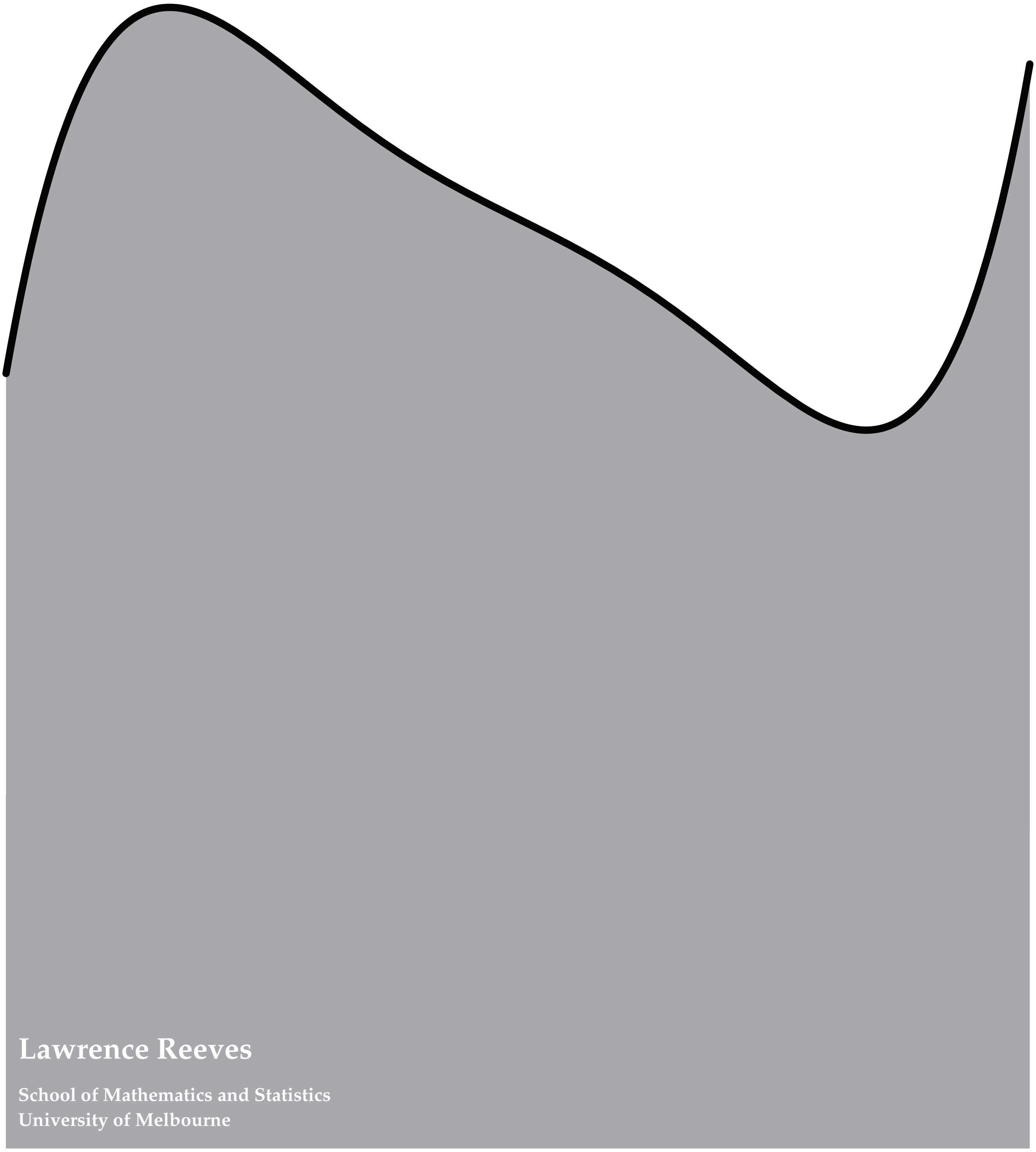
Lecture Notes on

Rings, Modules and Fields



Lawrence Reeves

School of Mathematics and Statistics
University of Melbourne

Version 18, 2016

The curve on the cover is a plot of the quintic $X^5 - 6X^3 - 27X - 3 \in \mathbb{R}[X]$, which is not solvable by radicals.

Contents

Preface

We assume some familiarity with the notion of a group (in particular cosets and quotients), modular arithmetic, and the Euclidean algorithm for integers. The book [Hun96] is very good reference for such things.

We do not, at least initially, insist that multiplication in a ring should be commutative. For much of the time we shall be considering properties of polynomial rings and their quotients (in which case multiplication is commutative). There are exercises embedded in the text as well as at the end of each section. Some introduce important concepts that are used subsequently.

Some notation

This is just a short list, given here for reference, of *some* of the notation used.

C_n the cyclic group of order n

$[E : F]$ the degree of E over F, where $E \supseteq F$ are fields

$\mathbb{F}_q$ the field with q elements

$F[a]$ the smallest subring of E that contains a and F, where $a \in E \supseteq F$

$F(a)$ the smallest subfield of E that contains a and F, where $a \in E \supseteq F$

$\mathrm{Gal}(E/F)$ the Galois group of E over F, where $E \supseteq F$

$\mathrm{irr}(a, F)$ the irreducible polynomial of a over F

$M_n(R)$ the ring of $n \times n$ matrices over R

${}_RM$ an R-module

$\mathbb{N}$ the natural numbers: $\{0, 1, 2, 3, \dots\}$

$\mathbb{N}^+$ the positive natural numbers: $\{1, 2, 3, \dots\}$

$R^\times$ the group of units in a unital ring R

$R[X]$ the ring of polynomials with coefficients from R

S_n the symmetric group on n elements

$\mathbb{Z}[i]$ the Gaussian integers: $\{x + iy \mid x, y \in \mathbb{Z}\} \subset \mathbb{C}$

$\mathbb{Z}/n\mathbb{Z}$ the ring of integers modulo n. (Sometimes also denoted $\mathbb{Z}_n$ or $\mathbb{Z}/n$)

Chapter I

Rings

Is this chapter we investigate the structure of rings: general algebraic structures in which there are two operations defined. A good example to keep in mind is the ring of polynomials $\mathbb{R}[X]$. After some general considerations, such as subrings and quotients, we'll look at particular properties that a ring may (or may not) possess. For example, we know that in the integers every element can be written as a product of primes. This turns out to not be true in every ring, but is true of $\mathbb{R}[X]$. We consider various classes of rings: integral domains, unique factorisation domains, principal ideal domains and Euclidean domains.

1 Fundamental definitions

1.1 Definition of a ring

Many familiar mathematical structures consist of a set on which two binary operations can be performed. You probably recognise all the following examples:

Examples 1.1.

Number systems:	$\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$
Polynomials:	$\mathbb{R}[X] = \{a_0 + a_1X + \cdots + a_nX^n : n \in \mathbb{N}, a_i \in \mathbb{R}\}$
Integers modulo n:	$\mathbb{Z}/n\mathbb{Z}$ (also denoted by $\mathbb{Z}_n$ or $\mathbb{Z}/n$ or $\mathbb{Z}/\langle n\rangle$)
Square matrices:	$M_n(\mathbb{R})$

All have the property that there are two binary operations, 'addition' and 'multiplication' and that the two obey some modest and natural conditions such as the distributive law. Writing down a list of their common properties leads us to the following:

Definition 1.2. A **ring** is a (non-empty) set R together with two binary operations $+$ and $\times$ satisfying the following conditions:

1. $(R, +)$ forms an abelian group
2. The operation $\times$ is associative: $\forall\, x, y, z \in R$, we have $x \times (y \times z) = (x \times y) \times z$
3. The left and right distributive laws hold: $\forall\, x, y, z \in R,\ \ x \times (y + z) = (x \times y) + (x \times z)$
   $\forall\, x, y, z \in R,\ \ (y + z) \times x = (y \times x) + (z \times x)$

The ring will be denoted $(R, +, \times)$ or simply R if the operations are clear from the context.

The operations $+$ and $\times$ will be referred to as addition and multiplication respectively. Multiplication will often be represented by concatenation, so we write ab in place of $a \times b$. The (unique) element of R that is the identity in the abelian group $(R, +)$ will be denoted by 0. The additive inverse of an element a is denoted $-a$.

Exercise 1. Suppose that R is a ring and $e_1, e_2 \in R$ each satisfy

$$\forall a \in R,\ e_i a = a = a e_i \tag{$*$}$$

Show that $e_1 = e_2$.

Definition 1.3. An element satisfying $(*)$ is called a **multiplicative identity** and is denoted by 1_R or simply 1.

Exercise 2. Let R be a ring, and $a, b \in R$ any two elements. Show that

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

Justify every step using the axioms from the definition of a ring.

Exercise 3. Let R be a ring in which there is a multiplicative identity. Show that if $1 = 0$ (i.e., the additive and multiplicative identities coincide), then R consists of a single element.

Remember that we are not insisting that multiplication should be commutative.

Definition 1.4. A ring $(R, +, \times)$ is said to be **commutative** if multiplication is commutative. (Addition is always commutative according to the definition.) It is **unital** if there is an identity for multiplication and $1 \neq 0$.

Remark. Some authors insist, as part of the definition, that a ring should have a multiplicative identity. They would then call a ring that has no multiplicative identity a 'pseudo-ring'.

Examples 1.5 (Some rings)**.**

1. Let X be any set and denote by $\mathcal{P}(X)$ the power set of X. Define operations on $\mathcal{P}(X)$ by

$$\begin{aligned} A + B &= (A \cup B) \setminus (A \cap B) \\ A \times B &= A \cap B \end{aligned}$$

   Then $(\mathcal{P}(X), +, \times)$ is a commutative, unital ring.

2. Let R be the set of all functions from $\mathbb{R}$ to $\mathbb{R}$. Defining operations in the usual pointwise way,

$$\begin{aligned} (f \times g)(x) &= f(x)g(x) \\ (f + g)(x) &= f(x) + g(x) \end{aligned}$$

   makes R into a commutative, unital ring.

3. Consider the following subset $\mathcal{H}$ of $M_2(\mathbb{C})$:

$$\mathcal{H} = \left\{ \begin{bmatrix} a & b \\ -\bar{b} & \bar{a} \end{bmatrix} \mid a, b \in \mathbb{C} \right\}$$

   With the usual matrix operations, $\mathcal{H}$ forms a (non-commutative) unital ring, called the **quaternions**.

4. The subset of the complex numbers given by $\mathbb{Z}[i] := \{m + in :\ m, n \in \mathbb{Z}\}$ with the operations from $\mathbb{C}$ forms a commutative unital ring. It is called the **Gaussian integers**.

5. Consider the set $(\mathbb{Z}/6\mathbb{Z})[\sqrt{5}] := \{a + b\sqrt{5} : \ a, b \in \mathbb{Z}/6\mathbb{Z}\}$. The operations

$$(a + b\sqrt{5}) + (\alpha + \beta\sqrt{5}) = (a + \alpha) + (b + \beta)\sqrt{5}$$
$$(a + b\sqrt{5}) \times (\alpha + \beta\sqrt{5}) = (a\alpha + 5b\beta) + (a\beta + b\alpha)\sqrt{5}$$

make R into a commutative, unital ring.

6. Let $R = \{0, 2, 4\} \subset \mathbb{Z}/6\mathbb{Z}$. With the operations coming from $\mathbb{Z}/6\mathbb{Z}$, R forms a commutative unital ring. What is the multiplicative identity ?

7. Let $R = \{0, 2, 4, 6\} \subset \mathbb{Z}/8\mathbb{Z}$. With the operations coming from $\mathbb{Z}/8\mathbb{Z}$, R forms a commutative but not unital ring.

1.2 Units and zero-divisors

Let R be a unital ring. An element $a \in R$ is a **unit** if there exists $b \in R$ such that $ab = ba = 1$. The element b is called the **multiplicative inverse** of a and is denoted a^{-1}. The set of units, together with the operation of multiplication, forms a group called the **group of units**. We denote it $R^{\times}$. The ring R is called a **division ring** if every non-zero element is a unit. We know from Exercise 2(a) that the zero element is never a unit. A **field** is a commutative division ring.

Examples 1.6.

1. $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are all fields.

2. For any prime integer p, $\mathbb{Z}/p\mathbb{Z}$ is a field. We will use the notation $\mathbb{F}_p$ to denote the field $\mathbb{Z}/p\mathbb{Z}$. We will see later that any field having p elements is isomorphic to $\mathbb{F}_p$, and that there are other finite fields. Finite fields are used extensively in cryptography and coding theory.

3. The following addition and multiplication tables define a field having four elements. It is not isomorphic to the ring $\mathbb{Z}/4\mathbb{Z}$ (which is not even a field).

$+$	$\mathbf{0}$	$\mathbf{1}$	$\boldsymbol{x}$	$\boldsymbol{y}$
$\mathbf{0}$	0	1	x	y
$\mathbf{1}$	1	0	y	x
$\boldsymbol{x}$	x	y	0	1
$\boldsymbol{y}$	y	x	1	0

$\times$	$\mathbf{0}$	$\mathbf{1}$	$\boldsymbol{x}$	$\boldsymbol{y}$
$\mathbf{0}$	0	0	0	0
$\mathbf{1}$	0	1	x	y
$\boldsymbol{x}$	0	x	y	1
$\boldsymbol{y}$	0	y	1	x

This field is isomorphic to that given in example 1.14. We will see later that this field is also isomorphic to a quotient ring of a polynomial ring, namely $\mathbb{F}_2[X]/\langle 1 + X + X^2 \rangle$. The notion of a quotient ring will be discussed shorty, but essentially this is the the ring of all polynomials with coefficients from $\mathbb{F}_2$, modulo the condition that $1 + X + X^2 = 0$.

4. The only units in $\mathbb{Z}$ are $1, -1$. (Therefore $\mathbb{Z}$ is not a field.)

5. $\mathbb{Z}/6\mathbb{Z}$ is not a field. The only units are 1 and 5.

6. The units in $\mathbb{R}[X]$ are the non-zero constant polynomials.

7. The ring of quaternions $\mathcal{H}$ is a division ring, but is not a field.

Definition 1.7. If $a, b \in R$ are non-zero elements in a ring R satisfying $ab = 0$ then they are called **zero-divisors**.

Remark. This is not quite the same as being a 'divisor of zero.' According to exercise 2(a), everything divides zero.

Example 1.8. In $\mathbb{Z}/6\mathbb{Z}$ the zero-divisors are 2, 3, 4. There are no zero-divisors in $\mathbb{R}$, $\mathbb{Z}$ or $\mathbb{R}[X]$.

Exercise 4. Show that a unit cannot be a zero-divisor.

Lemma 1.9 (Cancellation Law)**.** *Let R be a ring. Then R has no zero-divisors if and only if the following condition holds for all $a, b, c \in R$ with $a \neq 0$*

$$\begin{aligned} ab = ac &\implies b = c \\ ba = ca &\implies b = c \end{aligned}$$

Remark. We are not assuming that a is a unit, merely that it is non-zero.

Proof (of Lemma 1.9). First note that $ab = ac \iff ab - ac = 0 \iff a(b - c) = 0$. Suppose there are no zero-divisors. Then $ab = ac \implies a(b - c) = 0 \implies a = 0$ or $b - c = 0$. If $a \neq 0$, we therefore have $ab = bc \implies b = c$. The second condition follows in exactly the same way.

Now suppose that both conditions hold. If $a \neq 0$ and $ab = 0$, then $ab = a0 \implies b = 0$. Similarly if $ba = 0$. □

1.3 Integral domains and fields

Definition 1.10. An **integral domain** is a commutative unital ring in which there are no zero-divisors.[1]

Examples 1.11. $\mathbb{Z}, \mathbb{Z}/3\mathbb{Z}, \mathbb{R}, \mathbb{R}[X]$ are integral domains. $\mathbb{Z}/6\mathbb{Z}, M_2(\mathbb{R})$ are not integral domains.

Proposition 1.12. *Every field is an integral domain.*

Proof. Every field is commutative and unital. It follows from exercise 4 that there are no zero-divisors. □

The converse of this proposition is false: $\mathbb{Z}$ is an example of an integral domain that is not a field. If we add the condition that the ring be finite, then the converse does hold. Of course, although $\mathbb{Z}$ is not itself a field, it can be embedded into the field $\mathbb{Q}$. It is true in general that every integral domain can be embedded in a field called its **field of quotients**, see exercise 30.

Theorem 1.13. *Every finite integral domain is a field.*

Proof. Let R be a finite integral domain, and let $a \in R$ be a non-zero element. Define a map $f_a : R \to R$ by $f_a(b) = ab$. Since R is an integral domain, f_a is injective: $f_a(b) = f_a(b') \implies ab = ab' \implies b = b'$ by Lemma 1.9. An injective map from a finite set to itself is necessarily bijective. Therefore, since f_a is surjective, there is an element $b \in R$ such that $f_a(b) = 1$. Since $ab = 1$, a is a unit. Having shown that every non-zero element of R is a unit, we conclude that R is a field. □

Note. It's possible to adapt the above proof to remove the hypothesis that the ring be commutative. In that slightly stronger form it's called Wedderburn's Little Theorem.

If p is prime, then $\mathbb{Z}/p\mathbb{Z}$ is easily shown to be an integral domain, and therefore (as already noted) $\mathbb{Z}/p\mathbb{Z}$ (which we will often denote $\mathbb{F}_p$) is a field. These are not the only finite fields (as we will see in Section 26).

Example 1.14. Let

$$F_4 = \left\{ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \right\} \subseteq M_2(\mathbb{Z}/2\mathbb{Z})$$

[1] Sometimes the term 'domain' is used rather than 'integral domain'.

With the usual matrix operations, and remembering that the entries are from $\mathbb{Z}/2\mathbb{Z}$, this set forms a field. This is definitely not the same as the ring $\mathbb{Z}/4\mathbb{Z}$ (which has zero-divisors). We will see later that for any prime p and any $n \in \mathbb{N}^+$ there is a field having p^n elements, and that it is unique up to isomorphism. We will consider finite fields in more detail in a later section.

1.4 Exercises

5. Let $\xi \in \mathbb{C}$ be the root of the polynomial $X^2 + X + 1$ given by $\xi = (-1 + \sqrt{-3})/2$. Define the **Eisenstein Integers** as $\mathbb{Z}[\xi] = \{a + b\xi \mid a, b \in \mathbb{Z}\}$. Show that $\mathbb{Z}[\xi]$ is a ring (where the operations are those inherited from $\mathbb{C}$).

6. Let R be an integral domain such that $x^2 = x$ for all $x \in R$. Show that R has exactly two elements.

7. List all units in the following rings:

   (a) $\mathbb{Z}$

   (b) $\mathbb{Z} \times \mathbb{Z}$

   (c) $\mathbb{Z}/5\mathbb{Z}$

   (d) $\mathbb{Z}/15\mathbb{Z}$

   (e) $\mathbb{Q}$

   (f) $\mathbb{R}[X]$

8. True or false?

   (a) Every field is also a ring.

   (b) Every ring has a multiplicative identity.

   (c) Every ring with a multiplicative identity has at least two elements.

   (d) The non-zero elements in a field form a group under multiplication.

   (e) Addition in a ring is always commutative.

9. Give the multiplication table for the multiplicative group of units in $\mathbb{Z}/12\mathbb{Z}$. To which group of order 4 is it isomorphic?

10. Determine all the units of $\mathbb{Z}[i]$. (Hint: Use the absolute value.)

11. Consider the ring $\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} \mid a, b \in \mathbb{Z}\} \subset \mathbb{R}$ (with the operations come from $\mathbb{R}$).

    (a) Find a unit in $\mathbb{Z}[\sqrt{2}]$ other than ± 1.

    (b) Use your answer from (a) to produce infinitely many units in $\mathbb{Z}[\sqrt{2}]$.

12. Show that if R is an integral domain, then $R[X]$ is an integral domain.

2 Subrings and Ideals

2.1 Subrings

Subrings are defined as for any algebraic structure:

Definition 2.1. A **subring** of a ring R is a subset $S \subseteq R$ which, when equipped with the operations from R, forms a ring.

Lemma 2.2. *A non-empty subset $S \subseteq R$ of a ring $(R, +, \times)$ is a subring if and only if it satisfies the following for all $a, b \in S$:*

1. $a - b \in S$

2. $a \times b \in S$

Proof. Exercise 14. □

Remark. The first condition is equivalent to requiring that $(S,+)$ is a subgroup of $(R,+)$. In particular, $0 \in S$. However, even if R contains a multiplicative identity, we are *not* insisting that $1 \in S$.

Example 2.3 (Subrings of $\mathbb{Z}$)**.** For any $n \in \mathbb{N}$ the set $n\mathbb{Z} = \{\ldots, -2n, -n, 0, n, 2n, \ldots\}$ is a subring of $\mathbb{Z}$. Conversely, suppose that $S \subseteq \mathbb{Z}$ is a subring. We will show that $S = n\mathbb{Z}$ for some $n \in \mathbb{N}$. If $S = \{0\}$, then we can take $n = 0$. So we assume that $S \neq \{0\}$, which implies that $S \cap \mathbb{N}^+ \neq \emptyset$. Let n be the least element in $S \cap \mathbb{N}^+$. Since $n \in S$ and S is a subring of $\mathbb{Z}$, we have that $n\mathbb{Z} \subseteq S$. For the reverse inclusion, let $s \in S$ be any element of S. Then by the division algorithm for $\mathbb{Z}$ we have $s = nq + r$ for some $q, r \in \mathbb{Z}$ with $0 \leqslant r < n$. But then, since $nq \in S$, we have $r = s - nq \in S$. By the minimality of n, it must be the case that $r = 0$, and therefore $s = nq \in S$.

2.2 Ideals

The subrings $n\mathbb{Z}$ of $\mathbb{Z}$ satisfy a stronger property than just being a subring: they are closed under multiplication by *any* element of $\mathbb{Z}$.

Definition 2.4. An **ideal** in a ring R is a non-empty subset $I \subseteq R$ that satisfies the following for all $a, b \in I$ and $r \in R$:

1. $a - b \in I$
2. $r \times a \in I$ and $a \times r \in I$

We denote this by $I \lhd R$ (meaning I is an ideal in R).

Remark. If R is commutative, the two parts of the second condition are equivalent. It is possible to consider **left ideals** (or **right ideals**) that satisfy only the first (second, respectively) part of this condition. We will rarely do so.

Example 2.5. Show that the set $I = \{\sum_1^n \alpha_i X^i : n \in \mathbb{N}^+, \alpha_i \in \mathbb{R}\} \subseteq \mathbb{R}[X]$, of polynomials having constant term equal to zero, is an ideal in $\mathbb{R}[X]$.

Exercise 13. Give an example of a ring R and a subset $I \subseteq R$ such that I is left ideal but not a right ideal.

Comparing this definition with Lemma 2.2, we see that every ideal is a subring. The converse is false.

Example 2.6. $\mathbb{Z} \subseteq \mathbb{R}$ is a subring of $\mathbb{R}$, but not an ideal in $\mathbb{R}$.

Remark. If S is a subring of R and R is a subring of T, then S is a subring of T. However the corresponding statement is not true of ideals.

2.3 Exercises

14. Prove Lemma 2.2.

    Lemma. *Let R be a ring, and $S \subseteq R$ a non-empty subset. Then S is a subring of R iff it satisfies:*

    (a) $\forall a, b \in S$, $a - b \in S$

    (b) $\forall a, b \in S$, $ab \in S$

15. Let R be a unital ring, and I an ideal in R. Show that if I contains a unit from R, then $I = R$.

16. Show that a field F has only two ideals, namely F and $\{0\}$. Conversely, show that if a commutative unital ring has exactly two ideals, then it is a field.
17. Give an example of a ring R with multiplicative identity 1_R that has a proper subring $S \leqslant R$ with multiplicative identity $1_S \neq 1_R$.

3 Homomorphisms and Quotients

As with groups, or any other algebraic structure, it is natural consider maps that preserve the underlying structure. For a ring, this means that the maps preserve products and sums.

Definition 3.1. A **ring homomorphism** (or simply a homomorphism if the context is clear), is a map $\varphi : R \to S$ between rings such that for all $a, b \in R$:

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

An **isomorphism** of rings is a bijective homomorphism. If there exists an isomorphism between two rings, they are said to be isomorphic.

Remark. The first condition is equivalent to requiring that φ be a homomorphism of the underlying abelian groups $(R, +)$ and $(S, +)$.

Exercise 18. Is it true that the ring $\mathbb{Z}$ is isomorphic to the ring $2\mathbb{Z}$? What about the underlying abelian groups?

Lemma 3.2. *Let $\varphi : R \to S$ be a homomorphism.*

1. *The* ***kernel*** *of φ, $\ker(\varphi) = \{r \in R : \varphi(r) = 0\}$, is an ideal in R.*
2. *The* ***image*** *of φ, $\mathrm{Im}(\varphi)$, is a subring of S. (But not necessarily an ideal.)*

Proof. Let $a, b \in \ker(\varphi)$ and $r \in R$. Then $a - b \in \ker(\varphi)$ since

$$\begin{aligned} \varphi(a-b) &= \varphi(a) + \varphi(-b) && \text{(since } \varphi \text{ is a homomorphism)} \\ &= \varphi(a) - \varphi(b) && \text{(since } \varphi \text{ is a homomorphism)} \\ &= 0 - 0 = 0 \end{aligned}$$

For the second condition in the definition of an ideal we note that

$$\begin{aligned} \varphi(ra) &= \varphi(r)\varphi(a) && \text{(since } \varphi \text{ is a homomorphism)} \\ &= \varphi(r) \times 0 = 0 \end{aligned}$$

and

$$\begin{aligned} \varphi(ar) &= \varphi(a)\varphi(r) && \text{(since } \varphi \text{ is a homomorphism)} \\ &= 0 \times \varphi(r) = 0 \end{aligned}$$

Now to show that $\mathrm{Im}(\varphi)$ is a subring of S. Let $s, t \in \mathrm{Im}(\varphi)$ be two elements in the image. Then $s = \varphi(c)$ and $t = \varphi(d)$ for some $c, d \in R$. It follows that

$$\begin{aligned} s - t &= \varphi(c) - \varphi(d) = \varphi(c-d) \in \mathrm{Im}(\varphi) \\ st &= \varphi(c)\varphi(d) = \varphi(cd) \in \mathrm{Im}(\varphi) \end{aligned}$$

□

Exercise 19. Give an example of a homomorphism whose image is not an ideal in the codomain.

