Lecture Notes For

Mathematics 150C

Dr. Tyler J. Evans

Spring 2001

Contents

Chapter 1

Rings and Fields

1.1 Definition of a Ring

Up until now, we have focused our attention primarily on sets with a single binary operation. That is, we have been studying groups. However, we have also considered sets with two binary operations, namely fields. The next topic of study, rings, deals with sets with two binary operations, but we will weaken the field axioms slightly with respect to the "multiplication". Here is the main definition.

Definition 1.1.1 (Ring) *A set R together with two binary operations $+$ and $\cdot$ is a* **ring** *if*

R1. *$(R, +)$ is an abelian group. (We write 0 for the identity.)*

R2. *The operation $\cdot$ is associative.*

R3l. *We have $a(b + c) = ab + ac$ for all $a, b, c \in R$.*

R3r. *We have $(a + b)c = ac + bc$ for all $a, b, c \in R$.*

Before we give examples, we invite the reader to think about the properties of a field that have been omitted from the previous definition. For example, if R is a ring, the non-zero elements $R^\times$ do not necessarily form a group under $\cdot$.

Example 1.1.2 Every field is a ring. As we have noted, the converse is false. In particular, the familiar fields $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all rings with the usual addition and multiplication.

Example 1.1.3 The integers $\mathbb{Z}$ form a ring under usual addition and multiplication. This ring is not a field however, since $n \in Z$ has no multiplicative inverse if $n \neq \pm 1$.

Example 1.1.4 If R is a ring, the set

$$M_n(R) = \{A = [a_{ij}] : a_{ij} \in R, 1 \leq i, j \leq n\}$$

of $n \times n$ matrices with entries in R is a ring with the usual addition and multiplication of $n \times n$ matrices over R. This ring is also not a field.

Example 1.1.5 If X is a set, then the set of all functions $f : X \to \mathbb{R}$ (or $\mathbb{C}$ or any ring R!) is a ring under pointwise addition and multiplication of functions. That is, for $f, g : X \to \mathbb{R}$, we define $f + g$ and fg by

$$(f + g)(x) = f(x) + g(x)$$

and

$$(fg)(x) = f(x)g(x).$$

We will denote this ring by $F(X, \mathbb{R})$.

We leave it as an exercise for the reader to verify that each of the above examples is indeed a ring.

Example 1.1.6 We define a multiplication in the group $\mathbb{Z}_n$ by $\overline{a} \cdot \overline{b} = \overline{ab}$. We will see that this multiplication is well defined and hence $\mathbb{Z}_n$ is a ring, called the **ring of integers modulo** n. We will also see that $\mathbb{Z}_n$ is a field if and only if n is a prime.

We now want to see what elementary properties of rings that we can deduce only from the definition. For notation, recall that 0 will denote the additive identity of R, and if $a \in R$, $-a$ denotes the additive inverse for a.

Proposition 1.1.7 *If R is a ring, then for all $a, b \in R$, we have*

1. $0a = a0 = 0.$

2. $a(-b) = (-a)b = -(ab).$

3. $(-a)(-b) = ab.$

Proof. Let $a, b \in R$ be arbitrary.

(1) Using the left distributive law, we have $a0 = a(0 + 0) = a0 + a0$, and therefore $0 = a0$ by the cancellation law in the group $(R, +)$. Similarly we have $0 = 0a$.

(2) Recalling that the additive inverse for an element in R is unique, we note that

$$ab + a(-b) = a(b + (-b)) = a0 = 0$$

so that $a(-b) = -(ab)$. Similarly we have $(-a)b = -(ab)$.

(3) Now, using (2), we have $(-a)(-b) = -(a(-b)) = -(-(ab))$. Now $-(-(ab))$ is the element that when added to $-(ab)$ gives 0. Clearly ab satisfies this property so that $(-a)(-b) = ab$ as desired. ■

After our extensive work with groups and vector spaces, the following definition should feel very natural to the reader (does it?).

Definition 1.1.8 (Ring homomorphism) *If R and S are rings, a map $\varphi : R \to S$ is called a* **ring homomorphism** *if for all $a, b \in R$, we have both*

1. *$\varphi(a + b) = \varphi(a) + \varphi(b)$,*
2. *$\varphi(ab) = \varphi(a)\varphi(b)$.*

This definition says that a ring homomorphism is a homomorphism of abelian groups R and S that also preserves the multiplication. In particular, we can speak of the kernel and image of φ as a group homomorphism. Recall that the homomorphism $\varphi : R \to S$ (as groups) gives rise to a quotient group. We will see soon that we also have a “quotient ring” in this situation as well.

Example 1.1.9 Let $F(\mathbb{R}, \mathbb{R})$ be the ring of all functions $f : \mathbb{R} \to \mathbb{R}$ as defined above. If $a \in \mathbb{R}$, then a determines a ring homomorphism $\varphi_a : F(\mathbb{R}) \to \mathbb{R}$ via the formula

$$\varphi_a(f) = f(a).$$

We can verify that φ_a is a homomorphism immediately. If $f, g \in F(\mathbb{R})$, then by definition we have

$$\varphi_a(f + g) = (f + g)(a) = f(a) + g(a) = \varphi_a(f) + \varphi_a(g)$$

and

$$\varphi_a(fg) = (fg)(a) = f(a)g(a) = \varphi_a(f)\varphi_a(g).$$

This homomorphism is called the **evaluation (at a) homomorphism**. It will be very important when we study polynomial rings in detail. Indeed, solving a polynomial equation $p(x) = 0$ amounts to finding $a \in \mathbb{R}$ such that $p \in \ker \varphi_a$.

Again, to the experienced MAT 150AB student, the following definition comes as no surprise.

Definition 1.1.10 (Ring isomorphism) *A ring homomorphism* $\varphi : R \to S$ *is an* **isomorphism** *if* φ *is bijective as a function. If there exists an isomorphism* $\varphi : R \to S$*, we say that* R *and* S *are* **isomorphic**.

Example 1.1.11 We note easily that the set $2\mathbb{Z}$ of even integers form a ring under the usual addition and multiplication of integers. Moreover, the map $\varphi : \mathbb{Z} \to 2\mathbb{Z}$ defined by $\varphi(n) = 2n$ is easily seen to be a isomorphism of the abelian groups $\mathbb{Z}$ and $2\mathbb{Z}$. However, if $n, m \in \mathbb{Z}$, then

$$\varphi(nm) = 2nm \neq (2n)(2m) = \varphi(n)\varphi(m)$$

so that φ is **not** an isomorphism of rings. In fact, the rings $\mathbb{Z}$ and $2\mathbb{Z}$ are not isomorphic. (Did we prove this last statement here?)

In many of the examples we have considered so far (all but one of them, in fact), our rings have had a multiplicative identity as well as an additive identity. That is, we have seen an element $1 \in R$ with $1a = a1 = a$ for all $a \in R$. We note that the singleton set $\{0\}$ is a ring with $0 + 0 = 0$ and $0 \cdot 0 = 0$, and in this case, 0 is **both** and additive and multiplicative identity. This is the only case where this can happen though. Indeed, if 0 is a multiplicative identity, then by our proposition above, $a = 0a = 0$ for all $a \in R$. We call this singleton ring the **trivial ring**. We **always** exclude this example when we speak of rings with a multiplicative identity. That is, if $1 \in R$ is a multiplicative identity, then $1 \neq 0$.

Definition 1.1.12 (Commutative ring, unity) *If* R *is a ring such that* $ab = ba$ *for all* $a, b \in R$*, then we say* R *is a* **commutative ring**. *If* R *has a multiplicative identity* $1 \in R$*, then we say* R *is a* **ring with unity**. *The multiplicative identity is called* **unity**. *A ring with unity is also called a* **unital ring**.

Proposition 1.1.13 *If* R *is a ring with unity* $1 \in R$*, then* 1 *is the only unity.*

Proof. This follows from the general result that an identity for an associative binary operation is always unique. ∎

If $R_1, R_2, \ldots, R_n$ are all rings, then we can multiply elements of the product $R_1 \times \cdots \times R_n$ componentwise making the set $R = R_1 \times \cdots \times R_n$ into a ring called the **product ring**. Clearly the product R is commutative iff. each factor is commutative and if each R_i has unity 1_i, then $(1_1, \ldots, 1_n) \in R$

is unity for R. We leave the precise proofs of these statements as exercises for the reader. Here is one more definition.

Definition 1.1.14 (Unit, field,skew-field) *Let R be a ring with unity. An element $u \in R$ is a* **unit** *if $uv = vu = 1$ for some $v \in R$. If every non-zero element of a ring R is a unit, then R is called a* **division ring**. *A* **field** *is a commutative division ring. A non-commutative division ring is sometimes called a* **skew-field**.

We end the lecture with the notion of a **subring**. Again, the reader should be able to supply her or his own definition. In particular, a subset S of a ring R is a subring if S is itself a ring under the same operations. Similarly you can define subfield and sub-anything for that matter!

1.2 Integral Domains

One of the many useful algebraic facts about the ring of complex numbers $\mathbb{C}$ and its subrings $\mathbb{R}, \mathbb{Q}$ and $\mathbb{Z}$ is the so called **zero product rule**. It states that the only way a product of two elements in the ring is equal to zero is for one of the elements to be zero. We formalize this in the following definition.

Definition 1.2.1 (Divisor of zero) *If R is a ring and $0 \neq a \in R$, then a is a* **(left) divisor of zero** *if there exists an element $b \neq 0$ such that $ab = 0$. In this case, the element b is a* **(right) divisor of zero**.

If R is a commutative ring, then every left divisor of zero is also a right divisor of zero and vice versa. In this case we simply say **divisor of zero**. Our remarks above state that $\mathbb{C}$ has no divisors of zero.

Proposition 1.2.2 *In the ring $\mathbb{Z}_n$, the divisors of zero are precisely those elements $m \in \mathbb{Z}_n$ such that $(m, n) > 1$.*

Proof. Let $d = (m, n)$ and note that

$$m \cdot \frac{n}{d} = \frac{m}{d} \cdot n \equiv 0 \pmod{n}.$$

If $d > 1$, then $0 \not\equiv n/d \pmod{n}$ so that m is a zero divisor. Conversely, if $d = 1$ and $ma \equiv 0 \pmod{p}$, then $n|ma$ so that necessarily $n|a$. This implies $a \equiv 0 \pmod{p}$ so that m is not a zero divisor. ∎

Corollary 1.2.3 *If p is a prime, the ring $\mathbb{Z}_p$ has no divisors of zero.* ■

Definition 1.2.4 (Cancellation law for rings) *A ring R has the* **left cancellation law** *if for all $a \neq 0$ and all $b, c \in R$, $ab = ac$ implies $b = c$. Similarly, R has the* **right cancellation law** *if for all $a \neq 0$ and all $b, c \in R$, $ba = ca$ implies $b = c$. If R has both the left and right cancellation laws, we say the* **cancellation law holds** *for R.*

Theorem 1.2.5 *The cancellation law holds for R if and only if R has no left or right divisors of zero.*

Proof. Suppose R has the cancellation law. If $a \neq 0$ and $ab = 0$, then $ab = a0$ so that $b = 0$. It follows that R has no left divisors of zero. Similarly one can show that R has no right divisors of zero. Conversely, if R has no left or right divisors of zero and $ab = ac$ for some $0 \neq a$ and $b, c \in R$, then $0 = ab - ac = a(b - c)$. It follows that $b - c = 0$ so that $b = c$, and hence R has the left cancellation law. Similarly R has the right cancellation law. ■

This brings us to the main definition of the lecture.

Definition 1.2.6 (Integral domain) *An* **integral domain** *is a commutative unital ring with no divisors of zero.*

Example 1.2.7 All of the familiar rings $\mathbb{C}, \mathbb{R}, \mathbb{Q}$ and $\mathbb{Z}$ are integral domains.

Theorem 1.2.8 *Every field is an integral domain.*

Proof. Suppose F is a field and $a, b \in F$ satisfy $ab = 0$. If $a \neq 0$, then $a^{-1} \in F$ and $0 = a^{-1}ab = b$ so that a is not a divisor of zero. ■

We remark that the proof of this theorem actually shows that if $u \in R$ is a unit, then u is not a divisor of zero.

Theorem 1.2.9 *Every finite integral domain is a field.*

Proof. Let $1 = a_0, a_1, \ldots, a_n$ be all the non-zero elements of a finite integral domain D. If a_j is any one of the elements then each element in the list $a_j a_0, a_j a_1, \ldots, a_j a_n$ is distinct by the cancellation law that holds in an integral domain. It follows that $1 = a_j a_k$ for some k so that a_j is a unit. Therefore D is a commutative unital ring in which every non-zero element is a unit and hence D is a field. ■

Corollary 1.2.10 *If p is prime, then $\mathbb{Z}_p$ is a field.* ■

Definition 1.2.11 (Characteristic of a ring) *If R is a ring, the* **characteristic of** *R is the smallest positive integer n such that $n \cdot a = 0$ for all $a \in R$. (Here, $n \cdot a = a + \cdots + a$ n times.) If no such positive integer exists, we say R has* **characteristic zero***.*

Example 1.2.12 The characteristic of $\mathbb{Z}_n$ is n. The rings $\mathbb{C}, \mathbb{R}, \mathbb{Q}$ and $\mathbb{Z}$ all have characteristic zero.

Theorem 1.2.13 *If R is a unital ring, the R has characteristic n if and only if n is the smallest positive integer such that $n \cdot 1 = 0$.*

Proof. Of course if R has characteristic n, then $n \cdot 1 = 0$. Conversely, suppose n is the smallest positive integer that satisfies $n \cdot 1 = 0$ and let $a \in R$ be arbitrary. Then

$$n \cdot a = a + \cdots + a = a(1 + \cdots + 1) = a(n \cdot 1) = a0 = 0$$

so that R has characteristic n. ■

1.3 Field of Quotients

We saw in the last lecture that there is no difference between finite integral domains and finite fields. The integers $\mathbb{Z}$ provide an example of a integral domain that is not a field. The purpose of this lecture is to show that every integral domain can be regarded as being contained in some field called the *field of quotients of the integral domain.* We will see that the field $\mathbb{Q}$ of rational numbers is the field of quotients for the integers. Indeed, the construction given below is just an abstract version of the construction of the rational numbers from the integers. Nothing we do here is too difficult, but to be completely careful, the construction is quite long. We therefore will give a brief outline of what we propose to do. Our goal is to construct the smallest field F that contains a given integral domain D. We will proceed in four steps:

1. Define the elements of F.

2. Define two binary operations $+$ and $\cdot$ on F.

3. Show that $(F, +, \cdot)$ is a field.

4. Show that there is an injective ring homomorphism $D \to F$.

We will use the result of step 4 to identify D with its image in F so that we can think of D as a subdomain of F. We will leave some of the details to the reader as we proceed.

Step 1. Let D be an integral domain and let S be the subset of $D \times D$ defined by

$$S = \{(a,b) \in D \times D : b \neq 0\}.$$

The set S is too big to be our field F, so we cut it down a bit with the following lemma.

Lemma 1.3.1 *The relation $\sim$ defined on S by $(a,b) \sim (c,d)$ iff. $ad = bc$ is an equivalence relation on S.*

Proof. First, since D is commutative we have $ab = ba$ for all $a, b \in D$ so that $(a,b) \sim (a,b)$. Also, if $(a,b) \sim (c,d)$, then $ad = bc$ so that $cb = da$ and hence $(c,d) \sim (a,b)$. Finally if $(a,b) \sim (c,d)$ and $(c,d) \sim (e,f)$, then $ad = bc$ and $cf = de$. Now, using this and the commutativity in D, we compute

$$afd = adf = bcf = bde = bed.$$

But $d \neq 0$ so that the cancellation law in D implies $af = be$ and hence $(a,b) \sim (e,f)$. ■

We will denote the equivalence class containing (a,b) by $[(a,b)]$. We now complete step 1 by defining

$$F = S/\sim$$

to be the set of all equivalence classes in S under the relation $\sim$.

Step 2. We will now give the definitions of $+$ and $\cdot$ in F. We will define these operations in terms of representatives of the class, so that we will need to show that they are well defined. We state the precise result in the form of a lemma.

Lemma 1.3.2 *The operations $+$ and $\cdot$ defined on F by the formulas*

$$[(a,b)] + [(c,d)] = [(ad + bc, bd)]$$

and

$$[(a,b)][(c,d)] = [(ac, bd)]$$

are well defined binary operations on F.

Proof. We begin by noting that since $[(a,b)], [(c,d)] \in F$, the pairs $(a,b), (c,d) \in S$ so that $b \neq 0$ and $d \neq 0$. Since D is an integral domain, it follows that $bd \neq 0$ and hence $(ad + bc, bd), (ac, bd) \in S$.

Therefore the right-hand sides of the defining equations above both lie in F. It remains to show that the operations are well defined.

Suppose that $(a', b') \in [(a, b)]$ and $(c', d') \in [(c, d)]$. We must show that $(a'd' + b'c', b'd') \in [(ad + bc, bd)]$ and $(a'c', b'd') \in [(ac, bd)]$. We will do the first one, and leave the second to the reader.

We have, by hypothesis, $a'b = b'a$ and $c'd = d'c$ so that multiplying the first by $d'd$ and the second by $b'b$ and adding gives

$$a'bd'd + c'db'b = b'ad'd + d'cb'b.$$

Using various properties in the integral domain, we have

$$(a'd' + b'c')bd = b'd'(ad + bc)$$

so that $(a'd' + b'c', b'd') \in [(ad + bc, bd)]$ as desired. Similarly one shows that $(a'c', b'd') \in [(ac, bd)]$ and hence the operations are well defined. ■

Step 3. In this step, we must simply verify that the operations defined in step 2 make F into a field. The details are largely boring. Except for multiplicative inverses, each field axiom follows directly from the corresponding property that holds in the integral domain D. We leave the details to the reader.

Step 4. It remains to show that D is isomorphic to a subdomain of F. We accomplish this with the following lemma.

Lemma 1.3.3 *The map $i : D \to F$ defined by $i(a) = [(a, 1)]$ is an injective ring homomorphism.*

Proof. For any two elements $a, b \in D$, we have $i(a + b) = [((a + b), 1)]$ and

$$i(a) + i(b) = [(a, 1)] + [(b, 1)] = [(a \cdot 1 + 1 \cdot a, 1 \cdot 1)] = [(a + b, 1)]$$

so that $i(a + b) = i(a) + i(b)$. Similarly, $i(ab) = [(ab, 1)]$ and

$$i(a)i(b) = [(a, 1)][(b, 1)] = [(ab, 1 \cdot 1)] = [(ab, 1)]$$

and hence i is a ring homomorphism. Finally, if $i(a) = i(b)$, then $[(a, 1)] = [(b, 1)]$ so that $a \cdot 1 = 1 \cdot b$ and hence $a = b$. ■

This completes the construction of the field of quotients for the integral domain D. We summarize what we have proved, as well as clarify in which sense F is the *smallest* field containing D in the following theorem.

Theorem 1.3.4 *If D is an integral domain, then there exists a field F and an injective ring homomorphism $i : D \to F$ with the property that if E is any field containing D, then there exists an injective ring homomorphism $\psi : F \to E$ with $\psi(i(a)) = a$ for all $a \in D$.*

Proof. We let F be the field of quotients constructed above and let $i : D \to F$ be the map $i : a \mapsto [(a, 1)]$. Then we have seen that i is an injective ring homomorphism.

For notation, if $0 \neq b \in D$, then since $D \subseteq E$ and E is a field, b has a multiplicative inverse in E which we denote by b^{-1}. Note that the element b^{-1} may not be in D. Now we define $\psi : F \to E$ by $\psi([(a, b)]) = ab^{-1}$. To show that ψ is well defined, we note that if $(a', b') \in [(a, b)]$, then $ab' = ba'$ so that $ab^{-1} = a'b'^{-1}$ in E. Therefore ψ is well defined. We leave the proof that ψ is an injective ring homomorphism to the reader. ■

Corollary 1.3.5 *Every field E containing an integral domain D contains the field of quotients F of D.*

Proof. Referring to the proof of theorem (1.3.4), if E contains D, then E contains $\psi(F)$, and $\psi(F)$ is isomorphic to F. ■

Corollary 1.3.6 *Any two fields of quotients for an integral domain D are isomorphic.*

Proof. If E is another field of quotients of D, then the proof of theorem (1.3.4) can be suitably modified to show that there is an injective ring homomorphism $\varphi : E \to F$ with $\varphi(a) = a$ for all $a \in D$. You can check that $\varphi \circ \psi = 1_F$ and $\psi \circ \varphi = 1_E$ so that E is isomorphic to F. ■

1.4 Polynomial Rings

Based on your previous algebraic experience, you are probably completely willing to accept the idea that we can form polynomials with coefficients from an arbitrary ring R. Moreover, you are most likely willing to believe that we can add and multiply such polynomials using the usual rules so that in fact, the set of all polynomials with coefficients in R form a ring $R[X]$. With that said, we want to emphasize that we will be working with such polynomials from a slightly different point of view, and there are many details about the constructions involved that we wish to discuss carefully.

To begin, we will call X an **indeterminate** rather than a variable. If our ring is the ring $\mathbb{Z}$ of integers, one polynomial in $\mathbb{Z}[X]$ is $1X$ which we write simply as X. In solving polynomial equations, the reader is probably used to writing expressions like $X = 1$ or $X = 2$. However, we will *never* write such things because $1, 2 \in \mathbb{Z}$ and $X \notin \mathbb{Z}$. Similarly, we will never write $X + 4 = 0$ because the polynomial $X + 4$ is not the additive identity in the ring $\mathbb{Z}[X]$. At this point, the reader may feel we are being too formal in our discussion. What we are actually trying to do is to develop the theory of "solving polynomial equations" purely algebraically, and we want to avoid saying two things are equal in one context, and not equal in another.

