISBN: 0-387-90518-9.

Isaacs, I. Martin

(2009) *Algebra: a graduate course*. Vol. 100. Graduate Studies in Mathematics. Reprint of the 1994 original. Providence, RI: American Mathematical Society, pp. xii+516. ISBN: 978-0-8218-4799-2.

Jacobson, Nathan

(1975a) *Lectures in abstract algebra*. Volume II: Linear algebra, Reprint of the 1953 edition [Van Nostrand, Toronto, Ont.], Graduate Texts in Mathematics, No. 31. New York: Springer-Verlag, pp. xii+280.

(1975b) *Lectures in abstract algebra. III*. Theory of fields and Galois theory, Second corrected printing, Graduate Texts in Mathematics, No. 32. New York: Springer-Verlag, pp. xi+323.

Jacobson, Nathan

(1975c) *Lectures in abstract algebra. Vol. I*. Basic concepts, Reprint of the 1951 edition, Graduate Texts in Mathematics, No. 30. New York: Springer-Verlag, pp. xii+217.

(1985) *Basic algebra. I*. Second. New York: W. H. Freeman and Company, pp. xviii+499. ISBN: 0-7167-1480-9.

(1989) *Basic algebra. II*. Second. New York: W. H. Freeman and Company, pp. xviii+686. ISBN: 0-7167-1933-9.

Lam, T. Y.

(1999) *Lectures on modules and rings*. Vol. 189. Graduate Texts in Mathematics. New York: Springer-Verlag, pp. xxiv+557. ISBN: 0-387-98428-3. DOI: 10.1007/978-1-4612-0525-8. URL: http://dx.doi.org/10.1007/978-1-4612-0525-8.

(2001) *A first course in noncommutative rings*. Second. Vol. 131. Graduate Texts in Mathematics. New York: Springer-Verlag, pp. xx+385. ISBN: 0-387-95183-0. DOI: 10.1007/978-1-4419-8616-0. URL: http://dx.doi.org/10.1007/978-1-4419-8616-0.

Lang, Serge

(2002) *Algebra*. third. Vol. 211. Graduate Texts in Mathematics. New York: Springer-Verlag, pp. xvi+914. ISBN: 0-387-95385-X. DOI: 10.1007/978-1-4613-0041-0. URL: http://dx.doi.org/10.1007/978-1-4613-0041-0.

Mac Lane, Saunders and Garrett Birkhoff

(1979) *Algebra*. Second. New York: Macmillan Inc., pp. xv+586. ISBN: 0-02-374310-7.

Noether, Emmy

(1983) *Gesammelte Abhandlungen*. Edited and with an introduction by Nathan Jacobson, With an introductory address by P. S. Alexandrov [P. S. Aleksandrov]. Berlin: Springer-Verlag, viii+777 pp. (1 plate). ISBN: 3-540-11504-8.

Rotman, Joseph J.

(2010) *Advanced modern algebra*. Vol. 114. Graduate Studies in Mathematics. Second edition [of MR2043445]. Providence, RI: American Mathematical Society, pp. xvi+1008. ISBN: 978-0-8218-4741-1.

Waerden, B. L. van der

(1943) *Moderne Algebra. Parts I and II*. G. E. Stechert and Co., New York, pp. 272+224.

(1991a) *Algebra. Vol. I*. Based in part on lectures by E. Artin and E. Noether, Translated from the seventh German edition by Fred Blum and John R. Schulenberger. New York: Springer-Verlag, pp. xiv+265. ISBN: 0-387-97424-5. DOI: 10.1007/978-1-4612-4420-2. URL: http://dx.doi.org/10.1007/978-1-4612-4420-2.

(1991b) *Algebra. Vol. II*. Based in part on lectures by E. Artin and E. Noether, Translated from the fifth German edition by John R. Schulenberger. New York: Springer-Verlag, pp. xii+284. ISBN: 0-387-97425-3.

INDEX