Example 3.3. Fix $a \in \mathbb{R}$ and define a map $\varphi_a : \mathbb{R}[X] \to \mathbb{R}$ by $\varphi_a(\sum_0^n \alpha_i X^i) = \sum_0^n \alpha_i a^i$, that is, the image of a polynomial is given by evaluating at $X = a$. Then φ_a is a surjective ring homomorphism with kernel $\ker(\varphi_a) = \{p \in \mathbb{R}[X] \mid a \text{ is a root of } p\}$. Choosing $a = 0$ gives the ideal of example 2.5.

Lemma 3.4. *A homomorphism φ is injective if and only if* $\ker(\varphi) = \{0\}$.

Proof. Recall that, by definition, φ is injective if, for all a and b in its domain

$$\varphi(a) = \varphi(b) \implies a = b$$

Clearly, if φ is injective, then $\ker(\varphi) = \{0\}$.

For the converse, suppose that $\varphi(a) = \varphi(b)$. Then $\varphi(a - b) = \varphi(a) - \varphi(b) = 0$, which implies that $a - b \in \ker(\varphi)$. Since $\ker(\varphi) = \{0\}$. we conclude that $a - b = 0$. □

The lemma that the kernel of a ring homomorphism is an ideal can be compared to the statement that the kernel of a group homomorphism is a normal subgroup. For groups we can form the quotient of a group by a normal subgroup. Similarly, we can quotient a ring by an ideal.

Let $I \lhd R$ be an ideal in a ring R. Denote by R/I the set of (additive) cosets of I in R

$$R/I = \{a + I : a \in R\}$$

Define operations on this set by

$$\begin{aligned}(a + I) + (b + I) &= (a + b) + I \\ (a + I) \times (b + I) &= ab + I\end{aligned}$$

Let's check that the second operation is well-defined (meaning that it is independent of the choice of coset representative). Suppose that $(a + I) = (a' + I)$ and $(b + I) = (b' + I)$. Then $a' = a + x$ and $b' = b + y$ for some $x, y \in I$. Therefore

$$a'b' + I = (a + x)(b + y) + I = ab + xb + ay + xy + I = ab + I$$

We used that I is an ideal implies $xb, ay, xy \in I$ and hence $xb + ay + xy \in I$. Notice that we need that I is an ideal and not merely a subring.

Exercise 20. Check that the first operation above is also well-defined, and that with these operations R/I is a ring. What is the zero element in R/I ?

Definition 3.5. The ring R/I is called the **quotient ring**.

Examples 3.6.

1. For any $m \in \mathbb{Z}$ we can form the quotient $\mathbb{Z}/m\mathbb{Z}$.

2. The rings $\mathbb{Z}/4\mathbb{Z}$ and $2\mathbb{Z}/8\mathbb{Z}$ are not isomorphic. To see this observe that the former has an identity and the later does not.

3. $\mathbb{R}[X]/\langle X^2 + 1\rangle \cong \mathbb{C}$ (where $\langle X^2 + 1\rangle = \{f(X)(X^2 + 1) \mid f(X) \in \mathbb{R}[X]\} \lhd \mathbb{R}[X]$)

There is a direct relationship between ideals in R, quotients of R and kernels of homomorphisms from R onto another ring. We have already seen that the kernel of a homomorphism is an ideal. The following can be regarded as a kind of converse.

Lemma 3.7. *Let R be a ring. Given an ideal $I \lhd R$, the (**natural projection**) map*

$$\varphi : R \to R/I, \quad \varphi(a) = a + I$$

is a (surjective) ring homomorphism with $\ker(\varphi) = I$.

Proof. Let $a, b \in R$ be two elements of R. Then

$$\begin{aligned} \varphi(a+b) &= (a+b) + I = (a+I) + (b+I) = \varphi(a) + \varphi(b) \\ \varphi(ab) &= ab + I = (a+I)(b+I) = \varphi(a)\varphi(b) \end{aligned}$$

So φ is a homomorphism, and $\ker(\varphi) = I$ since

$$\varphi(a) = 0_{R/I} \iff a + I = 0_R + I \iff a \in I$$

□

3.1 Isomorphism theorems

We know that the kernel of a homomorphism is an ideal in the domain, and that the image is a subring of the codomain. They are related by the following

Theorem 3.8 (First Isomorphism Theorem). *Let $\varphi : R \to S$ be a ring homomorphism. Then*

$$R/\ker(\varphi) \cong \mathrm{Im}(\varphi)$$

An explicit isomorphism is given by $a + \ker(\varphi) \mapsto \varphi(a)$.

Proof. Denote by K the kernel $\ker(\varphi)$. Define a map $f : R/K \to \mathrm{Im}(\varphi)$ by $f(a+K) = \varphi(a)$. This is well-defined since

$$\begin{aligned} a + K = a' + K &\implies a' = a + k \quad \text{(for some } k \in K) \\ &\implies \varphi(a') = \varphi(a+k) = \varphi(a) + \varphi(k) = \varphi(a) + 0 = \varphi(a) \end{aligned}$$

We will show that f is an isomorphism. That f is a homomorphism follows from the fact that φ is a homomorphism, and the way in the which the operations in R/K are defined. It is clear that f is surjective. For injectivity,

$$f(a+K) = 0 \iff \varphi(a) = 0 \iff a \in K \iff a + K = 0_{R/K}$$

□

The First Isomorphism Theorem can be used to prove the following, which we give here for completeness.

Theorem 3.9 (Second and Third Isomorphism Theorems). *Let R be a ring.*

1. *Suppose $I \lhd R$ is an ideal and $S \leqslant R$ is a subring. Then*

$$(S+I)/I \cong S/(S \cap I)$$

2. *Suppose that $I, J \lhd R$ are ideals in R, and $I \subseteq J$. Then*

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

Where it is understood that part of the assertion being made is that each expression makes sense, e.g., that J/I is an ideal in R/I.

Proof. Exercise 25 □

3.2 Correspondence Theorem

Given a homomorphism, there is a correspondence between subrings (ideals) of the image and subrings (ideals) in the domain that contain the kernel of the homomorphism. This innocuous looking result is surprisingly useful.

We noted in Lemma 3.2 that the image of a homomorphism is always a subring of the codomain. We start by giving an extension of that result to all subrings and ideals of the domain.

Lemma 3.10. *Let $\varphi : R \to R'$ be a ring homomorphism.*

1. *If S is a subring (or ideal) in R, then $\varphi(S)$ is a subring (ideal) in $\operatorname{Im}(\varphi)$.*
2. *If S' is a subring (or ideal) in $\operatorname{Im}(\varphi)$, then $\varphi^{-1}(S')$ is a subring (ideal) in R.*

Proof. Given $a', b' \in \varphi(S)$, we have that $a' = \varphi(a)$ and $b' = \varphi(b)$ for some $a, b \in S$. Since S is a subring $a - b \in S$, and $a' - b' = \varphi(a) - \varphi(b) = \varphi(a - b) \in \varphi(S)$. Also $ab \in S$ implies that $a'b' = \varphi(a)\varphi(b) = \varphi(ab) \in \varphi(S)$. Noting that $\varphi(S)$ is non-empty given that S is, we conclude that $\varphi(S)$ is a subring of R'. If, further, S is an ideal in R and $r' \in \operatorname{Im}(\varphi)$, then $r' = \varphi(r)$ for some $r \in R$ and $r'a' = \varphi(r)\varphi(a) = \varphi(ra) \in \varphi(S)$. Similarly $a'r' \in \varphi(S)$, and we conclude that $\varphi(S)$ is an ideal in $\operatorname{Im}(\varphi)$.

For the second part, let $a, b \in \varphi^{-1}(S')$. Then $\varphi(a), \varphi(b) \in S'$, which implies that $\varphi(a - b) = \varphi(a) - \varphi(b) \in S'$ and $\varphi(ab) = \varphi(a)\varphi(b) \in S'$. As $\varphi^{-1}(S')$ is non-empty (it contains 0_R since $\varphi(0_R) = 0_{R'}$), we conclude that it is a subring of R. If, further, S' is an ideal in $\operatorname{Im}(\varphi)$ and $r \in R$, then $\varphi(ra) = \varphi(r)\varphi(a) \in S'$, which implies that $ra \in \varphi^{-1}(S')$. The argument that $ar \in \varphi^{-1}(S')$ is exactly the same. □

Remark. Of course, in the first part of the lemma, we can conclude that the image of a subring in R is a subring of R'. However, the image of an ideal in R is not always an ideal in R'.

Different subrings in the domain can have the same image in the codomain. However, if we restrict to only those subrings in the domain that contain the kernel, then we get a correspondence.

Theorem 3.11 (Correspondence Theorem)**.** *Let $\varphi : R \to R'$ be a ring homomorphism. The maps*

$$\Phi : \{S \leqslant R : \ \ker(\varphi) \subseteq S\} \to \{S' \leqslant R' : \ S' \subseteq \operatorname{Im}(\varphi)\}, \qquad \Phi(S) = \varphi(S)$$

$$\Psi : \{I \lhd R : \ \ker(\varphi) \subseteq I\} \to \{I' \subseteq R' : \ I' \lhd \operatorname{Im}(\varphi)\}, \qquad \Psi(I) = \varphi(I)$$

are inclusion-preserving bijections.

Proof. We give the argument for Ψ and leave the other case as an exercise. Given $I' \lhd \operatorname{Im}(\varphi)$, we know from the preceding lemma that $\varphi^{-1}(I')$ is an ideal in R that contains the kernel of φ. It follows that Ψ is surjective, since $\Psi(\varphi^{-1}(I')) = \varphi(\varphi^{-1}(I')) = I'$. For injectivity first note that if $I \lhd R$ contains the kernel of φ, then

$$\begin{aligned} \varphi(a) \in \varphi(I) &\implies \varphi(a) = \varphi(i) \quad \text{for some } i \in I \\ &\implies \varphi(a - i) = 0 \\ &\implies a - i \in \ker(\varphi) \\ &\implies a \in I \quad (\text{since } \ker(\varphi) \subseteq I) \end{aligned}$$

Now suppose that I and J are ideals in R that contain the kernel of φ, and that $\Psi(I) = \Psi(J)$. Then

$$a \in I \iff \varphi(a) \in \varphi(I) \iff \varphi(a) \in \Psi(I) \iff \varphi(a) \in \Psi(J) \iff a \in J$$

We have shown then that Ψ is bijective. Its inverse is the map $I' \mapsto \varphi^{-1}(I')$. That Ψ preserves inclusions then follows from the fact that $I \subseteq J \implies \varphi(I) \subseteq \varphi(J)$ and $I' \subseteq J' \subseteq \operatorname{Im}(\varphi) \implies \varphi^{-1}(I') \subseteq \varphi^{-1}(J')$. □

3.3 Exercises

21. The **direct product** $R \times S$ of two rings is a ring given by the set $\{(r, s) \mid r \in R, s \in S\}$ with operations defined by

$$(r_1, s_1) + (r_2, s_2) = (r_1 +_R r_2, s_1 +_S s_2)$$
$$(r_1, s_1) \times (r_2, s_2) = (r_1 \times_R r_2, s_1 \times_S s_2)$$

    (a) Is the map $r \mapsto (r, 0)$ from R to $R \times S$ a ring homomorphism ?

    (b) What about the **diagonal map** $r \mapsto (r, r)$ from R to $R \times R$?

22. (a) Is $\mathbb{Z}/8\mathbb{Z}$ isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$ (as rings)?

    (b) Is $\mathbb{Z}/15\mathbb{Z}$ isomorphic to $\mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z}$ (as rings)?

23. If I, J are ideals in R, the **sum** of I and J denoted $I + J$ is defined by

$$I + J = \{x + y \mid x \in I, y \in J\} \subseteq R$$

    (a) Show that $I + J$ is again an ideal in R.

    (b) Show that if $I + J = R$, then $R/(I \cap J) \cong R/I \times R/J$.

24. Using the above exercise 23 show that $\mathbb{Z}_{mn} \cong \mathbb{Z}_m \times \mathbb{Z}_n$ if and only if $\gcd(m, n) = 1$.

25. Use the First Isomorphism Theorem 3.8 to prove the Second and Third Isomorphism Theorems 3.9.

26. Prove the 'correspondence theorem', Theorem 3.11.

27. The **characteristic** of a unital ring R is the smallest $n \in \mathbb{N}^+$ such that

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

    if such an n exists; otherwise the characteristic is defined to be 0.

    (a) Show that the characteristic of an integral domain is either zero or a prime.

    (b) Let R be a unital ring with characteristic n. Verify that the map from $\mathbb{Z} \to R$ that sends $1_{\mathbb{Z}}$ to 1_R and m to $(1_R + 1_R + \cdots + 1_R)$ (m times) is a homomorphism with kernel equal to $n\mathbb{Z}$, and that R therefore contains a subring isomorphic to $\mathbb{Z}/n\mathbb{Z}$.

    (c) Conclude that every integral domain either contains a subring isomorphic to $\mathbb{Z}$, or contains a subring isomorphic to the field $\mathbb{F}_p$. (For some prime $p \in \mathbb{N}$.)

28. A **prime field** is a field with no proper subfields. Show that a prime field is isomorphic to either $\mathbb{Q}$ or $\mathbb{F}_p$ for some prime p (corresponding to the characteristic of the field being 0 or p).

29. Let R be a commutative unital ring of prime characteristic p. Show that, for all $x, y \in R$ and $n \in \mathbb{N}$, the following holds:

$$(x + y)^{p^n} = x^{p^n} + y^{p^n}$$

    Notice that this shows that the map $F : R \to R$ given by $F(x) = x^p$ is a ring homomorphism (called the **Frobenius map**).

30. In this exercise we will prove that every integral domain can be embedded in a field. The construction mimics the way in which $\mathbb{Q}$ is built from $\mathbb{Z}$. Let D be an integral domain and define

$$F = \{(a, b) \mid a, b \in D, b \neq 0\}/\sim \quad \text{where} \quad (a, b) \sim (c, d) \quad \text{if} \quad ad = bc.$$

    Define operations on F by:

$$\overline{(a, b)} + \overline{(c, d)} = \overline{(ad + bc, bd)}$$
$$\overline{(a, b)(c, d)} = \overline{(ac, bd)}$$

    (Where $\overline{(a, b)}$ denotes the equivalence class of $(a, b) \in D^2$ with respect to $\sim$.)

    Show that:

(a) These operations on F are well-defined;

(b) F, with these operations, is a ring;

(c) F is a field;

(d) The map $\varphi : D \to F$, given by $\varphi(a) = \overline{(a, 1)}$ is an injective homomorphism (and therefore its image is isomorphic to D).

(e) Show that any field that contains a subring D' that is isomorphic to D contains a subfield isomorphic to F (and containing D').

The field F is called the **field of quotients** of the integral domain D.

4 Constructions

We mention here some standard ways of producing new rings from old. This will be used extensively in the sequel.

4.1 Direct product

Given two rings R and S, their direct product is ring given by the set

$$R \times S = \{(r, s) : \ r \in R, s \in S\} \quad \text{(the usual cartesian product of two sets)}$$

equipped with the operations

$$\begin{aligned}(r, s) + (r', s') &= (r + r', s + s') \\ (r, s)(r', s') &= (rr', ss')\end{aligned}$$

These operations are sometimes said to be defined 'pointwise' or 'coordinatewise'. The operation of taking direct products is (up to isomorphism) associative and commutative; that is, $R\times(S\times T) \cong (R\times S)\times T$ and $R\times S \cong S\times R$. We use the usual convention of denoting $R \times R$ by R^2. Similarly we will speak about R^n, the direct product of n copies of R. If we take infinitely many rings R_i, then we can form the **direct sum** or the **direct product**.

4.2 Polynomial rings

Let R be a commutative unital ring. The underlying set is actually the same as the infinite direct sum. Precise notation is a bit cumbersome. Elements are of the form

$$a_0 + a_1X + a_2X^2 + \cdots + a_nX^n \qquad \text{where } n \geqslant 0, a_i \in R \text{ and } a_n \neq 0$$

The degree of such a polynomial f is equal to n and is denoted $\deg(f)$. The ring R embeds in $R[X]$ as the degree zero polynomials, and we will make this identification without comment. If R is commutative/unital/ID then $R[X]$ is commutative/unital/ID. The units in $R[X]$ are precisely the degree zero polynomials that are units in R. Since $R[X]$ is a ring, this construction can be iterated to give $R[X, Y] = (R[X])[Y]$ and $R[X_1, \dots, X_n]$.

4.3 Matrix rings

Let R be a commutative, unital ring, and $n \in \mathbb{N}^+$. An $n \times n$ matrix over R is a square array of elements from R. With addition and multiplication defined as usual, this forms a ring which we denote $M_n(R)$. The standard definition of determinant works in $M_n(R)$, and the determinant is an element of R. If $A, B \in M_n(R)$ are two matrices, then $\det(AB) = \det(A)\det(B)$. A matrix $A \in M_n(R)$ is invertible if and only if $\det(A)$ is a unit in R.

4.4 Ring of endomorphisms

Let R be a ring. The set of all ring homomorphisms from R to itself forms a ring. The operations are pointwise addition and composition, i.e.,

$$\begin{aligned}(f + g)(a) &= f(a) + g(a) \\ (fg)(a) &= f \circ g(a)\end{aligned}$$

4.5 Group rings

Let G be a group, and R a commutative unital ring. The **group ring (of G over R)** is the set

$$R(G) = \{a_1g_1 + \cdots + a_ng_n : n \in \mathbb{N}^+, a_i \in R, g_i \in G \text{ distinct } \}$$

of all finite formal sums, with addition defined in the obvious way, and multiplication given by

$$(\sum_i a_ig_i)(\sum_j b_jh_j) = \sum_{i,j}(a_ib_j)(g_ih_j)$$

Exercise 31. Check that, with these operations, $R(G)$ forms a ring.

4.6 Exercises

32. Let $\varphi_1 : R \to S_1$ and $\varphi_2 : R \to S_2$ be ring homomorphisms. Show that the map $\varphi : R \to S_1 \times S_2$ given by $\varphi(a) = (\varphi_1(a), \varphi_2(a))$ is a ring homomorphism.

33. Let $\varphi : R \to S$ be a homomorphism, and define a map $\Phi : R[X] \to S[X]$ by
$$\Phi(a_0 + a_1X + \cdots + a_nX^n) = \varphi(a_0) + \varphi(a_1)X + \cdots + \varphi(a_n)X^n$$
Show that Φ is a homomorphism.

34. Show that the units in $F[X]$, where F is a field, are the elements of $F \setminus \{0\}$.

35. Suppose that R is a commutative ring and $a \in R$ a fixed element. Show that the map from $R[X]$ to itself defined by
$$a_0 + a_1X + \cdots + a_nX^n \mapsto a_0 + a_1(X - a) + \cdots + a_n(X - a)^n$$
is an isomorphism of rings. Deduce that if $f(X) \in R[X]$, then $f(X)$ can be expressed in the form $f(X) = \sum b_i(X - a)^i$ for suitable $b_i \in R$.

36. Let R b a commutative unital ring and $r \in R$ a fixed element. Show that there is exactly one homomorphism $\varphi : R[X] \to R$ satisfying $\varphi(a) = a$ for all $a \in R$ and $\varphi(X) = r$.

37. Are the following matrices invertible?

(a) $\begin{bmatrix} 2 & 2 \\ 1 & 2 \end{bmatrix} \in M_2(\mathbb{Z}/3\mathbb{Z})$

(b) $\begin{bmatrix} 2 & 2 \\ 1 & 2 \end{bmatrix} \in M_2(\mathbb{Z}/6\mathbb{Z})$

(c) $\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \in M_2(\mathbb{Z})$

(d) $\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \in M_2(\mathbb{Q})$

(e) $\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \in M_2(\mathbb{Z})$

(f) $\begin{bmatrix} X & 2 \\ 0 & 1 \end{bmatrix} \in M_2(\mathbb{R}[X])$

(g) $\begin{bmatrix} 1 & X^2 + 1 \\ 0 & 2 \end{bmatrix} \in M_2(\mathbb{R}[X])$

5 Generating sets and principal ideals

5.1 Generating sets

Noting that the intersection of two subrings (or ideals) is a subring (ideal) enables us to make the following definition.

Definition 5.1. Let R be a ring, and $X \subseteq R$ be a subset. The **subring generated** by X is the intersection of all subrings of R that contain X. Similarly, the **ideal generated** by X is the intersection of all ideals in R that contain X.

In the case that $X = \{a_1, \dots, a_k\}$ we denote the ideal generated by X by $\langle a_1, \dots, a_k \rangle$.

Definition 5.2. An ideal $I \lhd R$ satisfying $I = \langle a \rangle$ for some $a \in R$ is called a **principal ideal**.

Notice that if R is unital, then $\langle 1 \rangle = R$. The following lemma states that the ideal generated by $X \subseteq R$ is the set of all R-linear combinations of elements from X.

Lemma 5.3. *Let R be a commutative unital ring and $X \subseteq R$. Then*

$$\langle X \rangle = \{r_1 a_1 + \dots + r_n a_n : n \in \mathbb{N}, r_i \in R, a_i \in X\}$$

Proof. As the set $I = \{r_1 a_1 + \dots + r_n a_n : n \in \mathbb{N}, r_i \in R, a_i \in X\}$ is clearly an ideal, we know that $\langle X \rangle \subseteq I$. Conversely, any R-linear combination of elements from X will lie in every ideal that contains X. It follows that $I \subseteq \langle X \rangle$. □

Examples 5.4.

1. We saw above that all ideals in $\mathbb{Z}$ are principal.
2. All ideals in $\mathbb{R}$ are principal as $\{0\}$ and $\mathbb{R}$ itself are the only ideals.
3. The ideal $\langle 2, X \rangle \lhd \mathbb{Z}[X]$ is *not* principal.

   Proof. Let $I = \langle 2, X \rangle$ and suppose that $I = \langle f \rangle$ for some $f \in \mathbb{Z}[X]$. Using Lemma 5.3, since $2 \in I$ we know that $2 = fg$ for some $g \in \mathbb{Z}[X]$. It follows that $\deg(f) = 0$ and that either $f = \pm 1$ or $f = \pm 2$. If $f = \pm 1$, then $\langle f \rangle = \mathbb{Z}[X]$. This can not be the case if $\langle f \rangle = \langle 2, X \rangle$ since (for example) $1 \notin I$. Similarly $I \neq \langle \pm 2 \rangle$ since $2 + X \in I$, but $2 + X \notin \langle 2 \rangle$. □

Exercise 38. Show that the ideal $\langle X, Y \rangle \lhd \mathbb{R}[X, Y]$ is *not* principal.

5.2 Principal ideal domains

We finish this section with the definition of an important class of rings. We have observed that all ideals in $\mathbb{Z}$ are principal, and we shall shortly see that the same is true in other rings such as $\mathbb{R}[X]$ and $\mathbb{Z}[i]$.

Definition 5.5. A **principal ideal domain** (PID for short) is an integral domain in which all ideals are principal.

Examples 5.6.

1. Any field F is (trivially) a PID, as there are only two ideals, $\{0\}$ and F, both of which are principal: $\{0\} = \langle 0 \rangle$ and $F = \langle 1 \rangle$.
2. The ring of polynomials $\mathbb{R}[X]$ is a PID, as we shall see shortly.
3. $\mathbb{Z}[X]$ is not a PID since the ideal $\langle 2, X \rangle$ is not principal.

5.3 Exercises

39. Show that every ideal in $\mathbb{Z}/12\mathbb{Z}$ is principal. Is $\mathbb{Z}/12\mathbb{Z}$ a PID?

6 Division and factorization in integral domains

Continuing to abstract properties from the integers, we will define divisors in an integral domain. This will lead to two versions of what a 'prime' is. We will then consider the use of the Euclidean algorithm in this context. In this section the ring R will always be an integral domain. Many of the definitions make sense in a more general setting.

6.1 Divisors

Definition 6.1. Let $a, b \in R$. We say that a **divides** b (or a is a **divisor** of b) if there exists $c \in R$ such that $b = ac$. We write $a \mid b$ to mean that a divides b. We say that a and b are **associates** if both $a \mid b$ and $b \mid a$. This will sometimes be denoted by $a \sim b$.

Notice that $a \mid b$ is the same as $b \in \langle a \rangle$.

Examples 6.2.

1. In $\mathbb{R}[X]$, $(X-1) \mid (X^5 - 1)$.
2. $2, -3 \in \mathbb{Z}$ are not associates.
3. $2, -2 \in \mathbb{Z}$ are associates.
4. $2, -3 \in \mathbb{Q}$ are associates.

Exercise 40.

a) Show that if $a \mid b$ and $b \mid c$, then $a \mid c$. (That is, it is a transitive relation.)

b) Show that if a divides a unit, then a is a unit.

c) Show that if a is a unit, then $a \mid b$ for all b.

d) Show that $a \mid b$ if and only if $\langle b \rangle \subseteq \langle a \rangle$

Exercise 41. Check that the relation of being associates defines an equivalence relation on R.
That is, show that for all $a, b, c \in R$:

1. $a \sim a$
2. $a \sim b \implies b \sim a$
3. $a \sim b$ and $b \sim c \implies a \sim c$

Exercise 42. Show that

$$a \sim b \iff \langle a \rangle = \langle b \rangle \iff \text{there is a unit } u \in R\text{, such that } a = bu$$

6.2 Irreducible elements

We now generalise the notion of a prime integer. A prime integer can be defined as one having no proper divisors, that is, it cannot be written as a product of two integers, unless one of the factors is 1 or -1. We make this a definition.