The first step we need to take is to give a formal definition of what a polynomial is. This may seem easy: a polynomial with coefficients in a ring R should be a formal sum

$$a_0 + a_1X + \cdots + a_nX^n.$$

Without saying something else, this definition is not good enough however. After all, surely we want the two distinct formal sums $1 + 2X$ and $1 + 2X + 0X^2$ to denote the same polynomial. Surprising as it may seem, we work around this difficulty by taking *infinite* sums! Here is the main definition.

Definition 1.4.1 (Polynomial) *Let R be a ring. A* **polynomial f with coefficients in** *R is an infinite formal sum*

$$f = \sum_{i=0}^{\infty} a_iX^i = a_0 + a_1X + \cdots + a_nX^n + \cdots$$

where $a_i \in R$ for all $i \in \mathbb{N}$ and $a_i = 0$ for all but finitely many values of i. The elements $a_i \in R$ are called the **coefficients** *of the polynomial. If for some $i > 0$, $a_i \neq 0$ but $a_j = 0$ for all $j > i$, then i is the* **degree of** *f. If no such $i > 0$ exists, then we say f has* **degree zero***.*

To simplify our notations for polynomials, we agree that if $a_i = 0$ for $i > n$, then we will not write the terms a_iX^i so that a polynomial is written

$$a_0 + a_1X + \cdots + a_nX^n.$$

Moreover, if R is unital, we write X^i in place of $1X^i$. Finally, we omit any term with $a_i = 0$ so that the polynomial $1 + 0X + X^2$ becomes $1 + X^2$. An element of the ring R is called a **constant polynomial**.

Definition 1.4.2 *Let R be a ring and let $f = \sum_{i=0}^{\infty} a_iX^i$ and $g = \sum_{i=0}^{\infty} b_iX^i$ be two polynomials with coefficients in R. We define the* **sum** *$f + g$ and the* **product** *fg by the formulas*

$$f + g = \sum_{i=0}^{\infty} (a_i + b_i)X^i,$$

and

$$fg = \sum_{i=0}^{\infty} \left(\sum_{k=0}^{i} a_k b_{i-k} \right) X^i.$$

We remark that if a_i and b_i are zero for all but finitely many values of i, then same is true for $c_i = a_i + b_i$ and $d_i = \sum_{k=0}^{i} a_k b_{i-k}$ so that $f + g$ and fg are polynomials with coefficients in R. We remark that if R is not commutative, we should not expect $\sum_{k=0}^{i} a_k b_{i-k}$ to equal $\sum_{k=0}^{i} b_k a_{i-k}$. The following theorem is routine, although the notations involved make it seen difficult!

Theorem 1.4.3 *If R is a ring, the set $R[X]$ of all polynomials with coefficients in R is a ring under polynomial addition and multiplication. If R is commutative, then $R[X]$ is commutative. If R is unital, then $R[X]$ is unital.*

Proof. We leave the verification that $(R[X], +)$ is an abelian group to the reader. The verification of associativity of $\cdot$ and the distributive laws is straight forward, if not a little cumbersome. We will write out associativity. Applying the ring axioms to $a_i, b_j, c_k \in R$, we compute

$$\begin{aligned}
&\left[\left(\sum_{i=0}^{\infty} a_i X^i \right) \left(\sum_{j=0}^{\infty} b_j X^j \right) \right] \left(\sum_{k=0}^{\infty} c_k X^k \right) \\
&= \left[\sum_{n=0}^{\infty} \left(\sum_{i=0}^{n} a_i b_{n-i} \right) X^n \right] \left(\sum_{k=0}^{\infty} c_k X^k \right) \\
&= \sum_{s=0}^{\infty} \left[\sum_{n=0}^{s} \left(\sum_{i=0}^{n} a_i b_{n-i} \right) c_{s-n} \right] X^s \\
&= \sum_{s=0}^{\infty} \left(\sum_{i+j+k=s} a_i b_j c_k \right) X^s \\
&= \sum_{s=0}^{\infty} \left[\sum_{m=0}^{s} a_{s-m} \left(\sum_{j=0}^{m} b_j c_{m-j} \right) \right] X^s \\
&= \left(\sum_{i=0}^{\infty} a_i X^i \right) \left[\sum_{m=0}^{\infty} \left(\sum_{j=0}^{m} b_j c_{m-j} \right) X^m \right] \\
&= \left(\sum_{i=0}^{\infty} a_i X^i \right) \left[\left(\sum_{j=0}^{\infty} b_j X^j \right) \left(\sum_{k=0}^{\infty} c_k X^k \right) \right],
\end{aligned}$$

and therefore multiplication is associative. Similarly you can show that the distributive laws hold and hence $R[X]$ is a ring. It is clear that if R, is commutative, then $\sum_{k=0}^{i} a_k b_{i-k} = \sum_{k=0}^{i} b_k a_{i-k}$ so that $fg = gf$ and hence $R[X]$ is commutative. Finally, if $1 \in R$ is unity, the constant polynomial 1 is obviously unity for $R[X]$. ∎

Example 1.4.4 The ring $\mathbb{Z}[X]$ is the ring of polynomials with integer coefficients. The ring $\mathbb{Q}[X]$ is the ring of polynomials with rational coefficients. Both of these rings are familiar from high school algebra!

Example 1.4.5 In the ring $\mathbb{Z}_2[X]$, we have $(X+1)^2 = X^2+1$! Moreover, $(X+1)+(X+1) = 0$.

If R is a ring, and X and Y are two indeterminates, then we can form the ring $(R[X])[Y]$ of polynomials in Y with coefficients in $R[X]$. It is fairly obvious, but tedious to prove carefully, that the ring $(R[X])[Y]$ is canonically isomorphic to the ring $(R[Y])[X]$. That is, every polynomial in Y whose coefficients are polynomials in X can be re-written as a polynomial in X with coefficients in $R[Y]$. We use this isomorphism to identify $(R[X])[Y]$ and $(R[Y])[X]$, and we denote this ring by $R[X,Y]$. Similarly we can define the **ring of polynomials in n indeterminates** $X_1,\ldots,X_n$ $R[X_1,X_2,\ldots,X_n]$. We will not work with polynomials with more than one indeterminate.

It is a nice exercise to show that if D is an integral domain, then so is $D[X]$. In particular, if F is a field, then the polynomial ring $F[X]$ is an integral domain. The field of fractions for $F[X]$ is called the **field of rational functions in** X and is denoted by $F(X)$. Similarly, $F(X_1,\ldots,X_n)$ is the field of fractions for $F[X_1,\ldots,X_n]$. The field $F(X_1,\ldots,X_n)$ plays an important role in modern algebraic geometry.

The next goal of the lecture is to show how the problem of "solving polynomial equations" can be cast in the language of homomorphisms. We begin with the a fundamental theorem about evaluation homomorphisms.

Theorem 1.4.6 *Let F be a subfield of a field E, and let $\alpha \in E$ be arbitrary. Then there is a unique ring homomorphism $\varphi_\alpha : F[X] \to E$ such that $\varphi_\alpha(X) = \alpha$ and $\varphi_\alpha(a) = a$ for all $a \in F$.*

Proof. Given $\alpha \in E$, we define $\varphi_\alpha : F[X] \to E$ by

$$\varphi_\alpha\left(\sum_{i=0}^{\infty} a_i X^i\right) = \sum_{i=0}^{\infty} a_i \alpha^i.$$

The right hand side of this defining equation is well defined since $a_i = 0$ for all but finitely many values of $i \in \mathbb{N}$ and $F \subset E$. The ring homomorphism property of φ_α is an immediate consequence of our definitions of polynomial addition and multiplication. We leave the details of the computation to the reader. Now, if $a \in F$, then a is a constant polynomial and by the definition of φ_α, we have $\varphi_\alpha(a) = a$. Finally, again by definition, we have $\varphi_\alpha(X) = \alpha$.

To show uniqueness, we note that if $\psi : F[X] \to E$ satisfies $\psi(a) = a$ for all $a \in F$ and $\psi(X) = \alpha$, then for all polynomials $f = \sum a_i X^i \in F[X]$, we have

$$\psi(f) = \psi\left(\sum a_i X^i\right) = \sum \psi(a_i)\psi(X)^i = \sum a_i \alpha^i = \varphi_\alpha(f)$$

so that $\psi = \varphi_\alpha$. ■

We remark here that the previous theorem is valid (with the same proof!) if F and E are commutative unital rings. We will primarily be interested in the case where F and E are fields. Although you would never guess from the simplicity of the proof of this theorem, it is difficult to over estimate the importance of this theorem in field theory. It forms the basis for nearly every result in Galois theory - the study of solving polynomial equations. We complete the connection with solving polynomial equations with the following definition.

Definition 1.4.7 (Zero of a polynomial) *Let F be a subfield of a field E and let $\alpha \in E$. If $f \in F[X]$, we say that α is a* **zero of** *f if $\varphi_\alpha(f) = 0$.*

We conclude this lecture with a remark. It may seem to the reader that all we have done in this lecture is to take a simple idea - solving polynomial equations - and make it unnecessarily complicated. In fact, what we have done is to cast a familiar problem in the language of mappings (homomorphisms). We can now use all the machinery we have developed, and will continue to develop, about mappings to solve these problems. We will see that this point of view is very useful indeed.

1.5 Factorization of Polynomials over a Field

One of the main problems in algebra is finding zeros of a given polynomial $f \in F[X]$ where F is a field. Suppose that F is a subfield of a field E and that $f \in F[X]$ factors in $F[X]$, so that $f = gh$ with $g, h \in F[X]$. If $\alpha \in E$, then using the evaluation homomorphism φ_α, we have

$$\varphi_\alpha(f) = \varphi_\alpha(gh) = \varphi_\alpha(g)\varphi_\alpha(h).$$

Since $F[X]$ is an integral domain, we see that α is a zero for f iff. α is a zero for g or α is a zero for h. Therefore the problem of finding a zero for the polynomial f can be reduced to the problem of finding a zero for a factor of f. This is one reason why it is useful to understand factorization in $F[X]$.

The following theorem is the backbone for all of the work we will do in this lecture. The reader is advised to compare this result with the division algorithm for $\mathbb{Z}$, whose importance ws established back in MAT 150A. Before we state the theorem, we remark that if $f \in F[X]$ is a polynomial, then $\deg f$ denotes the degree of f.

Theorem 1.5.1 (Division algorithm for polynomials) *Let F be a field and let*

$$f = a_nX^n + a_{n-1}X^{n-1} + \cdots + a_0$$

and

$$g = b_mX^m + b_{m-1}X^{m-1} + \cdots + b_0$$

be two elements of $F[X]$ with $a_n, b_m \neq 0$ and $m > 0$. Then there are unique elements $q, r \in F[X]$ such that $f = gq + r$ and $\deg r < m = \deg g$.

Proof. First we will show existence. Consider the set

$$S = \{f - gs : s \in F[X]\}.$$

Using the Well Ordering Principal, we can find $r \in S$ with minimal degree so that $f = gq + r$ for some $q \in F[X]$. We claim that $\deg r < m$. If $\deg r = 0$, then we are done since $m > 0$. Otherwise we have

$$r = c_tX^t + c_{t-1}X^{t-1} + \cdots + c_0$$

with $c_i \in F$ and $c_t \neq 0$ and $t \geq 1$. If $\deg r = t \geq m$, then

$$f - qg - (c_t/b_m)X^{t-m}g = r - (c_t/b_m)X^{t-m}g \tag{1.1}$$

and the second expression in (1.1) is of the form

$$r - (c_tX^t + \text{terms of lower degree}),$$

which is a polynomial of degree lower than t, the degree of r. However, the polynomial in equation (1.1) can be written in the form

$$f - g(q + (c_t/b_m)X^{t-m}),$$

so it is an element of S. But this contradicts the minimality of the degree of r. Therefore we must have $t = \deg r < m$. This shows existence.

For uniqueness, suppose that $f = gq + r$ and $f = gq' + r'$ so that subtracting we have

$$g(q - q') = r - r'.$$

Since $\deg(r - r') < \deg g$, this can happen iff. $q - q' = 0$ and hence $r - r' = 0$. ■

We remark here that if F is any field, you can compute the polynomials q and r using polynomial long division just as you did for the ring $\mathbb{R}[X]$ in high school. The following corollaries are also familiar from high school algebra courses. We state them as corollaries to emphasize that they follow immediately from the division algorithm, but they are important results on their own.

Corollary 1.5.2 *If F is a field, and element $\alpha \in F$ is a root of a polynomial $f \in F[X]$ if and only if $X - \alpha$ is a factor of f.*

Proof. If $f = g(X - \alpha)$, then $\varphi_\alpha(f) = \varphi_\alpha(g)(\alpha - \alpha) = 0$ so that α is a root of f.

Conversely, suppose $\alpha \in F$ is a root and use the division algorithm (1.5.1) to write $f = (X - \alpha)q + r$ with $\deg r < 1$. Again, applying the evaluation homomorphism φ_a to $f = (X - \alpha)q + r$ gives $\varphi_\alpha(r) = 0$. But $\deg r = 0$ implies $r \in F$ is a constant polynomial so that $r = 0$ and hence X_α is a factor of f. ■

We leave the proof of the next corollary to the reader.

Corollary 1.5.3 *If F is a field and $f \in F[X]$ has degree n, then f has at most n roots in F.* ■

Example 1.5.4 *Let $f, g \in \mathbb{Z}_5[X]$ be defined by $f = X^4 - 3X^3 + 2X^2 + 4X - 1$ and $g = X^2 - 2X + 3$. Find $q, r \in \mathbb{Z}_5[X]$ such that $f = gq + r$.*

We will encounter polynomials that cannot be factored (in a non-trivial manner) at all.

Definition 1.5.5 (Irreducible polynomial) *If F is a field, a non-constant polynomial $f \in F[X]$ is* **irreducible over** *F if $f = gh$ and $g, h \in F[X]$ implies either g or h is a constant.*

You will show in the homework that an element $f \in F[X]$ is a unit if and only if it is a constant. Therefore a polynomial $f \in F[X]$ is irreducible iff. f is not a unit and $f = gh$ implies that either g or h is a unit. This is taken as a definition of irreducible in any ring for an abstract theory of factorization. We will content ourselves to polynomial rings over fields.

We leave it as an exercise to show that degree 1 polynomials are irreducible.

Note that we have defined the notion of a polynomial being *irreducible over a field F*, not just irreducible. Indeed, a polynomial $f \in F[X]$ may be irreducible over F, but reducible (= not irreducible) over E if $F \leq E$.

Example 1.5.6 The polynomial $X^2 - 2$ is irreducible over $\mathbb{Q}$, but over $\mathbb{R}$ we have $X^2 - 2 = (X + \sqrt{2})(X - \sqrt{2})$ so that it is reducible over $\mathbb{R}$.

The problem of determining if a polynomial $f \in F[X]$ is irreducible can be quite difficult. The following theorem states that for low degree polynomials, the problem is equivalent to finding zeros.

Theorem 1.5.7 *If $f \in F[X]$ and $\deg f \leq 3$, then f is reducible over F if and only if f has a root in F.*

Proof. If f is reducible over F, then $f = gh$ with $g, h \in F[X]$ and $\deg g, \deg h < \deg f$. It follows, without loss of generality, that $\deg g = 1$ so that, up to a factor in F, $g = X - \alpha$ for some $\alpha \in F$. It follows that α is a root of f.

Conversely, if $\alpha \in F$ is a root of f, then $X - \alpha$ is a factor so that f is reducible over F. ■

The theory of factorization of polynomials over a field is very similar to the theory of factorization of ordinary integers in the sense that the analog of the fundamental theorem of arithmetic holds in $F[X]$. There is a wider class of rings, *unique factorization domains*, in which all elements can be factored uniquely into a product of irreducible elements. Historically, such rings were first investigated in attempts to prove the famous “Fermat’s Last Theorem”. We will not undertake the study of factorization in rings in this course. Rather, we will content ourselves with the results of that theory as they apply in polynomial rings over fields. As we proceed, the reader is urged to keep in mind that irreducible polynomials are the analogs of prime integers. This should make the following theorem come as no surprise. Although we could give a proof of the theorem now, it will be awkward with out some results about homomorphism so that we delay the proof until chapter 2 of these notes. Before we state the theorem, let us say that for $f, g \in F[X]$, that f **divides** g if $g = fh$ for some $h \in F[X]$.

Theorem 1.5.8 *Let $p \in F[X]$ be an irreducible polynomial. If p divides the product rs with $r, s \in F[X]$, then p divides r or p divides s.* ■

Corollary 1.5.9 *If $p \in F[X]$ is irreducible and p divides $r_1 \cdots r_n$, $r_i \in F[X]$, then p divides r_i for at least one i.*

Proof. We induct on n, the case $n = 1$ being trivial. Suppose the theorem holds for some $n \geq 1$ and suppose that p divides $r_1 \cdots r_n r_{n+1}$. Let $s = r_1 \cdots r_n$ and $r = r_{n+1}$ so that p divides sr. Therefore p

divides s or p divides r by theorem (1.5.8). If p divides $r = r_{n+1}$, we are done. Otherwise p divides $s = r_1 \cdots r_n$ and hence p divides r_i for some $i \leq n$ by induction. ■

Here is the main theorem we are after - the analog of the fundamental theorem of arithmetic for polynomials over a field. We remind the reader that the units in the ring $F[X]$ are precisely the non-zero constants.

Theorem 1.5.10 *If F is a field, then every non-constant polynomial $f \in F[X]$ can be written as a finite product of irreducible polynomials, and this product is unique up to the order of the irreducible factors and multiplication by a unit in F.*

Proof. We prove the existence of such a factorization by induction on $\deg f$. If $\deg f = 1$, then f is irreducible. Suppose that the theorem is true for all polynomials of degree less than or equal to n for some $n \geq 1$ and suppose that $\deg f = n + 1$. If f is irreducible, we are done. Otherwise we have $f = gh$ with $\deg g, \deg h \leq n$. By induction, both g and h factor into a product of irreducibles and hence f factors into a product of irreducibles. This show existence.

It remains to show the uniqueness statement. Suppose then that

$$p_1 p_2 \cdots p_r = f = q_1 q_2 \cdots q_s$$

are two factorizations of f into irreducible polynomials. Now, p_1 obviously divides f and hence, by the corollary to theorem (1.5.8), p_1 divides q_i for some i. Without loss of generality, we say p_1 divides q_1 so that

$$q_1 = u_1 p_1$$

for some $u_1 \in F[X]$. Since q_1 is irreducible, we must have $u_1 \in F$ a unit. Substituting $u_1 p_1$ for q_1 and canceling gives

$$p_2 \cdots p_r = u_1 q_2 \cdots q_s.$$

Continuing (inducting), we have

$$1 = u_1 u_2 \cdots u_r q_{r+1} \cdots q_s.$$

Looking at the degree of each side, we see that we must have $r = s$ so that the irreducible factors p_i and q_i were the same up to a unit. ■

Example 1.5.11 You can show that the polynomial $f = X^4 + 3X^3 + 2X + 4 \in \mathbb{Z}_5[X]$ factors into irreducibles as $f = (X - 1)^3(X + 1)$. This is the same factorization, up to units, as $f = (X - 1)^2(2X - 2)(3X + 3)$.

1.6 Non-commutative Examples

In this lecture, we discuss some non-commutative rings that will be very important in the rest of our course.

In MAT 150B, we saw that groups "arise naturally" as groups of permutations. In the same sense, our first example can be thought of as a natural example of a ring. Let A be any abelian group, and recall that a group homomorphism $\varphi : A \to A$ is also called an **endomorphism of** A. We let

$$\mathrm{End}(A) = \{\varphi : A \to A : \varphi \text{ is an endomorphism of } A\}$$

denote the set of all endomorphisms of A.

Theorem 1.6.1 *If A is an abelian group, the set* $\mathrm{End}(A)$ *of all endomorphisms of A is a unital ring under the operations $+$ and $\cdot$ defined for all $\varphi, \psi \in \mathrm{End}(A)$ and all $a \in A$ by*

$$(\varphi + \psi)(a) = \varphi(a) + \psi(a)$$

and

$$(\varphi\psi)(a) = \varphi(\psi(a)).$$

Proof. We will leave the verification that $(\mathrm{End}(A), +)$ is an abelian group under addition to the reader. Moreover, it is well know that function composition is an associative operation and obviously the identity map $1_A : A \to A$ is an identity for this operation. It remains to show that the distributive laws hold. If $\varphi, \psi, \theta \in \mathrm{End}(A)$ and $a \in A$, then using the homomorphism property and the definitions of $+$ and $\cdot$, we have

$$\theta(\varphi + \psi)(a) = \theta(\varphi(a) + \psi(a)) = \theta(\varphi(a)) + \theta(\psi(a)) = (\theta\varphi)(a) + (\theta\psi)(a)$$

and therefore $\theta(\varphi + \psi) = \theta\varphi + \theta\psi$. Similarly you can show the right distributive law. ■

Example 1.6.2 If V is a finite dimensional vector space over a field F, then the set

$$\mathrm{End}_F(V) = \{\varphi : V \to V : \varphi \text{ is } F\text{-linear}\}$$

is a subring of $\mathrm{End}(V)$. We leave it as an exercise for the reader to show that choosing a basis $\mathcal{B}$ for V determines a ring isomorphism from $\mathrm{End}_F(V)$ to $\mathrm{Mat}_n(F)$ where $n = \dim_F(V)$.

Example 1.6.3 Let F be a field, and consider the abelian group $(F[X], +)$ which we denote (with some possible confusion) by $F[X]$. We consider three special elements of $\mathrm{End}(F[X])$.
First, let $A : F[X] \to F[X]$ be defined by

$$A : \sum_{i=0}^{\infty} a_i X^i \mapsto \sum_{i=0}^{\infty} a_i X^{i+1}.$$

We will call A the "shift operator". We leave it to the reader to show that A is a group endomorphism $A : F[X] \to F[X]$. Note that A is not a ring homomorphism.
Second, let $B : F[X] \to F[X]$ be defined by

$$B : \sum_{i=0}^{\infty} a_i X^i \mapsto \sum_{i=1}^{\infty} i a_i X^{i-1}.$$

Again, the reader can easily verify that B is a group homomorphism. Note that B is just formal differentiation of polynomials. In your homework, you will show that $AB - BA = 1$ where 1 is the identity endomorphism $1 : F[X] \to F[X]$. Finally, if $\alpha \in F$, multiplication by α defines a group homomorphism $F[X] \to F[X]$ $(f \mapsto \alpha f)$.
The subring W of $\mathrm{End}(F[X])$ generated by A, B and the multiplications by all α in F is called the **Weyl algebra**.

We will study endomorphism rings $\mathrm{End}(A)$ extensively.
Another important example of non-commutative rings are group rings. Here is the definition.

Definition 1.6.4 (Group ring) *Let $G = \{1 = g_1, g_2, \ldots, g_n\}$ be a finite group and R a commutative unital ring. Let $R(G)$ denote the set of all formal sums*

$$\sum_{i=1}^{n} a_i g_i.$$

$R(G)$ is called the **group ring of** *G* **over** *R. If F is a field, $F(G)$ is called the* **group algebra of** *G* **over** *F.*

As usual, we need to justify the terminology "ring" in group ring. We will give the operations in the statement of the following theorem.

Theorem 1.6.5 *If G is a finite group and R is a commutative unital ring, then the operations $+$ and $\cdot$ defined on $R(G)$ defined for all $a, b \in R(G)$ by*

$$\left(\sum_{i=1}^{n} a_i g_i\right) + \left(\sum_{i=1}^{n} b_i g_i\right) = \sum_{i=1}^{n} (a_i + b_i) g_i,$$

and

$$\left(\sum_{j=1}^{n} a_j g_j\right)\left(\sum_{k=1}^{n} b_k g_k\right) = \sum_{i=1}^{n}\left(\sum_{g_j g_k = g_i} a_j b_k\right) g_i$$

make $(R(G), +, \cdot)$ *a unital ring.*

Proof. By now, the participating reader can see immediately that $(R(G), +)$ is an abelian group with identity $\sum 0g_i$. The distributive laws follows at once since the multiplication is defined by formally distributing and collecting like terms. The associativity law is not difficult to prove, but it is tedious to type! We leave it as an exercise. If $g_1 = e$ is the identity element, then the element of $R(G)$ defined by $a_1 = 1$ and $a_i = 0$ for $i > 1$ is unity. ∎

The unity element of $R(G)$ is a member of the family of elements of $R(G)$ defined by $a_i = 1$ and $a_j = 0$ for $j \neq i$. This implies that $(R(G), \cdot)$ contains an isomorphic copy of G. In particular, we see that $R(G)$ will not be commutative if G is non-abelian.

It can be shown that every unitary representation of a finite group G corresponds to a unitary representation of the group algebra $\mathbb{C}(G)$. This is the last bit of theory that we were missing to prove the orthogonality relations last quarter.

Our last example will be our first example of a skew-field. Let $\mathbb{H} = \mathbb{R}^4$ as an additive abelian group and let

$$1 = (1,0,0,0), \quad i = (0,1,0,0), \quad j = (0,0,1,0), \quad \text{and} \quad k = (0,0,0,1)$$

so that we can write every element of $\mathbb{H}$ as a sum

$$a = a_1 + a_2 i + a_3 j + a_4 k$$

with $a_1, a_2, a_3, a_4 \in \mathbb{R}$. We define a multiplication $\cdot$ on $\mathbb{H}$ by formally distributing and using the relations

$$1a = a1 = a \text{ for all } a \in \mathbb{H},$$

$$i^2 = j^2 = k^2 = -1,$$

and

$$ij = k, \quad jk = i, \quad ki = j, \quad ji = -k, \quad kj = -i, \quad \text{and} \quad ik = -j.$$

These relations are easy to remember if you think of the so called "cross product" of the usual unit vectors $i, j, k \in \mathbb{R}^3$. Or, if you prefer, in the sequence

$$i, j, k, i, j, k,$$

the product from left to right of two adjacent terms in the next term to the right, and the product from right to left of two adjacent terms is the negative of the next term to the left.

Once again, the distributive laws will hold in $\mathbb{H}$ by definition, and 1 is clearly unity. The verification of the associativity of $\cdot$ is a tedious chore that we leave to the reader. If we put all of this together, then we see tat $(\mathbb{H}, +, \cdot)$ is a non-commutative unital ring. To show $\mathbb{H}$ is a skew field, we first define, for all $a \in \mathbb{H}$, the **length** of a by

$$|a|^2 = a_1^2 + a_2^2 + a_3^2 + a_4^2$$

and the **conjugate** of a by

$$\overline{a} = a_1 - a_2 i - a_3 j - a_4 k.$$

We leave it as an exercise to show that $|a| = 0$ iff. $a = 0$ and $a\overline{a} = |a|^2$. It follows that if $a \neq 0$, then $a^{-1} = \overline{a}/|a|^2$. Therefore every non-zero element of $\mathbb{H}$ is a unit so that $\mathbb{H}$ is a skew field. We call $\mathbb{H}$ the **quaternions**. The symbol $\mathbb{H}$ is used in reference to Sir William Rowan Hamilton, who discovered this division ring in 1843.

We will look at some interesting properties of $\mathbb{H}$ in the exercises. In particular, you will show that the set

$$U = \{a \in \mathbb{H} : |a| = 1\}$$

is isomorphic to the special unitary group SU_2, and the conjugacy class defined by $\operatorname{tr} A = 0$ (the equator) is the unit 2-sphere in the space $\operatorname{Span}(i, j, k)$ of “purely imaginary” quaternions.

Chapter 2

Ideals and Quotient Rings

2.1 Ring Homomorphisms

Some of our discussion on ring theory up until now has been awkward due to the absence of a careful investigation of ring homomorphisms. As we have said, any investigation of an algebraic object must include an investigation of the functions that preserve the algebraic structures, whatever they may be. Also, we must look at the space of fibers (quotients) of these maps because it also has the same structure! Examples of this scheme of ideas include groups, with group homomorphisms and quotient groups and vector spaces, with linear maps and quotient spaces. We will now investigate these notions for rings. As we will see, half the work is already done from what we know about the abelian groups $(R, +)$.
We had to define the notion of ring homomorphism and isomorphism to discuss polynomial evaluation homomorphisms. We repeat the definition here for easy reference.

Definition 2.1.1 (Ring homomorphism) *If R and S are rings, a map $\varphi : R \to S$ is called a* **ring homomorphism** *if for all $a, b \in R$, we have both*

1. $\varphi(a + b) = \varphi(a) + \varphi(b)$,

2. $\varphi(ab) = \varphi(a)\varphi(b)$.

We have seen several example of ring homomorphisms, so lets get right to some theory. In the proofs, note how we use the fact that $\varphi : (R, +) \to (S, +)$ is a group homomorphism.

Theorem 2.1.2 *Let $\varphi : R \to S$ be a ring homomorphism.*

1. *If $R' \leq R$ is a subring, then $\varphi(R') \leq S$ is a subring.*
2. *If $S' \leq S$ is a subring, then $\varphi^{-1}(S') \leq R$ is a subring.*
3. *If R is unital and $\varphi(1_R) \neq 0_S$, then $\varphi(S)$ is unital and $\varphi(1_R) = 1_S$ is unity.*

Proof. (1) We know that $\varphi(R')$ is a subgroup of S since φ is a group homomorphism. Moreover, the equation $\varphi(a)\varphi(b) = \varphi(ab)$ holds for all $a, b \in R'$, and $ab \in R'$ so that $\varphi(R')$ is closed under multiplication and hence is a subring of S.
(2) Exercise.
(3) Note that for all $a \in R$, then writing $1 = 1_R$, we have

$$\varphi(a) = \varphi(1a) = \varphi(1)\varphi(a) \text{ and } \varphi(a) = \varphi(a1) = \varphi(a)\varphi(1).$$

Therefore, if $\varphi(1) \neq 0$, it is a multiplicative identity and hence $\varphi(1) = 1_S$ by uniqueness of such an identity. ■

Roughly speaking, the theorem states that ring homomorphisms send subrings to subrings, and unity to unity. Here is a long over due definition.

Definition 2.1.3 (Kernel of a ring homomorphism) *If $\varphi : R \to S$ is a ring homomorphism, then the* **kernel of** *φ is the subset*

$$\ker \varphi = \{a \in R : \varphi(a) = 0\}.$$

Note that the kernel of a ring homomorphism φ is just the kernel of φ considered as a homomorphism of abelian groups. In particular, we have the following useful fact.

Proposition 2.1.4 *A ring homomorphism $\varphi : R \to S$ is injective if and only if* $\ker \varphi = \{0\}$. ■

Recall that a ring homomorphism φ is an **isomorphism** if φ is bijective. It should come as no surprise that the class of all rings is partitioned into isomorphism classes by the usual equivalence relation. The same techniques that we used to show two groups are not isomorphic can be applied to rings. In addition, we can look for structural properties in rings such as being unital, commutative, etc....

Example 2.1.5 The ring $\mathrm{Mat}_2(\mathbb{R})$ of 2×2 real matrices is not isomorphic to $\mathbb{C}$. Note that $\mathrm{Mat}_2(\mathbb{R})$ has zero divisors and the field $\mathbb{C}$ does not.

Example 2.1.6 The field $\mathbb{R}$ is not isomorphic to the field $\mathbb{C}$. Any isomorphism $\varphi : \mathbb{C} \to \mathbb{R}$ maps 1 to 1 and hence -1 to -1. But then $\varphi(i) \in \mathbb{R}$ satisfies

$$\varphi(i)^2 = \varphi(i^2) = \varphi(-1) = -1$$

which is absurd.

Here is the main theorem of the lecture.

Theorem 2.1.7 *Let $\varphi : R \to S$ be a ring homomorphism and let $I = \ker \varphi$. The operation defined on the quotient group R/I defined by*

$$(a + I)(b + I) = ab + I$$

is a well defined binary operation making R/I into a ring. If R is commutative, then R/I is commutative. If R is unital, then R/I is unital provided $I \neq R$.

Proof. The whole issue here is to show that the operation is well defined. That is, the associative and distributive laws will follow immediately from those laws in R. Suppose then that $a' - a \in I$ and $b' - b \in I$ so that $a' = a + i_1$ and $b' = b + i_2$ for some $i_1, i_2 \in I$. Then we have

$$a'b' = (a + i_1)(b + i_2) = ab + ai_2 + i_1b + i_1i_2.$$

Now, $\varphi(ai_2) = \varphi(a)\varphi(i_2) = \varphi(a)0 = 0$ and similarly $\varphi(i_1b) = \varphi(i_1i_2) = 0$ so that $ai_2, i_1b, i_1i_2 \in I$. It follows that $a'b' - ab \in I$ and hence $ab + I = a'b' + I$. The reader can now easily verify the remaining ring axioms and that R/I is commutative if R is commutative. Finally, $1 + I$ is clearly unity if $1 \in R$ is unity. ■

The ring R/I is called the **quotient ring**.

2.2 Ideals

The purpose of this lecture is to abstract the properties of the kernel of a ring homomorphism and investigate quotient rings further. Recall from the proof of the last theorem in the previous lecture that if $\varphi : R \to S$ is a ring homomorphism, $a \in \ker \varphi$ and $b \in R$, then $ab \in \ker \varphi$ and $ba \in \ker \varphi$. Moreover, our work in group theory shows that $\ker \varphi$ is a subgroup of the additive group $(R, +)$. We abstract these ideas in the following definition.

Definition 2.2.1 (Ideal) *If R is a ring, a subset I of R is an* **ideal in** *R if*

1. *$(I,+) \leq (R,+)$ is an additive subgroup of $(R,+)$,*
2. *For all $a \in I$ and all $b \in R$, $ab \in I$ and $ba \in I$.*

If we let $bI = \{ba : a \in I\}$ and similarly define Ib, then an ideal in R is a subgroup I that satisfies $bI \subset I$ and $Ib \subset I$ for all $b \in R$.

Example 2.2.2 If $\varphi : R \to S$ is a ring homomorphism, then $I = \ker \varphi$ is an ideal in R.

Example 2.2.3 In the ring of integers, every subgroup $\langle n \rangle$ is an ideal. This follows since if $a \in \langle n \rangle$, then $a = nk$ for some $k \in \mathbb{Z}$ and hence for all $b \in \mathbb{Z}$, we have $ab = nbk = ba$ so that $ab, ba \in \langle n \rangle$. Conversely, if I is an ideal in $\mathbb{Z}$, then since I is a subgroup of the cyclic group $(\mathbb{Z},+)$, we must have $I = \langle n \rangle$ for some $n \in \mathbb{Z}$. To emphasize we are dealing with rings, we will denote the ideal $\langle n \rangle$ by (n). We have shown in this example that every ideal in the ring $\mathbb{Z}$ has the form (n) for some n. Such rings are called principal ideal rings, and we will study them in some detail.

In ring theory, ideals play the same role as normal subgroups in the theory of groups as the following theorem shows.

Theorem 2.2.4 *Let R be a ring and let I be an ideal in R, then the operation defined on the quotient group R/I defined by*

$$(a+I)(b+I) = ab + I$$

is a well defined binary operation making R/I into a ring. If R is commutative, then R/I is commutative. If R is unital, then R/I is unital provided $I \neq R$.

Proof. We can copy the proof for the case $I = \ker \varphi$, a ring homomorphism without any changes at all! That is, once again the whole issue here is to show that the operation is well defined. Suppose then that $a' - a \in I$ and $b' - b \in I$ so that $a' = a + i_1$ and $b' = b + i_2$ for some $i_1, i_2 \in I$. Then we have

$$a'b' = (a+i_1)(b+i_2) = ab + ai_2 + i_1b + i_1i_2.$$

Now, $ai_2, i_1b, i_1i_2 \in I$ since I is an ideal. It follows that $a'b' - ab \in I$ and hence $ab + I = a'b' + I$. The reader can now easily verify the remaining ring axioms and that R/I is commutative if R is commutative. Finally, $1 + I$ is clearly unity if $1 \in R$ is unity. ■

Continuing to parallel the theory of quotient groups, we remark that at this point, we know that kernels of ring homomorphisms are ideals, and we can form quotient rings by any ideal. The next step is to show that every ideal is the kernel of a ring homomorphism by showing that the quotient map $R \to R/I$ is in fact a ring homomorphism. We state the precise result as a theorem.

Theorem 2.2.5 *If I is an ideal in R, then the quotient map $\eta : R \to R/I$ defined by*

$$\eta : a \mapsto a + I$$

is a ring homomorphism with $\ker \eta = I$.

Proof. Given what we know from group theory, it suffices to show that η preserves the multiplication operation. So let $a, b \in R$ and compute

$$\eta(ab) = ab + I = (a + I)(b + I) = \eta(a)\eta(b).$$

■

We will look at the isomorphism theorems for quotient rings in the next lecture. Here are some useful facts about ideals.

Proposition 2.2.6 *If I and J are ideals in a ring R, then*

1. *$I \cap J$ is an ideal.*
2. *The set $I + J = \{a + b : a, \in I, b \in J\}$ is an ideal.*
3. *If R is unital, $I = R$ if and only if I contains a unit.*

Proof. (1) We know $I \cap J$ is a subgroup from MAT 150A. If $a \in I \cap J$ and $b \in R$, then since both I and J are ideals, we have $ab \in I$ and $ab \in J$. Similarly, $ba \in I$ and $ba \in J$ so that $ab, ba \in I \cap J$ and hence $I \cap J$ is an ideal.

(2) Since R is abelian as a group under $+$, we know $I + J = J + I$ and hence $I + J$ is a subgroup from our work in MAT 150A. (See the section on product of groups in the 150A notes.) Now, since both I and J are ideals, if $a + b \in I + J$ and $c \in R$, then $ca \in I$ and $cb \in J$ so that $c(a + b) = ca + cb \in I + J$. Similarly $(a + b)c \in I + J$ so that $I + J$ is an ideal.

(3) Suppose that R is unital. If $I = R$, the $1 \in I$ and 1 is a unit so I contains a unit. Conversely, suppose that $u \in I$ and u is a unit. Then there exists an element $v \in R$ with $vu = 1$ and hence $1 = vu \in I$ since I is an ideal. But then for all $a \in r$, $a = a1 \in I$ and hence $I = R$. ■

We note that for any ring R, the subgroups R and 0 are always ideals. Part (3) of this proposition gives us the following corollary.

Corollary 2.2.7 *If F is a field, then F and 0 are the only ideals in F.* ■

Definition 2.2.8 (Simple ring) *A non-trivial ring is called* **simple** *if R and 0 are the only ideals in R.*

The corollary states that all fields are simple. There are simple rings that are not fields, as the following example shows.

Example 2.2.9 Let F be a field and let $R = \mathrm{Mat}_n(F)$ be the ring of $n \times n$ matrices over F for some $n > 1$. We claim that R is simple. To show this, suppose that $0 \neq I$ is an ideal in R. Since $I \neq 0$, there exists a non-zero element $A \in I$. Since $A \neq 0$, we must have at least one entry, say $a_{i_0 j_0} \neq 0$. Using the matrix units e_{ij}, we note that for any $k \in \{1, 2, \ldots, n\}$, we have

$$e_{ki_0} A e_{j_0 k} = a_{i_0 j_0} e_{kk}$$

and hence $a_{i_0 j_0} e_{kk} \in I$ since $A \in I$ and I is an ideal. It follows that

$$a_{i_0 j_0} I_n = \sum_{k=1}^{n} a_{i_0 j_0} e_{kk} \in I.$$

But $a_{i_0 j_0} \neq 0$ so that $a_{i_0 j_0} I_n$ is invertible (a unit). It follows that $I = R$ and hence R is simple.

Definition 2.2.10 (Principal ideal) *Let R be a commutative unital ring and let $a \in R$. The reader can show that the set*

$$(a) = \{ra : r \in R\}$$

is an ideal of R. We call (a) the **principal ideal generated by** *a. If R has the property that every ideal in R is principal, then R is called a* **principal ideal ring (PIR)**.

Example 2.2.11 Since every ideal in $\mathbb{Z}$ is of the form (n), $\mathbb{Z}$ is a PIR.

Example 2.2.12 If R is a simple commutative unital ring, then R is a PIR. This follows since R only has two ideal, 0 and R, and $0 = (0)$ and $R = (1)$ so that both are principal.

We end this lecture with an important result about polynomial rings over a field.

Theorem 2.2.13 *If F is a field, the polynomial ring $F[X]$ is a PIR.*

Proof. We know $F[X]$ is a commutative unital ring so that it suffices to show that every ideal I in $F[X]$ is principal. The zero ideal is always principal, so let $0 \neq I$ be an ideal in $F[X]$ and choose a non-zero polynomial $f \in I$ with minimum degree. Without loss of generality, we may assume the leading coefficient of f is 1, otherwise we multiply f by the inverse of its leading coefficient getting another polynomial in I with the same degree and leading coefficient 1. We claim $(f) = I$. Clearly $(f) \subset I$ since $f \in I$ and I is an ideal. If $\deg f = 0$, then $f \in F[X]$ is a unit and hence $I = F[X] = (1)$ is principal. Otherwise, $\deg f > 1$ and given any $g \in I$, we can use the division algorithm to write

$$g = fq + r$$

where $q, r \in F[X]$ and either $r = 0$ or $\deg r < \deg f$. Now, $f \in I$ so that $fq \in I$ since I is an ideal. Therefore $r = g - gq \in I$ and hence $r = 0$ by the minimality of the degree of f. It follows that $g = fq \in (f)$ and hence $I \subset (f)$. We have already shown that $(f) \subset I$ and hence $(f) = I$ and I is principal. ■

2.3 Quotient Rings

In this lecture, we will state and prove the ring theoretic analog of the first isomorphism theorem. We will also give some examples to show how this theorem is used analyze the structure of quotient rings. We begin with a general lemma about lifting homomorphisms.

Lemma 2.3.1 (Lifting lemma) *Suppose that R, S and R' are rings, $\varphi : R \to S$ and $\pi : R \to R'$ are ring homomorphisms and π is surjective. Then there exists a unique ring homomorphism $\psi : R' \to S$ such that $\psi \circ \pi = \varphi$ if and only if $\ker \pi \subseteq \ker \varphi$. Moreover, ψ is surjective if and only if φ is surjective and ψ is injective if and only if $\ker \pi = \ker \varphi$.*

Proof. By hypothesis, we are given the diagram

$$\begin{tikzcd}
R \arrow[r, "\varphi"] \arrow[dr, "\pi"'] & S \\
& R'
\end{tikzcd}$$

with $\pi : R \to R'$ surjective and $\ker \pi \subseteq \ker \varphi$. We define $\psi : R' \to S$ as follows. Given $a' \in R'$, choose $a \in R$ with $\pi(a) = a'$ (π is onto) and define $\psi(a') = \varphi(a)$. We must show that ψ is well defined, independent of the choice of the preimage of a'. Suppose then that $b \in R$ satisfies $\pi(b) = a'$. Then $a - b \in \ker \pi$ and hence $a - b \in \ker \varphi$ so that $\varphi(a) = \varphi(b)$ and hence ψ is well defined.