Definition 6.3. An element $a \in R$ is called **irreducible** if a is not a unit and the following holds

$$a = bc \implies b \text{ is a unit or } c \text{ is a unit}$$

This is the same as saying that all divisors of a are either units or associates of a.

Example 6.4.

1. The irreducibles in $\mathbb{Z}$ are exactly the prime integers (where we allow negative primes).
2. Any linear polynomial in $F[X]$ is irreducible, where here F can be any field.
3. Both X^2+1 and X^2+X+1 are irreducible in $\mathbb{R}[X]$.
4. X^2+1 is not irreducible in $\mathbb{F}_2[X]$, since $X^2+1=(X+1)(X+1)$ and $X+1$ is not a unit. The polynomial X^2+X+1 is irreducible in $\mathbb{F}_2[X]$.
5. Neither X^2+1 nor X^2+X+1 is irreducible in $\mathbb{C}[X]$ since $X^2+1=(X-i)(X+i)$ and $X^2+X+1=(X-\frac{1}{2}(1+i\sqrt{3}))(X-\frac{1}{2}(1-i\sqrt{3}))$.

6.3 Prime elements

Another characterisation of the prime integers is that if p is prime and $p \mid ab$ then p divides one of a or b. Let's make this a definition.

Definition 6.5. An element $a \in R \setminus \{0\}$ is called **prime** if a is not a unit and the following holds for all $b, c \in R$:

$$a \mid bc \implies a \mid b \quad \text{or} \quad a \mid c$$

Example 6.6.

1. The primes in $\mathbb{Z}$ are exactly the 'usual' primes: $\pm 2, \pm 3, \pm 5, \pm 7, \ldots$
2. The element $X-1 \in \mathbb{R}[X]$ is prime.

We've generalised the notion of a prime integer in two ways. The next result says that one implies the other.

Proposition 6.7. *In an integral domain, prime elements are irreducible.*

Proof. Let p be a prime, and suppose that $p=bc$. Then $p \mid bc$ and so we have that either $p \mid b$ or $p \mid c$. Suppose that $p \mid b$. Then $p = pdc$ for some $d \in R$, and therefore $1 = dc$, since R is an integral domain. It follows that c is a unit. Thus p is irreducible.

□

The next example demonstrates that the converse to Proposition 6.7 does not hold.

Example 6.8 (Irreducible but not prime)**.** Consider the subring $\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\}$ of $\mathbb{C}$. We show that the element 2 is irreducible in $\mathbb{Z}[\sqrt{-5}]$, but not prime. To do this we will use a function that, in some sense, measures complexity.

Define a function $N : \mathbb{Z}[\sqrt{-5}] \to \mathbb{Z}$ by $N(a+b\sqrt{-5}) = a^2+5b^2$. Notice that N is simply the square of the magnitude of the complex number $a+b\sqrt{-5}$. It follows that N is multiplicative: $N(xy)=N(x)N(y)$. If x is a unit, then $N(x)=\pm 1$, since it must divide 1. It follows that 1 and -1 are the only units in $\mathbb{Z}[\sqrt{-5}]$.

To see that 2 is irreducible, note that it is not a unit, and suppose that $2=(a+\sqrt{-5}b)(\alpha+\sqrt{-5}\beta)$. Then, applying N we get that $4=(a^2+5b^2)(x^2+5y^2)$. So (a^2+5b^2) divides 4 in $\mathbb{Z}$, which implies that $(a^2+5b^2) \in \{1,2,4\}$. This can only occur if $b=0$ and $a \in \{\pm 1, \pm 2\}$, and therefore $(a+\sqrt{-5}b) \in \{\pm 1, \pm 2\}$. Thus 2 is irreducible.

To see that 2 is not prime, note that $2 \mid (1+\sqrt{-5})(1-\sqrt{-5})$. However 2 divides neither of the factors, since if it did we would obtain $N(2) \mid N(1 \pm \sqrt{-5})$, that is $4 \mid 6$.

The idea of a function that measures the complexity of elements is something we shall return to when we consider *Euclidean domains*.

6.4 Unique Factorization Domains

In the integers the 'fundamental theorem of arithmetic' states that every integer can be written as a product of irreducibles and that this factorization is essentially unique. Not all integral domains have this property, for example the integral domain $\mathbb{Z}[\sqrt{-5}]$.

Definition 6.9. An integral domain R is called a **unique factorization domain** (UFD for short) if the following hold:

1. *Existence of factorization:*
   Every element $a \in R$ that is nonzero and not a unit can be written as a product of irreducibles:
   $$a = a_1 a_2 \cdots a_n$$
2. *Uniqueness of factorization:*
   If $a = b_1 \dots b_m$ is another factorization of a into a product of irreducibles, then $m = n$ and there is a permutation π of $\{1, 2, \dots, n\}$, such that $b_i \sim a_{\pi(i)}$. That is, the two factorizations differ only by re-ordering and replacing each factor by an associate.

Examples 6.10.

1. $\mathbb{Z}$ is a UFD, as we already knew.
2. $\mathbb{Q}$, $\mathbb{R}$ are UFDs since there are no elements that are nonzero and non unit.
3. $\mathbb{Z}/5\mathbb{Z}[X]$, $\mathbb{R}[X]$ are all UFDs, as we'll see in section 9.
4. $\mathbb{Z}[X]$, $\mathbb{R}[X, Y]$ are all UFDs, as we'll see in section 10.

Exercise 43. Show that in a UFD irreducible elements are prime.

It follows from this and Example 6.8, that $\mathbb{Z}[\sqrt{-5}]$ is not a UFD. More explicitly, $2 \times 3 = (1-\sqrt{-5}) \times (1+\sqrt{-5})$ and all four elements are irreducible.

Example 6.11. The ring $\mathbb{Z}[\sqrt{-5}]$ is not a UFD because it fails the second part of the definition (uniqueness), although it does satisfy the first part (existence). Here is an example in which the first part fails to hold. Let $R = \mathbb{R}[X_1, X_2, \dots]$, and let $I \lhd R$ be the ideal generated by the set $\{X_2^2 - X_1, X_3^2 - X_2, X_4^2 - X_3, \dots\} \subset R$. Then in R/I the element $X_1 + I$ has no factorization as a product of irreducibles (and is not a unit):

$$X_1 + I = (X_2 + I)(X_2 + I) = (X_3 + I)(X_3 + I)(X_3 + I)(X_3 + I) = \cdots$$

6.5 Exercises

44. Let D be an integral domain, and $p, q \in D$ with $q|p$. Show that:
    (a) If p is a unit, then q is a unit.
    (b) If p is irreducible, then either q is a unit or p and q are associates.
    (c) If p and q are associates, then p is irreducible iff q is irreducible.

45. Let $d \in \mathbb{Z}$ be square-free, and $R = \mathbb{Z}[\sqrt{d}]$. Show that every (non-zero, non-unit) element of R can be written as a product of irreducibles.

46. Show that the following is equivalent to the definition of Unique Factorization Domain. R is an integral domain is which

    1. Every element $a \in R$ that is nonzero and not a unit can be written as a product of irreducibles:

    $$a = a_1 a_2 \dots a_n$$

    2′. Every irreducible element of R is prime.

7 Prime and maximal ideals

Definition 7.1. Let R be a commutative ring, and $I \neq R$ an ideal in R.

1. I is said to be **prime** if it satisfies the condition: $\forall a, b \in R,\ \ ab \in I \implies a \in I$ or $b \in I$
2. I is said to be **maximal** if it satisfies the condition: $\forall J \lhd R,\ \ I \subseteq J \implies J = I$ or $J = R$

Proposition 7.2. *Let R be a commutative unital ring and $I \lhd R$ an ideal in R. Then*

1. *I is prime $\iff$ R/I is an integral domain;*
2. *I is maximal $\iff$ R/I is a field.*

Proof. Note that since R is commutative and unital, so too is R/I for any ideal $I \lhd R$, provided only that $I \neq R$. Denote by $\varphi : R \to R/I$ the natural projection map.

1. Suppose I is prime. We need to show that R/I has no zero-divisors. Let x, y be two elements in R/I with $x \neq 0$. There are $a, b \in R$ with $\varphi(a) = x$ and $\varphi(b) = y$, and since $x \neq 0$, we have $a \notin I$. Then,

   $$xy = 0 \implies \varphi(a)\varphi(b) = 0 \implies \varphi(ab) = 0 \implies ab \in I \implies b \in I \implies \varphi(b) = 0 \implies y = 0$$

   Now suppose that R/I has no zero-divisors. Let $a, b \in R$ be such that $ab \in I$ and $a \notin I$. Then $\varphi(a) \neq 0$, and

   $$ab \in I \implies \varphi(ab) = 0 \implies \varphi(a)\varphi(b) = 0 \implies \varphi(b) = 0 \implies b \in I$$

2. Since R/I is unital, it is a field if and only if its only ideals are itself and $\{0\}$ (Exercise 16).

   $$\begin{aligned} I \text{ is maximal } &\iff R/I \text{ contains only two ideals} && \text{(Correspondence Theorem 3.11)} \\ &\iff R/I \text{ is a field} && \text{(Exercise 16)} \end{aligned}$$

□

Corollary 7.3. *Every maximal ideal is prime.* □

Lemma 7.4. *Let R be an integral domain and $a \in R \setminus \{0\}$.*

1. *If the ideal $\langle a \rangle$ is maximal, then a is irreducible.*
2. *Suppose that R is a PID. If a is irreducible, then the ideal $\langle a \rangle \lhd R$ is maximal.*

Proof. Suppose that $\langle a \rangle$ is maximal. Then $\langle a \rangle \neq R$, so a is not a unit. Now

$$a = bc \implies \langle a \rangle \subseteq \langle b \rangle \implies \langle a \rangle = \langle b \rangle \quad \text{or} \quad \langle b \rangle = R \qquad \text{(since } \langle a \rangle \text{ is maximal)}$$

If $\langle b\rangle = R$, then b is a unit. On the other hand

$$\begin{aligned}\langle b\rangle = \langle a\rangle &\implies b = au \text{ for some } u \in R\\ &\implies a = auc \implies 1 = uc \qquad\qquad \text{(since } R \text{ is an ID and } a \neq 0\text{)}\\ &\implies c \in R^*\end{aligned}$$

It follows that a is irreducible.

Now suppose that R is a PID and that a is an irreducible. Since a is not a unit we have that $\langle a\rangle \neq R$. Let $J \lhd R$ be an ideal satisfying $\langle a\rangle \subseteq J \subseteq R$. Since R is a PID, $J = \langle b\rangle$ for some $b \in R$. Then

$$\begin{aligned}\langle a\rangle \subseteq \langle b\rangle &\implies a = bc \text{ for some } c \in R\\ &\implies b \in R^* \quad \text{or} \quad c \in R^*\\ &\implies \langle b\rangle = R \quad \text{or} \quad \langle b\rangle = \langle a\rangle\end{aligned}$$

□

Exercise 47.

a) The hypothesis that R is an ID is necessary in the first part of the above lemma. To demonstrate this, find an element $a \in \mathbb{Z}/6\mathbb{Z}$ such that $\langle a\rangle$ is maximal and a is *not* irreducible.

b) Give an example of an ID R, and an element $a \in R$ such that a is irreducible but $\langle a\rangle$ is not maximal.

Lemma 7.5. *Let R be an integral domain and $a \in R \setminus \{0\}$. The ideal $\langle a\rangle \lhd R$ is a prime ideal if and only if a is a prime element.*

Proof. Suppose that $\langle a\rangle$ is prime. Then

$$a \mid bc \implies bc \in \langle a\rangle \implies (b \in \langle a\rangle \quad \text{or} \quad c \in \langle a\rangle) \implies (a \mid b \quad \text{or} \quad a \mid c)$$

Conversely, if a is prime, then

$$bc \in \langle a\rangle \implies a \mid bc \implies (a \mid b \quad \text{or} \quad a \mid c) \implies (b \in \langle a\rangle \quad \text{or} \quad c \in \langle a\rangle) \quad \square$$

7.1 Exercises

48. Let R be a PID, S and integral domain and $\varphi : R \to S$ a surjective homomorphism. Show that either φ is an isomorphism or S is a field.

49. Let I, J, and P be ideals in R, with P prime. Show that if $IJ \subseteq P$, then either $I \subseteq P$ or $J \subseteq P$.

50. Determine the maximal ideals in the following rings:

    (a) $\mathbb{R}$ (b) $\mathbb{Z}$ (c) $\mathbb{Z}/11\mathbb{Z}$ (d) $\mathbb{Z}/12\mathbb{Z}$

51. Show that $\langle 2\rangle \subseteq \mathbb{Z}[\sqrt{-5}]$ is not prime. Show that $\langle 11\rangle \subseteq \mathbb{Z}[\sqrt{-5}]$ is prime.

52. Given two ideals $I, J \subseteq R$ we define their product IJ to be the ideal generated by the set $\{ij \mid i \in I, j \in J\} \subseteq R$. Consider the ring $\mathbb{Z}[\sqrt{-5}]$.

    (a) Show that $\langle 2\rangle = \langle 2, 1+\sqrt{-5}\rangle\langle 2, 1-\sqrt{-5}\rangle$.

    (b) Show that $\langle 2, 1+\sqrt{-5}\rangle$ is prime.

8 $\mathbb{R}[X]$ is a PID

We will show that for *any* field F, the ring of polynomials $F[X]$ is a PID. For the rest of this section, F denotes a field.

8.1 Division with remainder in $F[X]$

We state and prove a direct analogue of the 'division algorithm' in $\mathbb{Z}$. Both the statement and the proof of the theorem follow closely the situation for $\mathbb{Z}$.

Theorem 8.1. *Given $f, g \in F[X]$ with $g \neq 0$, there exist polynomials $q, r \in F[X]$ such that $f = qg + r$ and either $\deg(r) < \deg(g)$ or $r = 0$. Moreover, the polynomials q and r are unique.*

Proof. Let $S = \{f - gs \mid s \in F[X]\}$, and let $r \in S$ be an element having the minimum degree possible amongst elements of S. Since r is in S, it is clear that $f = gq + r$ for some $q \in F[X]$. We need to show that either $\deg(r) < \deg(g)$ or $r = 0$. If $0 \in S$, then we can take $r = 0$. Suppose that $0 \notin S$. Let $t = \deg(r)$ and let $c \in F$ be the coefficient of X^t in r. Similarly let $m = \deg(g)$ and let $b \in F$ be the coefficient of X^m in g. Note that b (and c) is nonzero and therefore a unit. If it were the case that $t \geqslant m$, then the polynomial

$$f - g(q + X^{t-m}cb^{-1}) = r - gX^{t-m}cb^{-1}$$

is an element of S and has degree strictly less than that of r. (The only way the we could have $\deg(r) = \deg(r - gX^{t-m}cb^{-1})$ is if $m = t = 0$ which would imply that $r - gX^{t-m}cb^{-1} = 0 \in S$.) Since this contradicts the choice of r, we conclude that $t < m$.

To see that q and r are uniquely determined by f and g, suppose that q', r' are polynomials in $F[X]$ that satisfy the conclusion of the theorem. Then $gq + r = gq' + r'$ which implies that $g(q - q') = r' - r$. If $r = r' = 0$, then we must also have $q - q' = 0$, as $g \neq 0$. If at least one of r and r' is nonzero, then $\deg(r' - r) < \deg(g)$ and it must be the case that $q - q' = 0$, and therefore also $r - r' = 0$. □

Remark. The condition that F is a field can be relaxed. It is enough to insist that it be an ID and that b, the leading coefficient of g, be a unit.

Corollary 8.2. *Let $f \in F[X]$. Then $a \in F$ is a root of f if and only if $(X - a) \mid f$.*

Proof. If $f = (X - a)q$, it is clear that a is a root of f. Conversely, suppose that a is a root of f. Let $g = (X - a)$ and apply the theorem to conclude that $f = (X - a)q + r$, where either $r = 0$ or $\deg(r) < 1$. The expression for f gives $f(a) = r$ since $\deg(r) = 0$, and therefore $r = 0$ and $f = (X - a)q$. □

An immediate consequence is the following.

Theorem 8.3 (Vandermonde's Theorem)**.** *A polynomial equation of degree n over a field has at most n roots.* □

8.2 $F[X]$ is a PID

Theorem 8.4. *Let F be a field. The polynomial ring $F[X]$ is a PID.*

Remark. The polynomial ring in two (or more) variables, is *not* a PID, although it is, as we shall see shortly, a UFD. The ideal $\langle X, Y \rangle \lhd \mathbb{R}[X, Y]$ is not principal.

Proof. Let I be an ideal in $F[X]$. We need to show that I is principal. If $I = \{0\}$, then we are done as $I = \{0\} = \langle 0 \rangle$. So assume that $I \neq 0$, and let $g \in I - \{0\}$ be an element of minimal degree amongst all elements of $I - \{0\}$. If $\deg(g) = 0$, then g is a unit (since F is a field) and $I = (g) = R$, so we are done. We may assume then that $\deg(g) \geqslant 1$. Let $f \in I$. By the previous theorem, 8.1, we know that $f = qg + r$ with $\deg(r) < \deg(g)$. But since $r = f - qg$, $f, g \in I$ and I is an ideal, we know that $r \in I$. We conclude that $r = 0$, since g has minimal degree in $I - \{0\}$. Having shown that any element $f \in I$ can be written as a multiple of g, we know that $I = \langle g \rangle$, and therefore I is principal. □

Remark. The above proof is entirely analogous to the proof that $\mathbb{Z}$ is a PID.

Example 8.5. Since $\mathbb{R}[X]$ is a PID and $X^2 + X + 1 \in \mathbb{R}[X]$ is irreducible, the ideal it generates $\langle X^2 + X + 1\rangle$ is maximal, and therefore the quotient ring $\mathbb{R}[X]/\langle X^2 + X + 1\rangle$ is a field.

Similarly the quotient ring $\mathbb{F}_2[X]/\langle X^2 + X + 1\rangle$ is a field. It has exactly 4 elements. The ring $\mathbb{F}_3[X]/\langle X^2 + X - 1\rangle$ is also a field. It has 9 elements.

In general, if $\mathbb{F}$ is a field and $f \in \mathbb{F}[X]$ is irreducible, then the quotient ring $\mathbb{F}[X]/\langle f\rangle$ is a field. Moreover, if $\mathbb{F}$ is a finite field, then $\mathbb{F}[X]/\langle f\rangle$ is finite and has $|\mathbb{F}|^{\deg(f)}$ elements.

8.3 Exercises

53. Let $f, g \in \mathbb{Z}/5\mathbb{Z}[X]$ be given by $f(x) = X^4 - 3X^3 + 2X^2 + 4X - 1$ and $g(x) = X^2 - 2X + 3$. Use 'long division' to find $q, r \in \mathbb{Z}/5\mathbb{Z}[X]$ such that $\deg(r) < \deg(g)$ and $f = qg + r$.
54. Using Corollary 8.2, show that $X^2 - 2$ is irreducible in $\mathbb{Q}[X]$. Show that $X^2 - 2$ *is not* irreducible in $\mathbb{R}[X]$.
55. Show that $f(X) = X^3 + 3X + 2$ is irreducible in $(\mathbb{Z}/5\mathbb{Z})[X]$.
56. Show that $R[X]/\langle X - a\rangle$ is isomorphic to R for any $a \in R$.
57. If we regard the reals $\mathbb{R}$ as a subring of the complex numbers $\mathbb{C}$, we can extend the inclusion to a homomorphism $\varphi : \mathbb{R}[X] \to \mathbb{C}$ by defining $\varphi(X) = i \in \mathbb{C}$. Show that φ induces an isomorphism $\mathbb{R}[X]/\langle X^2 + 1\rangle \cong \mathbb{C}$.
58. Let R be an integral domain. If $f, g \in R[X]$ and if the highest order coefficient of g is a unit, show that $\exists\, q, r \in R[X]$ such that
(a) $f = gq + r$, and
(b) either $r = 0$ or $\deg(r) < \deg(g)$.
59. Show that if $R[X]$ is a PID, then R is a field. (This is the converse of Theorem 8.4.)
60. Which are fields? (a) $\mathbb{Q}[X]/\langle X^2 - 5X + 6\rangle$ (b) $\mathbb{Q}[X]/\langle X^2 - 6X + 6\rangle$
61. (Rational Root Test) Show that if the reduced fraction r/s is a root of $f(X) = a_0 + a_1X + \cdots + a_nX^n \in \mathbb{Z}[X]$, then $r|a_0$ and $s|a_n$. Deduce that if f is monic and has a rational root, then it has an root that is an integer that divides a_0.
62. List all the maximal ideals in the following rings:

(a) $\mathbb{R}[X]/\langle X^2\rangle$ (b) $\mathbb{R}[X]/\langle X^2 + 1\rangle$ (c) $\mathbb{C}[X]/\langle X^2 + 1\rangle$

9 Every PID is a UFD

To show that an integral domain is a unique factorization domain we need to establish that every (non-unit, non-zero) element can be written as a product of (finitely many) irreducibles, and that this factorization is essentially unique. For existence we use the 'ascending chain condition', and for uniqueness the property that every irreducible element is prime (in a PID).

9.1 Ascending chain condition

Definition 9.1. Let R be a commutative ring. Then we say that R satisfies the **ascending chain condition** (ACC for short) if for every chain of ideals in R

$$I_1 \subseteq I_2 \subseteq \cdots \subseteq I_i \subseteq \cdots$$

there is a $N \in \mathbb{N}$ such that $I_i = I_N$ for all $i \geqslant N$.

Rings that satisfy the ACC (or equivalent) are called **Noetherian**.

Remark. The famous *Hilbert Basis Theorem* states that if R is Noetherian, then so too is $R[X]$. See, for example, [Art91, p. 469].

Examples 9.2.

1. $\mathbb{Z}$ satisfies the ACC. Every ideal is of the form $\langle m\rangle$ for some $m \in \mathbb{Z}$.
2. $\mathbb{R}[X]$ satisfies the ACC, as we shall see shortly.
3. $\mathbb{R}[X_1, X_2, \ldots]$, the polynomial ring on infinitely many variables, is not Noetherian. The chain of ideals
$$\langle X_1\rangle \subset \langle X_1, X_2\rangle \subset \langle X_1, X_2, X_3\rangle \subset \cdots$$
never stabilizes.
4. Another example of a non Noetherian ring is $C(\mathbb{R})$ the ring of all continuous functions from $\mathbb{R}$ to itself. Defining $I_i = \{f : \mathbb{R} \to \mathbb{R} \mid f(x) = 0 \quad \text{for all} \quad |x| \leqslant i\}$ gives a chain of ideals that does not stabilize.

There is a natural process by which we can try to decompose an element as a product of irreducibles: just keep writing each factor as a product. When we do this in the integers, the process must eventually terminate because each factor has strictly smaller magnitude. The following proposition says that in a ring that satisfies the ACC, the process always eventually halts.

Proposition 9.3. *Let R be an integral domain. If R satisfies the ascending chain condition, then every non-unit, non-zero element of R can be written as a product of irreducibles.*

Proof. Let $a \in R$ be non-zero and not a unit. Suppose that we had an infinite sequence of non-trivial factorizations. Then we would have elements $a_0 = a, a_1, a_2 \ldots$ such that $a_{i+1} \mid a_i$ and $a_{i+1} \not\sim a_i$. But this would give an infinite ascending chain of ideals
$$\langle a\rangle \subsetneq \langle a_1\rangle \subsetneq \langle a_2\rangle \subsetneq \cdots$$
contradicting the hypothesis that R satisfies the ACC. □

Proposition 9.4. *Let R be a PID. Then R satisfies the ascending chain condition.*

Proof. Given ideals $I_i \lhd R$ with $I_1 \subseteq I_2 \subseteq \ldots$, let $I = \cup_{i=1}^{\infty} I_i$. Since I is an ideal, it is given by $I = \langle a\rangle$ for some $a \in R$. Then $a \in \cup_{i=1}^{\infty} I_i$ implies that $a \in I_N$ for some N, which implies that $\langle a\rangle \subseteq I_N$. Then for any $i \geqslant N$ we have $I \subseteq I_N \subseteq I_i \subseteq I$, which implies that $I_i = I_N = I$. □

Exercise 63. Adapt the above proof to show that if R is a commutative unital ring in which all ideals are finitely generated, then R satisfies the ACC. Then prove the converse!

9.2 Prime versus irreducible

We saw in Proposition 6.7 that, in an integral domain, prime elements are irreducible. The converse does not hold in general, but does in a PID. We saw in Exercise 43 that it holds in any UFD. Of course, we can't use that result here, as we have not yet shown that every PID is a UFD.

The following lemma says that if irreducible elements are prime, then factorizations are essentially unique.

Lemma 9.5. *Let R be an integral domain in which all irreducible elements are prime. Suppose that $a_1, \ldots, a_n, b_1, \ldots, b_m \in R$ are irreducible elements such that*
$$a_1 a_2 \ldots a_m \sim b_1 b_2 \ldots b_n$$
Then $m = n$ and there is a permutation π of $\{1, 2, \ldots, n\}$, such that $b_i \sim a_{\pi(i)}$.

Proof. If either m or n is equal to 1, then the result holds by the definition of irreducible element.

Suppose then that $m,n \geqslant 2$. Clearly, $a_1 \mid a_1a_2\ldots a_m$, so we must have that $a_1 \mid b_1b_2\ldots b_n$. Since a_1 is prime, this implies that $a_1 \mid b_i$ for some $i \in \{1,\ldots,n\}$. By re-ordering, we can assume that $i = 1$. Since $a_1 \mid b_1$ and b_1 is irreducible, we have that $a_1 \sim b_1$. The cancellation law then tells us that

$$a_2\ldots a_m \sim b_2\ldots b_n$$

and, by induction, we are done. □

Lemma 9.6. *In a PID, irreducible elements are prime.*

Proof. Applying Lemma 7.4, Corollary 7.3 and Lemma 7.5 gives

$$p \quad \text{irreducible} \quad \Longrightarrow \langle p\rangle \quad \text{is maximal} \quad \Longrightarrow \langle p\rangle \quad \text{is prime} \quad \Longrightarrow p \quad \text{is prime}$$

□

Remark. Once we have established that all PIDs are UFDs the above lemma follows from Exercise 43. However, we need the lemma in order to prove that PIDs are UFDs.