Clearly ψ is a ring homomorphism since it is defined by φ and φ and π and both of these maps are ring homomorphisms. Specifically, if $a', b' \in R'$ and $a, b \in R$ satisfy $\pi(a) = a'$ and $\pi(b) = b'$, then $\pi(a + b) = a' + b'$ and $\pi(ab) = a'b'$ and hence we have, by definition,

$$\psi(a' + b') = \varphi(a + b) = \varphi(a) + \varphi(b) = \psi(a') + \psi(b')$$

and

$$\psi(a'b') = \varphi(ab) = \varphi(a)\varphi(b) = \psi(a')\psi(b').$$

Now, if $\psi' : R' \to S$ also satisfies $\psi' \circ \pi = \varphi$, then for all $a' \in R'$, we have

$$\psi'(a') = \psi'(\pi(a)) = \varphi(a) = \psi(\pi(a)) = \psi(a')$$

so that $\psi = \psi'$ and hence ψ is unique.

Now, if φ is surjective and $c \in S$, then choose $a \in R$ with $\varphi(a) = c$ so that $\psi(\pi(a)) = \varphi(a) = c$ and hence ψ is surjective.

If $\ker \pi = \ker \varphi$ and $\psi(a') = \psi(b')$, then if $\pi(a) = a'$ and $\pi(b) = b'$, then $\varphi(a - b) = \varphi(a) - \varphi(b) = \psi(a') - \psi(b') = 0$ so that $a - b \in \ker \varphi$. Therefore $a - b \in \ker \pi$ so that $a' = b'$ and hence ψ is injective. ■

Corollary 2.3.2 (First isomorphism theorem) *If $\varphi : R \to S$ is a ring homomorphism and $I = \ker \varphi$, then R/I is isomorphic to $\operatorname{im} \varphi$.*

Proof. We apply the lifting lemma to the diagram

$$\begin{tikzcd}
R \arrow[r, "\varphi"] \arrow[dr, "\pi"'] & \operatorname{im} \varphi \arrow[r, hook] & S \\
 & R/I &
\end{tikzcd}$$

where $\pi : R \to R/I$ is the quotient map. That is, φ is onto its image and $\ker \varphi = I = \ker \pi$ so that the lifting lemma (2.3.1) gives an isomorphism $\psi : R/I \to \operatorname{im} \varphi$. ■

Corollary 2.3.3 (Second isomorphism theorem) *Let R be a ring and let I and J be two ideals in R. Then $I \cap J$ is an ideal in J, I is an ideal in $I + J$ and*

$$(I + J)/I \cong J/(I \cap J).$$

Proof. Clearly $I \cap J$ is an ideal in J and I is an ideal in $I + J$. Let φ be the composition

$$\begin{tikzcd} J \arrow[r, hook] & I + J \arrow[r, "\eta"] & (I + J)/I. \end{tikzcd}$$

Clearly $\ker \varphi = I \cap J$ so that we may use the first isomorphism theorem (2.3.2) to conclude $J/(I \cap J) \cong \operatorname{im} \varphi$. Now, if $(a + b) + I \in (I + J)/I$, then since $a \in I$, we have

$$(a + b) + I = b + I$$

and hence $\varphi(b) = b + I = (a + b) + I$ so that φ is surjective. This completes the proof. ■

Corollary 2.3.4 (Third isomorphism theorem) *Let R be a ring and let I and J be two ideals in R such that $I \leq J \leq R$. Then J/I is an ideal in R/I and*

$$(R/I)/(J/I) \cong R/J.$$

Proof. First, consider the diagram

$$\begin{tikzcd} R \arrow[r, "\eta_I"] \arrow[dr, "\eta_J"'] & R/I \\ & R/J \end{tikzcd}$$

Since $I = \ker \eta_I$ and $J = \ker \eta_J$, the lifting lemma implies that we have a surjective map $\varphi : R/I \to R/J$ commuting with the quotient maps η_I and η_J. Moreover, if $a + I \in R/I$, then $\varphi(a + I) = a + J$ by construction. Therefore we know that

$$\ker \varphi = \{a + I \in R/I : a \in J\} = \eta_I(J) = J/I.$$

It follows that J/I is an ideal in R/I (it is a kernel), and we have the diagram

$$\begin{tikzcd} R \arrow[r, "\eta_I"] \arrow[dr, "\eta_J"'] & R/I \arrow[r, "\eta"] \arrow[d, "\varphi"] & (R/I)/(J/I) \\ & R/J & \end{tikzcd}$$

where $\eta : R/I \to (R/I)/(J/I)$ is the quotient map. Since φ is surjective and $\ker \varphi = \ker \eta$, the first isomorphism theorem (2.3.2) implies that $(R/I)/(J/I) \cong R/J$. ■

We conclude the lecture with some examples of identifying quotient rings.

Example 2.3.5 If F is a field, then $F[X]/(X)$ is isomorphic to F. To see this, we note that the evaluation homomorphism $\varphi_0 : F[X] \to F$ is trivially surjective since $\varphi_0(a) = a$ for all $a \in F$. Moreover, we see that $f \in \ker \varphi_0$ iff. $f(0) = 0$ iff. $f \in (X)$ by Corollary 1.5.2 so that $\ker \varphi_0 = (X)$. The result now follows directly from the first isomorphism theorem.

Example 2.3.6 The ring $\mathbb{R}[X]/(X^2+1)$ is isomorphic to $\mathbb{C}$. To see this, we consider the evaluation homomorphism $\varphi_i : \mathbb{R}[X] \to \mathbb{C}$. Again φ_i is clearly surjective since for all $a + ib \in \mathbb{C}$ $(a, b \in \mathbb{R})$, we have $\varphi_i(a + bX) = a + bi$. Moreover, $X^2 + 1 \in \ker \varphi_i$ since $i^2 + 1 = 0$, and hence $(X^2 + 1) \subseteq \ker \varphi_i$. Now, since $\mathbb{R}[X]$ is a PIR, we have $\ker \varphi_i = (f)$ for some $f \in \mathbb{R}[X]$. Using the division algorithm in $\mathbb{R}[X]$, we can write

$$f = (X^2 + 1)q + r$$

where $q, r \in \mathbb{R}[X]$ and either $r = 0$ or $\deg r < 2$. Applying φ_i to this equality shows that $\varphi_i(r) = 0$. If $\deg r = 1$, this implies that $i \in \mathbb{R}$, a contradiction. Therefore $r \in \mathbb{R}$ must be a constant and hence $\varphi_i(r) = 0$ implies $r = 0$. This shows that $f \in (X^2 + 1)$ and hence $\ker \varphi_i = (f) \subseteq (X^2 + 1)$.

We have shown that $\ker \varphi_i = (X^2 + 1)$ and hence the first isomorphism theorem implies that $\mathbb{R}[X]/(X^2 + 1) \cong \mathbb{C}$.

We will see that both of these examples are quite general.

2.4 Prime and Maximal Ideals

The purpose of this lecture is to look more closely at the lattice of ideals in a ring R. Historically, many of these notions were first investigated abstractly in the study of factorization in rings. The terminology involved here reflects these origins.

We recall here that every ring R has at least two ideals, the **improper ideal** R and the **trivial ideal** $\{0\}$. A ring is simple if these are the only two ideals. The quotients by these ideals are not interesting at all: R/R has only one element, and $R/(0)$ is obviously isomorphic to R. We will therefore turn our attention to ideals I in R such that $\{0\} \neq I \neq R$. Such an ideal is called a **proper nontrivial ideal**.

Let us investigate the conditions under which the quotient ring R/I is a integral domain or a field. The answer is not as straight forward as you might expect.

Definition 2.4.1 (Maximal ideal) *If R is a ring, and ideal M is* **maximal** *if $M \neq R$, and if N is any ideal of R satisfying $M \subseteq N$, then either $N = M$ or $N = R$.*

Another way to state this (very important) definition is as follows. An ideal M is maximal if M is a proper ideal in R, and R is the only ideal of R that properly contains M. One useful fact about maximal ideals in commutative unital rings is given in the following theorem.

Theorem 2.4.2 *If R is a commutative unital ring, and M is an ideal in R, then M is maximal iff. R/M is a field.*

Proof. Suppose that M is maximal. We know R/M is a commutative unital ring by theorem 2.2.4. It therefore suffices to show that every non-zero element of R/M is a unit. To this end, let $a + M \in R/M$ with $a + M \neq M$ (i.e. $a + M$ is non-zero in R/M). Now, let $N = (a) + M$ so that N is an ideal in R by proposition 2.2.6(2). Moreover, we clearly have $M \subseteq N$ and $a \in N$. But $a + M \neq M$ so that $a \notin M$. It follows that N properly contains M and hence $N = R$ by maximality. Since $1 \in R$, we can write $1 = ra + m$ for some $r \in R$ and $m \in M$ and hence we have

$$1 + M = (ra + m) + M = ra + M = (r + M)(a + M).$$

This shows that $a + M$ is a unit in R/M and hence R/M is a field.

Conversely, suppose that R/M is a field. Then R/M is unital and hence $M \neq R$. If N is an ideal of R such that $M \subset N \subset R$ (proper inclusions), then the third isomorphism theorem implies that N/M is an ideal of R/M. But R/M is a field so that the only two ideals of R/M are R/M (improper) and M/M (trivial). In the first case, we have $N = R$, and in the second we have $N = M$. Therefore M is maximal. ■

This theorem is false if you omit commutativity (i.e. you do not even get a division ring). For example, we have shown that (0) is a maximal ideal in the matrix ring $R = \mathrm{Mat}_n(F)$, but $R/(0) \cong R$ has zero divisors. We now turn to the question: when is R/I an integral domain. The zero divisor condition for cosets reads: $(a + I)(b + I) = I$ implies that $a + I = I$ or $b + I = I$. This motivates the following definition.

Definition 2.4.3 (Prime ideal) *An ideal $N \neq R$ in a commutative ring R is* **prime** *if for all $a, b \in R$, $ab \in N$ implies that either $a \in N$ or $b \in N$.*

Theorem 2.4.4 *If R is a commutative unital ring, and $N \neq R$ is a proper ideal in R, then N is prime iff. R/N is an integral domain.*

Proof. Again, since $N \neq R$, we know R/N is a commutative unital ring by theorem 2.2.4. It therefore remains to show that R/N has no divisors of zero. Suppose then that $(a+N)(b+N) = N$

in R/N. Then $ab + N = N$ so that $ab \in N$. But N is a prime ideal so that either $a \in N$ or $b \in N$. It follows that $a + N = N$ or $b + N = N$ so that R/N has no divisors of zero.

Conversely, if R/N is an integral domain, then it is a unital ring so that $N \neq R$. Moreover, if $a, b \in R$ and $ab \in N$, then $(a + N)(b + N) = ab + N = N$ so that $a + N = N$ or $b + N = N$. It follows that $a \in N$ or $b \in N$ and hence N is a prime ideal. ■

Corollary 2.4.5 *If R is a commutative unital ring, then every maximal ideal is prime.*

Proof. If M is maximal, then R/M is a field, hence R/M is an integral domain, and hence M is prime. ■

Definition 2.4.6 (Divides, prime element, irreducible element) *Let R be a commutative unital ring. If $a, b \in R$, we say a* **divides** *b if $b = ar$ for some $r \in R$. We write $a|b$ if a divides b. An element $p \in R$ is called* **prime** *if p is not a unit and $p|ab$ implies $p|a$ or $p|b$. An element r is* **irreducible** *if r is not a unit and $r = ab$ implies that either a or b is a unit.*

We now specialize to the class of PIRs that have no divisors of zero. Such rings are called **principle ideal domains (PID)**. We remark that if F is a field, then $F[X]$ is a PID.

Theorem 2.4.7 *If R is a PID and $0 \neq p \in R$, then (p) is a maximal ideal if and only if $p \in R$ is irreducible.*

Proof. Suppose that (p) is maximal. Then $(p) \neq R$ and hence p is not a unit by proposition 2.2.6(3). If $p = ab$, then $ab \in (p)$. Now, (p) is a maximal ideal, and hence (p) is a prime ideal by corollary (2.4.5) so that we have $a \in (p)$ or $b \in (p)$. If $a \in (p)$, then $a = pr$ for some $r \in R$ and hence

$$p = ab = prb.$$

This implies that $1 = rb$ since R is a integral domain, and hence b is a unit. Similarly, if $b \in (p)$, then a is a unit and hence p is irreducible.

Conversely, if p is irreducible, then p is not a unit and hence $(p) \neq R$. To show that (p) is maximal, suppose that $(p) \subset (a) \subseteq (1)$ where the first inclusion is proper. Then $p \in (a)$ so that $p = ab$ for some $b \in R$. But p is irreducible so that either a or b is a unit. If b is a unit, then $bu = 1$ for some $u \in R$ and hence $pu = abu = a$. This implies that $a \in (p)$ and hence $(a) \subseteq (p)$ contrary to assumption. Therefore we must have a a unit, and hence $(a) = R$ by proposition 2.2.6(3). This shows that (p) is a maximal ideal. ■

Theorem 2.4.8 *If R is a PID and $0 \neq p \in R$, then (p) is a prime ideal if and only if $p \in R$ is a prime element.*

Proof. If $0 \neq (p)$ is a prime ideal, then $(p) \neq R$ so that p is not a unit. If $a, b \in R$ and $p|ab$, then $ab \in (p)$ so that either $a \in (p)$ or $b \in (p)$ and hence $p|a$ or $p|b$. This shows that p is a prime element. Conversely, if p is a prime element, then p is not a unit and hence $(p) \neq R$. If $a, b \in R$ and $ab \in (p)$, then $p|ab$ and hence $p|a$ or $p|b$. It follows that $a \in (p)$ or $b \in (p)$ and hence (p) is a prime ideal. ■

Corollary 2.4.9 *If R is a PID, then every irreducible element is prime.*

Proof. If R is a PID and p is irreducible, then theorem (2.4.7) implies that (p) is a maximal ideal, and hence (p) is a prime ideal by corollary (2.4.5). It then follows from theorem (2.4.8) that p is a prime element. ■

This corollary patches a hole in our study of polynomial rings. Namely, recall that we gave no proof of the following theorem (theorem (1.5.8)).

Theorem 2.4.10 *Let $p \in F[X]$ be an irreducible polynomial. If p divides the product rs with $r, s \in F[X]$, then p divides r or p divides s.* ■

Proof. Note that if F is a field, then $F[X]$ is a PIR by theorem 2.2.13, and $F[X]$ is an integral domain by problem #4 of homework #2. Therefore $F[X]$ is a PID and we can apply the previous corollary to the irreducible element $f \in F[X]$ to conclude that f is a prime element. ■

We conclude this lecture with an interpretation of theorem (2.4.7) for polynomial rings.

Theorem 2.4.11 *If F is a field and $f \in F[X]$, then f is irreducible over F if and only if $F[X]/(f)$ is a field.*

Proof. If f is irreducible, then (f) is maximal by theorem (2.4.7) and hence $F[X]/(f)$ is a field by theorem (2.4.2).

Conversely, if $F[X]/(f)$ is a field, then (f) is a maximal ideal by (2.4.2) and hence f is irreducible, again by theorem (2.4.7). ■

2.5 Manufacturing Roots of Polynomials

We have now developed enough machinery tackle the problem of finding roots for polynomials. As we will see, if we are willing to "enlarge" the field of coefficients, we can always find a root for any non-constant polynomial. As you read this section, notice how much use we make of the evaluation homomorphism. We begin by establishing the terminology commonly used in this business.

Definition 2.5.1 (Field extension) *If E and F are fields with $F \subseteq E$, we say that E* **is an extension of** *F. A sequence of extensions*

$$F \subseteq E_1 \subseteq \cdots \subseteq E_n$$

is called a **tower***.*

We leave it as an exercise for the reader to show that if E is an extension of F, then E is a vector space over F. The dimension $\dim_F(E)$ is called the **degree of the extension** and is written $[E : F]$. If $[E : F] < \infty$, then we say that E **is a finite extension of** F. We are ready for our first main theorem. Before we give it, we caution the reader that we are going to do some things that, for the purest, will be entirely beyond the pale. Specifically, if $\varphi : F \to E$ is an injective homomorphism from one field into another, then we will often choose to identify F with its image $\varphi(F) \subseteq E$, and hence we will consider E to be an extension of F. There are some set theoretic dangers with such shenanigans, but the economics gained are worth the risks. Of course, if the danger is acute, we will avoid such identifications.

Theorem 2.5.2 *If F is a field and $f \in F[X]$ is a non-constant polynomial, then there exists an extension field E of F and an element $\alpha \in E$ such that α is a root of f. That is, $f(\alpha) = 0$.*

Proof. Given a non-constant $f \in F[X]$, theorem (1.5.10) implies that f is a product of irreducible polynomials. If p is one such irreducible factor, and $E \supseteq F$ is an extension of F with a root α for p, then clearly α is also a root of f. Therefore we may assume that $f \in F[X]$ is irreducible over F. Now, theorem (2.4.2) implies that (f) is a maximal ideal so that $E = F[X]/(f)$ is a field by theorem (2.4.11). Let $\varphi : F \to E$ be the restriction of the quotient map $F[X] \to F[X]/(f)$ to the constant polynomials. Then φ is a ring homomorphism since the restriction of a homomorphism to a subring is a homomorphism. Now, if $a \in F$, then $\varphi(a) = a + (f) = (f)$ iff. $a \in (f)$ iff. $a = fg$ for some $g \in F[X]$. But f is irreducible so that $\deg f \geq 1$ and hence $\deg(fg) \geq 1$ unless $g = 0$. But $\deg a = 0$ unless $a = 0$ so that we must have $a = 0$ and hence φ is injective. If we identify F with its image $\varphi(F) \leq E$, then E is an extension of F.

Now let $\alpha \in E$ be defined by $\alpha = X + (f)$. If $f = a_0 + a_1X + \cdots + a_nX^n$, then we compute using the evaluation homomorphism $\varphi_\alpha : F[X] \to E$:

$$\begin{aligned}
\varphi_\alpha(f) &= \varphi_\alpha(a_0 + a_1X + \cdots + a_nX^n) \\
&= a_0 + a_1\varphi_\alpha(X) + \cdots + a_n\varphi_\alpha(X)^n \\
&= a_0 + a_1\alpha + \cdots + a_n\alpha^n \\
&= a_0 + a_1(X + (f)) + \cdots + a_n(X + (f))^n \\
&= a_0 + (a_1X + (f)) + \cdots + (a_nX^n + (f)) \\
&= (a_0 + a_1X + \cdots + a_nX^n) + (f) \\
&= f + (f) \\
&= (f).
\end{aligned}$$

Therefore $f \in \ker\varphi_\alpha$ and hence $\alpha \in E$ is a root of f as desired. ■

Example 2.5.3 Let $f = X^2 + 1 \in \mathbb{R}[X]$. Then we have seen that $\mathbb{R}[X]/(f) \cong \mathbb{C}$ and the isomorphism is given explicitly by $g + (f) \mapsto g(i)$. In particular, we have $X + (f) \mapsto i$ so that the construction in the previous theorem, when applied to $\mathbb{R}[X]$ and $X^2 + 1$ produces the extension $\mathbb{C}$ of $\mathbb{R}$ and the root $i \in \mathbb{C}$.

Definition 2.5.4 (Algebraic and transcendental elements) *If E is an extension field of a field F, and element $\alpha \in E$ is called* **algebraic over** F *if* $\ker\varphi_\alpha \neq 0$ *where* $\varphi_\alpha : F[X] \to E$ *is the evaluation homomorphism. If* $\ker\varphi_\alpha = 0$*, then we say α is* **transcendental over** F*. If $E = \mathbb{C}$ and $F = \mathbb{Q}$, then an algebraic element $\alpha \in \mathbb{C}$ over $\mathbb{Q}$ is called an* **algebraic number***. Similarly we have* **transcendental numbers***.*

In other words, $\alpha \in E$ is algebraic over F if α is the root of some non-zero polynomial $f \in F[X]$. Since $F[X]$ is a PID (theorem 2.2.13), we know that $\ker\varphi_\alpha = (f)$ for some $f \in F[X]$. Moreover, the proof of theorem (2.2.13) shows that we may assume that f is the smallest degree monic polynomial in $F[X]$ for which α is a root. If $\alpha \in E$ is transcendental over F, then $\varphi_\alpha : F[X] \to E$ is injective and hence an isomorphism onto its image in E. We summarize all of these ideas in the following theorem.

Theorem 2.5.5 *Let E be an extension field of a field F and let $\alpha \in E$. Then α is transcendental over F if and only if the evaluation homomorphism $\varphi_\alpha : F[X] \to E$ is an isomorphism onto its* *image. If α is algebraic over F, then there exists a unique, monic irreducible polynomial $p \in F[X]$ such that* $\ker \varphi_\alpha = (p)$.

Proof. First, as we have already remarked, $\alpha \in E$ is transcendental over F iff. $\ker \varphi_\alpha = 0$ iff. φ_α is injective iff. φ_α is an isomorphism onto its image in E. This proves the first statement of the theorem.

Now, if α is algebraic, then we have also remarked that since $F[X]$ is a PID, we have $0 \neq \ker \varphi_\alpha = (p)$ for some monic $p \in F[X]$ of minimal degree. Moreover, $\ker \varphi_\alpha \neq F[X]$ since $F \not\subset \ker \varphi_\alpha$ so that $\deg p \geq 1$. It remains to show that p is irreducible. If $p = rs$ for some $r, s \in F[X]$, then $p(\alpha) = 0$ implies that $r(\alpha)s(\alpha) = 0$ so that either $r(\alpha) = 0$ or $s(\alpha) = 0$ since E is an integral domain (it is a field). By the minimality of the degree of p, one of r or s must be a constant and hence p is irreducible and the proof is complete. ■

Definition 2.5.6 (Irreducible polynomial for α) *If E is an extension of F and $\alpha \in E$ is algebraic over F, the* **irreducible polynomial for α over** *F is the unique monic polynomial p of minimal degree satisfying $p(\alpha) = 0$. We denote the irreducible polynomial for α over F by* $\operatorname{irr}(\alpha, F)$. *The degree of* $\operatorname{irr}(\alpha, F)$ *is the* **degree of α over** *F and is denoted by* $\deg(\alpha, F)$.

Example 2.5.7 If $i = \sqrt{-1} \in \mathbb{C}$, then clearly $\operatorname{irr}(i, \mathbb{R}) = X^2 + 1$ since i is not a root of an linear polynomial over $\mathbb{R}$ and $X^2 + 1$ is monic. Therefore $\deg(i, \mathbb{R}) = 2$.

Let us try to summarize what we have done so far. If E is an extension field of a field F and $\alpha \in E$, we consider the evaluation homomorphism $\varphi_\alpha : F[X] \to E$. We have two cases:

Algebraic case. If $\alpha \in E$ is algebraic over F, then the kernel of φ_α is the maximal (2.4.7) ideal $(\operatorname{irr}(\alpha, F))$ in $F[X]$ and hence $F[X]/(\operatorname{irr}(\alpha, F))$ is a field (2.4.2). The first isomorphism theorem implies that this field is isomorphic to the image $\varphi_\alpha(F[X]) \leq E$, which we denote by $F[\alpha]$. We leave it as an exercise to show that the image $F[\alpha] = \varphi_\alpha(F[X])$ is the smallest subfield of E containing F and α. We denote this field by $F(\alpha)$ and call it F **adjoin** α. Therefore if $\alpha \in E$ is algebraic over F, $F[\alpha] = F(\alpha)$.

Transcendental case. This time $\varphi_\alpha : F[X] \to F[\alpha]$ is an isomorphism (2.5.5) and hence $F[\alpha]$ is *not* a subfield of E. It is a subring, and hence an integral domain. Now, theorem (1.3.4) implies that E contains the field of fractions of $F[\alpha]$. We leave it as an exercise to show that this field of fractions is the smallest subfield of E containing F and α. That is, the field of fractions of $F[\alpha]$ is $F(\alpha)$.

Definition 2.5.8 (Simple extension) *An extension E of F is* **simple** *if $E = F(\alpha)$ for some $\alpha \in E$.*

We conclude this lecture with a theorem about the dimension of simple extensions $F(\alpha)$ in the case that α is algebraic.

Theorem 2.5.9 *Suppose $E = F(\alpha)$ is a simple extension of F with $\alpha \in E$ algebraic over F. If $\deg(\alpha, F) = n$, then $[F(\alpha) : F] = n$ so that in particular, simple algebraic extensions are finite.*

Proof. Let $p = \operatorname{irr}(\alpha, F)$. Since α is algebraic over F, we know $F(\alpha) = F[\alpha] = \operatorname{im} \varphi_\alpha$. We claim that $(1, \alpha, \ldots, \alpha^{n-1})$ is a basis for $F[\alpha]$ over F, and the result follows immediately. First, if $\beta \in F[\alpha]$, then $\beta = \varphi_\alpha(g)$ for some $g \in F[X]$. Using the division algorithm, we write $g = pq + r$ where $\deg r < \deg p$ or $r = 0$. This shows that $\beta = \varphi_\alpha(g) = \varphi_\alpha(pq + r) = \varphi_\alpha(r)$. If $r = 0$, then $\beta = 0 \in \operatorname{Span}_F(1, \alpha, \ldots, \alpha^{n-1})$. Otherwise we have $r = a_0 + a_1 X + \cdots + a_{n-1} X^{n-1}$ with all $a_j \in F$ and hence

$$\beta = \varphi_\alpha(r) = a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1}$$

so that $(1, \alpha, \ldots, \alpha^{n-1})$ spans $F[\alpha]$. Moreover, if

$$a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} = 0$$

with all $a_j \in F$, then $f = a_0 + a_1 X + \cdots + a_{n-1} X^{n-1} \in (p)$ so that $f = 0$ by the minimality of the degree of p. This shows that $(1, \alpha, \ldots, \alpha^{n-1})$ is linearly independent over F and hence a basis for $F[\alpha]$ over F. ■

Chapter 3

Modules

3.1 Introduction to modules

This lecture introduces our next topic - modules. Loosely speaking, a module is like a vector space, except that the scalars will only be assumed to come from a commutative ring. As we will see, this will change some familiar results from linear algebra dramatically. From now on, R will always be a commutative unital ring unless stated otherwise. Here is the main definition.

Definition 3.1.1 (R-module) *If R is a commutative unital ring, an abelian group M (written additively) is an R-**module** if we are given a function $R \times M \to M$ denoted $(r, m) \mapsto rm$ satisfying for all $m, n \in M$, $r, s \in R$:*

M1. $1m = m$.

M2. $r(m + n) = rm + rn$.

M3. $(r + s)m = rm + sm$.

M4. $(rs)m = r(sm)$.

In particular, we note that the axioms **M1-M4** are precisely the axioms for M to be a vector space over R in the case that R is a field. In fact, we see at once that if R is a field, then an R-module M is just a vector space over R.

Example 3.1.2 Every ring R is an R-module under the map $R \times R \to R$ given by multiplication in the ring. The axioms **M1-M4** are readily verified (they are the ring axioms). This module is called the **regular R-module**.

Example 3.1.3 If R is a ring and $n \geq 1$ is an integer, the product group R^n is an R-module under the map $(r, (x_1, \ldots, x_n)) \mapsto (rx_1, \ldots, rx_n)$. We leave the proof of this as an exercise.

One way to get started thinking about modules is to look for some over familiar rings. We have remarked that modules over fields are vector spaces, so we know all about those. The "next" familiar ring is the ring of integers $\mathbb{Z}$. As we will see, we know all about $\mathbb{Z}$-modules as well.

Theorem 3.1.4 *Every abelian group M is a $\mathbb{Z}$-module in a unique way.*

Proof. Define $\mathbb{Z} \times M \to M$ by $(a, m) \mapsto a \cdot m$. Then **M1** is trivial, and **M3** and **M4** are the laws of exponents in the abelian group M (written additively of course). Finally, **M2** is the identity $(mn)^a = m^a n^a$ which is valid in an abelian group. This shows that M is a $\mathbb{Z}$-module. Furthermore, since the module axioms are all exponent laws that *must be valid* in any abelian group, we see that this module structure is unique. ■

We will make the last statement of the theorem more precise when we have defined the notion of isomorphic R-modules.

OK, by now the notation $(r, m) \mapsto rm$ must be driving your group action sensors wild! It is no accident! As we have remarked before, just as groups were meant to act on sets, rings were meant to act as endomorphisms of abelian groups. We have looked at ideas like this two times now, so we will leave the proof of the following theorem to you. Go do it!

Theorem 3.1.5 *Let M be an abelian group and let R be a commutative unital ring. Then M is an R-module if and only if there exists a unital ring homomorphism $\rho : R \to \mathrm{End}(M)$.* ■

With this theorem at hand, we see that what we are doing is studying the representation theory of (commutative unital) rings.

3.2 Module homomorphisms

Since a module is similar to a vector space in the sense that there is a ring of "scalars" that operates on an abelian group, the reader can probably formulate his or her own definition of a module homomorphism.

Definition 3.2.1 (Module homomorphism) *If R is a commutative unital ring, and M and N are R-modules, then a mapping $\varphi : M \to N$ is an* R-**module homomorphism** *if for all $x, y \in M$ and all $r \in R$ we have*

1. $\varphi(x + y) = \varphi(x) + \varphi(y)$.

2. $\varphi(rx) = r\varphi(x)$.

That is, an R-module homomorphism is a homomorphism of abelian groups (1) that is compatible with the action of R (2). Note that is R is a field (so that M and N are just vector spaces over R), then an R-module homomorphism is what we have called a linear transformation. Similarly, homomorphisms of $\mathbb{Z}$-modules are simply homomorphisms of abelian groups.

Example 3.2.2 Consider the abelian group $\mathbb{Z}$ as a $\mathbb{Z}$-module. The map $f : \mathbb{Z} \to \mathbb{Z}$ given by $f(n) = 2n$ is easily seen to be a $\mathbb{Z}$-module homomorphism. Note that f is not a ring homomorphism.

Along with homomorphisms, the "sub-objects" in this business are called submodules. You can guess that they are defined to subsets of the module, that are themselves modules under the same operations. This amounts to requiring that we have a subgroup of M that is invariant under the action of R. Here is the definition.

Definition 3.2.3 (Submodule) *Let be a R commutative unital ring and let M be an R-module. A subset $N \subseteq M$ of M is a* **submodule of** M *if*

1. *N is a subgroup of the group M.*

2. *$rn \in N$ for all $r \in R$ and all $n \in N$.*

We often refer to condition (2) of the definition by saying that N is **stable** under R or that N is R-**invariant**.

Example 3.2.4

1. If M is an R-module, then 0 and M are submodules called the **trivial submodule** and the **improper submodule** respectively.

2. The submodules of the regular module R are just the ideals of R.

3. The submodules of a $\mathbb{Z}$-module M are just the subgroups of the abelian group M.

Now, the rest of the lecture can be summed up in one sentence! Namely, all of the usual results about homomorphisms, sub-objects and their interaction hold for R-modules. We will list the results here, but we omit the proofs. You will be asked to supply the detailed proofs in your homework exercises.

Theorem 3.2.5 *Let be a R commutative unital ring, M and N be R-modules and let $\varphi : M \to N$ be a R-module homomorphism.*

1. *If $M' \leq M$ is a submodule, then $\varphi(M') \leq N$ is a submodule.*
2. *If $N' \leq N$ is a submodule, then $\varphi^{-1}(N') \leq M$ is a submodule.*
3. *The kernel $\ker\varphi$ and image $\operatorname{im}\varphi$ are submodules of M and N respectively where $\ker\varphi$ denotes the kernel of φ as a group homomorphism.*
4. *φ is injective iff. $\ker\varphi = \{0\}$.*

■

Of course, an **isomorphism of R-modules** is a bijective module homomorphism. We say that two R-modules M and N are **isomorphic** if there exists an isomorphism $\varphi : M \to N$.

Theorem 3.2.6 *Let R be a commutative unital ring and let M be an R-module. Then if N is a submodule of M, the operation defined on the quotient group M/N defined by*

$$r(x + N) = rx + N$$

is a well defined operation $R \times M/N \to M/N$ making M/N into an R-module. ■

The module M/N is called the **quotient module of M by N**. The canonical map $M \to M/N$ is a surjective R-module homomorphism with kernel equal to N.

Proposition 3.2.7 *If K and N are submodules of an R-module M, then*

1. *$K \cap N$ is a submodule of M.*
2. *The set $K + N = \{k + n : k \in K, n \in N\}$ is a submodule of M.*

We say that K and N are **independent** if $K \cap N = 0$. We say that K and M **generate** M if $K + N = M$. We write $M = K \oplus N$ if K and N are independent and generate M. Note that in this case, every element of M cam be written uniquely as a sum $k + n$ with $k \in K$ and $m \in M$. *cf.* Proposition 2.9.6 of the 150A notes.

Definition 3.2.8 (Simple module) *A non-trivial R-module M is called* **simple** *if M and 0 are the only submodules in M.*

Lemma 3.2.9 (Lifting lemma) *Suppose that M, N and M' are R-modules, $\varphi : M \to N$ and $\pi : M \to M'$ are module homomorphisms and π is surjective. Then there exists a unique module homomorphism $\psi : M' \to M$ such that $\psi \circ \pi = \varphi$ if and only if $\ker \pi \subseteq \ker \varphi$. Moreover, ψ is surjective if and only if φ is surjective and ψ is injective if and only if $\ker \pi = \ker \varphi$.* ■

Corollary 3.2.10 (First isomorphism theorem) *If $\varphi : M \to N$ is an R-module homomorphism and $K = \ker \varphi$, then M/K is isomorphic to $\operatorname{im} \varphi$ as an R-module.* ■

Corollary 3.2.11 (Second isomorphism theorem) *Let M be an R-module and let K and N be two submodules of M. Then $K \cap N$ is a submodule of N, K is a submodule of $K + N$ and*

$$(K + N)/K \cong N/(K \cap N).$$

■

Corollary 3.2.12 (Third isomorphism theorem) *Let M be an R-module and let K and N be two submodules of M such that $K \leq N \leq M$. Then N/K is a submodule of M/K and*

$$(M/K)/(N/K) \cong M/N.$$

■

We conclude this lecture with a remark on logical efficiency. Theoretically, we could have developed the notion of module as soon as we knew that $\operatorname{End}(M)$ is a ring for all abelian groups M. We could have then stated and proved the lifting lemma and the isomorphism theorems for modules. The same results for rings would then follow at once by applying the module results to the regular R-module R. In fact, many of the notions we have looked at for rings can be defined for R-modules, and the particular case of the regular module coincides with the ring theoretic notions. For example, a ring R is simple iff. it is simple as the regular R-module and so on.

3.3 Free modules and bases

We continue to assume that R is a commutative unital ring throughout this section. The reader will no doubt agree that matrices play an invaluable role in linear algebra. The same is true for the theory of R-modules for any commutative ring R (not just fields). In particular, one can go back to the definition of a matrix over $\mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ given in section 1.1.2 of the MAT 150A notes (still online) and replace the entries with entries in any ring R. Such a matrix is called an **R-matrix**. The definitions of matrix addition and multiplication are unchanged. Moreover the proof of theorem 1.2.1 in the same section goes through for R-matrices unchanged. We can also take the same definition of determinant for an $n \times n$ R-matrix. The same proof used over fields shows that $\det(AB) = \det(A)\det(B)$ for $n \times n$ R-matrices A and B. Therefore A is invertible if and only if $\det A = \delta$ is a unit in the ring R.

Now, the concepts of spanning sets and independent sets transfer immediately to R-modules. Namely, we have the following definition.

Definition 3.3.1 *Let M be an R-module and let $(m_1, \dots, m_n)$ be an ordered set of elements of M. Then an* **R-linear combination of the** *m_i is an element $m \in M$ of the form*

$$m = r_1m_1 + r_2m_2 + \cdots + r_nm_n$$

where each $r_i \in R$. The elements $r_i \in R$ are called the **coefficients** *of the linear combination. If $S = (m_1, m_2, \dots, m_n)$ is an ordered set of vectors in M, the set of all linear combinations of the vectors in S is called the* **span of** *S and is denoted RS. Therefore*

$$RS = \{m : m = r_1m_1 + r_2m_2 + \cdots + r_nm_n, r_i \in R\}.$$

If $S \subset M$ and $RS = M$, then we say S **generates** *M. If S is finite and generates M, we say that M is finitely generated.*

The reader should show that if S is any subset of an R-module M, then RS is always a submodule of M. More generally, if I is any ideal in R, then IS is a submodule of M. Finally S is a submodule of M iff. $RS = S$. (All nice exercises!) All of the modules that we will encounter will be finitely generated. If R is a field so that an R-module M is just a vector space, then M is finitely generated if and only if it is finite dimensional. Now we turn to independence. Again, the definition transfers to R-modules without change.

Definition 3.3.2 *A subset $S = \{m_1, m_2, \ldots, m_n\}$ of an R-module M is called* **(linearly) independent** *if the equation*

$$r_1 m_1 + r_2 m_2 + \cdots + r_n m_n = 0,$$

with $r_i \in R$, implies that $r_i = 0$ for all $i = 1, 2, \ldots, n$. If S is not linearly independent, we say it is **dependent**.

Definition 3.3.3 (Basis) *If M is an R-module, a subset $S \subseteq M$ is called a* **basis** *if S is a linearly independent spanning set.*

Example 3.3.4 The "standard basis" $(e_1, \ldots, e_k)$ is easily seen to be a basis for the module R^n.

Just as with vector spaces, if $(m_1, \ldots, m_n)$ is a basis for an R-module M, then every element $m \in M$ can be written as an R-linear combination of the m_i in a unique way. One place we we see that the theory of arbitrary R-modules differs dramatically from vector spaces, is that not every R-module will have a basis. In fact, "most" R-modules will not have a basis if R is not a field.

Definition 3.3.5 (Free module) *A finitely generated R-module M is* **free** *if there exists an R-module isomorphism $\varphi : R^n \to M$ for some $n \in \mathbb{N}$. The integer n is called the* **rank** *of the free module M. To avoid some pathology in the future, we say that the trivial module 0 is free of rank 0 and the empty set $\emptyset$ is a basis.*

Proposition 3.3.6 *A finitely generated R-module M has a basis if and only if it is free.*

Proof. Let $\mathcal{B} = (m_1, \ldots, m_n)$ denote an ordered set of elements of M and define a map $\psi : R^n \to M$ by

$$\psi(X) = \mathcal{B}X = (m_1, \ldots, m_n) \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 m_1 + \cdots + x_n m_n.$$

This map is easily seen to be an R-module homomorphism. Moreover, ψ is surjective iff. $\mathcal{B}$ generates M and ψ is injective iff. $\mathcal{B}$ is independent. Therefore $\mathcal{B}$ is a basis for M iff. $\psi : R^n \to M$ is an isomorphism iff. M is free. ■

Example 3.3.7 The proposition shows that a free $\mathbb{Z}$-module is isomorphic to the $\mathbb{Z}$-module $\mathbb{Z}^n$ for some n. Therefore every free $\mathbb{Z}$-module is infinite. It follows that no finite $\mathbb{Z}$-module (finite abelian group) is free. So we see at last an example of a finitely generated module that is not free, any finite abelian group will do!

Definition 3.3.8 *An R-module M is* **cyclic** *if it is generated by a single element. That is, $M = Rx = \{rx : r \in R\}$ for some $x \in M$. If N is any R-module and $x \in N$, then $Rx \leq N$ is the* **cyclic submodule generated by** *x.*

Proposition 3.3.9 *If F is a free R-module with basis $(x_1, \dots, x_n)$, then $F = Rx_1 \oplus \cdots \oplus Rx_n$.*

Proof. We know that $Rx_i \cap (Rx_1 + \cdots Rx_{i-1} + Rx_{i+1} + \cdots + Rx_n) = 0$ since $(x_1, \dots, x_n)$ is independent. Moreover, $Rx_1 + \cdots + Rx_n = F$ since $(x_1, \dots, x_n)$ generates F. Therefore $F = Rx_1 \oplus \cdots \oplus Rx_n$. ■

It is a surprising fact that for some rings, the cyclic submodule Rx for $x \in M$ is not free even if M is free of rank 1! This is not the case for PIDs, as the following proposition shows. We will use this lemma later on as a base step for an induction.

Proposition 3.3.10 *If R is a PID and M is a free R-module of rank 1, then every non-zero submodule of M is free of rank 1.*

Proof. By hypothesis, there exists $x \in M$ such that the map $\psi : R \to M$ given by $\psi(r) = rx$ is an isomorphism. Let $0 \neq N \leq M$ be a non zero submodule of M so that $\psi^{-1}(N)$ is a non zero submodule of R. That is, $\psi^{-1}(N)$ is a non zero ideal of R, and hence $\psi^{-1}(N) = (a)$ for some $0 \neq a \in R$ since R is a PID. Therefore

$$N = \psi((a)) = \{rax : r \in R\}.$$

Therefore the map $\varphi : R \to N$ given by $\varphi(r) = rax$ is an onto homomorphism of R-modules. Moreover, $\varphi(r) = 0$ iff. $rax = 0$ iff. $\psi(ra) = 0$ iff. $ra = 0$ since ψ is injective. But R is an integral domain and $a \neq 0$ so that we must have $r = 0$ and hence φ is injective as well. This shows that $\varphi : R \to N$ is an isomorphism of R-modules and hence N is free of rank 1 by definition. ■

We remark here that, just as for vector spaces, R-homomorphisms $M \to N$ of free R-modules M and N can be given by multiplication by R-matrices once basis are chosen. Moreover, change of basis matrices work just as they do for vector spaces, and a change of a basis matrix is necessarily invertible. These ideas are discussed in detail on page 455 in Artin's text. We will make little use of these facts, so we do not expound on them here.

We can now state the main goal of the current chapter: We want to classify all finitely generated modules M over a PID R. To put the problem in perspective, consider the case that R is a field so that an R-module is a vector space over R. Since every finitely generated vector space has a basis (see Lemma 3.3.11 of the 150A notes), every finitely generated vector space is free. Moreover, Theorem 3.3.14 of the 150A notes implies that any two vector spaces over R of the same dimension are isomorphic. Therefore, if R is a field, we see that every finitely generated R-module is isomorphic to R^n, and thus we have classified finitely generated modules over fields. We want to do the same thing for PIDs. We already know that not every R-module is free, so the answer will be different. However, since a change of basis matrix is invertible, any two free R-modules of the same rank n are isomorphic. We end this lecture with an important general fact about free modules as well as the first step toward the structure theorem we desire.

Theorem 3.3.11 *If M is an R-module and $x_1, \ldots, x_n \in M$, then there is a unique homomorphism $\varphi : R^n \to M$ with $\varphi(e_i) = x_i$ for all i. In particular, if M is generated by the x_i, φ is surjective.*

Proof. Let $x_1, \ldots, x_n \in M$ and define $\varphi : R^n \to M$ by

$$\varphi : (r_1, \ldots, r_n) \mapsto r_1x_1 + \cdots + r_nx_n.$$

Then the reader can check that φ is a R-module homomorphism and that $\varphi(e_i) = x_i$ for all i. If $\psi : R^n \to M$ also satisfies $\psi(e_i) = x_i$, then for all $r = \sum r_ie_1 \in R^n$, we have

$$\psi(r) = \sum_{i=1}^{n} r_i\psi(e_i) = \sum_{i=1}^{n} r_ix_i = \sum_{i=1}^{n} r_i\varphi(e_i) = \varphi(r)$$

and hence $\varphi = \psi$ is unique. Finally, if the x_i generate M, then every $x \in M$ has the form $x = \sum r_ix_i$ so that if $r = \sum r_ie_i$ we see that $\varphi(r) = x$ and hence φ is surjective. ■

The theorem implies that every finitely generated R-module M is the image of a free module of finite rank.