9.3 PID implies UFD

Assembling the results of the previous sections we have the following

Theorem 9.7. *Every principal ideal domain is a unique factorization domain.*

Proof. Let R be a PID and $a \in R$ a non-zero non-unit element. By Proposition 9.4, R satisfies the ascending chain condition, and therefore a can be written as a product of irreducibles by Proposition 9.3. That the second part of the definition of UFD is satisfied, is precisely the statement of Lemma 9.5, which applies by Lemma 9.6. □

Corollary 9.8. *For any field F, the polynomial ring $F[X]$ is a unique factorization domain.* □

9.4 Exercises

64. If R is a PID and $0 \neq p \in R$, then the following are equivalent:
    (a) the ideal $\langle p\rangle$ is prime;
    (b) p is an irreducible element;
    (c) $\langle p\rangle$ is a maximal ideal in R;
    (d) $R/\langle p\rangle$ is a field;
    (e) $R/\langle p\rangle$ is an integral domain.

    This statement collects the results of several earlier exercises and results. For this exercise you should write out a proof of these implications in the indicated order: each implies the next and the last implies the first. Note that this result applies to the case $R = F[X]$ where F is a field and p is a non-constant polynomial.

65. Factor the following into irreducibles in $\mathbb{Z}[i]$: (a) 5 (b) 7 (c) 4+3i

10 $\mathbb{R}[X_1,\ldots,X_n]$ is a UFD

In this section we show that if R is a UFD, then $R[X]$ is a UFD. It follows that $R[X,Y] = (R[X])[Y]$ is a UFD, and $R[X_1,\ldots,X_n]$ is a UFD. The tools we will use in the proof are greatest common divisors and the Gauss Lemma.

10.1 Greatest common divisors

In $\mathbb{Z}$ the greatest common divisor of two elements is often defined to be the largest amongst all common divisors. In other rings we do not have an ordering, and so can't use the same definition.

Definition 10.1. Let R be a commutative ring. A **greatest common divisor** (or gcd) of a finite number of elements $a_1, \dots, a_n \in R$ is an element $d \in R$ satisfying:

1. d is a common divisor: $d \mid a_i$ for all $i \in \{1, \dots, n\}$;
2. If d' is another common divisor, then $d'|d$.

It is clear from the second part of the definition that any two gcds are associates, but they need not be equal. In $\mathbb{Z}$, both 2 and -2 are gcds of the elements 4 and 6. In general, there does not necessarily exist a gcd. For example, the elements 6 and $2 + 2\sqrt{-5}$ have no gcd in $\mathbb{Z}[\sqrt{-5}]$. However, in a UFD any collection of elements has a gcd.

Lemma 10.2. *In a UFD, a greatest common divisor of any finite collection of elements (at least one of which is non-zero) exists.*

Proof. We first show that for any two elements $a, b \in R$, a gcd exists. If either element is a unit, then 1 is a greatest common divisor since anything that divides a unit divides 1. (Any other unit will also be a gcd.) If one of the elements is zero, then the other is a gcd. So suppose that both a and b are non-zero, and not units. Since R is a UFD, we have factorizations of a and b as products of irreducibles. Rearranging, we can write these factorizations as

$$\begin{aligned} a &= p_1^{m_1} \dots p_k^{m_k} \\ b &= p_1^{n_1} \dots p_k^{n_k} u \end{aligned}$$

where each p_i is irreducible, $m_i, n_i \geqslant 0$, u is a unit and no two of the p_i are associates (i.e., $p_i \sim p_j$ implies $i = j$). Let $d = p_i^{\min\{m_1, n_1\}} \dots p_k^{\min\{m_k, n_k\}}$. It is clear that d divides both a and b. To see that is a gcd, suppose that d' is another common divisor. We have an irreducible factorization $d' = c_1 \dots c_l$, which can be rewritten as $d' = q_1^{l_1} \dots q_{k'}^{l_k'}$ where no two of the irreducibles q_i are associates. Since d' is a common divisor of a and b, we must have that for all $i \in \{1, \dots k'\}$ there is a $j \in \{1, \dots k\}$ such that $q_i \sim p_j$ and $l_i \leqslant \min\{m_j, n_j\}$. It follows that $d' \mid d$, and d is a gcd of a and b.

The case in which there are three or more elements follows by induction and the observation that if d_m is a gcd of $\{a_1, \dots, a_m\}$ and d_{m+1} is a gcd of $\{d_m, a_{m+1}\}$, then d_{m+1} is a gcd of $\{a_1, \dots, a_m, a_{m+1}\}$. □

Example 10.3. The polynomial $X - 1$ is a gcd of $X^2 - 1, X^3 - 2X^2 - 5X + 6 \in \mathbb{C}[X]$

Definition 10.4. A collection of elements in a UFD is called **relatively prime** if its gcd is a unit.

Example 10.5. The polynomials $2X - 2$ and $2X^2 - 2X - 4$ are relatively prime in $\mathbb{C}[X]$.

10.2 Primitive polynomials and the Gauss Lemma

Definition 10.6. A polynomial $a_0 + a_1X + \cdots + a_nX^n \in R[X]$ is called **primitive** if it is non-constant and $\{a_0, \dots, a_n\}$ is relatively prime in R.

Notice that an element in $R[X]$ that is non-constant and irreducible is necessarily primitive.

Exercise 66. Prove the following lemma.

Lemma 10.7. *Let R be a UFD. Let $f \in R[X]$ be a non-constant polynomial. Then there exist $a \in R$ and a primitive polynomial $\hat{f} \in R[X]$ such that $f = a\hat{f}$. Moreover, a and $\hat{f}$ are unique up to associates.* □

The following lemma allows us to relate factorization in $\mathbb{Q}[X]$ and factorization in $\mathbb{Z}[X]$.

Lemma 10.8 (Gauss Lemma)**.** *Let R be a UFD. If $f, g \in R[X]$ are primitive, then so too is their product fg.*

Proof. Let $h = fg$. Suppose that $p \in R$ is an irreducible that divides all the coefficients of h. The natural projection homomorphism $R \to R/\langle p\rangle$ induces a homomorphism $\varphi : R[X] \to (R/\langle p\rangle)[X]$ (as in Exercise 33). Since R is a UFD and p is irreducible, p is prime (Exercise 43), which in turn implies that $R/\langle p\rangle$ is an integral domain (Lemma 7.5 and Proposition 7.2). Since p divides every coefficient in h, $\varphi(h) = 0$, which implies that $\varphi(f)\varphi(g) = 0$. Therefore, one of $\varphi(f)$ or $\varphi(g)$ must equal zero, which contradicts the hypothesis that they are primitive. □

Lemma 10.9. *Let R be a UFD and F its field of quotients. Let $f \in R[X]$, and $g_1, g_2 \in F[X]$ be such that $f = g_1g_2$. Then there exist $g_1', g_2' \in R[X]$ with $f = g_1'g_2'$ and $g_i' \sim g_i$. Moreover, if g_1 is in $R[X]$ and is primitive, we can take $g_1' = g_1$.*

Proof. Note first that if any of f, g_1 or g_2 is degree zero, then the result holds. We assume then that each has degree at least 1.

For each i there is a non-zero element $d_i \in R$ such that $h_i = d_ig_i$ is in $R[X]$. (This is sometimes referred to as 'clearing denominators.') Write each of f, h_1 and h_2 as a constant multiple of a primitive polynomial: $f = c\hat{f}$, $h_i = c_i\hat{h_i}$. Then

$$\begin{aligned}
f = g_1g_2 &\implies d_1d_2f = h_1h_2 \\
&\implies d_1d_2c\hat{f} = c_1c_2\hat{h_1}\hat{h_2} \\
&\implies \hat{f} \sim \hat{h_1}\hat{h_2} && \text{by Lemmas 10.7 and 10.8} \\
&\implies \hat{f} = u\hat{h_1}\hat{h_2} && \text{for some unit } u \in R \\
&\implies f = cu\hat{h_1}\hat{h_2} \\
&\implies f = g_1'g_2' && \text{where } g_1' = \hat{h_1} \text{ and } g_2' = cu\hat{h_2}
\end{aligned}$$

Note that if g_1 is in $R[X]$ and is primitive, then we may choose $d_1 = c_1 = 1$ and $\hat{h_1} = h_1 = g_1$. □

Corollary 10.10. *If $f \in R[X]$ is irreducible in $R[X]$ and $\deg(f) \geqslant 1$, then f is irreducible in $F[X]$.* □

To see that the hypothesis that $\deg(f) \geqslant 1$ is necessary consider $f = 2 \in \mathbb{Z}[X]$.

Corollary 10.11. *Let $f, g \in R[X]$ with f primitive. If f divides g in $F[X]$, then f divides g in $R[X]$.* □

10.3 R a UFD implies $R[X]$ a UFD

We will show that every polynomial in $R[X]$ is a product of irreducibles, and that in $R[X]$, irreducible elements are prime. As in Section 9, this is enough to show that $R[X]$ is a UFD.

To show that an element can be written as a product of irreducibles we will think if it as an element of $F[X]$, where F is the field of quotients of R. We know that $F[X]$ is a UFD, and so we have a factorization as a product of irreducibles in $F[X]$. In order to obtain irreducibles in $R[X]$ we use the following technical lemma.

Lemma 10.12. *If $f \in R[X]$ is primitive in $R[X]$ and irreducible in $F[X]$, then it is irreducible in $R[X]$.*

Proof. Suppose that we have $f = gh$ in $R[X]$. Considering this equation as being in $F[X]$ we conclude that one of g or h must be a unit in $F[X]$. Suppose g is a unit in $F[X]$. Then $\deg(g) = 0$, and since f is primitive and $f = gh$ it follows that g is a unit in R. □

Proposition 10.13. *Every non-zero, non-unit element in $R[X]$ can be written as a product of irreducibles.*

Proof. Let $f \in R[X]$ be non-zero, non-unit. If $\deg(f) = 0$ then, since R is a UFD, we can factorize as a product of irreducibles in R. Note that if $a \in R$ is irreducible, then it is also irreducible in $R[X]$. So assume that $\deg(f) \geqslant 1$. As an element of $F[X]$, $\hat{f}$ is non-zero and non-unit and can therefore be written as a product of elements that are irreducible in $F[X]$. We have

$$\begin{aligned} & \hat{f} = f_1 \cdots f_k \qquad \text{(where each } f_i \in F[X] \text{ is irreducible in } F[X]) \\ \Longrightarrow\ & \hat{f} = f_1' \cdots f_k' \qquad (f_i' \in R[X] \text{, Lemma 10.9, irreducible in } F[X]) \\ \Longrightarrow\ & f_i' \text{ is primitive (since } \hat{f} \text{ is) and irreducible in } R[X] \text{ for all } i \text{ (by Lemma 10.12)} \end{aligned}$$

□

Proposition 10.14. *Irreducible elements in $R[X]$ are prime.*

Proof. Let $f \in R[X]$ be irreducible, and suppose that $f \mid g_1 g_2$. The case in which $\deg(f) = 0$ follows from the fact that R is a UFD, and therefore irreducible elements in R are prime in R. So we assume that $\deg f \geqslant 1$. Then f is irreducible in $F[X]$ by Corollary 10.10 and therefore prime (in $F[X]$) as $F[X]$ is a PID. Therefore, in $F[X]$, f divides one of the g_i. It follows that f divides one of the the g_i in $R[X]$ by Corollary 10.11. □

Theorem 10.15. *If R is a unique factorization domain, then so too is $R[X]$.*

Proof. Follows from Lemma 9.5 and Propositions 10.13 and 10.14. □

Examples 10.16. $\mathbb{Z}[X], \mathbb{Z}[X, Y], \mathbb{R}[X, Y], \mathbb{R}[X_1, \ldots, X_n]$ are all UFDs. (None of them are PIDs.)

10.4 Exercises

67. True or false:
    (a) Every field is a UFD.
    (b) Every field is a PID.
    (c) Every PID is a UFD.
    (d) Every UFD is a PID.
    (e) In a UFD, any two irreducibles are associates.

(f) If D is a PID, then $D[X]$ is a PID.

(g) If D is a UFD, then $D[X]$ is a UFD.

(h) Irreducible elements in an integral domain are prime.

(i) In a UFD, if p is irreducible and $p|a$, then p appears in every factorisation of a.

68. Express the following as the product of a constant polynomial and a primitive polynomial:

    (a) $18X^2 - 12X + 48$ in $\mathbb{Z}[X]$;

    (b) $18X^2 - 12X + 48$ in $\mathbb{Q}[X]$;

    (c) $2X^2 - 3X + 6$ in $\mathbb{Z}/7\mathbb{Z}[X]$.

69. Factor $4X^2 - 4X + 8$ into a product of irreducibles in:

    (a) $\mathbb{Z}[X]$

    (b) $\mathbb{Q}[X]$

    (c) $\mathbb{Z}/11\mathbb{Z}[X]$

70. Prove that if R is a PID and $a, b \in R$, then any gcd of a, b can be written as an R-linear combination of a, b.

71. Let R be a PID and let S be an integral domain containing R. Let $a, b, d \in R$. If d is a gcd of a, b in R, show that d is a gcd of a, b in S.

72. If p, q are relatively prime in $\mathbb{Z}$, show that they are relatively prime in $\mathbb{Z}[i]$.

73. Show that in a UFD a gcd of da, db is d times a gcd of a, b.

74. Let R be an integral domain, and $a, b, d, d' \in R$. Show that if $d \sim d'$ and d is a gcd of a and b, then d' is a gcd of a and b.

75. Show that if $a = qb + r$, then d is a gcd of a and b if and only if d is a gcd of b and r.

76. Consider the homomorphism φ from $\mathbb{Z}[X]$ to the real numbers which is the identity on $\mathbb{Z} \subset \mathbb{Z}[X]$, and takes X to $(1 + \sqrt{2})$. Show that the kernel of φ is a principal ideal and find a generator for this ideal.

77. Show that $\mathbb{Z}[X]/\langle 2X - 1\rangle \cong \mathbb{Z}[1/2]$, where $\mathbb{Z}[1/2]$ denotes the smallest subring of $\mathbb{Q}$ that contains $\mathbb{Z}$ and $\frac{1}{2}$. Note that $\mathbb{Z}[1/2] = \{m/2^k \mid m \in \mathbb{Z}, k \in \mathbb{N}\}$.

11 Irreducible polynomials

Later we will be interested in rings of the form $F[X]/\langle f\rangle$. This is a field if and only if $f \in F[X]$ is irreducible. Deciding whether or not a polynomial in $F[X]$ is irreducible is not trivial.

Any linear polynomial in $F[X]$ is irreducible. Suppose $f \in F[X]$ has degree at least 2. Then if it has a root in F, it is not irreducible since it has a linear factor. This follows from Corollary 8.2. The *converse is, in general, false:* the polynomial $X^4 + 2X^2 + 1$ is not irreducible in $\mathbb{R}[X]$, but has no roots in $\mathbb{R}$. For low degree polynomials, however, the converse does hold.

Exercise 78. Let $f \in F[X]$ have degree 2 or 3. Show that f is irreducible if and only if it has no roots in F.

11.1 Eisenstein's Irreducibility Criterion

This gives a sufficient condition for an element in $\mathbb{Z}[X]$ to be irreducible in $\mathbb{Q}[X]$, and hence in $\mathbb{Z}[X]$ if it is primitive.

Although the results of this sections are stated for $\mathbb{Z}$ and $\mathbb{Q}$, they apply equally well to any UFD (in place of $\mathbb{Z}$) and its field of quotients (in place of $\mathbb{Q}$).

Theorem 11.1. *Let $f = a_0 + a_1X + \cdots + a_nX^n \in \mathbb{Z}[X]$ with $n \geqslant 1$. Suppose there is a prime integer $p \in \mathbb{Z}$ such that:*

(a) p divides a_i for all $i \in \{0, \ldots, n-1\}$

(b) p does not divide a_n

(c) p^2 does not divide a_0

Then f is irreducible in $\mathbb{Q}[X]$.

Proof. Suppose, for a contradiction, that f is reducible in $\mathbb{Q}[X]$. It follows from Lemma 10.9 that $f = gh$ for some $g, h \in \mathbb{Z}[X]$ with g and h of degree at least 1. Let F_p be the field $\mathbb{Z}/\langle p\rangle$, and consider the homomorphism $\varphi : \mathbb{Z}[X] \to F_p[X]$ induced by the projection $\mathbb{Z} \to \mathbb{Z}/\langle p\rangle$, $a \mapsto \overline{a} = a + \langle p\rangle$. The conditions of the theorem ensure that $\overline{a}_n \neq 0$ and $\varphi(f) = \overline{a}_n X^n$. Since φ is a homomorphism, $\varphi(g)\varphi(h) = \overline{a}_n X^n$. This implies that $\varphi(g) = \alpha X^k$ and $\varphi(h) = \beta X^m$ with $k + m = n$. Note that $k = \deg(g) \geqslant 1$ and $m = \deg h \geqslant 1$. It follows that both the constant term of g and the constant term of h are divisible by p. This contradicts (c). □

Example 11.2. Using Eisenstein's criterion we conclude that the polynomial $X^4 + 50X^2 + 30X + 20$ is irreducible in $\mathbb{Q}[X]$.

Corollary 11.3. *Let $p \in \mathbb{N}$ be prime. The polynomial $X^{p-1} + X^{p-2} + \cdots + X + 1$ is irreducible in $\mathbb{Q}[X]$.*

Proof. Let $f(X) = (X^{p-1} + X^{p-2} + \cdots + X + 1)$. Substituting $X = Y + 1$ (cf., Exercise 35) gives

$$\begin{aligned}
(X-1)f(X) &= X^p - 1\\
\Longrightarrow\ Yf(Y+1) &= (Y+1)^p - 1\\
&= Y^p + \binom{p}{1}Y^{p-1} + \cdots + \binom{p}{p-1}Y\\
\Longrightarrow\ f(Y+1) &= Y^{p-1} + \binom{p}{1}Y^{p-2} + \cdots + \binom{p}{p-1}
\end{aligned}$$

The last polynomial is irreducible by Theorem 11.1. To see that it satisfies the hypotheses, note that since $\binom{p}{i} = p(p-1)\dots(p-i+1)/(i!)$ is an integer, if $i < p$, then $i!$ divides $(p-1)\dots(p-i+1)$. It follows that $\binom{p}{i}$ is divisible by p whenever $1 \leqslant i < p$. Also, $\binom{p}{p-1} = p$ is not divisible by p^2.

Having shown that $f(Y+1)$ is irreducible, we conclude that $f(X)$ is irreducible, since otherwise the isomorphism of Exercise 35 would give a contradiction. □

The factorization $X^p - 1 = (X-1)(X^{p-1} + X^{p-2} + \cdots + X + 1)$, is therefore a factorization into irreducibles.

11.2 Computation modulo p

Proposition 11.4. *Let $f = a_0 + a_1X + \cdots + a_nX^n \in \mathbb{Z}[X]$ with $n \geqslant 1$ and $p \in \mathbb{N}$ a prime that does not divide a_n. If $\overline{f} = \overline{a}_0 + \overline{a}_1X + \cdots + \overline{a}_nX^n \in (\mathbb{Z}/\langle p\rangle)[X]$ is irreducible, then f is irreducible in $\mathbb{Q}[X]$.*

Proof. Suppose f is reducible in $\mathbb{Q}[X]$. Then $f = gh$ with $g, h \in \mathbb{Z}[X]$ each of degree at least 1, and $\overline{f} = \overline{g}\overline{h}$ which, since $\overline{f}$ is irreducible, implies that one of $\overline{g}$ or $\overline{h}$ is a unit in $(\mathbb{Z}/\langle p\rangle)[X]$. Say $\overline{g}$ is a unit. It follows that the highest order coefficient in g is divisible by p. This contradicts the fact that the highest order coefficient of f is not divisible by p. □

11.3 A factorization algorithm for $\mathbb{Z}[X]$

There is a systematic, though possibly very long, method to factorize any polynomial in $\mathbb{Z}[X]$. We outline it here for interest. Given $f \in \mathbb{Z}[X] \setminus \{0, 1, -1\}$ with $\deg(f) = n$ we can proceed as follows:

1. If $n = 0$, then factorize in $\mathbb{Z}$.
2. Otherwise, let $m = \lfloor \frac{n}{2} \rfloor \in \mathbb{N}$, and calculate $f(0)$, $f(1)$,..., $f(m)$.

(a) If $f(a) = 0$ for some $0 \leqslant a \leqslant m$, then $(X - a)$ is a factor of f. If $f = \pm(X - a)$, then f is irreducible. If not, f is reducible. Write $f = (X - a)f'$ and start again.

(b) If $f(a) \neq 0$ for all $a \in \{0, 1, \ldots, m\}$, let $D = \{(d_0, d_1, \ldots, d_m) \in \mathbb{Z}^{m+1} \mid d_i$ is a divisor of $f(i)\}$. This is a finite set. For each $d = (d_0, d_1, \ldots, d_m) \in D$ let $g_d \in \mathbb{Q}[X]$ be the unique polynomial with $\deg(g_d) \leqslant m$ and $g(i) = d_i$ for all $i \in \{0, 1, \ldots, m\}$.

i) If there is a $d \in D$ such that g_d is a proper factor of f in $\mathbb{Z}[X]$, then we write $f = g_d f'$ and start again.

ii) If no g_d is a proper factor of f, then f is irreducible.

It is left to the reader to convince themselves that this procedure works.

11.4 Exercises

79. Show that the following are irreducible in $\mathbb{Q}[X]$:

(a) $X^2 - 12$ (b) $8X^3 + 6X^2 - 9X + 24$ (c) $2X^{10} - 25X^3 + 10X^2 - 30$

80. Determine which of the following is irreducible in $\mathbb{Q}[X]$:

(a) $X^4 - 16X^2 + 4$ (b) $X^4 - 32X^2 + 4$

81. Test for irreducibility the following polynomials $\mathbb{Q}[X]$:

(a) $X^4 - X^3 - X^2 - X - 2$ (c) $7X^3 + 6X^2 + 4X + 4$

(b) $2X^4 - 5X^3 + 3X^2 + 4X - 6$ (d) $9X^4 + 4X^3 - X + 7$

82. Test each of the following for irreducibility in $\mathbb{Q}[X]$:

(a) $X^5 - 4X + 22$ (c) $X^4 + 1$

(b) $2X^5 + 12X^4 - 15X^3 + 18X^2 - 45X + 3$

83. Let $n \geqslant 1$.

(a) Show that there is an irreducible polynomial of degree n in $\mathbb{Q}[X]$.

(b) Show that there are infinitely many (non associate) irreducible polynomials of degree n in $\mathbb{Q}[X]$.

84. Factor $X^5 + 5X + 5$ into irreducible factors in $\mathbb{Q}[X]$ and in $\mathbb{F}_2[X]$.

85. Factorize $X^3 + X^2 + 1$ in $\mathbb{F}_p[X]$, for $p = 2, 3$.

86. List all monic polynomials of degree ≤ 2 in $\mathbb{F}_3[X]$. Determine which of these polynomials are irreducible.

87. Determine all irreducible polynomials of degree ≤ 4 in $\mathbb{F}_2[X]$.

88. By considering images in $\mathbb{F}_2[X]$, show that the following are irreducible in $\mathbb{Q}[X]$:

(a) $X^2 + 2345X + 125$ (b) $X^3 + 5X^2 + 10X + 5$

12 Euclidean Domains

The greatest common divisor of two integers can be efficiently calculated using the well-known Euclidean Algorithm. The essential tool is the idea of division with remainder, with the remainder being ‘simpler’ than the original. In $\mathbb{Z}$ ‘simpler’ means it has smaller absolute value. In $\mathbb{R}[X]$ ‘simpler’ means it has lower degree (see section 8.1). We make a definition of this property.

12.1 Definition of Euclidean domain

Definition 12.1. A **Euclidean Domain** (or ED for short) is an integral domain R such that there exists a function $\sigma : R \setminus \{0\} \to \mathbb{N}$ satisfying

$$\forall\, a, b \in R \text{ with } b \neq 0,\ \exists\, q, r \in R \text{ such that } a = bq + r \text{ and either } r = 0 \text{ or } \sigma(r) < \sigma(b)$$

The function σ is called a **norm function**.

Remark.

1. The definition does *not* require that q and r be unique.
2. There can be many different maps σ that show that R is a Euclidean domain.
3. Given such a σ, define a new function $\sigma' : R \setminus \{0\} \to \mathbb{N}$ by $\sigma'(a) = \min\{\sigma(ab) \mid b \in R \setminus \{0\}\}$. Then σ' satisfies the above property plus the additional property:

$$\forall\, a, b \in R \setminus \{0\},\ \sigma'(a) \leqslant \sigma'(ab)$$

This additional property can be useful when considering units in R.

To see that σ' is a norm function: Let $a, b \in R$ with $b \neq 0$ and suppose that b does not divide a. Let $c \in R \setminus \{0\}$ be such that $\sigma'(b) = \sigma(bc)$. Then $\exists\, q, r' \in R$ such that $(ac) = q(bc) + r'$ and $\sigma(r') < \sigma(bc)$. Letting $r = a - qb$ we have: $r' = rc$ and $a = qb + r$ and $\sigma'(r) \leqslant \sigma(rc) = \sigma(r') < \sigma(bc) = \sigma'(b)$.

Example 12.2. It follows from Theorem 8.1 that, for any field F, $F[X]$ is a ED, with a suitable function being $\sigma(f) = \deg(f)$.