Theorem 3.3.12 *Let R be a PID and let F be a free R-module of rank n. Then if $G \leq F$ is a submodule of F, then G is free of rank m with $m \leq n$.*

Proof. Let $(x_1, \ldots, x_n)$ be a basis for F and for each $1 \leq k \leq n$, let $F_k = \mathrm{Span}_R(x_1, \ldots, x_k)$ and $G_k = G \cap F_k$. By induction, we will prove that each G_k is free of rank less than or equal to k. If $G_1 = 0$, then G_1 is free of rank 0. If $G_1 \neq 0$, then G_1 is a submodule of Rx_1 and hence G_1 is free of rank 1 by proposition (3.3.10). Let $k > 1$ and suppose that G_{k-1} is free of rank less than or equal to $k-1$. If $G_k = G_{k-1}$, then we are done. Otherwise let

$$I = \{b \in R : \text{there exists } x' \in F_{k-1} \text{ and } x' + bx_k \in G_k\}.$$

If $\pi'_k : F \to Rx_k$ is the usual coordinate projection restricted to G_k, then $I = \{b \in R : bx_k \in \operatorname{im} \pi'_k\}$. Therefore I is an ideal in R, say $I = (a)$. Since $G_k \neq G_{k-1}$, we must have $a \neq 0$. Therefore there is some $y \in G_k$ and $y' \in F_{k-1}$ with $y = y' + ax_k$. We claim that $G_k = G_{k-1} \oplus Ry$.

Let $z \in G_k$. Then there exist $z' \in F_{k-1}$ and $c \in R$ with $z = z' + cx_k$. Therefore $c \in (a)$ and hence $c = ad$ for some $d \in R$. This means that $z - dy = z' - dy' \in G_{k-1}$ and hence $G_k = G_{k-1} + Ry$. Now, $ry \in G_{k-1} \subseteq F_{k-1}$ iff. $rax_k = 0$ iff. $ra = 0$ iff. $r = 0$. Therefore $G_{k-1} \cap Ry = 0$ and hence $G_k = G_{k-1} \oplus Ry$. By induction, G_{k-1} is free of rank less than or equal to $k-1$ so that G_k is free of rank less than or equal to k. ■

Corollary 3.3.13 *If R is a PID and M is an R-module generated by m elements, then if N is a submodule of M, N is generated by less than or equal to m elements.*

Proof. Suppose that M is generated by m elements and take a surjective homomorphism $\varphi : R^m \to M$ (theorem (3.3.11). Now, $\varphi^{-1}(N)$ is a submodule of the free module R^m by theorem 3.2.5 so that $\varphi^{-1}(N)$ is free of rank $n \leq m$ by theorem (3.3.12). If $(y_1, \ldots, y_n)$ is a basis for $\varphi^{-1}(N)$, then easily $(\varphi(y_1), \ldots, \varphi(y_n))$ is a generating set for N and hence N is generated by less than or equal to m elements. ■

3.4 Some odds and ends about PIDs

In this lecture, we list and prove some useful facts about PIDs that we will need in our classification theorem. We could have covered this material before we began studying modules. Until further notice, let R be a PID.

Definition 3.4.1 *If R is a commutative unital ring and $a_1, \ldots, a_n \in R$, a* **greatest common divisor (GCD)** *of the a_i is an element $d \in R$ such that $d|a_i$ for all i and if $e \in R$ is another element satisfying $e|a_i$ for all i, then $e|d$. A* **least common multiple (LCM)** *of the a_i is an element $m \in R$ such that $a_i|m$ for all i and if $n \in R$ is another element in R with $a_i|n$ for all i, then $m|n$. We say that the elements $a_1, \ldots, a_n$ are* **relatively prime** *if the GCD of the a_i is 1.*

Before we state the following proposition, we recall that for a commutative ring R and $a, b \in R$, $a \in (b)$ if and only if $b|a$.

Proposition 3.4.2 *Let R be a PID and let $a_1, \ldots, a_n \in R$.*

1. *The generator of the ideal $(a_1, \dots, a_n)$ is a GCD of the a_i.*

2. *The generator of the ideal $(a_1) \cap \cdots \cap (a_n)$ is a LCM of the a_i.*

3. *If d and d' are GCDs of the a_i, then $d = d'u$ for some unit $u \in R$.*

4. *If m and m' are LCMs of the a_i, then $m = m'u$ for some unit $u \in R$.*

5. *$(a_1, \dots, a_n) = (d)$ if and only if d is a GCD of the a_i.*

6. *$(a_1, \dots, a_n) = R$ if and only if the a_i are relatively prime.*

7. *If $a, b \in R$ are relatively prime and $(b) \subseteq (a)$, then $(a) = R$.*

Proof.

1 Since R is a PID, we know $(a_1, \dots, a_n) = (d)$ for some $d \in R$. Now, each $a_i \in (a_1, \dots, a_n)$ and hence each $a_i \in (d)$ so that $d|a_i$ for all i. If $e \in R$ and $e|a_i$ for all i, then $a_i = er_i$ for some $r_i \in R$ for all i. Now, $d \in (a_1, \dots, a_n)$ so that $d = s_1a_1 + \cdots + s_na_n$ for some $s_i \in R$ and hence

$$d = s_1a_1 + \cdots + s_na_n = s_1r_1e + \cdots + s_nr_ne = (s_1r_1 + \cdots + s_nr_n)e,$$

so that $e|d$. This shows that d is a GCD of the a_i.

2 Again, R is a PID so that the ideal $(a_1) \cap \cdots \cap (a_n) = (m)$ is principle generated by m for some $m \in R$. Moreover, $m \in (a_i)$ for all i so that $a_i|m$ for all i. Now, if $a_i|n$ for all i, then for all i we have $n = a_ir_i$ for some $r_i \in R$ and hence $n \in (a_1) \cap \cdots \cap (a_n) = (m)$. It follows that $n \in (m)$ and hence $m|n$. This shows that m is a LCM of the a_i.

3 If d and d' are GCDs of the a_i, then $d|d'$ and $d'|d$. Therefore $d' \in (d)$ and $d \in (d')$ so that $(d) = (d')$. It follows (from you midterm exam problem) that $d = d'u$ for some unit $u \in R$.

4 Proceed as in (3).

5 If $(d) = (a_1, \dots, a_n)$, then d is a GCD of the a_i by (1). Now let d' be any GCD of the a_i. Then by (3), $d' = du$ for some unit $u \in R$ and hence $(d') = (du) = (d)$. Therefore $d' = (a_1, \dots, a_n)$.

6 This is immediate: the a_i are relatively prime iff. the GCD is 1 iff. $(a_1, \dots, a_n) = (1) = R$.

7 If a and b are relatively prime, then by (6), we have $1 = ax + by$ for some $x, y \in R$. Since $(b) \subseteq (a)$, $ax, by \in (a)$ and hence $(1) \in (a)$ so that $(a) = R$.

■

Now, since R is commutative, it is easy to see that $(a_1, \ldots, a_n) = (a_1) + \cdots + (a_n)$ so that we have the following corollary.

Corollary 3.4.3 *If R is a PID and $a_1, \ldots, a_n \in R$, then*

$$(a_1) \cap \cdots \cap (a_n) \subseteq (a_1) + \cdots + (a_n).$$

Proof. We know that if m and d are an LCM and GCD for the a_i respectively, then $(a_1) \cap \cdots \cap (a_n) = (m)$ and $(a_1, \ldots, a_n) = (d)$. Now, $d | a_i$ for all i and $a_i | m$ for all i so that $d | m$ and hence $m \in (d)$ so that $(m) \subseteq (d)$. ■

Now, if we let $a = a_1 \cdots a_n$ be the product of all the a_i, then $a_i | a$ for all i and hence $m | a$. This implies that $a \in (m)$ and hence $(a) \subseteq (m)$. This proves the following corollary.

Corollary 3.4.4 *If R is a PID and $a_1, \ldots, a_n \in R$, then*

$$(a_1 \cdots a_n) \subseteq (a_1) \cap \cdots \cap (a_n).$$

■

We do not want to go into the details here, but for completeness, we remark that one can show that every PID has the unique factorization into products of irreducible elements (such as the integers $\mathbb{Z}$ and polynomials $F[X]$ over a field). Such rings are called **Unique factorization domains (UFD)**. Every PID is a UFD. Using this, it is not hard to show that the inclusion in the previous corollary is an equality if and only if the a_i are relatively prime. Let use briefly sketch the details for the case $n = 2$, and leave the general (induction) proof to the reader.

Lemma 3.4.5 *If R is a UFD and $a, b \in R$ are relatively prime, then $(ab) = (a) \cap (b)$.*

Proof. Since R is a UFD, there are irreducible elements $p_1, \ldots, p_n \in R$ and natural numbers $e_1, \ldots, e_n, f_1, \ldots, f_n \geq 0$ such that

$$a = p_1^{e_1} \cdots p_n^{e_n} \text{ and } b = p_1^{f_1} \cdots p_n^{f_n}.$$

If we let $\alpha_i = \min(e_i, f_i)$ and $\beta_i = \max(e_i, f_i)$ for all i, then the reader can easily show that

$$d = p_1^{\alpha_1} \cdots p_n^{\alpha_n}$$

is a GCD for a and b and

$$m = p_1^{\beta_1} \cdots p_n^{\beta_n}$$

is an LCM for the a_i. Moreover, since (trivially) $\alpha_i + \beta_i = e_i + f_i$ for all i, we see that $ab = dm$. If a and b are relatively prime, then d is a unit and hence $(m) = (ab)$. The result now follows. ■

Since every PID is a UFD, lemma (3.4.5) is valid for any PID R.

3.5 Torsion modules and order ideals

Recall that our present goal is to classify all finitely generated modules over a PID. We have remarked that not all such modules are free, so the situation is more complicated than that of modules over a field (vector spaces). As a first step in our classification, we will see that every finitely generated module M over a PID R does have a (possibly 0) direct summand which is free. The complementary summand consists of elements of "finite order". More precisely, we make the following definition.

Definition 3.5.1 *An R-module M is* **torsion free** *if for all $x \in M$ and all $a \in R$, $ax = 0$ implies either $a = 0$ or $x = 0$.*

Note that any (finitely generated) free module over an integral domain is torsion free. This follows since if M is finitely generated and free, then $M \cong R^n$ so that $x \in M$ has the form $(r_1, \ldots, r_n) \in R^n$. Now, if R is an integral domain, then $ar_i = 0$ implies that $a = 0$ or $r_i = 0$. But $ax = 0$ iff. $ar_i = 0$ for all i so that we see that either $a = 0$ or $r_i = 0$ for all i and hence $x = 0$. The converse of this fact is only partially true.

Proposition 3.5.2 *If R is a PID, then a finitely generated R-module M is free if and only if M is torsion free.*

Proof. We have just remarked that a finitely generated free module over an integral domain is torsion free. For the converse, assume that $0 \neq M$ is torsion free. (If $M = 0$, then it is trivially free.) Let $\{x_1, \ldots, x_n\}$ be a generating set for M so that $M = Rx_1 + \cdots + Rx_n$ and each $x_i \neq 0$. Since M is torsion free, if $0 \neq x \in M$, then $\{x\}$ is independent so that Rx is free of rank 1. If we choose a maximal independent subset from $\{x_1, \ldots, x_n\}$, say $\{x_1, \ldots, x_m\}$ (re-enumerating if necessary), then the submodule

$$F = Rx_1 \oplus \cdots \oplus Rx_m$$

is free of rank m since each Rx_i is free of rank 1 and the x_i are independent.

Now, for each $i > m$, since $\{x_1, \ldots, x_m\}$ is *maximal* independent, there exists a non-zero element $a_i \in R$ with $a_i x_i \in F$. Let $a = a_{m+1}a_{m+2}\cdots a_n$ and note that $ax \in F$ for all $x \in M$ (homework exercise). Now, R is an integral domain so that $a \neq 0$. Therefore, since M is torsion free, the map $\lambda_a : M \to F$ given by $\lambda_a(x) = ax$ is injective and hence an isomorphism onto its image $\operatorname{im}\lambda_a = aM = \{ax : x \in M\}$. But λ_a is a R-module homomorphism (homework exercise) so that aM is a submodule of the finitely generated free module F. Therefore theorem (3.3.12) implies that aM is finitely generated and free. But $aM \cong M$ so that M is finitely generated and free. ■

Now let M be any R-module and let $x \in M$. The map $R \to M$ defined by $a \mapsto ax$ is an R-module homomorphism. Its image is the cyclic module generated by x. Define the **order ideal $\mathcal{O}_x$ of x** to be the set

$$\mathcal{O}_x = \{a \in R : ax = 0\}.$$

Note the order ideal of x is just the kernel of the map $a \mapsto ax$ and hence it is an an ideal. If R is a PID and $x \in M$ has a non zero order ideal, then $0 \neq \mathcal{O}_x = (a)$ for some non zero $a \in R$ which is unique up to multiplication by a unit. We call a the **order of $x \in M$**.

Definition 3.5.3 *If M is an R-module, an element $x \in M$ is* **torsion** *if $\mathcal{O}_x \neq 0$. We let*

$$M_t = \{x \in M : x \text{ is torsion}\}.$$

A module M is a **torsion module** *if $M = M_t$.*

We note that a module is torsion free iff. $M_t = 0$. This is because M is torsion free iff. $0 \neq a \in R$ and $0 \neq x \in M$ implies $ax \neq 0$ iff. $0 \neq x$ implies $\mathcal{O}_x = 0$ iff. $0 \in M$ is the only torsion element.

Lemma 3.5.4 *For every R module M, M_t is a submodule of M that is a torsion module. Moreover, the quotient M/M_t is torsion free.*

Proof. Suppose that $x, y \in M_t$ and $0 \neq a, b \in R$ satisfy $ax = by = 0$. Since R is a PID, we have $ab \neq 0$. Moreover, $ab(x + y) = abx + aby = 0 + 0 = 0$ so that $x + y \in M_t$. Also, for all $r \in R$, $a(rx) = (ar)x = (ra)x = r(ax) = r0 = 0$ so that $rx \in M_t$ as well. This shows that M_t is a submodule of M. By definition, each element of M_t is torsion so that $(M_t)_t = M_t$ and hence M_t is a torsion module. Finally, suppose that $x + M_t \in M/M_t$ and $0 \neq a \in R$ satisfies $ax + M_t = M_t$. Then $ax \in M_t$ so that $b(ax) = 0$ for some $0 \neq b \in R$. Now, R is a PID so that $ab \neq 0$ and $(ab)x = 0$ so that $x \in M_t$. This shows that M_t is the only torsion element of M/M_t and hence M/M_t is torsion free. ■

Not surprisingly, we call M_t the **torsion submodule of** M. We can now take the next step in out classification.

Theorem 3.5.5 *If M is a finitely generated module over a PID and M_t is the torsion submodule of M, then there is a finitely generated free submodule $M_f \leq M$ such that*

$$M = M_t \oplus M_f.$$

That is, every finitely generated module over a PID is a direct sum of a torsion module and a free module of finite rank.

Proof. Let M_t be the torsion submodule and let $\pi : M \to M/M_t$ be the quotient map. By lemma (3.5.4), M/M_t is torsion free. Since M is finitely generated, M/M_t is finitely generated and hence M/M_t is free by proposition (3.5.2). This means that we can choose a basis

$$(x_1 + M_t, x_2 + M_t, \ldots, x_m + M_t)$$

for M/M_t. We can then define a R-module homomorphism $\psi : M/M_t \to M$ by defining $\psi(x_i + M_t) = x_i$ for $1 \leq i \leq m$ and extending linearly (theorem 3.3.11). Then clearly $\pi \circ \psi = 1_{M/M_t}$ so that ψ is injective. If we let $M_f = \operatorname{im} \psi$, then M_f is a submodule of M and $\psi : M/M_t \to M_f$ is an isomorphism. It follows that M_f is finitely generated and free.

We claim that $M = M_t \oplus M_f$. If $x \in M_t \cap M_f$, then $\pi(x) = M_t$ and $x = \psi(y + M_t)$ for some $y = \sum_{i=1}^{m} a_i x_i$ with $a_i \in R$. But

$$M_t = \pi(x) = \pi(\psi(y + M_t)) = y + M_t$$

so that $y \in M_t$. But $(x_1 + M_t, x_2 + M_t, \ldots, x_m + M_t)$ is a basis in M/M_t so that we must have $a_i = 0$ for all i and hence $y = 0$ so that $x = 0$. This shows that $M_t \cap M_f = 0$.

Finally, if $x \in M$, then $\pi(x) = \sum a_i x_i + M_t$. Let $f = \sum a_i x_i$ so that $f \in \operatorname{im} \psi = M_f$. Moreover, if $t = x - f$, then

$$\pi(t) = \pi(x) - \pi(f) = \pi(x) - \pi\left(\psi\left(\sum a_i x_i + M_t\right)\right) = \pi(x) - \left(\sum a_i x_i + M_t\right) = \pi(x) - \pi(x) = 0$$

and hence $t \in \ker \pi = M_t$. Clearly $x = t + f$ and hence $M = M_t + M_f$ and the proof is complete. ■

Why is this the first step in our classification? We know that a finitely generated free R-module is isomorphic to R^n where n is the rank of the module. Just as for vector spaces, any two free modules of the same rank are isomorphic. (We alluded earlier to change of basis matrices which are still valid over commutative rings. These are the desired isomorphisms between two R-modules of a given rank.) Therefore finite rank free modules are classified by their rank. Having this in our hands, the previous theorem implies that we can complete out classification of finitely generated modules over PIDs if we can classify all finitely generated torsion modules over a PID. This is what we now set out to do.

3.6 Structure theorem for finitely generated modules over PIDs

Let R be a PID and let $\mathcal{P}$ be a complete set of representatives of all prime elements of R. (That is, we consider two primes p, p' in R to be equivalent if $p = p'u$ for some unit. This is an equivalence relation on the set of all prime elements (nice exercise) and we choose one prime from each class. For the ring $\mathbb{Z}$, this amounts to choosing say the positive primes. In $F[X]$, we choose irreducible *monic* polynomials.) For each prime $p \in \mathcal{P}$ and each R-module M, we let

$$M(p) = \{x \in M : \mathcal{O}_x = (p^n) \text{ for some } n \geq 0\}.$$

That is, $M(p)$ is the set of all elements of M that have order a power of p.

Lemma 3.6.1 *$M(p)$ is a submodule of M for all primes $p \in \mathcal{P}$.*

Proof. Exercise. ■

The following decomposition theorem holds for arbitrary torsion modules, not just finitely generated ones.

Lemma 3.6.2 *If R is a PID and M is a torsion module over R, then*

$$M = \bigoplus_{p \in \mathcal{P}} M(p).$$

Proof. Let $0 \neq x \in M$. Since $1x = x$, $\mathcal{O}_x \neq R$. But R is a PID so that $\mathcal{O}_x = (a)$ for some $0 \neq a \in R$. (We know $a \neq 0$ since M is a torsion module.) Now, $(a) \neq R$ so that a is not a unit and hence we can write

$$a = p_1^{e_1} \cdots p_n^{e_n}$$

where p_i is irreducible for each i. (This is because R is a PID and hence a UFD). Since R is a PID, corollary 2.4.9 implies that each p_i is a prime. For each $i = 1, \ldots, n$, let

$$q_i = p_1^{e_1} \cdots p_{i-i}^{e_{i-1}} p_{i+1}^{e_{i+1}} \cdots p_n^{e_n}$$

so that $a = q_i p_i^{e_i}$. Now, since $ax = 0$, we have $q_i x \in M(p_i)$ for all i. Moreover, we leave it as an exercise to show that $\{q_1, \ldots, q_n\}$ are relatively prime. Using this, we can write

$$1 = r_1 q_1 + \cdots + r_n q_n$$

for some $r_i \in R$ and hence

$$x = 1x = r_1 q_1 x + \cdots + r_n q_n x \in M(p_1) + \cdots + M(p_n).$$

This shows that the $M(p)$, $p \in \mathcal{P}$, generate M.

Now suppose that $p_1, \ldots, p_n, q \in \mathcal{P}$ are distinct and that

$$x \in (M(p_1) + \cdots + M(p_n)) \cap M(q).$$

Since $x \in M(q)$, then $\mathcal{O}_x = (q^m)$ for some $m \geq 0$. Also, since $x \in M(p_1) + \cdots + M(p_n)$, $x = r_1 x_1 + \cdots + r_n x_n$ where $r_i \in R$ and there exist $e_i \geq 0$ such that $p_i^{e_i} x_i = 0$ for all i. If we let

$$a = p_1^{e_1} \cdots p_n^{e_n},$$

then clearly $ax = 0$ so that $a \in (q^m)$ and hence $(a) \subseteq (q^m)$. Since the $p_1, \ldots, p_n, q \in \mathcal{P}$ are distinct primes, they are relatively prime and hence a and q^m are relatively prime. By proposition 3.4.2(7), this means that $(q^m) = R$ and hence $\mathcal{O}_x = R$ so that $x = 0$. Therefore the $M(p)$ are independent and the proof is complete. ■

If we combine theorem (3.5.5) and lemma (3.6.2), then we see that if M is a finitely generated module over a PID R, then

$$M = \left(\bigoplus_{p \in \mathcal{P}} M(p) \right) \oplus M_f$$

where $\mathcal{P}$ is a complete set of all primes in R and M_f is a free module over R of finite rank. Therefore, to complete our classification, we need only classify finitely generated torsion modules of the form $M(p)$. That is, finitely generated modules M with the property that $p^n x = 0$ for all $x \in M$ for some prime $p \in R$ and $n \in \mathbb{N}$.

For such a module M, given $x \in M$, we have $\mathcal{O}_x = (a)$ for some $a \in R$. But $p^n x = 0$ so that $p^n \in (a)$ and hence $a | p^n$. But since $p \in R$ is prime, this implies that $a = p^r$ for some $0 \leq r \leq n$. The element

p^r is called the **order of** x. Note that if p^r is the order of x, then $p^r x = 0$ but $p^{r-1}x \neq 0$. Also p^r is the order of x if and only $R/(p^r) \cong Rx$.

Definition 3.6.3 *Let M be a finitely generated module with $p^n M = 0$ for some $p \in \mathcal{P}$ and some $n \in \mathbb{N}$. The* **socle of** *M is*

$$\mathrm{Soc}(M) = \{x \in M : px = 0\}.$$

We leave it as an exercise for the reader to show that $\mathrm{Soc}(M)$ is a submodule of M.

Lemma 3.6.4 *If M is a finitely generated module with $p^n M = 0$ for some $p \in \mathcal{P}$ and some $n \in \mathbb{N}$ and $M = M_1 \oplus M_2$, then $\mathrm{Soc}(M) = \mathrm{Soc}(M_1) \oplus \mathrm{Soc}(M_2)$.*

Proof. Let $x \in \mathrm{Soc}(M)$ so that $px = 0$. Since $M = M_1 \oplus M_2$, there exist unique $y \in M_1$ and $z \in M_2$ with $x = y + z$. Now, $0 = px = py + pz$ and $py \in M_1$ and $pz \in M_2$. It follows that $py = pz = 0$ by uniqueness so that $y \in \mathrm{Soc}(M_1)$ and $z \in \mathrm{Soc}(M_2)$ so that $\mathrm{Soc}(M) = \mathrm{Soc}(M_1) + \mathrm{Soc}(M_2)$.
Now, for any module M, $\mathrm{Soc}(M) \subset M$ so that if $x \in \mathrm{Soc}(M_1) \cap \mathrm{Soc}(M_2)$, then $x \in M_1 \cap M_2$ so that $x = 0$. Therefore $\mathrm{Soc}(M_1)$ and $\mathrm{Soc}(M_2)$ are independent and the proof is complete. ■

Lemma 3.6.5 *If M is a finitely generated module with $p^n M = 0$ for some $p \in \mathcal{P}$ and some $n \in \mathbb{N}$, then $\mathrm{Soc}(M)$ is a $R/(p)$-module via the operation*

$$(a + (p), x) \mapsto ax.$$

Proof. First, if $x \in \mathrm{Soc}(M)$, then $px = 0$ so that for all $a \in R$, $p(ax) = a(px) = a0 = 0$ and hence $ax \in \mathrm{Soc}(M)$. The map defined in the statement of the lemma clearly satisfies the module axioms provided it is well defined. If $a + (p) = b + (p)$, then $a - b \in (p)$ so that $ax - bx = (a - b)x = pkx = 0$. Therefore the map is well defined and $\mathrm{Soc}(M)$ is a module over $R/(p)$. ■

Now, R is a PID so that the prime element p is irreducible and hence (p) is a maximal ideal. It follows that $R/(p)$ is a field and hence $\mathrm{Soc}(M)$ is a vector space over $R/(p)$. Since $\mathrm{Soc}(M)$ is a submodule of M, and M is finitely generated, corollary (3.3.13) implies that $\mathrm{Soc}(M)$ is finitely generated as an R-module and hence as an $R/(p)$-module. We have shown the following **important fact:**

$$\mathrm{Soc}(M) \text{ is a finite dimensional vector space over the field } R/(p).$$

Definition 3.6.6 *If M is an R-module, the elements $y_1, \ldots, y_m$ are* **weakly independent** *if*

$$r_1 y_1 + \cdots + r_m y_m = 0$$

implies that $r_i y_i = 0$ for all i.

Lemma 3.6.7 *If M is an R-module and $y_1, \ldots, y_m$ are weakly independent, then $R\{y_1, \ldots, y_m\} = Ry_1 \oplus \cdots \oplus Ry_m$.*

Proof. We have $R\{y_1, \ldots, y_m\} = Ry_1 + \cdots + Ry_m$ by definition. If

$$x \in Ry_i \cap (Ry_1 + \cdots + Ry_{i-1} + R_{i+1}y_{i+1} + \cdots + Ry_m),$$

then there exist elements $a_j \in R$, $1 \leq j \leq m$ with

$$a_i y_i = a_1 y_1 + \cdots + a_{i-1} y_{i-1} + a_{i+1} y_{i+1} + \cdots + a_m y_m$$

and hence

$$a_1 y_1 + \cdots - a_i y_i + \cdots + a_m y_m = 0.$$

Therefore $a_j y_j = 0$ for all j since the y_j are weakly independent and hence $x = a_i y_i = 0$. ■

OK, we need one more technical lemma before we can give our next big step in the decomposition theorem. I know it is easy to feel fatigued by now, but stick with me - it will all pay off!

Lemma 3.6.8 *Suppose that M is a torsion module with $p^n M = 0$ and $p^{n-1} M \neq 0$ for some $n \geq 1$. Let $0 \neq x \in M$ with $p^{n-1} x \neq 0$ and let $\overline{M} = M/Rx$. Let $\overline{y}_1, \ldots, \overline{y}_m \in \overline{M}$ be weakly independent elements of $\overline{M}$. Then for each i, there exists a representative y_i of $\overline{y}_i$ such that y_i has the same order as $\overline{y}_i$. Moreover, the elements $x, y_1, \ldots, y_m$ are weakly independent.*

Proof. If $\overline{M} = 0$, there is nothing to prove, so let $\overline{y} \in \overline{M}$ have order p^e for some $e \geq 1$. If y is a representative of $\overline{y}$ in M, then $p^e y \in Rx$ and hence $p^e y = ax$ for some $a \in R$. Since p is a prime, we can write $a = p^s c$ where $p \not| c$. If $s \geq n$, then $p^e y = 0$ so that the order of y is also p^e (if $p^r y = 0$ for some $r < e$, then $p^r \overline{y} = 0$, a contradiction). If $s < n$, then $p^s cx$ has order p^{n-s} and hence y has order p^{e+n-s}. We must have $e + n - s \leq n$ (why?) and hence $e \leq s$. Moreover, the element

$$y - p^{s-e} cx$$

has order p^e and

$$y - (y - p^{s-e} cx) = p^{s-e} cx \in Rx$$

so that $y - p^{s-e} cx$ also represents $\overline{y}$. We have shown that if $\overline{y} \in \overline{M}$, we can find a representative $y \in M$ of $\overline{y}$ with the same order as $\overline{y}$. Therefore, for each $i = 1, \ldots, m$, we can choose a representative y_i of $\overline{y}_i$ with the same order as $\overline{y}_i$.

We will now show that $x, y_1, \ldots, y_m$ are weakly independent. Suppose that $a, a_1, \ldots, a_m \in R$ and

$$ax + a_1y_1 + \cdots + a_my_m = 0.$$

Therefore in the quotient $\overline{M}$, we have

$$a_1\overline{y}_1 + \cdots + a_m\overline{y}_m = \overline{0}.$$

But the $\overline{y}_i$ are weakly independent so that $a_i\overline{y}_i = \overline{0}$ for all i. If p^{e_i} is the order of $\overline{y}_i$, then $a_i\overline{y}_i = \overline{0}$ implies $a_i \in (p^{e_i})$ and hence p^{e_i} divides a_i. That is, $a_i = p^{e_i}k_i$ for all i with $k_i \in R$. Now $a_iy_i = p^{e_i}k_iy_i = 0$ since p^{e_i} is also the order of y_i. This in turn implies that $ax = 0$ and hence $x, y_1, \ldots, y_m$ are weakly independent as desired. ∎

We now have the machinery in place to prove the next step in out decomposition. Recall that we are working on modules M over PIDs with the property that $p^nM = 0$ and $p^{n-1}M \neq 0$ for some prime $p \in R$.

Lemma 3.6.9 *If M is a finitely generated module with $p^nM = 0$ for some $p \in \mathcal{P}$ and some $n \in \mathbb{N}$, then there exist unique natural numbers*

$$n_1 \geq n_2 \geq \cdots \geq n_k \geq 1$$

and

$$M \cong R/(p^{n_1}) \oplus \cdots \oplus R/(p^{n_k}).$$

Proof. We will induct on $d = \dim_{R/(p)}(\mathrm{Soc}(M))$, the case $d = 0$ being trivial. Given $0 \neq x \in M$, the remarks before definition (3.6.3) imply that the order of x has the form p^r with $1 \leq r \leq n$. Pick $0 \neq x_1 \in M$ so that the order of x_1 is p^{n_1}, with n_1 maximal. Note that n_1 is uniquely determined by M. Let $\overline{M} = M/Rx_1$ denote the quotient module. Since M is finitely generated, $\overline{M}$ is finitely generated. Moreover, $p^nM = 0$ implies $p^n\overline{M} = 0$. We claim that $\dim_{R/(p)}(\mathrm{Soc}(\overline{M})) < \dim_{R/(p)}(\mathrm{Soc}(M))$. To show this suppose that $\overline{y}_1, \ldots, \overline{y}_m \in \mathrm{Soc}(\overline{M})$ is a basis for $\mathrm{Soc}(\overline{M})$ over $R/(p)$. We claim that the $\overline{y}_i$ are weakly independent over R. For suppose that

$$a_1\overline{y}_1 + \cdots a_m\overline{y}_m = \overline{0} = Rx_1.$$

Then

$$(a_1 + (p))\overline{y}_1 + \cdots (a_m + (p))\overline{y}_m = \overline{0}$$

which implies that $a_i \in (p)$ for all i because the $\overline{y}_i$ are linearly independent over $R/(p)$. Therefore, for all i, $a_i = pk_i$ for some $k_i \in R$ and hence

$$a_i\overline{y}_i = k_ip\overline{y}_i = \overline{0}$$

for all i since $\overline{y}_i \in \mathrm{Soc}(\overline{M})$. This shows that the $\overline{y}_i$ are weakly independent over R, establishing our claim. Now $p^{n_1-1}x_1 \neq 0$ and clearly $p^{n_1-1}x_1 \in \mathrm{Soc}(M)$ so that we have an element $x' = p^{n_1-1}x_1 \in Rx_1 \cap \mathrm{Soc}(M)$. Now, lemma (3.6.8) implies that we can find representatives $y_i \in M$ for each $\overline{y}_i$ with the same order (respectively). That is $y_i \in \mathrm{Soc}(M)$ for each $1 \leq i \leq m$. Moreover, the same lemma (3.6.8) implies that the set $x', y_1, \ldots, y_m$ is weakly independent over R. We leave it as an exercise for the reader to show that this implies that $x', y_1, \ldots, y_m$ are linearly independent over $R/(p)$ and hence $\dim_{R/(p)}(\mathrm{Soc}(M)) \geq m+1$ so that $\dim_{R/(p)}(\mathrm{Soc}(\overline{M})) < \dim_{R/(p)}(\mathrm{Soc}(M))$ as claimed. Now, by induction, there exist unique natural numbers $n_2 \geq \cdots \geq n_k \geq 1$ with

$$\overline{M} \cong R/(p^{n_2}) \oplus \cdots \oplus R/(p^{n_k}).$$

For each $i = 2, \ldots, k$, we have $R/(p^{n_i}) \cong R\overline{x}_i$ for some $\overline{x}_i \in \overline{M}$. Therefore the order of $\overline{x}_i$ is p^{n_i} and the $\overline{x}_i$ are weakly independent by lemma (3.6.7). Therefore, by lemma (3.6.8), there are representatives $x_2, \ldots, x_k \in M$ with the same orders p^{n_i} (respectively) and $x_1, x_2, \ldots, x_k$ are weakly independent. Therefore lemma (3.6.7) implies that

$$R\{x_1, \ldots, x_k\} = Rx_1 \oplus \cdots \oplus Rx_k.$$

But given $x \in M$, we know $\overline{x} \in \overline{M}$ is a R-linear combination of $\overline{x}_2, \ldots, \overline{x}_k$. Say $\overline{x} = \sum_{j=2}^m r_j\overline{x}_j$. Then if $y = \sum_{j=2}^m r_jx_j$, it is easy to see that $x - y \in \ker(M \to \overline{M}) = Rx_1$ so that $x - y = r_1x_1$ for some $r_1 \in R$ and hence $x_1, \ldots, x_k$ generate M. Therefore

$$M = Rx_1 \oplus \cdots \oplus Rx_k.$$

Also, we know n_1 is unique and $n_1 \geq n_2$ by the maximality of n_1 so that

$$n_1 \geq n_2 \geq \cdots \geq n_k \geq 1.$$

Finally, since the order of x_i is p^{n_i}, we have $Rx_i \cong R/(p^{n_i})$ so that, finally,

$$M \cong R/(p^{n_1}) \oplus \cdots \oplus R/(p^{n_k}).$$

■

This brings us to our first *major* characterization of finitely generated modules over a PID R. The statement of those theorem may seem a bit overwhelming, but we have already done all of the work to prove it, so the proof will be a nice recap of the material above.

Theorem 3.6.10 *If M is a finitely generated module over a PID R, then there exist unique primes $p_1, \ldots, p_m \in \mathcal{P}$ and for each such prime p_i there exist unique natural numbers*

$$n_{i1} \geq n_{i2} \geq \cdots \geq n_{ik_i} \geq 1$$

and there is a unique integer $r \geq 0$ such that

$$M \cong M_f \oplus \left(\bigoplus_{i=1}^{m} \left(\bigoplus_{j=1}^{k_i} R/(p_i^{n_{ij}}) \right) \right)$$

where $M_f \leq M$ is free of rank r.

Proof. Given M, we know by theorem (3.5.5) that

$$M = M_f \oplus M_t$$

where M_f is a free module of rank r for some unique $r \geq 0$ and M_t is the torsion submodule of M. Moreover, lemma (3.6.2) implies that we have

$$M_t = \bigoplus_{p \in \mathcal{P}} M_t(p).$$

Since M is finitely generated, M_t is finitely generated by corollary (3.3.13). Therefore $M_t(p) = 0$ for all but a finite number of distinct primes $p_1, \ldots p_m \in \mathcal{P}$ so that

$$M_t = \bigoplus_{i=1}^{m} M_t(p_i).$$

This shows that

$$M = M_f \oplus \left(\bigoplus_{i=1}^{m} M_t(p_i) \right).$$

Now, for all $i = 1, \ldots, m$, we have $p^{l_i} M(p_i) = 0$ for some $l_i \in \mathbb{N}$ and hence, for each $1 \leq i \leq m$, we apply lemma (3.6.9) to conclude that there exist unique natural numbers $n_{i1} \geq n_{i2} \geq \cdots \geq n_{ik_i} \geq 1$ with

$$M_t(p_i) \cong R/(p_i^{n_{i1}}) \oplus \cdots \oplus R/(p_i^{n_{ik_i}}).$$

Putting all of this together, we then have that

$$M = M_f \oplus \left(\bigoplus_{i=1}^{m} \left(\bigoplus_{j=1}^{k_i} R/(p_i^{n_{ij}}) \right) \right)$$

as desired. ■

The uniqueness statement of the theorem implies that the prime powers $p_i^{n_{ij}}$ together with the rank r completely characterize M. We call the prime powers $p_i^{n_{ij}}$ the **elementary divisors of** M. Therefore theorem (3.6.10) states that a finitely generated module over a PID is completely determined by the rank of the free part $M_f \leq M$ and the elementary divisors of the torsion submodule M_t.

Next, we want to reassemble the p-power modules into another valuable decomposition for M. The key fact is contained in the following lemma.

Lemma 3.6.11 *Let R be a PID and let $p_1, \ldots, p_n \in \mathcal{P}$ be distinct primes. If $e_1, \ldots, e_n \in \mathbb{N}$ and $x = p_1^{e_1} \cdots p_n^{e_n}$ then*

$$R/(x) \cong R/(p_1^{e_1}) \oplus \cdots \oplus R/(p_n^{e_n}).$$

Proof. Let $\pi_i : R \to R/(p_i^{e_i})$ denote the quotient map and define $\varphi = \pi_1 \times \cdots \times \pi_n : R \to R/(p_1^{e_1}) \times \cdots R/(P_n^{e_n})$ by $\varphi(a) = (\pi_1(a), \ldots, \pi_n(a))$. Then φ is a surjective ring homomorphism with $\ker \varphi = (p_1^{e_1}) \cap \cdots \cap (p_n^{e_n})$. Therefore the first isomorphism theorem (3.2.10) implies that

$$R/(p_1^{e_1}) \cap \cdots \cap (p_n^{e_n}) \cong R/(p_1^{e_1}) \oplus \cdots \oplus R/(p_n^{e_n}).$$

Since the p_i are distinct primes, they are relatively prime in pairs so that $(p_1^{e_1}) \cap \cdots \cap (p_n^{e_n}) = (p_1^{e_1} \cdots p_n^{e_n})$ by lemma (3.4.5). Therefore $R/(x) \cong R/(p_1^{e_1}) \oplus \cdots \oplus R/(p_n^{e_n})$. ■

OK, if M is finitely generated and $p_1, \ldots, p_m \in \mathcal{P}$ are the primes present in the elementary divisors of M, we organize the exponents n_{ij} in the following (not necessarily square) array:

$$\begin{gathered} p_1 : n_{11} \geq \cdots \geq n_{1k_1} \\ p_2 : n_{21} \geq \cdots \geq n_{2k_2} \\ \vdots \\ p_m : n_{m1} \geq \cdots \geq n_{mk_m} \end{gathered}$$

Then for each $h = 1, \ldots, k = \max\{k_1, \ldots, k_m\}$, let

$$q_h = p_1^{n_{1h}} p_2^{n_{2h}} \cdots p_m^{n_{mh}}$$

where we put a 1 for any $p_i^{n_{ih}}$ with n_{ih} not in the array. Easily we see that for $2 \leq h \leq k$, we have $q_h | q_{h-1}$ and hence $(q_{h-1}) \leq (q_h)$. This proves all but the uniqueness statement in the follow theorem: our final goal!

Theorem 3.6.12 (Fundamental Theorem of Finitely Generated Modules over a PID) *Let M be a finitely generated module over a PID R. Then there is a unique integer $r \geq 0$ and a unique chain of non trivial ideals*

$$(q_1) \leq (q_2) \leq \cdots \leq (q_k)$$

in R with

$$M \cong M_f \oplus R/(q_1) \oplus R/(q_2) \oplus \cdots \oplus R/(q_k)$$

where M_f is a free R module of rank r.

Proof. Given M, we use theorem (3.6.10) to write

$$M \cong M_f \oplus \left(\bigoplus_{i=1}^{m} \left(\bigoplus_{j=1}^{k_i} R/(p_i^{n_{ij}}) \right) \right)$$

where $r \geq 0$ is unique, $p_1, \ldots, p_m \in \mathcal{P}$ are distinct and for each such prime p_i there exist unique natural numbers

$$n_{i1} \geq n_{i2} \geq \cdots \geq n_{ik_i} \geq 1.$$

Using the notation given before the statement of the theorem, lemma (3.6.11) implies that

$$R/(q_h) \cong R/(p_1^{n_{1h}}) \oplus R/(p_2^{n_{2h}}) \oplus \cdots \oplus R/(p_m^{n_{mh}}).$$

Therefore we have that

$$\bigoplus_{i=1}^{m} \left(\bigoplus_{j=1}^{k_i} R/(p_i^{n_{ij}}) \right) = \bigoplus_{j=1}^{\max\{k_i\}} \left(\bigoplus_{i=1}^{m} R/(p_i^{n_{ij}}) \right) \cong \bigoplus_{j=1}^{k_i} R/(q_j).$$

We have already remarked that $(q_{h-1}) \leq (q_h)$ for $2 \leq h \leq k = \max\{k_i\}$. This shows that

$$M \cong M_f \oplus R/(q_1) \oplus R/(q_2) \oplus \cdots \oplus R/(q_k)$$

with $(q_1) \leq (q_2) \leq \cdots \leq (q_k)$ and each (q_h) non trivial.

It remains to establish the uniqueness. For this, we may assume that M is a torsion module. We induct on the number of primes that appear as factors in the decomposition (3.6.2). If there is only one prime p, then $M = M(p)$ and hence the uniqueness is settled by lemma (3.6.9). If there is more than one prime, with say p one of them, then by lemma (3.6.2) we have $M = M(p) \oplus N$ where N is the sum of the non zero $M(q)$ appearing in decomposition of M. Now, N is a finitely generated torsion module with one less prime in its factorization so that, by induction, there are unique ideals

$$(q_1') \leq (q_2') \leq \cdots \leq (q_k')$$

in N with

$$N \cong R/(q_1') \oplus R/(q_2) \oplus \cdots \oplus R/(q_k').$$

Moreover, by lemma (3.6.9), there are unique natural numbers $n_1 \geq \cdots \geq n_h \geq 1$ such that

$$M(p) \cong R/(p^{n_1}) \oplus R/(p^{n_2}) \oplus \cdots \oplus R/(p^{n_h}).$$

Then for M, it is clear that we have $q_1 = q_1' p^{n_1}$, $q_2 = q_2' p^{n_2}$ and so on. ■

The elements $q_i \in R$ are called the **invariant factors for** M.

Chapter 4

Two Applications of the Structure Theorem

4.1 Structure theorem for (finitely generated) abelian groups

Our first application of the structure theorem (3.6.12) will be the case $R = \mathbb{Z}$. Here, a finitely, generated $\mathbb{Z}$-module is just a finitely generated abelian group.