Example 12.3 (cf. Example 6.8)**.** We show that the Gaussian integers, $\mathbb{Z}[i] = \{x + iy \mid x, y \in \mathbb{Z}\}$, form a Euclidean domain. Let's define $\sigma : \mathbb{Z}[i] \to \mathbb{N}$ by $\sigma(x + iy) = |x + iy|^2 = x^2 + y^2$. Let $a, b \in \mathbb{Z}[i]$ be given by $a = a_i + ia_2$, $b = b_1 + ib_2$, $b \neq 0$. Define $w \in \mathbb{C}$ by $w = ab^{-1}$, where we regard $\mathbb{Z}[i]$ as a subset of $\mathbb{C}$ in the obvious way. Choose $q \in \mathbb{Z}[i]$ such that $|w - q| \leqslant 1/\sqrt{2}$. Then $a = bw = bq + b(w - q)$ and $\sigma(b(w - q)) = |b|^2|w - q|^2 = \sigma(b)|w - q|^2 \leqslant \sigma(b)/2 < \sigma(b)$. Setting $r = b(w - q)$, and noting that $b(w - q) = a - bq \in \mathbb{Z}[i]$, we are done. Notice that the choice of q is not, in general, unique.

Theorem 12.4. *Every Euclidean domain is a principal ideal domain.*

Proof. The argument is essentially the same as the one used to show that $F[X]$ is a PID (Theorem 8.4).

Suppose that R is a ED with $\sigma : R \setminus \{0\}$ as in the definition. Let $I \lhd R$ be an ideal. We need to show that I is principal. If $I = \{0\}$, there is nothing to show, so assume that $I \neq 0$. Choose $b \in I$ such that $b \neq 0$ and $\sigma(b) = \min\{\sigma(c) \mid c \in I \setminus \{0\}\}$. We will show that $I = \langle b \rangle$. For any $a \in I$, we have that there are $q, r \in R$ such that $a = bq + r$ and either $r = 0$ or $\sigma(r) < \sigma(b)$. Since $r = a - bq \in I$, it must be that $r = 0$. Therefore $a \in \langle b \rangle$. □

Remark. We have now shown the following implications:

$$ED \implies PID \implies UFD \implies ID$$

None of the reverse implications hold. The only counterexample we haven't seen is that of a PID that is not a ED. Such an example is given in the exercises.

12.2 The Euclidean algorithm

This algorithm for finding the greatest common divisor of two elements in a Euclidean Domain proceeds exactly as for the integers, with the usual 'division algorithm' replaced by the defining property of a ED. Our main application will be to polynomials over a field.

Euclidean Algorthm. Let R be a ED with norm function σ. Given two elements $a, b \in R$ with $b \neq 0$, proceed as follows:

0. Let $i = 0$, $a_0 = a$, $b_0 = b$.
1. Write $a_i = b_i q_i + r_i$ with $r_i = 0$ or $\sigma(r_i) < \sigma(b_i)$.
2. If $r_i = 0$, then stop with answer b_i.
3. Otherwise, let $a_{i+1} = b_i$ and $b_{i+1} = r_i$.
4. Increment i by one, and go to step 1.

Proof. We will prove that this procedure eventually terminates, and that the answer produced is a $\gcd(a, b)$. From Exercise 75 we know that $\gcd(a_{i+1}, b_{i+1}) = \gcd(a_i, b_i)$. Noting that a_i is a gcd of a_i and 0, we see that if the procedure stops, then the output is indeed a gcd of a and b. That the procedure stops follows from the fact that $0 < \sigma(b_{i+1}) < \sigma(b_i) < \sigma(b)$. □

By working back through the algorithm we can find an expression for the gcd as an R-linear combination of a and b.

Example 12.5. To illustrate, we use the algorithm to find a gcd of $X^3 + 2X^2 + 4X - 7, X^2 + X - 2 \in \mathbb{R}[X]$. Using 'long division' we obtain:

$$X^3 + 2X^2 + 4X - 7 = (X^2 + X - 2)(X + 1) + (5X - 5)$$
$$X^2 + X - 2 = (5X - 5)(\frac{1}{5}X + \frac{2}{5}) + 0$$

So a gcd is $(5X - 5)$ and

$$5X - 5 = (X^3 + 2X^2 + 4X - 7) - (X + 1)(X^2 + X - 2)$$

12.3 Exercises

89. Show that every field is a ED (you should give an explicit norm function).
90. Let $\xi \in \mathbb{C}$ be the root of the polynomial $X^2 + X + 1$ given by $\xi = (-1 + \sqrt{-3})/2$. Define the **Eisenstein Integers** as $\mathbb{Z}[\xi] = \{a + b\xi \mid a, b \in \mathbb{Z}\}$. Show that $\mathbb{Z}[\xi]$ is a Euclidean domain.
91. Find a gcd of $X^3 - 6X^2 + X + 4$ and $X^5 - 6X + 1$ in $\mathbb{Q}[X]$.
92. Consider the polynomials $f = X^3 - 6X^2 + X + 4$ and $g = X^4 - 6X^3 + 5$ in $\mathbb{Q}[X]$. Find a gcd d of f and g and then find polynomials a and b in $\mathbb{Q}[X]$ such that $d = af + bg$.
93. Use the Euclidean algorithm to calculate $\gcd(X^3 + 2X^2 + 4X - 7, X^2 + X - 2)$ in $\mathbb{Q}[X]$, and express it as a linear combination of the two polynomials.
94. This exercise is optional. Let $\eta = (1 + \sqrt{-19})/2$. Using the following steps, show that $\mathbb{Z}[\eta] = \{x + y\eta \mid x, y \in \mathbb{Z}\}$ is a PID but not a ED.
    (This is a hard question! Feel free to skip it. It is here mainly so that we have a concrete example to show that not every PID is a ED.)
    (a) Show that the only units in $\mathbb{Z}[\eta]$ are 1 and -1.
    (b) Show that 2 and 3 are irreducible in $\mathbb{Z}[\eta]$.
    (c) Now suppose the $\mathbb{Z}[\eta]$ is a ED with norm function σ satisfying $\sigma(a) \leqslant \sigma(ab)$. Show that the set of elements in $\mathbb{Z}[\eta] \setminus \{0\}$ that minimize σ is exactly $\{1, -1\}$.
    (d) Let m be an element of $\mathbb{Z}[\eta] \setminus \{0, 1, -1\}$ that achieves the minimum of σ on that set. By writing $2 = mq + r$ with $\sigma(r) < \sigma(m)$ or $r = 0$, show that $m \in \{-2, 2, -3, 3\}$.
    (e) By writing $\eta = mq + r$ with $\sigma(r) < \sigma(m)$ or $r = 0$, derive a contradiction.

    This establishes that $\mathbb{Z}[\eta]$ is not a ED. Now to show that it is a PID.

(f) Let $N : \mathbb{C} \to \mathbb{R}$ be given by $N(z) = z\overline{z}$ (i.e., the square of the absolute value). Show that given $a, b \in \mathbb{Z}[\eta]$ with $N(b) \geqslant N(a)$ and $a \nmid b$, there exist $c, d \in \mathbb{Z}[\eta]$ such that $0 < N(ad - bc) < N(a)$.

(g) Use the preceding part to show that every ideal in $\mathbb{Z}[\eta]$ is principal as follows: Given a non-zero ideal $I \lhd \mathbb{Z}[\eta]$, let $a \in I$ minimize N among nonzero elements of I. Show that any other element of I is a multiple of a.

Chapter II

Modules

13 Fundamental concepts

A module is a generalisation of a vector space in which the scalars do not necessarily form a field, but may be any (commutative unital) ring. Roughly speaking, an R-module is an abelian group on which the ring R acts linearly. A module in which the scalars are a field is the same as a vector space. A module in which the scalars are the integers is the same as an abelian group.

The main result we will obtain is a structure theorem for finitely generated modules in the case where the scalars are a PID. This is then used to obtain the structure theorem for finitely generated abelian groups. Using the same techniques we derive the Jordan Normal Form of a linear transformation of a complex vector space.

13.1 Definition

Definition 13.1. Let R be a commutative unital ring. An R-**module** M is an abelian group (whose operation we will denote by addition) together with a map $R \times M \to M$ satisfying the following for all $\rho, \sigma \in R$ and all $m, n \in M$:

(i) $1m = m$

(ii) $(\rho\sigma)m = \rho(\sigma m)$

(iii) $(\rho + \sigma)m = \rho m + \sigma m$

(iv) $\rho(m + n) = \rho m + \rho n$

We also call M a 'module over R', or simply a 'module'. The elements of the ring R and of M will often be referred to as **scalars** and **vectors** respectively. We will sometimes denote an R-module M by ${}_RM$.

Note. For the remainder of this chapter the ring R is always assumed to be commutative and unital.

Examples 13.2.

1. If R is a field then an R-module is simply a vector space over R, since the definition then becomes exactly that of a vector space.
2. A ring R is an R-module. If $I \lhd R$ is an ideal, then I is an R-module.
3. $R^n = \{(r_1, \dots, r_n) \mid r_i \in R\}$ is an R-module. The operations are the usual coordinatewise addition and scalar multiplication:

$$\begin{aligned}(r_1, \dots, r_n) + (s_1, \dots, s_n) &= (r_1 + s_1, \dots, r_n + s_n)\\ r(r_1, \dots, r_n) &= (rr_1, \dots, rr_n)\end{aligned}$$

4. Any abelian group can be regarded as a $\mathbb{Z}$-module, and vice-versa.
5. $R[X]$ forms a module over R.

13.2 Submodules, homomorphisms, quotients and products

Definition 13.3. A **submodule** of an R-module M is a subset that itself forms an R-module when using the operations inherited from M.

Exercise 95. Show that a subset $N \subseteq M$ is a submodule if and only if the following hold

(i) N is non-empty

(ii) $u, v \in N \implies u + v \in N$ (closed under vector addition)

(iii) $u \in N$, $\rho \in R \implies \rho u \in N$ (closed under scalar multiplication)

Example 13.4. 1. If $I \lhd R$ is an ideal, then ${}_R I$ is a submodule of ${}_R R$.

2. If S is a commutative unital ring and R is a subring of S (such that $1_S \in R$), then S is an R-module.

Definition 13.5. An R-module **homomorphism** is a map $\varphi : V \to W$ between R-modules such that for all $u, v \in V$ and all $\rho \in R$:

(i) $\varphi(u + v) = \varphi(u) + \varphi(v)$

(ii) $\varphi(\rho u) = \rho\varphi(u)$

A bijective homomorphism is called an **isomorphism**.

Exercise 96. Show that $\ker(\varphi)$ is a submodule of V and that $\mathrm{Im}(\varphi)$ is a submodule of W.

Definition 13.6. Given a submodule W of V, the **quotient module** V/W is given by the (additive) cosets $\{v+W \mid v \in V\}$ with the operations

(i) $(u + W) + (v + W) = (u + v) + W$

(ii) $\rho(v + W) = \rho v + W$

Definition 13.7. Let U and V be two R-modules. The **direct product** of U and V, denoted $U \oplus V$, is the R-module with underlying set $\{(u, v) \mid u \in U, v \in V\}$ and the operations given by

$$(u, v) + (x, y) = (u + x, v + y)$$
$$\rho(u, v) = (\rho u, \rho v)$$

The direct product of a finite number of R-modules is defined similarly.

13.3 Exercises

97. Let M be an R-module. Show that for all $\rho \in R$ and $m \in M$ we have:
(a) $0_R m = 0_M$ (b) $\rho 0_M = 0_M$ (c) $(-\rho)m = -(\rho m) = \rho(-m)$

98. State and prove module versions of the three isomorphism theorems, and the correspondence theorem.

99. (a) Let M be an R-module. Suppose that U and V are two submodules of M satisfying
i) $U \cap V = \{0\}$, and
ii) $U + V = M$.
Show that $M \cong U \oplus V$.

(b) Let U and V be R-modules and $M = U \oplus V$. Define submodules U' and V' of M by $U' = \{(u, 0) \mid u \in U\}$ and $V' = \{(0, v) \mid v \in V\}$. Show that

i) $U' \cap V' = \{0\}$,

ii) $U' + V' = M$, and

iii) $U' \cong U$, $V' \cong V$

100. Show that if $N_i \subseteq M_i$, $1 \leq i \leq 2$ are R- modules, then

$$\frac{M_1 \oplus M_2}{N_1 \oplus N_2} \cong \frac{M_1}{N_1} \oplus \frac{M_2}{N_2}$$

101. Let R be a PID, $p \in R$ an irreducible element, $k \geqslant 1$ and let M be the R-module $R/\langle p^k \rangle$. Let $N = p^{k-1}M$.

(a) Show that N is a submodule of M.

(b) Show that N is contained in every non-zero submodule of M.
(Hint: Consider the surjective homomorphism $R \to M$, $a \mapsto a + \langle p^k \rangle$.)

14 Free modules and bases

In the study of vector spaces, the notion of a basis is extremely useful. We now consider the corresponding notion in a module.

Definition 14.1. Let S be a subset of a module M. The **submodule generated** by S is the intersection of all submodules of M that contain S. This is easily seen to be a submodule of M and is denoted by $\langle S \rangle$. If $\langle S \rangle = M$ we say that S is a **generating set** for M.

Exercise 102. Show that

$$\langle S \rangle = \{\rho_1 u_1 + \cdots + \rho_k u_k \mid k \in \mathbb{N}, \rho_i \in R, u_i \in S\}$$

Definition 14.2. A subset $S \subseteq M$ is called **linearly dependent** if there exist $\rho_1, \ldots, \rho_k \in R$ at least one of which is non-zero, and $u_1, \ldots, u_k \in S$ such that $\rho_1 u_1 + \cdots + \rho_k u_k = 0$. A subset that is not linearly dependent is called **linearly independent**.

Definition 14.3. A subset $S \subseteq M$ that is linearly independent and which is a generating set for M is called a **basis** of M. If there exists a basis for M, M is called a **free module**.

Remark. 1. All modules over a field are free.

2. For any R, R^n is a free R-module. A basis is $\{(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)\}$.

3. To see that, in general, not all modules are free consider the $\mathbb{Z}$-module C_2. Since $2u = 0$ for all u, we see that every nonempty subset is linearly dependent. There can not be a basis because there are no (non-empty) linearly independent sets.

4. Here is another example of a non-free module. Let $R = \mathbb{Z}[X]$ and $I \lhd R$ the ideal generated by $\{2, X\}$. Then I can be regarded as an R-module. It is not free however. It follows from the fact that I is not a principal ideal, that any generating set of ${}_RI$ must contain at least two elements. But if $u, v \in {}_RI$ are two distinct elements, the identity $vu + (-u)(v) = 0$ implies that no linearly independent subset of ${}_RI$ can contain two or more elements.

5. The ring $\mathbb{Z}/m\mathbb{Z}$ is free when considered as a module over itself, but is not free when considered as a $\mathbb{Z}$-module.

Lemma 14.4. *Let M be an R-module. A subset S of M is a basis of M if and only if every element of M can be written uniquely as a linear combination of elements from S.*

Exercise 103. Prove Lemma 14.4

Module homomorphisms from a free module to another module are determined by their effect on the elements in a basis.

Lemma 14.5. *Let M be an R-module and $S \subseteq M$ a basis of M. Then any map from S to an R-module N extends uniquely to a homomorphism from M to N. That is, given a map $f : S \to N$, there is a unique R-module homomorphism $\varphi : M \to N$ such that $\varphi|_S = f$.*

Proof. An element $u \in M$ can be written uniquely as a linear combination $u = \sum_{s \in S} u_s s$, where $u_s \in R$ and only finitely many of them are non-zero. Define a map $\varphi : M \to N$ by $\varphi(\sum_{s\in S} u_s s) = \sum_{s\in S} u_s f(s)$. To see that this is a homomorphism, let $u, v \in M$ be such that $u = \sum_{s\in S} u_s s$, $v = \sum_{s\in S} v_s s$, then

$$\begin{aligned}
\varphi(u+v) &= \varphi(\sum_{s\in S} u_s s + \sum_{s\in S} v_s s) = \varphi(\sum_{s\in S}(u_s+v_s)s) \\
&= \sum_{s\in S}(u_s+v_s)f(s) = \sum_{s\in S} u_s f(s) + \sum_{s\in S} v_s f(s) \\
&= \varphi(\sum_{s\in S} u_s s) + \varphi(\sum_{s\in S} v_s s) = \varphi(u) + \varphi(v) \\
\varphi(\rho u) &= \varphi(\rho \sum_{s\in S} u_s s) = \varphi(\sum_{s\in S} \rho u_s s) \\
&= \sum_{s\in S} \rho u_s f(s) = \rho \sum_{s\in S} u_s f(s) = \rho\varphi(u)
\end{aligned}$$

Suppose that $\psi : M \to N$ were another homomorphism satisfying $\psi|_S = f$. Then

$$\begin{aligned}
\psi(u) &= \psi(\sum_{s\in S} u_s s) = \sum_{s\in S} u_s \psi(s) && \text{(since } \psi \text{ is a homomorphism)} \\
&= \sum_{s\in S} u_s f(s) && \text{(since } \psi|_S = f\text{)} \\
&= \varphi(u)
\end{aligned}$$

□

The following is a direct analogue of the result for vector spaces.

Lemma 14.6. *If M is a free R-module with basis $\{u_1, \ldots, u_n\}$, then $M \cong R^n$.*

Proof. The map $\varphi : M \to R^n$ given by $\varphi(\sum_1^n \rho_i u_i) = (\rho_1, \ldots, \rho_n)$ is readily seen to be an isomorphism. □

Remark. Free modules share many of the properties of vector spaces, but not all. For example, even if a module is free, not every generating set necessarily contains a basis. Consider, for example, the generating set $\{2, 3\}$ for the $\mathbb{Z}$-module $\mathbb{Z}$. No subset of $\{2, 3\}$ is a basis for $\mathbb{Z}$. Also, the subset $\{2\} \subseteq \mathbb{Z}$ is a linearly independent set that can not be extended to a basis.

Proposition 14.7. *Suppose that R is an integral domain and $m, n \in \mathbb{N}$. Then $R^m \cong R^n$ (as R-modules) if and only if $m = n$.*

Proof. We will show that any linearly independent set in R^m has at most m elements, from which it follows that if $R^m \cong R^n$ then $m = n$. We use induction on m. The identity $uv - vu = 0$ shows that any subset of R that contains at least two elements is linearly dependent.

Now suppose that any linearly independent subset of R^{m-1} contains at most $m-1$ elements, and let $S \subseteq R^m$ be linearly independent. We want to show that $|S| \leqslant m$. Let $\pi : R^m \to R$ be the module homomorphism given by projection onto the first factor, that is, $\pi(r_1, \ldots, r_m) = r_1$. Note that $\ker(\pi) \cong R^{m-1}$. If S is contained in $\ker(\pi)$, then we have that $|S| \leqslant m - 1$, so we may assume that there exists $s \in S \setminus \ker(\pi)$. To each element of $S \setminus \{s\}$ we add a multiple of s so that the result lies in $\ker(\pi)$. To this end, note that if $x \in S \setminus \{s\}$ then $\pi(s)x - \pi(x)s \in \ker(\pi)$. Now consider the set $S' = \{\pi(s)x - \pi(x)s \mid x \in S \setminus \{s\}\}$. Then $S' \subseteq \ker(\pi)$ and S' is linearly independent since

$$\begin{aligned}
&\sum_x \mu_x(\pi(s)x - \pi(x)s) = 0 && \text{(for elements } \mu_x \in R\text{)}\\
\Longrightarrow\ &\sum_x \mu_x \pi(s)x - (\sum_x \mu_x \pi(x))s = 0 &&\\
\Longrightarrow\ &\forall\, x,\ \mu_x \pi(s) = 0 && \text{(since } S \text{ is linearly independent)}\\
\Longrightarrow\ &\forall\, x,\ \mu_x = 0 && \text{(since } \pi(s) \neq 0 \text{ and } R \text{ is an ID)}
\end{aligned}$$

As S' is a linearly independent subset of $\ker(\pi)$, we have $|S'| \leqslant m - 1$ and therefore $|S| = |S'| + 1 \leqslant m$. □

Remark. The theorem is false without the hypothesis that R be an integral domain. On the other hand, if R is finite, then the result holds whether or not R is an integral domain.

It follows from the previous two results that, when R is an integral domain, any two bases of a free R-module have the same number of elements. The number of elements in a basis is called the **rank** of the free R-module.

Proposition 14.8. *Every finitely generated R-module is a homomorphic image of a free R-module of finite rank.*

Proof. Let M be an R-module, and $\{u_1, \ldots, u_m\} \subseteq M$ a generating set. Fix a basis $\{e_1, \ldots, e_m\}$ for R^m. Define a homomorphism $\varphi : R^m \to M$ by extending the map that sends e_i to u_i (Lemma 14.5). Since $\mathrm{Im}(\varphi)$ contains a generating set, φ is surjective. □

It follows that every (finitely generated) R-module is isomorphic to F/N for some free module F and submodule N of F.

14.1 Exercises

104. Let R be a ring (commutative and unital) and V a free module of finite rank over R.
Prove or disprove:
(a) Every set of generators of V contains a basis of V;
(b) Every linearly independent set in V can be extended to a basis of V.

105. Let R be an integral domain and I an ideal in R. Show that I is free, when considered as an R-module, if and only if it is principal.

106. Let F and G be two free R-modules of rank m and n respectively. Show that the R-module $F \oplus G$ is free of rank $m + n$.

107. Show that if N and M/N are finitely generated as R-modules, then M is also a finitely generated R-module.

108. Show that $\mathbb{Q}$ is not finitely generated as a $\mathbb{Z}$-module.

109. A module is called **cyclic** if it has a generating set with one element.

   (a) Is a quotient module of a cyclic module cyclic?

   (b) Is a submodule of a cyclic module cyclic?

110. In each case write the $\mathbb{Z}$-module M/N as a direct sum of cyclic submodules.

(a) $M = \mathbb{Z} \oplus \mathbb{Z}$ and N the submodule generated by $(0, 3)$.

(b) $M = \mathbb{Z} \oplus \mathbb{Z}$ and N the submodule generated by $(2, 0)$ and $(0, 3)$.

(c) $M = \mathbb{Z} \oplus \mathbb{Z}$ and N the submodule generated by $(2, 3)$.

(d) $M = \mathbb{Z} \oplus \mathbb{Z}$ and N the submodule generated by $(6, 9)$.

111. Let V be a two dimensional vector space over $\mathbb{Q}$ having basis $\{v_1, v_2\}$. Let T be the linear transformation on V defined by $T(v_1) = 3v_1 - v_2$, $T(v_2) = 2v_2$. Make V into a $\mathbb{Q}[X]$-module by defining $X \cdot u = T(u)$.

(a) Show that the subspace $U = \{av_2 \mid a \in \mathbb{Q}\}$ of V spanned by v_2 is actually a $\mathbb{Q}[X]$-submodule of V.

(b) Consider the polynomial $f = X^2 + 2X - 3$. Determine the vectors $f \cdot v_1$ and $f \cdot v_2$, that is, express them as linear combinations of v_1 and v_2.

15 Torsion

Definition 15.1. The **annihilator** of an element $u \in M$ in an R-module is

$$\mathrm{ann}_R(u) = \{\rho \in R \mid \rho u = 0\}$$

An element $u \in M$ is said to be **torsion** if $\mathrm{ann}_R(u) \neq \{0\}$. The **torsion submodule** T_M consists of all torsion elements in M, that is,

$$T_M = \{u \in M \mid \exists\, \rho \in R \setminus \{0\}, \rho u = 0\}$$

The module M is said to be a **torsion module** if all elements in M are torsion, and **torsion-free** if zero is the only torsion element.

Exercise 112.

a) Show that $\mathrm{ann}_R(u)$ is an ideal in R.

b) Show that if R is an integral domain, then T_M is a submodule of M.

c) Let M be a free module over an integral domain R. Show that M is torsion-free.

d) Give an example of a torsion-free module over an integral domain that is not free. (Hint: The ring can not be a PID.)

Proposition 15.2. *Let M be a module over an integral domain R. The quotient module M/T_M is torsion-free.*

Proof. Let $\rho \in R$ be non-zero.

$$\begin{aligned} \rho(u + T_M) = 0 + T_M &\implies \rho u + T_M = 0 + T_M \\ &\implies \rho u \in T_M \\ &\implies \text{there is a non-zero } \sigma \in R \text{ such that } \sigma(\rho u) = 0 \\ &\implies (\sigma\rho)u = 0 \\ &\implies u \in T_M \ \text{(since } \rho \text{ and } \sigma \text{ are non-zero and } R \text{ is an integral domain)} \\ &\implies u + T_M = 0 + T_M \end{aligned}$$

□

15.1 Exercises

113. Let I be an ideal in an integral domain R. Show that $\mathrm{ann}_R(R/I) = I$.

114. Let M_1 and M_2 be two R-modules. Show that $\mathrm{ann}_R(M_1 \oplus M_2) = \mathrm{ann}_R(M_1) \cap \mathrm{ann}_R(M_2)$.

115. Show that R considered as a module over itself is torsion-free if and only if R is an integral domain.

116. Show that $\mathbb{Q}$ as a $\mathbb{Z}$-module is torsion-free but not free.

16 Submodules of free modules

In general, a submodule of a free module need not be free. For example, let $I = \langle 2, X\rangle \lhd \mathbb{Z}[X]$ be the ideal generated by 2 and X. Then when considered as a $\mathbb{Z}[X]$-module, I is not free. It is a submodule of a free module, namely $\mathbb{Z}[X]$ itself considered as a $\mathbb{Z}[X]$ module.

In this section we show that every submodule of a free module over a PID is itself free.

16.1 The splitting lemma

Lemma 16.1. *Let R be a commutative unital ring. Let F be a free R-module and M an R-module. Let $\varphi : M \to F$ be a surjective homomorphism. Then there exists a submodule $F' \subseteq M$ such that $F' \cong F$ and $M = F' \oplus \ker(\varphi)$.*

Proof. Let $X = \{x_i \mid i \in I\}$ be a basis for F. Since φ is surjective, there exist elements $u_i \in M$ such that $\varphi(u_i) = x_i$. The map $f : X \to M$ given by $f(x_i) = u_i$ extends to a homomorphism $\psi : F \to M$. Since $\varphi \circ \psi(x_i) = x_i$ we have that $\varphi \circ \psi = \mathrm{Id}_F$. It follows that ψ is injective. Letting $F' = \mathrm{Im}(\psi)$, we have that $F' \cong F$.

It remains to show that $M = F' \oplus \ker(\varphi)$. For any $u \in M$ we have $\psi \circ \varphi(u) \in F'$ and $u - \psi \circ \varphi(u) \in \ker(\varphi)$. It follows that $M = F' + \ker(\varphi)$. Let $v \in F' \cap \ker(\varphi)$. Then $v = \psi(w)$ for some $w \in F$, and also $\varphi(v) = 0$. Therefore $\varphi \circ \psi(w) = 0$, which implies that $w = 0$ because $\varphi \circ \psi = \mathrm{Id}_F$. If follows that $v = 0$ and therefore that $F' \cap \ker(\varphi) = \{0\}$. □

16.2 Submodules of free modules over a PID are free

Theorem 16.2. *Let R be a PID, and F a free R-module of finite rank r. Then every submodule of F is free and has rank at most r.*

Proof. We use induction on the rank r of F. If F has rank 1, then $F \cong {}_RR$ (Lemma 14.6), any submodule N of F is an ideal in R. Since R is a PID, the ideal is generated by a single element. If $N = \{0\}$, then N is free of rank 0. Otherwise $N = \langle u\rangle$ for some non zero $u \in R$ and $N \cong R$ (as an R-module).

For the induction, suppose that the conclusion of the theorem is true for all free R-modules of rank at most $r - 1$. Let $\{x_1, \dots, x_r\}$ be a basis for F, and let $F' \subseteq F$ be the submodule generated by $\{x_1, \dots, x_{r-1}\}$. Then F' is free and $\{x_1, \dots, x_{r-1}\}$ is a basis for it. Let $N \subseteq F$ be a submodule. We want to show that N is free and has rank at most r. Let $\pi : F \to F/F'$ be natural projection, and note that $F/F' \cong R$. Consider the restriction $\pi|_N : N \to F/F'$. Since F/F' is free of rank 1, we know that $\mathrm{Im}(\pi|_N)$ is free of rank at most 1. Also, $\ker(\pi|_N) = N \cap F' \subseteq F'$ is free of rank at most $r - 1$. By Lemma 16.1 $N = L \oplus (N \cap F')$ where $L \cong \mathrm{Im}(\pi|_N)$. Since the direct sum of two free modules is free and rank adds, N is free of rank at most r. □

16.3 Exercises

117. Suppose that R is a principal ideal domain. Let M be a non-trivial R-module which has no proper submodules (that is, the only submodules are M itself and $\{0\}$). Show that either R is a field and $M \cong R$ or R is not a field and $M \cong R/\langle p\rangle$ for some prime $p \in R$.

118. Let $R = \mathbb{Z}/6\mathbb{Z}$, and let F be the R-module R^2. Write down a basis for F. Let $N = \{(0, 0), (3, 0)\} \subseteq F$. Show that N is a submodule of F, and that N is not free. Why does this not contradict Theorem 16.2?

119. Let $R = \mathbb{Z}$ and $F = \mathbb{Z}^3$. Let $N = \{(x, y, x) \in F \mid x + y + z = 0\}$. Show that N is a submodule of F. Find a basis for N.

17 Matrices

We want to analyse the structure of finitely generated modules. We have already noted that any such module is isomorphic to F/N for some free module F. Since N is a submodule of a free module, and assuming that R is a PID, N is also free. The inclusion map $N \to F$ is a homomorphism. Homomorphisms between free R-modules can be represented by matrices over R. By considering such matrices, we will be able to say a lot about the structure of F/N.

17.1 The matrix of a homomorphism

Let R be an integral domain, and F and G two finitely generated free R-modules. Fix bases for $\mathcal{B} = \{f_1, \dots, f_m\}$ and $\mathcal{C} = \{g_1, \dots, g_n\}$ for F and G, every R-module homomorphism $\varphi : G \to F$ is represented by a unique matrix in $M_{m\times n}(R)$ as follows. For each element g_j in the basis for G write $\varphi(g_j)$ in terms of the basis for F, that is,

$$\varphi(g_j) = \sum_{i=1}^{m} a_{ij} f_i$$

The matrix (a_{ij}) is called the **matrix of the homomorphism** φ with respect to the given bases, and will be denoted by $[\varphi]_{\mathcal{B},\mathcal{C}}$ or simply $[\varphi]$.

Exercise 120. Show that for all $u \in G$,

$$[\varphi(u)]_{\mathcal{B}} = [\varphi]_{\mathcal{B},\mathcal{C}}[u]_{\mathcal{C}}$$

where $[u]_{\mathcal{C}}$ is the **coordinate matrix** of u with respect to $\mathcal{C}$, that is $[u]_{\mathcal{C}} = (u_{j1}) \in M_{n\times 1}$ is determined by the equation $u = \sum_{j=1}^{n} u_{j1} g_j$.

Definition 17.1. Two matrices $A, B \in M_{m\times n}(R)$ are said to be **equivalent** if there exist invertible matrices $X \in M_{m\times m}$ and $Y \in M_{n\times n}$ such that $A = XBY$.

Equivalent matrices represent the same homomorphism, but with respect to different choices of bases.

17.2 Equivalent diagonal matrices

Proposition 17.2. *Let R be a PID and $A \in M_{m\times n}(R)$. Then A is equivalent to a diagonal $m\times n$ matrix* $\operatorname{diag}(d_1, d_2, \dots, d_{\min\{m,n\}})$ *satisfying $d_1|d_2|\cdots|d_{\min\{m,n\}}$.*

Definition 17.3. The diagonal matrix as in the above proposition is called the **invariant factor matrix** of A or the **Smith normal form** of A.

Outline of proof. We will show that A is equivalent to a matrix in the form

$$\left[\begin{array}{c|ccc} d & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array}\right] \qquad (*)$$

where

(a) $B \in M_{(m-1)\times(n-1)}(R)$

(b) d and all entries in B are R-linear combinations of the entries from A, and

(c) d divides all entries in B.

Repeated application then gives the required result.

We describe how this can be done algorithmically in the case in which R is actually a Euclidean Domain with norm function σ.

Note that if we apply an elementary row or column operation, the new matrix is equivalent to the old. We will apply a sequence of row and column operations to put A into the form given in $(*)$. If all entries in A are zero, then it is already in the required form, so we assume that there is at least one non-zero entry. Then, by swapping rows and columns we can ensure that the top left entry a_{11} is non-zero. Suppose that some other entry in the first row of A is non-zero. Then by swapping columns we can assume that a_{12} is non-zero. Applying the Euclidean algorithm (using column operations) to the entries a_{11} and a_{12} we obtain a new matrix in which the first two entries in the new matrix are $d = \gcd(a_{11}, a_{12})$ and 0. The other entries in the first two columns will also have changed.

$$\lceil a_{11}\ a_{12}\ \cdots \rceil \xrightarrow{\text{column operations}} \lceil d\ 0\ \cdots \rceil$$

Repeating a finite number of times we obtain a matrix whose first row is of the form $\lceil d\ 0 \cdots 0 \rceil$. A similar process along the first column enables us to obtain a matrix in the required form $(*)$. One then needs to ensure that d divides all entries in B.

Suppose that there is an entry in B that is not divisible by d. Then apply the row operation that adds that row to the first row. The top left entry is still d, but there are other non-zero entries in the first row, at least one of which is not divisible by d. Now simply begin the whole process again, to clear all entries in the first row and first column, aside from the top left entry. How do we know that this process will eventually terminate with a matrix in the form $(*)$? The point is that after each iteration, the value of $\sigma(d)$ has been strictly decreased. Since $\sigma(d)$ is a natural number, this can only happen a finite number of times.

The general case, in which R is merely a PID, is very similar. In addition to elementary matrices we need to multiply by another sort of invertible matrix. In place of the Euclidean algorithm we use the fact (see Exercise 70) that in a PID we have $d = \gcd(a, b) = xa + yb$ for some $x, y \in R$. We have

$$\begin{aligned} d &= xa + yb && \\ &= d(xa' + yb') && \text{(where } a = da' \text{ and } b = db'\text{)} \\ 1 &= xa' + yb' && \text{(since } d \neq 0 \text{ and } R \text{ is an ID)} \end{aligned}$$

So

$$\begin{bmatrix} a & b & \cdots \\ \vdots & \ddots & \\ & & \\ & & \end{bmatrix} \left[\begin{array}{cc|ccc} x & -b' & 0 & \cdots & 0 \\ y & a' & 0 & \cdots & 0 \\ \hline 0 & 0 & & & \\ \vdots & \vdots & & I & \\ 0 & 0 & & & \end{array}\right] = \begin{bmatrix} d & 0 & \cdots \\ \vdots & \ddots & \\ & & \\ & & \end{bmatrix}$$

The second matrix is invertible, since its determinant is 1. In place of a norm function σ, we define a function $\lambda : R \setminus \{0\} \to \mathbb{N}$ by

$$\lambda(r) = \begin{cases} 0 & \text{if } r \text{ is a unit} \\ k & \text{if } r \text{ is not a unit and } r = p_1 \dots p_k \text{ with } p_i \text{ irreducible} \end{cases}$$

This is not a norm function, but can still be used to justify that the process terminates. Notice that $\lambda(ab) = \lambda(a)\lambda(b)$, and that if d divides a, but is not an associate of a, then $\lambda(d) < \lambda(a)$. □

Example 17.4. Beginning with the following matrix $A \in M_{2\times 3}(\mathbb{Z})$, we apply row and column operations to obtain a matrix in the diagonal form described in Proposition 17.2.

$$A = \begin{bmatrix} 6 & 4 & 4 \\ 4 & 8 & 0 \end{bmatrix} \xrightarrow{C1 \mapsto C1 - C2} \begin{bmatrix} 2 & 4 & 4 \\ -4 & 8 & 0 \end{bmatrix} \xrightarrow[C3 \mapsto C3 - 2C1]{C2 \mapsto C2 - 2C1} \begin{bmatrix} 2 & 0 & 0 \\ -4 & 16 & 8 \end{bmatrix}$$

$$\xrightarrow{R2 \mapsto R2 + 2R1} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 16 & 8 \end{bmatrix} \xrightarrow{C2 \mapsto C2 - 2C3} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 8 \end{bmatrix} \xrightarrow{C2 \leftrightarrow C3} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 8 & 0 \end{bmatrix} = D$$

We can obtain invertible matrices X and Y such that $XAY = D$ by applying the row and column operations to the identity matrices of the appropriate size.

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \xrightarrow{R2 \mapsto R2 + 2R1} \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} = X$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{C1 \mapsto C1 - C2} \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[C3 \mapsto C3 - 2C1]{C2 \mapsto C2 - 2C1} \begin{bmatrix} 1 & -2 & -2 \\ -1 & 3 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow{C2 \mapsto C2 - 2C3} \begin{bmatrix} 1 & 2 & -2 \\ -1 & -1 & 2 \\ 0 & -2 & 1 \end{bmatrix} \xrightarrow{C2 \leftrightarrow C3} \begin{bmatrix} 1 & -2 & 2 \\ -1 & 2 & -1 \\ 0 & 1 & -2 \end{bmatrix} = Y$$

Example 17.5. Starting with the matrix $A \in M_{3\times 3}(\mathbb{Q}[X])$ below we use row and column operations to put it into the diagonal form described in Proposition 17.2.

$$A = \begin{bmatrix} 1-X & 1+X & X \\ X & 1-X & 1 \\ 1+X & 2X & 1 \end{bmatrix} \xrightarrow{\text{C1}\leftrightarrow\text{C3}} \begin{bmatrix} X & 1+X & 1-X \\ 1 & 1-X & X \\ 1 & 2X & 1+X \end{bmatrix} \xrightarrow{\text{R1}\leftrightarrow\text{R3}} \begin{bmatrix} 1 & 2X & 1+X \\ 1 & 1-X & X \\ X & 1+X & 1-X \end{bmatrix}$$

$$\xrightarrow[R3 \mapsto R3 - XR1]{R2 \mapsto R2 - R1} \begin{bmatrix} 1 & 2X & 1+X \\ 0 & 1-3X & -1 \\ 0 & 1+X-2X^2 & 1-2X-X^2 \end{bmatrix}$$

$$\xrightarrow[C3 \mapsto C3 - (1+X)C1]{C2 \mapsto C2 - 2XC1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1-3X & -1 \\ 0 & 1+X-2X^2 & 1-2X-X^2 \end{bmatrix} \xrightarrow{C2 \leftrightarrow C3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 1-3X \\ 0 & 1-2X-X^2 & 1+X-2X^2 \end{bmatrix}$$

$$\xrightarrow{C3 \mapsto C3 + (1-3X)C2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1-2X-X^2 & 2-4X+3X^2+3X^3 \end{bmatrix}$$

$$\xrightarrow{R3 \mapsto R3 + (1-2X-X^2)R2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2-4X+3X^2+3X^3 \end{bmatrix} = D$$

17.3 Exercises

121. Let G be the group of units in $M_{2\times 2}(\mathbb{Z})$, that is, $G = GL_2(\mathbb{Z})$).

Show that G is generated by the set

$$\left\{ \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

122. For the matrices $A, D \in M_{3\times 3}(\mathbb{Q}[X])$ from Example 17.5 find invertible matrices $L, R \in M_{3\times 3}(\mathbb{Q}[X])$ satisfying $LAR = D$.

123. Given the matrix A, find invertible matrices L, R and elements d_1, d_2, d_3 such that $LAR = \operatorname{diag}(d_1, d_2, d_3)$ and $d_1 | d_2 | d_3$.

$$\text{(a) } A = \begin{pmatrix} 7 & 8 & 9 \\ 4 & 5 & 6 \\ 1 & 2 & 3 \end{pmatrix} \in M_{3\times 3}(\mathbb{Z}) \qquad \text{(b) } A = \begin{pmatrix} 1-X & 1+X & X \\ X & 1-X & 1 \\ 1+X & 2X & 1 \end{pmatrix} \in M_{3\times 3}(\mathbb{Q}[X])$$

124. Find the invariant factor matrices over $\mathbb{Z}$ for the first three of the following matrices, and over $\mathbb{Q}[X]$ for the last two of the following matrices:

$$\text{(a)} \begin{bmatrix} 3 & 1 \\ -1 & 2 \end{bmatrix} \qquad \text{(b)} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \qquad \text{(c)} \begin{bmatrix} -4 & -6 & 7 \\ 2 & 2 & 4 \\ 6 & 6 & 15 \end{bmatrix}$$

$$\text{(d)} \begin{bmatrix} X & 1 & -2 \\ -3 & X+4 & -6 \\ -2 & 2 & X-3 \end{bmatrix} \qquad \text{(e)} \begin{bmatrix} X & 0 & 0 \\ 0 & 1-X & 0 \\ 0 & 0 & 1-X^2 \end{bmatrix}$$

125. Let F be a free module of rank m over an integral domain R. Let $\operatorname{End}_R(F)$ denote the ring of all homomorphisms from F to itself. The operations being given by

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

Show that $\operatorname{End}_R(F) \cong M_{m\times m}(R)$ as rings

126. Let F be a free module over an integral domain R, and $\varphi : F \to F$ a homomorphism. Let $\mathcal{B} = \{f_1, \ldots, f_m\}$ be a basis for F. Show that the following are equivalent:

(a) $\{\varphi(f_1),\dots,\varphi(f_m)\}$ is a basis for F;

(b) φ is an isomorphism;

(c) The matrix $[\varphi]_{\mathcal{B},\mathcal{B}}$ is invertible.

127. Show that an $n\times n$ matrix over a PID is invertible if and only if it is equivalent to the identity matrix.

128. Let $f_1, f_2, ..., f_s$ be a basis of a free module V over a PID R. Suppose that $f = r_1f_1 + r_2f_2 + \cdots + r_sf_s$ and that 1 is a gcd of $r_1, r_2, \dots, r_s$. Show that f is a part of a basis for V.

129. Let $\varphi : \mathbb{Z}^k \to \mathbb{Z}^k$ be a homomorphism given by multiplication by an integer matrix A. Show that the image of φ has finite index (in $\mathbb{Z}^k$) if and only if $\det A \neq 0$, and that in this case the index of $\varphi(\mathbb{Z}^k)$ in $\mathbb{Z}^k$ is equal to $|\det A\,|$.

18 The structure theorem

Theorem 18.1. *Let M be a finitely generated module over a principal ideal domain R. Then there exist elements $d_1, d_2, \dots, d_k \in R$ satisfying $d_1 \mid d_2 \mid \cdots \mid d_k$ such that*

$$M \cong R/\langle d_1\rangle \oplus R/\langle d_2\rangle \oplus \cdots \oplus R/\langle d_k\rangle \tag{$*$}$$

Proof. Since M if finitely generated (by k elements say), there is a surjective homomorphism $\varphi : R^k \to M$ (Proposition 14.8). Let $N = \ker(\varphi)$. By Theorem 16.2, N is free and of rank $s \leqslant k$. Fix bases for N and for R^k. The inclusion map $N \to R^k$ is a homomorphism between free modules and so can be represented by a matrix $A \in M_{k\times s}(R)$. By Proposition 17.2, A is equivalent to a matrix of the form

$$\begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ & & \ddots & 0 \\ 0 & 0 & \cdots & d_s \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

Where there are $k-s$ rows of zeros at the bottom, and $d_1 \mid d_2 \mid \cdots \mid d_s$. It follows that there is a basis $\{f_1, f_2, \dots, f_k\}$ of R^k such that $\{d_1f_1, d_2f_2, \dots, d_sf_s\}$ is a basis for $N \subseteq R^k$. If $s < k$ define $d_i = 0$ for all $s < i \leqslant k$. The map $\psi : R^k \to R/\langle d_1\rangle \oplus R/\langle d_2\rangle \oplus \cdots \oplus R/\langle d_k\rangle$ given by $\psi(\sum_{i=1}^k r_if_i) = (r_1 + \langle d_1\rangle, \dots, r_k + \langle d_k\rangle)$ is a homomorphism of R-modules. The result follows from the first isomorphism theorem, since ψ is surjective, and $\ker(\psi) = N$. □

Remark. The d_i might be zero. If $d_i = 0$, then $d_j = 0$ for all $j \geqslant i$. If d_i is a unit, then d_j is a unit for all $j \leqslant i$.

Corollary 18.2. *Let M be a finitely generated module over a PID.*

1. *If M is torsion-free, then M is free.*
2. *$M = T_M \oplus F$, where T_M is the torsion submodule of M and F is a free submodule of finite rank;*

Proof. If any of the d_i in $(*)$ are non-zero and non-unit, then the right-hand side of $(*)$ would contain non-zero torsion elements. This establishes the first part of the theorem.

The module M/T_M is torsion-free (Proposition 15.2) and finitely generated. Therefore, M/T_M is free (using the first part) and of finite rank. Consider the surjective homomorphism $\varphi : M \to M/T_M$. By Lemma 16.1 $M = \ker(\varphi) \oplus F$ where $F \cong M/T_M$. Note that $\ker(\varphi) = T_M$.

□

Remark. The submodule $F \leqslant M$ is not uniquely determined. For example if we take $R = \mathbb{Z}$ and $M = (\mathbb{Z}/2\mathbb{Z}) \oplus \mathbb{Z}$, then $T_M = \{(0 + 2\mathbb{Z}, 0), (1 + 2\mathbb{Z}, 0)\}$, but for F we can take either $\{(0 + 2\mathbb{Z}, a) \mid a \in \mathbb{Z}\}$ or $\{(a + 2\mathbb{Z}, a) \mid a \in \mathbb{Z}\}$.

Definition 18.3. If any of the d_i is a unit, $R/(d_i) \cong \{0\}$ so we can drop that summand from the decomposition of M. The decomposition $(*)$ (with all d_i non-unit) is called the **invariant factor decomposition** of M. The non-unit elements d_i are called the **invariant factors** of M. The non-zero, non-unit d_i are called the **torsion invariants**. The number of zero d_i is called the **torsion-free rank** of M.

Proposition 18.4. *For a given M as in the Structure Theorem* 18.1*, the invariant factors are all uniquely defined by M, up to associates. The torsion-free rank is uniquely determined by M.*

We postpone the proof of Proposition 18.4 until later.

Example 18.5. Let M be the $\mathbb{Z}$-module F/N where $F = \mathbb{Z}^2$ and $N = \langle (6,4),(4,8),(4,0)\rangle \leqslant \mathbb{Z}^2$. We write M as a direct sum of non-trivial cyclic $\mathbb{Z}$-modules.
Consider the homomorphism $\varphi : \mathbb{Z}^3 \to \mathbb{Z}^2$ given by $(a,b,c) \mapsto a(6,4) + b(4,8) + c(4,0)$. Then $N = \operatorname{Im}(\varphi)$, and with respect to the standard bases, the matrix of the homomorphism is $A = \begin{bmatrix} 6 & 4 & 4 \\ 4 & 8 & 0 \end{bmatrix}$. From Example 17.4 we know that $XAY = D$ where

$$X = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \quad D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 8 & 0 \end{bmatrix} \quad Y = \begin{bmatrix} 1 & -2 & 2 \\ -1 & 2 & -1 \\ 0 & 1 & -2 \end{bmatrix}$$

Since D represents φ and $N = \operatorname{Im}(\varphi)$, we conclude that

$$M = \mathbb{Z}^2/\operatorname{Im}(\varphi) \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/8\mathbb{Z}$$

The invariant factors of M are 2 and 8.
Let's justify our expression for M a little further. Consider the bases

$$\mathcal{B} = \{(1,-1,0),(-2,2,1),(2,-1,-2)\} \qquad \mathcal{C} = \{(1,-2),(0,1)\}$$

of $\mathbb{Z}^3$ and $\mathbb{Z}^2$ respectively. These bases correspond to the columns of Y and X^{-1}. Notice that

$$\begin{aligned} \varphi(b_1) &= \varphi(1,-1,0) = (2,-4) = 2c_1 \\ \varphi(b_2) &= \varphi(-2,2,1) = (0,8) = 8c_2 \\ \varphi(b_3) &= \varphi(2,-1,-2) = (0,0) \end{aligned}$$

So we have

$$F = \langle c_1\rangle \oplus \langle c_2\rangle \qquad N = \langle 2c_1\rangle \oplus \langle 8c_2\rangle$$

Since $\langle 2c_1\rangle \subseteq \langle c_1\rangle$ and $\langle 8c_2\rangle \subseteq \langle c_2\rangle$ we conclude (see Exercise 100 of Section 1) that

$$F/N \cong \langle c_1\rangle/\langle 2c_1\rangle \oplus \langle c_2\rangle/\langle 8c_2\rangle \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/8\mathbb{Z}$$

Example 18.6. Let M be the $\mathbb{Q}[X]$-module F/N where $F = \mathbb{Q}[X]^3$ and N is the submodule of F generated by $\{(1-X, X, 1+X),(1+X, 1-X, 2X),(X,1,1)\}$. We derive the invariant factor decomposition of M
Consider the homomorphism $\varphi : \mathbb{Q}[X]^3 \to \mathbb{Q}[X]^3$ whose matrix, with respect to the standard bases, is

$$A = \begin{bmatrix} 1-X & 1+X & X \\ X & 1-X & 1 \\ 1+X & 2X & 1 \end{bmatrix}$$

Then $N = \operatorname{Im}(\varphi)$ and $M = \mathbb{Q}[X]^3/N$ is the cokernel of φ. From Example 17.5, A is equivalent to

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2-4X+3X^2+3X^3 \end{bmatrix}$$

It follows that

$$\begin{aligned} M &\cong \mathbb{Q}[X]/(1) \oplus \mathbb{Q}[X]/(-1) \oplus \mathbb{Q}[X]/(2-4X+3X^2+3X^3) \\ &\cong \mathbb{Q}[X]/(2-4X+3X^2+3X^3) \end{aligned}$$

So M is a torsion module and $\operatorname{ann}_R(M) = (2-4X+3X^2+3X^3) \lhd \mathbb{Q}[X]$.

18.1 Exercises

130. Let V be the $\mathbb{Z}[i]$-module $(\mathbb{Z}[i])^2/N$ where $N = \langle(1+i, 2-i), (3, 5i)\rangle$. Write V as a direct sum of cyclic modules.

19 Primary decomposition

We will use the following result to rewrite the invariant factor decomposition in an alternative way.

Lemma 19.1. *Let R be a PID, and $a, b \in R$ two relatively prime elements. Then*

$$R/\langle ab\rangle \cong R/\langle a\rangle \oplus R/\langle b\rangle \qquad \textit{(as R-modules)}$$

Proof. Define $\varphi : R \to R/\langle a\rangle \oplus R/\langle b\rangle$ by $\varphi(u) = (u + \langle a\rangle, u + \langle b\rangle)$. Then $\ker(\varphi) = \langle a\rangle \cap \langle b\rangle$. Since R is a PID and a and b are relatively prime, $\langle a\rangle \cap \langle b\rangle = \langle ab\rangle$. So we have $\ker(\varphi) = \langle ab\rangle$.

Again, since R is a PID and a and b are relatively prime, there exist $x, y \in R$ such that $xa + yb = 1$. Given any element $(c+\langle a\rangle, d+\langle b\rangle) \in R/\langle a\rangle \oplus R/\langle b\rangle$, we have $\varphi(cyb+dxa) = (cyb+\langle a\rangle, dxa+\langle b\rangle) = (c+\langle a\rangle, d+\langle b\rangle)$. Therefore φ is surjective, and the required isomorphism then follows from the first isomorphism theorem (for modules). □

Theorem 19.2. *Let M be a finitely generated module over a principal ideal domain R. Then there exist prime elements $p_1, \ldots, p_s \in R$ and numbers $r, n_1, n_2, \ldots, n_s \in \mathbb{N}$ such that*

$$M \cong \quad R/\langle p_1^{n_1}\rangle \oplus R/\langle p_2^{n_2}\rangle \oplus \cdots \oplus R/\langle p_s^{n_s}\rangle \oplus\ R^r \tag{†}$$

Proof. From Theorem 18.1 we have

$$M \cong R/\langle d_1\rangle \oplus R/\langle d_2\rangle \oplus \cdots \oplus R/\langle d_k\rangle$$

where $d_1 \mid d_2 \mid \cdots \mid d_k$ and all d_i are non-unit. Each non-zero d_i has an irreducible factorisation

$$d_i = p_1^{n_1} p_2^{n_2} \ldots p_{m_i}^{n_{m_1}}$$

Lemma 19.1 then tells us that

$$R/\langle d_i\rangle \cong R/\langle p_1^{n_1}\rangle \oplus \cdots \oplus R/\langle p_1^{n_1}\rangle$$

□

Definition 19.3. The expression given in (†) is called the **primary decomposition** of M.