Theorem 4.1.1 *If A is a finitely generated abelian group, then there is a unique integer $r \geq 0$ and a unique (possibly empty) set of integers $m_1, \dots, m_k$ with $m_j > 1$ for all j and $m_k | m_{k-1} | \cdots | m_1$ such that*

$$A \cong \mathbb{Z}^r \oplus \mathbb{Z}_{m_1} \oplus \mathbb{Z}_{m_2} \oplus \cdots \oplus \mathbb{Z}_{m_k}.$$

Proof. Since $\mathbb{Z}$ is PID and a finitely generated abelian group A is a finitely generated $\mathbb{Z}$-module, the structure theorem (3.6.12) applies and hence there is a unique integer $r \geq 0$ and a unique chain of non trivial ideals

$$(m_1) \leq (m_2) \leq \cdots \leq (m_k)$$

in $\mathbb{Z}$ with

$$A \cong \mathbb{Z}^r \oplus \mathbb{Z}/(m_1) \oplus \mathbb{Z}/(m_2) \oplus \cdots \oplus \mathbb{Z}/(m_k).$$

An ideal (m_j) in $\mathbb{Z}$ is non-trivial if and only if the generator $m_j > 1$ and $(m_j) \leq (m_{j+1})$ if and only if $m_{j+1} | m_j$. ■

Note that if, in addition, the group A is finite, then necessarily $r = 0$ and hence A is a finite direct sum of cyclic groups, and the cyclic factors are uniquely determined (up to order) by the invariant factors m_j. Moreover, the product $m = m_1 \cdots m_k$ of the invariant factors is the order of A. We illustrate these ideas with an example.

Example 4.1.2 Find all abelian groups of order 24 up to isomorphism. There are exactly three possibilities for the invariant factors: $m_1 = 24$, $m_1 = 12, m_2 = 2$ and $m_1 = 6, m_2 = 2, m_3 = 2$. Therefore there are exactly three abelian groups of order 24 up to isomorphism. They are, respectively, $\mathbb{Z}_{24}$, $\mathbb{Z}_{12} \oplus \mathbb{Z}_2$ and $\mathbb{Z}_6 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$.

In practice, if the order of A is a large number, it can be difficult to determine the invariant factors m_j for A. However, we do have the alternate form of the structure theorem (3.6.10) which decomposes A using the elementary divisors. In practice, it is easy to determine all of the elementary divisors of A (in fact, if you look back at the proof of theorem (3.6.12), we used the elementary divisors to find the invariant factors!). We will illustrate this by example as well, but first we restate theorem (3.6.10) for finitely generated $\mathbb{Z}$-modules.

Theorem 4.1.3 *If A is a finitely generated abelian group, then there is a unique integer $r \geq 0$ and n (not necessarily distinct) prime integers $p_1, \ldots, p_n$ such that*

$$A \cong \mathbb{Z}^r \oplus \mathbb{Z}_{p_1^{e_1}} \oplus \cdots \oplus \mathbb{Z}_{p_n^{e_n}}.$$

The prime powers $p_i^{e_i}$ are uniquely determined by A and this direct sum decomposition is unique up to the order of the factors.

Proof. By theorem (3.6.10), there exist a unique $r \geq 0$ and unique (positive) primes $p_1, \ldots, p_m \in \mathbb{Z}$ such that for each such prime p_i there exist unique natural numbers

$$n_{i1} \geq n_{i2} \geq \cdots \geq n_{ik_i} \geq 1$$

such that

$$A \cong \mathbb{Z}^r \oplus \left(\bigoplus_{i=1}^{m} \left(\bigoplus_{j=1}^{k_i} \mathbb{Z}_{p_i^{n_{ij}}} \right) \right).$$

If we let $n = \sum_{i=1}^{m} k_i$, then

$$\bigoplus_{i=1}^{m} \left(\bigoplus_{j=1}^{k_i} \mathbb{Z}_{p_i^{n_{ij}}} \right)$$

is a sum of n cyclic groups of prime power order. The uniqueness follows from the uniqueness of the n_{ij}. ■

Since the product of the elementary divisors of a finite group must be the order of the group, we can quickly determine all possible elementary divisors if a finite abelian group A by looking at the various ways to form elementary divisors from the prime factorization of $|A|$. Here is an example.

Example 4.1.4 Determine all abelian groups of order 24 up to isomorphism. First note that $24 = 2^3 \cdot 3$. Therefore the only possible collections of elementary divisors are $\{2^3, 3\}, \{2^2, 2, 3\}$ and $\{2, 2, 2, 3\}$. The corresponding abelian groups are

$$\mathbb{Z}_8 \oplus \mathbb{Z}_3,\ \mathbb{Z}_4 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_3,\ \text{and } \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_3 \text{ respectively.}$$

Theorem (3.6.10) implies that any abelian group of order 24 is isomorphic to one of these 3 groups, and no two of these three are isomorphic. The reader should compare this example with the invariant factor decomposition given earlier. When reconciling the two decompositions, it will be useful to recall that for two integers a, b, $\mathbb{Z}_a \oplus \mathbb{Z}_b \cong \mathbb{Z}_{ab}$ iff. a and b are relatively prime.

Example 4.1.5 Determine all abelian groups of order 1500 up to isomorphism. First note that $1500 = 2^2 \cdot 3 \cdot 5^3$. Therefore the only possible collections of elementary divisors are $\{2^2, 3, 5^3\}$, $\{2, 2, 3, 5^3\}$, $\{2^2, 3, 5^2, 5\}$, $\{2, 2, 3, 5^2, 5\}$, $\{2^2, 3, 5, 5, 5\}$ and $\{2, 2, 3, 5, 5, 5\}$. Each of these families determines a group of order 1500. For example, $\{2, 2, 3, 5^3\}$ determines

$$\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_{125}.$$

Theorem (3.6.10) implies that any abelian group of order 1500 is isomorphic to one of these 6 groups, and no two of these 6 are isomorphic.

Example 4.1.6 From the proof of theorem (3.6.12), we can determine the possible invariant factors for an abelian group of order 1500 from the elementary divisors. For example, for the elementary divisors $2^2, 3, 5, 5, 5$, we form the array of exponents

$$\begin{array}{l} 2 : 2 \\ 3 : 1 \\ 5 : 1 \geq 1 \geq 1 \end{array}$$

so that the elementary divisors are $m_1 = 2^2 \cdot 3 \cdot 5, m_2 = 5$ and $m_3 = 5$. Using invariant factors, the corresponding abelian group of order 1500 is

$$\mathbb{Z}_{60} \oplus \mathbb{Z}_5 \oplus \mathbb{Z}_5$$

and $5|5|60$. Using the elementary divisors, the corresponding abelian group of order 1500 is

$$\mathbb{Z}_4 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_5 \oplus \mathbb{Z}_5 \oplus \mathbb{Z}_5.$$

But $4, 3, 5$ are relatively prime in pairs so that $\mathbb{Z}_4 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_5 \cong \mathbb{Z}_{60}$ and the decomposition is the same.

4.2 The Jordan canonical form

Our second application of the structure theorem (3.6.12) concerns a linear operator $T : V \to V$ on a finite dimensional vector space. At first, it is not obvious how the structure theorem applies to this situation at all. The next proposition will "tip our hand".

Proposition 4.2.1 *Let V be a finite dimensional vector space over a field $\mathbb{F}$ and let $T : V \to V$ be a linear operator on V. Then*

(1) V is an $\mathbb{F}[t]$-module with the operation $\mathbb{F}[t] \times V \to V$ given by

$$(f(t), v) \mapsto f(t)v = f(T)v.$$

(2) The restriction of the $\mathbb{F}[t]$ action on V to $\mathbb{F}$ is just the given vector space structure on V and hence V is finitely generated as a $\mathbb{F}[t]$-module.

(3) $tv = T(v)$ for all $v \in V$.

(4) An additive subgroup $W \leq V$ is a $\mathbb{F}[t]$-submodule if and only if W is a T-invariant $\mathbb{F}$-subspace of V.

Proof. (1) Exercise.

(2) The restriction to $\mathbb{F}$ is the scalar multiplication in V over $\mathbb{F}$ by definition. Now, $\dim_{\mathbb{F}}(V)$ is finite so that every element $v \in V$ is a $\mathbb{F}$-linear combination of some finite set of vectors $\{v_1, \ldots, v_n\} \subset V$. But every such combination is a $\mathbb{F}[t]$-linear combination as well so that V is finitely generated as a $\mathbb{F}[t]$-module.

(3) The definition.

(4) If $W \leq V$ is an $\mathbb{F}[t]$-submodule, then W is stable under $\mathbb{F}$ (constant polynomials) as well as t. By the definition of the $\mathbb{F}[t]$ action on V, this implies that W is a $\mathbb{F}$-subspace of V and $T(W) \subseteq W$.

Therefore W is a T-invariant $\mathbb{F}$-subspace of V. Conversely, if W is a T-invariant $\mathbb{F}$-subspace of V, then W is stable under $\mathbb{F}$ and t. It follows that W is invariant under any polynomial $f(t) \in \mathbb{F}[t]$ and hence W is a $\mathbb{F}[t]$-submodule of V. ■

Now, the polynomial ring $\mathbb{F}[t]$ is a PID, and the proposition states (among other things) that given a linear operator $T : V \to V$, we have a finitely generated $\mathbb{F}[t]$-module V, and hence the structure theorem (3.6.12) applies. Namely, we have the following interpretation of (3.6.12).

Theorem 4.2.2 *If V is a finite dimensional vector space over a field $\mathbb{F}$ and $T : V \to V$ is a linear operator on V, then there exist n (not necessarily distinct) irreducible monic polynomials $p_1, \ldots, p_n \in \mathbb{F}[t]$ such that*

$$_{\mathbb{F}[t]}V \cong \mathbb{F}[t]/(p_1^{e_1}) \oplus \cdots \oplus \mathbb{F}[t]/(p_n^{e_n})$$

as $\mathbb{F}[t]$-modules. The prime powers $p_i^{e_1}$ are uniquely determined by T and V.

Proof. We have remarked that the structure theorem (3.6.12) applies to the $\mathbb{F}[t]$-module V and so we have a decomposition of V into a direct sum of a free $\mathbb{F}[t]$-module and cyclic submodules of prime power order with the appropriate uniqueness statement. However, since $\dim_{\mathbb{F}}(V) < \infty$, the free part in the decomposition is necessarily 0 ($\mathbb{F}[t]$ is infinite dimensional as a $\mathbb{F}$-module). ■

Let $\psi : V \to \mathbb{F}[t]/(p_1^{e_1}) \oplus \cdots \oplus \mathbb{F}[t]/(p_n^{e_n})$ be the $\mathbb{F}[t]$-module isomorphism of theorem (4.2.2) and let

$$W_i = \psi^{-1}(\mathbb{F}[t]/(p_i^{e_i}))$$

for $1 \leq i \leq n$. Then proposition (4.2.1)(4) implies that W_i is a $\mathbb{F}$-subspace of V, necessarily finite dimensional. Moreover, since the W_i are independent as $\mathbb{F}[t]$-modules, they are independent as $\mathbb{F}$-subspaces as well. It follows that if we choose a basis $\mathcal{B}_i$ for W_i as an $\mathbb{F}$-subspace for all i, then $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_n)$ is a basis for V. That is,

$$_{\mathbb{F}}V = W_1 \oplus \cdots \oplus W_n$$

as $\mathbb{F}$-vector spaces. Moreover, proposition (4.2.1)(4) also implies that W_i is T-invariant for all i so that the matrix of T with respect to the basis $\mathcal{B}$ has block form.

Now, fix an i between 1 and n and for notational brevity let $W = W_i$ and

$$p_i^{e_i} = f = t^k + a_{k-1}t^{k-1} + \cdots + a_1 t + a_0.$$

Note that $1 + (f)$ generates $\mathbb{F}[t]/(f) \cong W$ as an $\mathbb{F}[t]$-module. Let $w_0 = \psi^{-1}(1 + (f)) \in W$ and let $w_i = T^i(w_0)$ for $i \geq 1$. Then $w_i \in W$ for all i and $\psi(w_i) = t^i + (f)$ for all i, the latter since ψ is an $\mathbb{F}[t]$-module homomorphism and $T^i(w_0) = t^i w_0$ by definition.

Now, given $w \in W$, we know $\psi(w) = g(t) + (f)$ for some $g \in \mathbb{F}[t]$ and since $\deg f = k$, we may assume that $\deg g \leq k-1$. Say

$$g(t) = b_{k-1}t^{k-1} + \cdots + b_1 t + b_0$$

with $b_j \in \mathbb{F}$. Therefore we have

$$w = \psi^{-1}(g(t) + (f)) = \sum_{j=0}^{k-1} b_j \psi^{-1}(t^j + (f)) = \sum_{j=0}^{k-1} b_j w_j$$

and hence the set $\{w_0, \ldots, w_{k-1}\}$ spans W as an $\mathbb{F}$-module. Moreover, if

$$b_0 w_0 + \cdots + b_{k-1} w_{k-1} = 0$$

for some scalars $b_j \in \mathbb{F}$. then $g(t) + (f) = (f)$ where

$$g(t) = b_0 + \cdots + b_{k-1}t^{k-1} \in \mathbb{F}[t].$$

But $\deg g(t) \leq k-1$ so that $g(t) \in (f)$ iff. $g = 0$ and hence $b_j = 0$ for all $1 \leq j \leq k-1$. This shows that $\{w_0, \ldots, w_{k-1}\}$ is a basis for W as an $\mathbb{F}$-module. We note that

$$\dim_{\mathbb{F}}(W) = k = \deg f = \deg p_i^{e_i}.$$

Note that by construction,

$$w_{i+1} = T(w_i)$$

for $0 \leq i \leq k-1$ and

$$\psi(w_k) = t^k + (f) = -\left(\sum_{j=0}^{k-1} a_j t^j\right) + (f)$$

and hence

$$w_k = -\left(\sum_{j=0}^{k-1} a_j w_j\right).$$

It follows that the matrix of $T|_W$ with respect to the basis $(w_0, \ldots, w_{k-1})$ for W has the form

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{k-1} \end{bmatrix}. \tag{4.1}$$

We have essentially proven the following theorem.

Theorem 4.2.3 *Let V be a finite dimensional vector space over a field $\mathbb{F}$ and let $T : V \to V$ be a linear operator on V. Then there is a basis $\mathcal{B}$ for V in which the matrix of T is in block form where each of the blocks has the form (4.1). Moreover, the scalars in the non-trivial column of each such block are uniquely determined by T.*

Proof. Given $T : V \to V$, we let W_i be as above so that $\mathcal{B}_i = (w_{i_0}, \ldots, w_{i_{k_i-1}})$ is a basis for W_i where $k_i = \deg p_i^{e_i}$. On W_i, the restriction of T has the form (4.1) and the non-trivial column consists of the non-leading coefficients of the (monic) polynomial $p_i^{e_i}$, and are therefore uniquely determined by theorem (4.2.2). Finally, we have seen that $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_n)$ is a basis for V, and the matrix for T with respect to $\mathcal{B}$ is in block form, with the ith block being the matrix of $T|_{W_i}$. This completes the proof. ■

When we write T in the above basis, the operator is said to be in **rational canonical form**. It is the best form available for an arbitrary operator over a $\mathbb{F}$-vector space when $\mathbb{F}$ is any field. If $\mathbb{F} = \mathbb{C}$ is the field of complex numbers, we can do better than the rational canonical form.

If $\mathbb{F} = \mathbb{C}$, then every irreducible polynomial has the form $t - a$ for some $a \in \mathbb{C}$. Therefore, using the notation above, we have $p_i^{e_i} = (t - a_i)^{e_i}$ so that

$$W_i \cong \mathbb{C}[t]/((t - a_i)^{e_i})$$

and $\dim_{\mathbb{F}}(W_i) = e_i$.

Again we fix an i between 1 and n and write $W = W_i, a = a_i, f = (t - a_i)^{e_i}$ and $e_i = k$ for brevity. We again define $w_0 = \psi^{-1}(1 + (f))$, but this time we set

$$w_i = (T - a \cdot I)^i w_0$$

for $i \geq 1$. Note that $\psi(w_i) = (t - a)^i + (f)$ because ψ is a $\mathbb{C}[t]$-module homomorphism. Moreover, $\psi(w_i) = 0$ iff. $w_i = 0$ since ψ is an isomorphism. We have

$$w_1 = (T - a \cdot I)w_0, \quad \ldots, \quad w_{k-1=}(T - a \cdot I)w_{k-2}, \quad w_k = (T - a \cdot I)w_{k-1} = (T - a \cdot I)^k w_0 = 0$$

where $w_k = 0$ since $\psi(w_k) = (f)$. Solving each equation gives

$$Tw_0 = w_1 + aw_0, \quad \ldots, \quad Tw_{k-2} = w_{k-1} + aw_{k-2}, \quad Tw_{k-1} = aw_{k-1}.$$

We leave it as an exercise to show that $(w_0, \ldots, w_{k-1})$ is a $\mathbb{C}$-linearly independent set. Since $\dim_{\mathbb{C}}(W) = k$, $(w_0, \ldots, w_{k-1})$ is a basis for W and the above computation shows that the matrix for the restriction of T to W in this basis is

$$\begin{bmatrix} a & 0 & \cdots & 0 & 0 \\ 1 & a & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & a \end{bmatrix}. \tag{4.2}$$

These matrices are called **Jordan blocks**. The previous remarks constitute a proof of the following theorem. We will write out the proof to tidy up.

Theorem 4.2.4 *Let V be a finite dimensional complex vector space and $T : V \to V$ an operator on V, then there is a basis $\mathcal{B}$ for V in which the matrix for T is block form with Jordan blocks (4.2).*

Proof. Since every irreducible polynomial over $\mathbb{C}$ is linear, theorem (4.2.2) implies that

$$V = W_1 \oplus \cdots \oplus W_n$$

where each W_i is isomorphic to $\mathbb{C}[t]/((t-a_i)^{e_i})$. For each i, we construct the basis for W_i as above. Then the matrix of T is block form and the blocks have the form (4.2). ■

The matrix of T in the basis of the theorem is referred to as the **Jordan form of** T. It is uniquely determined by T up to a permutation of the basis vectors since the elementary divisors of the $\mathbb{C}[t]$-module defined by T are uniquely determined by T.

Note that the Jordan form of T is lower triangular, and hence the diagonal entries, that is those complex numbers $a_i \in \mathbb{C}$ appearing in the decomposition of V are the eigenvalues of T. The number of appearances of any given eigenvalue a for T is exactly the multiplicity of the root a for the characteristic polynomial p of T.

Let J be a Jordan block with diagonal entry $a \in \mathbb{C}$. Clearly a is the only eigenvalue for J and the eigenspace for a is 1 dimensional ($(aI - T)$ has nullity 1.) Conversely, if an operator has all eigenvectors scalar multiples of some fixed vector, then it can only have one eigenvalue (else independent eigenvectors). Moreover, its Jordan form cannot have more than one Jordan block (else it has at least 2 eigenvectors in a basis).

It follows that for an operator T, the number of Jordan blocks with diagonal entry a in the Jordan form of T is exactly the dimension of the eigenspace for the eigenvalue $a \in \mathbb{C}$. This dimension is an integer between 1 and the multiplicity of a in the characteristic polynomial p of T. If these multiplicities are not too big, we can use this to find the Jordan form. Here is an example.

Example 4.2.5 Determine the Jordan form of the matrix

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

The characteristic polynomial is $(t-1)^3$, and by inspection the matrix $I_3 - A$ has rank 1. Therefore the eigenspace is 2 dimensional so that there are two Jordan blocks in the Jordan form. Necessarily the Jordan form is

$$J = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

If a is a k-fold root in the characteristic polynomial of T, the number of Jordan blocks in the Jordan form with diagonal entry a is an integer between 1 and k. Here are the possibilities for small k (we omit 0 entries).

$$k = 1 : [a];$$

$$k = 2 : \begin{bmatrix} a & \\ 1 & a \end{bmatrix}, \begin{bmatrix} a & \\ & a \end{bmatrix};$$

$$k = 3 : \begin{bmatrix} a & & \\ 1 & a & \\ & 1 & a \end{bmatrix}, \begin{bmatrix} a & & \\ 1 & a & \\ & & a \end{bmatrix}, \begin{bmatrix} a & & \\ & a & \\ & & a \end{bmatrix};$$

$$k = 4 : \begin{bmatrix} a & & & \\ 1 & a & & \\ & 1 & a & \\ & & 1 & a \end{bmatrix}, \begin{bmatrix} a & & & \\ 1 & a & & \\ & 1 & a & \\ & & & a \end{bmatrix}, \begin{bmatrix} a & & & \\ 1 & a & & \\ & & a & \\ & & 1 & a \end{bmatrix}, \begin{bmatrix} a & & & \\ 1 & a & & \\ & & a & \\ & & & a \end{bmatrix}, \begin{bmatrix} a & & & \\ & a & & \\ & & a & \\ & & & a \end{bmatrix}.$$

Note that for $k \leq 3$, the number of Jordan blocks with diagonal entry a, (i.e. the dimension of the null space of $T - a \cdot 1$) is completely determines the portion of the Jordan form with diagonal entry a. When $k = 4$ however, there are two possibilities for two 2 Jordan blocks. It can be shown that the operator $(T - a \cdot 1)^2$ distinguishes the cases. We won't worry about the details.

Example 4.2.6 What is the Jordan form of a matrix whose characteristic polynomial is

$$(t-3)^3(t-7)^4$$

and such that the space of eigenvectors for the eigenvalue 3 is 2 dimensional and the space of eigenvectors for the eigenvalue 7 is 3 dimensional.

The information given means that for $a = 3$, we have 2 Jordan blocks in a 3×3 matrix and for $a = 7$ we have 3 Jordan blocks in a 4×4 matrix. The only possibility is

$$\begin{bmatrix} 2 & & & & & & \\ 1 & 2 & & & & & \\ & & 2 & & & & \\ & & & 7 & & & \\ & & & 1 & 7 & & \\ & & & & & 7 & \\ & & & & & & 7 \end{bmatrix}.$$

Index