Example 19.4. Suppose that M is a $\mathbb{Q}[X]$-module such that

$$M \cong \mathbb{Q}[X]/\langle X^4 - 32X + 4\rangle \oplus \mathbb{Q}[X]/\langle X^5 - 1\rangle \oplus \mathbb{Q}[X]/\langle X^2 - 2X + 1\rangle \oplus \mathbb{Q}[X]^2$$

From the irreducible factorizations

$$\begin{aligned} X^4 - 32X + 4 &= (X^2 - 6X + 2)(X^2 + 6X + 2) \\ X^5 - 1 &= (X-1)(X^4 + X^3 + X^2 + X + 1) \\ X^2 - 2X + 1 &= (X-1)(X-1) \end{aligned}$$

We obtain the primary decomposition

$$\begin{aligned} M = \mathbb{Q}[X]/\langle X-1\rangle \oplus \mathbb{Q}[X]/\langle X-1\rangle^2 \oplus \mathbb{Q}[X]/\langle X^2 - 6X + 2\rangle \oplus \mathbb{Q}[X]/\langle X^2 + 6X + 2\rangle \\ \oplus\, \mathbb{Q}[X]/\langle X^4 + X^3 + X^2 + X + 1\rangle \oplus \mathbb{Q}[X]^2 \end{aligned}$$

19.1 Exercises

131. Show that the $\mathbb{Z}$-module $\mathbb{Z}_{p^n}$, where p is a prime and n a non-negative integer, is not a direct sum of two non-trivial $\mathbb{Z}$-modules.

132. Let $R = \mathbb{Q}[X]$ and suppose that the torsion R-module M is a direct sum of four cyclic modules whose annihilators are $(X-1)^3$, $(X^2+1)^2$, $(X-1)(X^2+1)^4$, and $(X+2)(X^2+1)^2$. Determine the primary decomposition of M and the invariant factor decomposition of M. If M is thought of as a vector space over $\mathbb{Q}$ on which X acts as a linear transformation denoted A, determine the minimum and characteristic polynomials of A and the dimension of M over $\mathbb{Q}$.

20 Application to abelian groups

Since abelian groups are $\mathbb{Z}$-modules, and $\mathbb{Z}$ is a PID, the above structure theorem applies to finitely generated abelian groups. We state the result in this special case.

Theorem 20.1 (Structure Theorem for Finitely Generated Abelian Groups).
Let G be a finitely generated abelian group. Then

$$G = \mathbb{Z}_{d_1} \oplus \cdots \oplus \mathbb{Z}_{d_k} \oplus (\mathbb{Z})^m$$

where $m \in \mathbb{N}$, $d_i \in \mathbb{N}$, $d_i \geqslant 2$, $d_1|d_2|\cdots|d_k$ □

20.1 Exercises

133. How many abelian groups of order 136 are there? Give the primary and invariant factor decompositions of each.

134. Determine the invariant factors of the abelian group $C_{100} \oplus C_{36} \oplus C_{150}$.

135. Find a direct sum of cyclic groups which is isomorphic to the abelian group $\mathbb{Z}^3/N$, where N is generated by $\{(2,2,2),(2,2,0),(2,0,2)\}$.

136. Find an isomorphic direct product of cyclic groups, where V is an abelian group generated by x, y, z and subject to relations:

    (a) $3x + 2y + 8z = 0, 2x + 4z = 0$

    (b) $x + y = 0, 2x = 0, 4x + 2z = 0, 4x + 2y + 2z = 0$

    (c) $2x + y = 0, x - y + 3z = 0$.

137. Suppose that the abelian group M is generated by three elements x, y, z subject to the relations $4x + y + 2z = 0, 5x + 2y + z = 0, 6y - 6z = 0$. Determine the invariant factors of M and hence exhibit M as a direct sum of cyclic groups.

21 Jordan normal form

This gives a canonical form for linear transformations $T : V \to V$ of a finite dimensional complex vector space. We want to find a basis for V such that the matrix of T is as simple as possible.

We can endow V with a $\mathbb{C}[X]$-module structure by defining scalar multiplication as follows

$$(\sum a_i X^i)v = \sum a_i T^i(v)$$

Since V is finite dimensional as a $\mathbb{C}$-module, it is finitely generated as a $\mathbb{C}[X]$-module. Indeed, any generating set for ${}_{\mathbb{C}}V$ will be a generating set for ${}_{\mathbb{C}[X]}V$. Since ${}_{\mathbb{C}[X]}V$ is finitely generated and $\mathbb{C}[X]$ is a PID, we can apply the structure theorem. Noting that the prime elements in $\mathbb{C}[X]$ are exactly the linear polynomials, from Theorem 19.2, V has a primary decomposition of the form

$$V \cong \mathbb{C}[X]/\langle (X-\lambda_1)^{m_1}\rangle \oplus \cdots \oplus \mathbb{C}[X]/\langle (X-\lambda_k)^{m_k}\rangle \oplus \mathbb{C}[X]^r$$

In fact it must be the case that $r = 0$, that is, that the torsion-free rank of ${}_{\mathbb{C}[X]}V$ is zero. The set $\{1, X, X^2, \dots\} \subseteq \mathbb{C}[X]$ is linearly independent over $\mathbb{C}$. Therefore if $r \geqslant 1$ then ${}_{\mathbb{C}}V$ would contain an infinite linearly independent set, which would contradict the fact that V is a finite dimensional complex vector space.

We thus have

$$V \cong \mathbb{C}[X]/\langle (X-\lambda_1)^{m_1}\rangle \oplus \cdots \oplus \mathbb{C}[X]/\langle (X-\lambda_k)^{m_k}\rangle$$

The summands are submodules of ${}_{\mathbb{C}[X]}V$ and therefore subspaces of ${}_{\mathbb{C}}V$ that are preserved by the linear transformation T. Let $W = \mathbb{C}[X]/\langle X-\lambda)^m\rangle$. We will analyse the restriction of T to W. Denote the element $f + (X-\lambda)^m \in W$ by $\bar{f}$.

Lemma 21.1. *The set $\{\bar{1}, \bar{X} - \bar{\lambda}, (\bar{X}-\bar{\lambda})^2, \dots, (\bar{X}-\bar{\lambda})^{m-1}\}$ is a basis for ${}_{\mathbb{C}}W$.*

Proof. Suppose that $\xi_i \in \mathbb{C}$, $1 \leqslant i \leqslant m-1$. Then

$$\begin{aligned} \sum_{i=1}^{m-1} \xi_i(\bar{X}-\bar{\lambda})^i = 0 &\implies \sum_{i=1}^{m-1} \xi_i(X-\lambda)^i \in (X-\lambda)^m \\ &\implies \sum_{i=1}^{m-1} \xi_i(X-\lambda)^i = 0 \quad \text{(since 0 is the only element of degree less than } m) \\ &\implies \forall i,\ \ \xi_i = 0 \end{aligned}$$

So the set is linearly independent.

From the division algorithm for $\mathbb{C}[X]$, we know that for any $f \in \mathbb{C}[X]$, there is a $q \in \mathbb{C}[X]$ such that $\deg(q) < m$ and $\bar{q} = \bar{f}$. It follows that $\bar{f} \in \operatorname{span}\{\bar{1}, \bar{X}, \dots, \bar{X}^{m-1}\}$. The result then follows from the fact that $\operatorname{span}\{\bar{1}, \bar{X}, \dots, \bar{X}^{m-1}\} = \operatorname{span}\{\bar{1}, \bar{X}-\bar{\lambda}, \dots, (\bar{X}-\bar{\lambda})^{m-1}\}$. □

Now we calculate the matrix, with respect to this basis, of the the linear transformation $T|_W : W \to W$. For this we calculate the images of the basis elements, noting that $T(\bar{f}) = X\bar{f}$.

$$\begin{aligned} T(\bar{1}) &= X\bar{1} = \bar{X} = \lambda\bar{1} + (\bar{X}-\bar{\lambda}) \\ T(\bar{X}-\bar{\lambda})^i &= X(\bar{X}-\bar{\lambda})^i = \lambda(\bar{X}-\bar{\lambda})^i + (\bar{X}-\bar{\lambda})^{i+1} \qquad \text{(for } 1 \leqslant i < m) \end{aligned}$$

The matrix of $T|_W$ is therefore

$$\begin{bmatrix} \lambda & 0 & 0 & & 0 & 0 \\ 1 & \lambda & 0 & \cdots & 0 & 0 \\ 0 & 1 & \lambda & & 0 & 0 \\ & & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & & \lambda & 0 \\ 0 & 0 & 0 & \cdots & 1 & \lambda \end{bmatrix} \in M_{m\times m}(\mathbb{C})$$

Definition 21.2. A matrix in the above form is called an **elementary Jordan matrix**.

Exercise 138. Let A denote the above matrix.

(a) Show that the characteristic polynomial of A is $(X-\lambda)^m$.

(b) Show that the minimal polynomial of A is $(X-\lambda)^m$.

(c) Show that the dimension of the eigenspace (corresponding to the only eigenvalue of A) is 1.

Theorem 21.3. *There is a basis of V with respect to which the linear transformation T has a matrix in the form*

$$\begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_k \end{bmatrix}$$

where each matrix A_i is an elementary Jordan matrix and all other entries are zero.

Proof. Follows directly from the decomposition of V into summands of the form W and the above matrix for $T|_W$. □

Definition 21.4. A matrix in the above form will be called a **Jordan normal form** matrix.

Example 21.5. The following are examples of matrices in Jordan Normal Form:

$$\begin{bmatrix}1&0&0\\1&1&0\\0&0&2\end{bmatrix} \qquad \begin{bmatrix}2&0&0&0&0&0&0&0&0\\0&2&0&0&0&0&0&0&0\\0&1&2&0&0&0&0&0&0\\0&0&0&2&0&0&0&0&0\\0&0&0&1&2&0&0&0&0\\0&0&0&0&0&3&0&0&0\\0&0&0&0&0&1&3&0&0\\0&0&0&0&0&0&1&3&0\\0&0&0&0&0&0&0&1&3\end{bmatrix}$$

21.1 Calculating the Jordan normal form of a matrix

Given a matrix $A \in M_n(\mathbb{C})$ we can calculate its Jordan normal form by considering the matrix $A-XI \in M_n(\mathbb{C}[X])$. Let V be a complex vector space with basis $\{v_1, \dots, v_n\}$. Let $T : V \to V$ be the linear transformation whose matrix (with respect to the the basis $\{v_1, \dots, v_n\}$) is A, and define ${}_{\mathbb{C}[X]}V$ as in the previous section. The following lemma tells us that ${}_{\mathbb{C}[X]}V \cong \mathbb{C}[X]^n/N$, where N is generated by the columns of the matrix $A - XI$. We then find $d_1, \dots, d_n \in \mathbb{C}[X]$ such that $A - XI \sim \operatorname{diag}(d_1, \dots, d_n)$ (Proposition 17.2) and obtain (as in the previous section) that ${}_{\mathbb{C}[X]}V \cong \mathbb{C}[X]/\langle d_1\rangle \oplus \cdots \oplus \mathbb{C}[X]/\langle d_n\rangle$. From this we then get the primary decomposition of ${}_{\mathbb{C}[X]}V$ and hence the Jordan Normal Form of A, as explained in the previous section.

Lemma 21.6. *Let $A \in M_n(\mathbb{C})$, let F be a free $\mathbb{C}[X]$-module of rank n, with basis $\mathcal{F} = \{f_1, \dots, f_n\}$. Let $\{v_1, \dots, v_n\}$ be a basis for a complex vector space V. Let π be the surjective $\mathbb{C}[X]$-module homomorphism $\pi : F \to {}_{\mathbb{C}[X]}V$ determined by $\pi(f_i) = v_i$. Let $\varphi : F \to F$ be the homomorphism whose matrix with respect to $\mathcal{F}$ is $A - XI$. Then* $\ker(\pi) = \operatorname{Im}(\varphi)$.

Proof. We first show that $\operatorname{Im}(\varphi) \subseteq \ker(\pi)$. Let $\mathcal{V} = \{v_1, \dots, v_n\}$. It is enough to show that for all $f_j \in \mathcal{F}$ we have $\pi \circ \varphi(f_j) = 0$. Let $a_{ij} \in \mathbb{C}$ be the entry in the i-th row and j-th column of A. Then $(A - XI)_{ij} = a_{ij} - \delta_{ij}X$ and

$$\begin{aligned}
\pi \circ \varphi(f_j) &= \pi(\sum_{i=1}^{n}(a_{ij} - \delta_{ij}X)f_i) \quad \text{(since } [\varphi]_{\mathcal{B}} = A - XI\text{)}\\
&= \pi((\sum_{i=1}^{n} a_{ij}f_i) - Xf_j)\\
&= \sum_{i=1}^{n} a_{ij}\pi(f_i) - X\pi(f_j)\\
&= \sum_{i=1}^{n} a_{ij}v_i - Xv_j \quad \text{(from the definition of } \pi\text{)}\\
&= \sum_{i=1}^{n} a_{ij}v_i - T(v_j) \quad \text{(from the way in which scalar multipn is defined in } {}_{\mathbb{C}[X]}V\text{)}\\
&= T(b_j) - T(b_j) \quad \text{since } [T]_{\mathcal{V}} = A\\
&= 0
\end{aligned}$$

Now for the reverse inclusion. Given any $f \in F$, we have $f = (\sum_i \alpha_i f_i) + \varphi(f')$ for some $f' \in F$ and $\alpha_i \in \mathbb{C}$. (Note that the α_i are in $\mathbb{C}$ not $\mathbb{C}[X]$.)
Then

$$\begin{aligned}
f \in \ker(\pi) &\implies \pi(\sum_i \alpha_i f_i) + \pi(\varphi(f')) = 0\\
&\implies \pi(\sum_i \alpha_i f_i) = 0 \qquad \text{(since } \mathrm{Im}(\varphi) \subseteq \ker(\pi)\text{)}\\
&\implies \sum_i \alpha_i v_i = 0 \qquad \text{(since } \pi(f_i) = v_i\text{)}\\
&\implies \alpha_i = 0 \qquad \text{for all } i\\
&\implies f = \varphi(f')\\
&\implies f \in \mathrm{Im}(\varphi)
\end{aligned}$$

□

Calculating the diagonal matrix equivalent to $A - XI$ then enables us to write down the primary decomposition of ${}_{\mathbb{C}[X]}V$ and hence the Jordan Normal Form of the matrix A.

Example 21.7. Calculate a Jordan normal form matrix that is similar to the matrix

$$A = \begin{bmatrix} 2 & 0 & 0 & 0\\ -1 & 1 & 0 & 0\\ 0 & -1 & 0 & -1\\ 1 & 1 & 1 & 2 \end{bmatrix}$$

We consider the matrix $XI - A \in M_{4\times 4}(\mathbb{C})$ and put it into diagonal form (as in Proposition 17.2).

$$XI - A = \begin{bmatrix} X-2 & 0 & 0 & 0\\ 1 & X-1 & 0 & 0\\ 0 & 1 & X & 1\\ -1 & -1 & -1 & X-2 \end{bmatrix} \xrightarrow{R1\leftrightarrow R4} \begin{bmatrix} -1 & -1 & -1 & X-2\\ X-2 & 0 & 0 & 0\\ 1 & X-1 & 0 & 0\\ 0 & 1 & X & 1 \end{bmatrix}$$

$$\xrightarrow[R3\mapsto R3+R1]{R2\mapsto R2+(X-2)R1} \begin{bmatrix} -1 & -1 & -1 & X-2\\ 0 & -X+2 & -X+2 & (X-2)^2\\ 0 & X-2 & -1 & X-2\\ 0 & 1 & X & 1 \end{bmatrix}$$

$$\xrightarrow[\substack{C3\mapsto C3-C1\\ C4\mapsto C4+(X-2)C1}]{C2\mapsto C2-C1} \begin{bmatrix} -1 & 0 & 0 & 0\\ 0 & -X+2 & -X+2 & (X-2)^2\\ 0 & X-2 & -1 & X-2\\ 0 & 1 & X & 1 \end{bmatrix}$$

$$\xrightarrow{R2\leftrightarrow R4} \begin{bmatrix} -1 & 0 & 0 & 0\\ 0 & 1 & X & 1\\ 0 & -X+2 & -X+2 & (X-2)^2\\ 0 & X-2 & -1 & X-2 \end{bmatrix}$$

$$\xrightarrow[R4\mapsto R4-(X-2)R2]{R3\mapsto R3+(X-2)R2} \begin{bmatrix} -1 & 0 & 0 & 0\\ 0 & 1 & X & 1\\ 0 & 0 & (X-2)(X-1) & (X-2)(X-1)\\ 0 & 0 & -(X-1)^2 & 0 \end{bmatrix}$$

$$\xrightarrow[C4\mapsto C4-C2]{C3\mapsto C3-XC2} \begin{bmatrix} -1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & (X-2)(X-1) & (X-2)(X-1)\\ 0 & 0 & -(X-1)^2 & 0 \end{bmatrix}$$

$$\xrightarrow{C4\mapsto C4-C3} \begin{bmatrix} -1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & (X-2)(X-1) & 0\\ 0 & 0 & -(X-1)^2 & (X-1)^2 \end{bmatrix} \xrightarrow{R4\mapsto R4-R3} \begin{bmatrix} -1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & (X-2)(X-1) & 0\\ 0 & 0 & -(X-1) & (X-1)^2 \end{bmatrix}$$

$$\xrightarrow{R3\leftrightarrow R4} \begin{bmatrix} -1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & -(X-1) & (X-1)^2\\ 0 & 0 & (X-2)(X-1) & 0 \end{bmatrix}$$

$$\xrightarrow{R4\mapsto R4+(X-2)R3}\begin{bmatrix}-1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & -(X-1) & (X-1)^2\\ 0 & 0 & 0 & (X-2)(X-1)^2\end{bmatrix}$$
$$\xrightarrow{C4\mapsto C4+(X-1)C3}\begin{bmatrix}-1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & -(X-1) & 0\\ 0 & 0 & 0 & (X-2)(X-1)^2\end{bmatrix}$$

From which it follows that the Jordan normal form of the matrix is

$$J=\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 1 & 1 & 0\\ 0 & 0 & 0 & 2\end{bmatrix}$$

21.2 Exercises

139. Suppose that the linear transformation T acts on an 8 dimensional complex vector space V. Using T we make V into a $\mathbb{C}[t]$-module (where t is an indeterminate) in the usual way. Suppose that as a $\mathbb{C}[t]$-module $V\cong\mathbb{C}[t]/\langle(t+5)^2\rangle\oplus\mathbb{C}[t]/\langle(t-3)^3(t+5)^3\rangle$. What is the Jordan (normal) form for the transformation T? What are the eigenvalues of T and how many eigenvectors does T have? What are the minimal and characteristic polynomials of T?

140. Determine the Jordan normal form of the matrix

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

(a) By deducing it from the characteristic and minimal polynomials;

(b) By calculating the invariant factor matrix of $XI-A\in M_{3\times 3}(\mathbb{C})$.

141. Find all possible Jordan normal forms for a matrix whose characteristic polynomial is $(t+2)^2(t-5)^3$

22 Uniqueness of invariant factors

We will make use of the following result, the proof of which can be found in [HH70, Lemma 9.4].

Lemma 22.1. *Let T be a finitely generated torsion module over a PID R, and let M and N be two R-modules. Then*

$$T\oplus M\cong T\oplus N\implies M\cong N$$

□

The proof of the above result uses the existence of the decomposition in Theorem 18.1, but does not use the uniqueness of the invariant factors, which we are now able to establish.

Proof of Proposition 18.4. Suppose we have two decompositions of the kind described in Theorem 18.1. That is, suppose that $d_1,\ldots d_k,e_1,\ldots,e_l\in R$ are non-zero, non-unit such that $d_1\mid d_2\mid\cdots\mid d_k$ and $e_1\mid e_2\mid\cdots\mid e_l$ and

$$M\cong R/\langle d_1\rangle\oplus R/\langle d_2\rangle\oplus\cdots\oplus R/\langle d_k\rangle\oplus R^m \tag{II.1}$$
$$M\cong R/\langle e_1\rangle\oplus R/\langle e_2\rangle\oplus\cdots\oplus R/\langle e_l\rangle\oplus R^n \tag{II.2}$$

We need to show that $m=n$, $k=l$ and $d_i\sim e_i$ for all $1\leqslant i\leqslant k$.

Let T_M be the torsion submodule of M. From the above isomorphisms, it follows that

$$T_M\cong R/\langle d_1\rangle\oplus R/\langle d_2\rangle\oplus\cdots\oplus R/\langle d_k\rangle \tag{II.3}$$
$$T_M\cong R/\langle e_1\rangle\oplus R/\langle e_2\rangle\oplus\cdots\oplus R/\langle e_l\rangle \tag{II.4}$$

and

$$T_M \oplus R^m \cong T_M \oplus R^n$$

From Lemma 22.1 it follows that $R^m \cong R^n$, which implies that $m = n$ by Proposition 14.7.

Notice that $\operatorname{ann}_R(T_M) = \langle d_k \rangle$ from (II.3) and $\operatorname{ann}_R(T_M) = \langle e_l \rangle$ from (II.4). It follows that $d_k \sim e_l$. Applying Lemma 22.1, we obtain

$$R/\langle d_1 \rangle \oplus R/\langle d_2 \rangle \oplus \cdots \oplus R/\langle d_{k-1} \rangle \cong R/\langle e_1 \rangle \oplus R/\langle e_2 \rangle \oplus \cdots \oplus R/\langle e_{l-1} \rangle$$

Applying the same reasoning yields $d_{k-1} \sim e_{l-1}$. Continuing in the same way, we obtain $k = l$, and $d_i \sim e_i$ for all i.

□

Chapter III

Fields

23 Field Extensions

We will want to enlarge a given field to, for example, ensure that a given polynomial has a root. Since we are thinking of extending a given field, we introduce an alternative notation to saying that the smaller is a subfield of the larger.

Definition 23.1. If E and F are fields with E a subfield of F, we say that F is an **extension** of E.

Example 23.2. The complex numbers $\mathbb{C}$ are an extension of the real numbers $\mathbb{R}$. The real numbers are an extension of the rational numbers $\mathbb{Q}$. The field $\mathbb{Q}(i) = \{x + iy \mid x, y \in \mathbb{Q}\}$ is a subfield of $\mathbb{C}$ and an extension of $\mathbb{Q}$.

Remark. An extension E of F can be regarded as a vector space over F.

We know that a polynomial in $f \in F[X]$ need not have any roots in F. For example, $X^2 + X + 1 \in \mathbb{F}_2[X]$ has no roots in $\mathbb{F}_2$. However, it is always possible to extend to a field $E \supseteq F$ such that f has a root in E. The polynomial $X^2 + X + 1$ does have a root in the field F_4 of Example 1.14 and F_4 contains a copy of $\mathbb{F}_2$.

Proposition 23.3. *Let F be a field and $f \in F[X]$ a non-constant polynomial. Then there is an extension field $E \supseteq F$ and an element $a \in E$ such that $f(a) = 0$.*

Proof. Since $F[X]$ is a UFD, the element f has a prime factorization $f = p_1 \dots p_n$. It is therefore sufficient to prove the result under the assumption that f is prime. Assuming that $p \in F[X]$ is prime, we know that $E = F[X]/\langle p\rangle$ is a field. The map $F \to E$ given by $f \mapsto f + \langle p\rangle$ is clearly a homomorphism. Since p has degree at least 1, this homomorphism is injective, and so we regard F as a subring of E. To complete the proof we note that the element $X + \langle p\rangle \in E$ is a root of p, since if $p = a_0 + \cdots + a_m X^m$ and $I = \langle p\rangle$ we have

$$\begin{aligned} p(X+I) &= (a_0 + I)(1 + I) + (a_1 + I)(X + I) + \cdots + (a_m + I)(X + I)^m \\ &= (a_0 + I)(1 + I) + (a_1 + I)(X + I) + \cdots + (a_m + I)(X^m + I) \\ &= (a_0 + I) + (a_1 X + I) + \cdots + (a_m X^m + I) \\ &= (a_0 + \cdots + a_m X^m) + I \\ &= 0 + I \qquad (\text{since } a_0 + \cdots + a_m X^m = p \in I) \\ &= 0_E \end{aligned}$$

□

Example 23.4. Consider the polynomial $X^2 - 2 \in \mathbb{Q}[X]$. This clearly has no roots in $\mathbb{Q}$. The field E constructed in the above proof is $\mathbb{Q}[X]/\langle X^2-2\rangle$ which is isomorphic to the subfield of $\mathbb{R}$ given by $\mathbb{Q}(\sqrt{2}) = \{x+y\sqrt{2} \mid x, y \in \mathbb{Q}\}$. Of course, we know that by extending all the way to $E = \mathbb{C}$ our polynomial would have a root. The point is that we don't need to go that far.

23.1 Exercises

142. Let F be a field and $D : F[X] \to F[X]$ the map given by

$$D(a_0 + a_1X + \cdots + a_nX^n) = a_1 + 2a_2X + \cdots + na_nX^{n-1}$$

(a) Verify that $D(fg) = D(f)g + fD(g)$.

An element $a \in E$ in an extension $E \supseteq F$ is called a **multiple root** of $f \in F[X]$ if $(X-a)^2$ divides f (in $E[X]$).

(b) Show that if $a \in E$ is a multiple root of $f \in F[X]$, then a is a root of $D(f)$.

(c) Suppose that $f \in F[X]$ is irreducible. Show that if $D(f) \neq 0$, then f has no multiple root in any extension field of F.

(d) Show that if F has characteristic 0 and $f \in F[X]$ is irreducible, then f has no multiple roots in any extension field of F.

(e) Give an example of fields $E \supset F$, an irreducible $f \in F[X]$ and an element $a \in E$ such that a is a multiple root of f.

24 Algebraic elements and algebraic extensions

24.1 Algebraic and transcendental elements

Definition 24.1. Let E be an extension of the field F. An element $a \in E$ is called **algebraic** over F if there is a non-zero element in $F[X]$ having a as a root. An element is called **transcendental** over F if it is not algebraic.

Example 24.2.

1. $\sqrt{2} \in \mathbb{R}$ is algebraic over $\mathbb{Q}$.
2. $i \in \mathbb{C}$ is algebraic over $\mathbb{Q}$.
3. $\pi \in \mathbb{R}$ is transcendental over $\mathbb{Q}$.[1]
4. $\pi \in \mathbb{R}$ is algebraic over $\mathbb{R}$.

Given an element $a \in E \supset F$ that is algebraic over F, the set $I = \{f \in F[X] \mid f(a) = 0\}$ is an ideal in $F[X]$. Since $F[X]$ is a PID, we have $I = \langle p\rangle$ for some $p \in F[X]$.

Exercise 143. Let $a \in E \supset F$ be algebraic over F.

(a) Show that $I = \{f \in F[X] \mid f(a) = 0\}$ is an ideal in $F[X]$.

Let $p \in F[X]$ be such that $I = \langle p\rangle$.

(b) Show that p is irreducible.

Definition 24.3. If a is algebraic over F, the unique monic irreducible polynomial having a as a root is called the **irreducible polynomial for a over F**. It will be denoted $\operatorname{irr}(a, F)$. It is also sometimes called the minimal polynomial for a. The degree of $\operatorname{irr}(a, F)$ will be called the **degree of a over F** and will be denoted $\deg(a, F)$.

[1] This was first proved by the German mathematician Ferdinand von Lindemann in 1882.

Example 24.4. Let $a = \sqrt{1+\sqrt{3}} \in \mathbb{R}$. Then

$$a = \sqrt{1+\sqrt{3}} \implies a^2 = 1+\sqrt{3} \implies (a^2-1)^2 = 3 \implies a^4 - 2a^2 - 2 = 0$$

Since the polynomial $X^4 - 2X^2 - 2$ is irreducible (by, for example, Eisenstein's criterion) we conclude that $\mathrm{irr}(a, \mathbb{Q}) = X^4 - 2X^2 - 2$ and $\deg(a, \mathbb{Q}) = 4$.

Example 24.5. Consider the element $a = \sqrt{2} + \sqrt{3} \in \mathbb{R}$. Let's calculate $\mathrm{irr}(a, \mathbb{Q})$. We are looking for a $\mathbb{Q}$-linear relationship between powers of a. Calculation gives $a^2 = 5 + 2\sqrt{6}$, $a^3 = 11\sqrt{2} + 9\sqrt{3}$ and $a^4 = 49 + 20\sqrt{6}$. The vectors $v_0 = (1, 0, 0, 0)$, $v_1 = (0, 1, 1, 0)$, $v_2 = (5, 0, 0, 2)$, $v_3 = (0, 11, 9, 0)$ and $v_4 = (49, 0, 0, 20)$ are linearly dependent in $\mathbb{Q}^4$. Since

$$\begin{bmatrix} 1 & 0 & 5 & 0 & 49 \\ 0 & 1 & 0 & 11 & 0 \\ 0 & 1 & 0 & 9 & 0 \\ 0 & 0 & 2 & 0 & 20 \end{bmatrix} \text{ is row-equivalent to } \begin{bmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 10 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

we observe that $v_4 = -v_0 + 10v_2$. It follows that $a^4 - 10a^2 + 1 = 0$. The polynomial $X^4 - 10X^2 + 1 \in \mathbb{Q}[X]$ is divisible by $\mathrm{irr}(a, \mathbb{Q})$. The polynomial $X^4 - 10X^2 + 1$ is irreducible in $\mathbb{Q}[X]$ (exercise!) so we conclude that $\mathrm{irr}(\sqrt{2} + \sqrt{3}, \mathbb{Q}) = X^4 - 10X^2 + 1$, and $\deg(\sqrt{2} + \sqrt{3}, \mathbb{Q}) = 4$.

The difference between algebraic and transcendental elements is reflected in the corresponding evaluation maps.

Definition 24.6. Let $a \in E \supseteq F$. Recall that $F[a]$ denotes the smallest subring of E that contains F and a. We denote by $F(a)$ the smallest subfield of E that contains F and a, that is, the intersection of all subfields containing F and a. Given $a_1, \dots a_m \in E$, $F[a_1, \dots, a_m]$ and $F(a_1, \dots, a_m)$ are defined similarly.

Remark. It follows from the definition that $F[a] \subseteq F(a)$ and that $F(a)$ is isomorphic to the field of quotients of $F[a]$.

Lemma 24.7. *Let $a \in E \supseteq F$, where E and F are fields. Let $\varphi_a : F[X] \to E$ be the homomorphism given by $\varphi_a(f) = f(a)$ (i.e., φ_a is evaluation at a). Then,*

1. $\mathrm{Im}(\varphi_a) = F[a]$
2. *If a is algebraic over F, then φ_a is not injective and $\ker(\varphi_a) = \langle \mathrm{irr}(a, F) \rangle$. The map φ_a induces (as in the first isomorphism theorem) an isomorphism*

$$F[X]/\langle \mathrm{irr}(a, F) \rangle \cong F[a] \quad \textit{and} \quad F[a] = F(a)$$

3. *If a is transcendental over F, then φ_a is injective and φ_a gives an isomorphism*

$$F[X] \cong F[a] \quad \textit{and} \quad F[a] \subsetneq F(a)$$

Proof. Since the image of φ_a is a subring of E and it contains F and a, it follows that $F[a] \subseteq \mathrm{Im}(\varphi_a)$. On the other hand, $\mathrm{Im}(\varphi_a)$ is contained in any subring that contains F and a. Therefore $\mathrm{Im}(\varphi_a) \subseteq F[a]$.

The element a is algebraic if and only if $\ker(\varphi_a) \neq \{0\}$. In the case in which a is algebraic, $\ker(\varphi_a) = \langle f \rangle$ for some non-zero polynomial f since $F[X]$ is a PID. Then ?THM? **??** tells us that f is an associate of $\mathrm{irr}(a, F)$. Since $\mathrm{irr}(a, F)$ is irreducible and $F[X]$ is a PID, $F[X]/\langle \mathrm{irr}(a, F) \rangle$ is a field and therefore $F[a] = F(a)$. If a is transcendental, $F[a] \neq F(a)$ as $F[a] \cong F[X]$ and $F[X]$ is not a field.

□

24.2 Algebraic and finite extensions

Since $F[a]$ is a ring, it forms an abelian group with respect to addition, and since $F \subseteq F[a]$ we can multiply an element of $F[a]$ by a scalar from F in a natural way. In other words, $F[a]$ forms a vector space over F.

Lemma 24.8. *Let $a \in E \supseteq F$, with a algebraic over F. Let $n = \deg(a, F)$. Then $\{1, a, \ldots, a^{n-1}\}$ is a basis for $F[a]$ as a vector space over F. Moreover, every element $b \in F[a]$ is algebraic over F and $\deg(b, F) \leqslant \deg(a, F)$.*

Remark. We will see shortly that in fact $\deg(b, F)$ divides $\deg(a, F)$.

Proof. Let $\mathcal{B} = \{1, a, \ldots, a^{n-1}\}$, and let $\alpha_i \in F$ be such that $\operatorname{irr}(a, F) = \sum_{i=0}^{n-1} \alpha_i X^i + X^n$. Since a is a root of this polynomial, we have $a^n = -\sum_{i=0}^{n-1} \alpha_i a^i$. It follows that for all $k \geqslant n$, $a^k \in \operatorname{span}(\mathcal{B})$, and therefore that for all $f \in F[X]$, $f(a) \in \operatorname{span}(\mathcal{B})$. We have shown that $\mathcal{B}$ is a spanning set for $F[a]$ (as a vector space over F). To show linear independence, note that $\sum_{i=0}^{n-1} \gamma_i a^i = 0$ implies that a is a root of the polynomial $g = \sum_{i=0}^{n-1} \gamma_i X^i \in F[X]$. But $\deg(g) < \deg(a, F)$, so we must have $g = 0$ (i.e., for all i, $\gamma_i = 0$).

Let $b \in F[a]$. The set $\{1, b, \ldots, b^n\} \subset F[a]$ is necessarily linearly dependent because it has more that $\dim_F(F[a])$ elements. Therefore there exist $\beta_i \in F$, not all of which are zero, such that $\sum_{i=0}^{n} \beta_i b^i = 0$. Letting $h = \sum_{i=0}^{n} \beta_i X^i \in F[X]$ and noting that $h(b) = 0$ we conclude that $\deg(b, F) \leqslant n = \deg(a, F)$.

□

Definition 24.9. An extension E of F is called an **algebraic extension** if every element of E is algebraic over F. It is called a **finite extension** (of degree n) if E is of finite dimension n as a vector space over F. In the case in which E is a finite extension of F we denote the degree by $[E : F]$.

Remark. From Lemma 24.8 we know that if $a \in E$ is algebraic over F, then $F(a)$ is a finite extension of F and $[F(a) : F] = \deg(a, F)$.

Exercise 144. Show that every finite extension is algebraic.

Example 24.10. Let $E = \{a \in \mathbb{R} \mid a \text{ is algebraic over } \mathbb{Q}\}$. Then E is an algebraic extension of $\mathbb{Q}$, but is not a finite extension of $\mathbb{Q}$. See Exercise 150.

Lemma 24.11. *Let E F and K be fields, with E a finite extension of F and K a finite extension of E. Then K is a finite extension of F and*

$$[K : F] = [K : E]\,[E : F]$$

Proof. Let $m = [E : F]$ and $n = [K : E]$. Let $\{\alpha_1, \ldots, \alpha_m\}$ be a basis for E over F and let $\{\beta_1, \ldots, \beta_n\}$ be a basis for K over E. We will show that $\{\alpha_i\beta_j \mid 1 \leqslant i \leqslant m,\ 1 \leqslant j \leqslant n\}$ is a basis for K over F. Given any $k \in K$ we have

$$\begin{aligned}
k &= \sum_{j=1}^{n} b_j \beta_j && \text{(for some } b_i \in E) \\
&= \sum_{j=1}^{n} \Big(\sum_{i=1}^{m} a_{ij} \alpha_i\Big) \beta_j && \text{(for some } a_{ij} \in F) \\
&= \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij} \alpha_i \beta_j \\
&\in \operatorname{span}\{\alpha_i\beta_j \mid 1 \leqslant i \leqslant m,\ 1 \leqslant j \leqslant n\}
\end{aligned}$$

For linear independence we have

$$\begin{aligned}\sum_{j=1}^{n}\sum_{i=1}^{m} c_{ij}(\alpha_i\beta_j)=0 &\implies \sum_{j=1}^{n}(\sum_{i=1}^{m} c_{ij}\alpha_i)\beta_j=0 \\ &\implies \forall\, j,\ \sum_{i=1}^{m} c_{ij}\alpha_i=0 &&\text{(since the } \beta_j \text{ are linearly independent)} \\ &\implies \forall\, j, \forall\, i, c_{ij}=0 &&\text{(since the } \alpha_i \text{ are linearly independent)}\end{aligned}$$

□

Corollary 24.12. *Let $a \in E \subseteq F$, with a algebraic over F. Then for all $b \in F(a)$, $\deg(b, F)$ divides $\deg(a, F)$.*

Proof. We have $F \subseteq F(b) \subseteq F(a)$ and

$$\deg(a,F) = [F(a):F] = [F(a):F(b)]\,[F(b):F] = [F(a):F(b)]\ \deg(b,F)$$

□

Example 24.13. Consider $a = 2^{\frac{1}{4}} \in \mathbb{R}$. Then $\operatorname{irr}(a, \mathbb{Q}) = X^4 - 2$ and therefore $\deg(a, \mathbb{Q}) = 4$. By the above corollary, any element of $\mathbb{Q}(2^{\frac{1}{4}})$ has degree that divides 4. So, for example, no element of $\mathbb{Q}(2^{\frac{1}{4}})$ is a root of $X^3 - 2$ (or any other irreducible cubic polynomial).

Remark. Finite extensions of $\mathbb{Q}$ are called **number fields** or **algebraic number fields** and are central to the study of Algebraic Number Theory.

24.3 Exercises

145. Find $\operatorname{irr}(a, \mathbb{Q})$ and $\deg(a, \mathbb{Q})$ for
(a) $a = \sqrt{3 - \sqrt{6}}$ (b) $a = \sqrt{(\frac{1}{3}) + \sqrt{7}}$ (c) $a = \sqrt{2} + i$
[You should justify why the polynomial is irreducible.]

146. Show that the following elements are algebraic over $\mathbb{Q}$ and find their minimal polynomials:
   (a) $2^{\frac{1}{3}}$
   (b) $\sqrt{3} + \sqrt{2}$
   (c) $\frac{(\sqrt{5}+1)}{2}$
   (d) $\frac{(i\sqrt{3}-1)}{2}$

147. Find the dimension and a basis for the following extensions:
   (a) $\mathbb{R}(\sqrt{2} + i)$ over $\mathbb{R}$;
   (b) $\mathbb{Q}(\sqrt{2} + i)$ over $\mathbb{Q}$;
   (c) $\mathbb{Q}(\sqrt{2} + \sqrt{3})$ over $\mathbb{Q}$;
   (d) $\mathbb{Q}(\sqrt{3}, i)$ over $\mathbb{Q}$.

148. Let F be a field and $k \in F$ an element which is not a square in F. Show that

$$K = \{\begin{pmatrix} a & kb \\ b & a \end{pmatrix} \mid a, b \in F\} \leqslant M_{2\times 2}(F)$$

is a field and that it is isomorphic to $F(\sqrt{k})$.

149. Show that the set of algebraic numbers (over $\mathbb{Q}$) in $\mathbb{R}$ forms a subfield of $\mathbb{R}$. (Use that $a \in \mathbb{R}$ is algebraic iff $[\mathbb{Q}(a) : \mathbb{Q}]$ is finite.)

150. Let $E = \{a \in \mathbb{R} \mid a \text{ is algebraic over } \mathbb{Q}\}$. Show that E is an algebraic extension of $\mathbb{Q}$, but is not a finite extension of $\mathbb{Q}$.

151. Suppose that E and K are two extensions of F, and let $a \in E$ and $b \in K$ be algebraic over F. Prove that $\operatorname{irr}(a, F) = \operatorname{irr}(b, F)$ if and only if there exists an isomorphism $\varphi : F(a) \to F(b)$ such that $\varphi(a) = b$ and $\varphi|_F = \mathrm{Id}_F$.

25 Constructions with straight-edge and compass

There are classical questions about whether certain lengths or angles can be constructed using a straight-edge and compass. We can establish that certain of these, such as being able to trisect an angle or to construct a nonagon, are impossible. The book [Sti05] by John Stillwell is a good place to read about these questions.

25.1 Constructible points in the Euclidean plane

We first formalise what kind of operations are allowed. Two points in the plane are given. We choose a coordinate system so that the two points are $(0, 0)$ and $(1, 0)$. Starting with these two points we inductively define a subset of the plane. The points so defined will be called **constructible**. Given two distinct points P and Q in the plane, denote by $L(P, Q)$ the straight line containing P and Q and by $C(P, Q)$ the circle with centre P that passes through Q. Suppose that P_1, Q_1, P_2, Q_2 are constructible points in the plane with $P_1 \neq Q_1$ and $P_2 \neq Q_2$. Then the points given by the sets $L(P_1, Q_1) \cap L(P_2, Q_2)$ and $L(P_1, Q_1) \cap C(P_2, Q_2)$ and $C(P_1, Q_1) \cap C(P_2, Q_2)$ are all defined to be constructible.

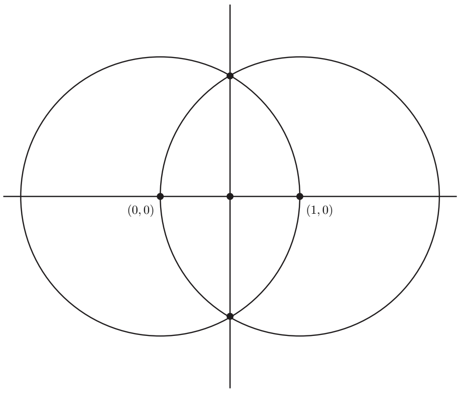


Figure III.1: Illustration demonstrating that the points $(\frac{1}{2}, \frac{\sqrt{3}}{2})$, $(\frac{1}{2}, -\frac{\sqrt{3}}{2})$ and $(\frac{1}{2}, 0)$ are constructible.

25.2 Constructible numbers

Definition 25.1. A real number $x \in \mathbb{R}$ is called **constructible** if $|x|$ is equal to the distance between two constructible points.

The connection with fields and field extensions is given by the next two results.

Proposition 25.2. *1. The constructible numbers form a subfield of* $\mathbb{R}$.

2. If $a > 0$ is constructible, then $\sqrt{a}$ is constructible.

Proof. We need to show that for any two constructible numbers $a, b > 0$, all of the numbers $a + b$, $a - b$, a^{-1}, ab and $\sqrt{a}$ are constructible. Each is shown by describing a construction and appealing to elementary geometry in the the Euclidean plane. We show that ab is constructible. The other cases are similar and the details can be found in the books [Sti05, Art91].

Given that a is constructible, we can construct a right triangle with non hypotenuse side lengths 1 and a as shown in Figure III.2. We then construct a similar triangle in which the side that had length 1 is now of length b. The other non-hypotenuse side will be on length ab. □

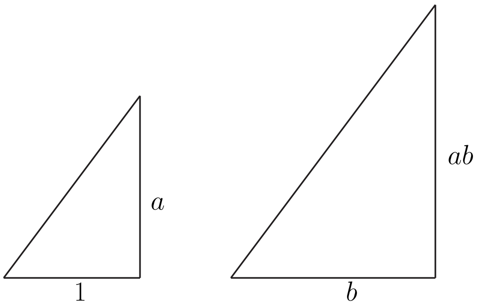


Figure III.2: Construction to show that ab is constructible.

Proposition 25.3. *Let a be a constructible real number. Then there is a chain of subfields of $\mathbb{R}$*

$$\mathbb{Q} = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_{n-1} \subseteq F_n$$

such that

1. *$a \in F_n$*

2. *For all i, there exists $a_i \in F_i$ such that $F_{i+1} = F_i(\sqrt{a_i})$.*

Proof. Suppose that $P_1 \neq Q_1, P_2 \neq Q_2$ are points in the plane all of whose coordinates lie in some subfield F of $\mathbb{R}$. The points of $L(P_1, Q_1) \cap L(P_1, Q_1)$ have coordinates that are given by the solution of a linear system of equations, and are therefore in F. Finding the points of $L(P_1, Q_1) \cap C(P_1, Q_1)$ involves solving a quadratic equation and the coordinates therefore lie in $F(\sqrt{d})$ for some $d \in F$. Solving for the points of $C(P_1, Q_1) \cap C(P_1, Q_1)$ involves solving two simultaneous quadratic equations. However, since both describe circles, taking the difference of the two equations produces a linear equation and we have reduced to the previous case.

Now consider a constructible number $a > 0$. It is the distance between two constructible points $P = (p_1, p_2)$ and $Q = (q_1, q_2)$. The point P is constructed from the points $(0, 0)$ and $(1, 0)$ by a finite sequence of constructions involving the intersections of lines and circles. From the previous paragraph we conclude that there is a finite sequence of subfields $\mathbb{Q} = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_k$ such that $F_{i+1} = F_i(\sqrt{d_i})$ for some $d_i \in F_i$ and $p_1, p_2 \in F_k$. Similarly, there is a finite sequence of subfields $\mathbb{Q} = G_0 \subseteq G_1 \subseteq \cdots \subseteq G_l$ such that $G_{i+1} = G_i(\sqrt{e_i})$ for some $e_i \in G_i$ and $q_1, q_2 \in G_l$. The result then follows by taking F_i as above for $0 \leqslant i \leqslant k$ and $F_{i+1} = F_i(e_{i-k})$ for $k \leqslant i \leqslant k + l - 1$. □

Theorem 25.4. *If $a \in \mathbb{R}$ is constructible, then a is algebraic over $\mathbb{Q}$ and $\deg(a, \mathbb{Q}) = 2^n$ for some $n \in \mathbb{N}$.*

Proof. Since $a_1 \in F_i$, $\deg(\sqrt{a_i}, F_i)$ is either 1 or 2. Note that $[F_{i+1}, F_i] = [F_i(a_i) : F_i] = \deg(\sqrt{a_i}, F_i)$ by Lemma 24.8. Apply Lemma 24.11 repeatedly to conclude that $[F_n : \mathbb{Q}] = 2^m$ for m. Then Corollary 24.12 says that $\deg(a, \mathbb{Q})$ divides 2^m. □

Remark. This result shows that while all constructible numbers are algebraic (over $\mathbb{Q}$), not all algebraic numbers are constructible. For example $2^{\frac{1}{3}}$ is algebraic, but not constructible.

25.3 Impossible constructions

Trisecting an angle

Given an angle θ we can bisect the angle, that is, we can construct the angle $\theta/2$. By constructing an angle we mean that we can construct points P_1, Q_1, P_2, Q_2 such that the lines $L(P_1, Q_1)$ and $L(P_2, Q_2)$ intersect at that angle. If θ is constructible in this sense, then the numbers $\sin(\theta)$ and $\cos(\theta)$ are constructible.

Given an angle θ is it possible (just with straight-edge and compass) to construct the angle $\theta/3$? The answer is no it is not, in general, possible. For suppose that it was. Noting that $\pi/3$ is constructible, it would therefore be possible to construct an angle of $\pi/9$ and hence the number $a = \cos(\pi/9)$ would be constructible. However, a is not constructible because $\deg(a, \mathbb{Q}) = 3$. To see this, use the standard trigonometric identities to show that $\cos(3\theta) = 4\cos^3(\theta) - 3\cos(\theta)$. Letting $\theta = \pi/9$ gives $1 + 6\cos(\pi/9) - 8(\cos(\pi/9))^3$. The polynomial $1 + 6X - 8X^3 \in \mathbb{Q}[X]$ is irreducible because it is degree 3 and its image in $\mathbb{F}_5[X]$ has no roots.

Squaring the circle

Given a circle (ie., given two points: the centre and a point on the circle), is it possible to construct a square having area equal to that of the circle ?

That this is not in general possible, follows from the fact that π and therefore $\sqrt{\pi}$ is not constructible.

Doubling a cube

Given a cube (i.e., given a side length), is it possible to construct a cube of twice the volume ?

The answer is again no, since $2^{\frac{1}{3}}$ is not constructible as $\deg(2^{\frac{1}{3}}, \mathbb{Q}) = 3$.

26 Finite Fields

We have seen examples of fields that have finitely many elements, namely for any prime $p \in \mathbb{N}$, $\mathbb{F}_p = \mathbb{Z}/\langle p \rangle$ is a field and has p elements. Another example of a finite field is $\mathbb{F}_2[X]/\langle X^3 + X + 1 \rangle$: since $\mathbb{F}_2[X]$ is a PID and $X^3 + X + 1 \in \mathbb{F}_2[X]$ is irreducible, $\mathbb{F}_2[X]/\langle X^3 + X + 1 \rangle$ is a field. It has 8 elements. Is there a finite field having, for example, 6 elements? We'll see that the answer is "no".

In this section we will investigate the size and structure of finite fields. Finite fields are sometimes called **Galois fields** and a field with q elements is sometimes denoted $GF(q)$. We'll stick with the notation $\mathbb{F}_q$ for a field of size q.

26.1 All finite fields have prime power order

We first recall the definition of the characteristic of a field. If F is any field (finite or not) there is a natural homomorphism $\varphi : \mathbb{Z} \to F$ that sends $m \in \mathbb{Z}$ to the element of F given by adding $1 \in F$ to itself m times.[2] If φ is injective, we say that F is of characteristic 0. Otherwise, as F is a field, the kernel of φ is a prime ideal in $\mathbb{Z}$. Let $p \in \mathbb{N}$ be the unique (positive) prime such that $\ker(\varphi) = \langle p \rangle \lhd \mathbb{Z}$. We say that F has **characteristic** p. If a field is of characteristic 0, then it is necessarily infinite. A finite field must therefore be of characteristic p for some prime $p \in \mathbb{N}$.

Exercise 152. Give an example of an infinite field whose characteristic is not zero.

Lemma 26.1. *A field F is of characteristic p if and only if F contains a subfield isomorphic to $\mathbb{F}_p$.*

Proof. Let $\varphi : \mathbb{Z} \to F$ be the homomorphism described above. If F is of characteristic p, then $\ker(\varphi) = \langle p \rangle$ and so from the first isomorphism theorem $\mathrm{Im}(\varphi) \cong \mathbb{Z}/\langle p \rangle$. Conversely, if $\psi : \mathbb{F}_p \to F$ is an injective homomorphism, then $\varphi(m) = 1_F + \cdots + 1_F = \psi(1_{\mathbb{F}_p}) + \cdots + \psi(1_{\mathbb{F}_p})$. This implies that the characteristic of $\mathbb{F}_p$ divides the characteristic of F. □

[2] If $m < 0$, add 1 to itself $|m|$ times and then take the additive inverse.

Remark. If F has characteristic p, then there is a *unique* subfield isomorphic to $\mathbb{F}_p$, and we will identify it with $\mathbb{F}_p$.

Proposition 26.2. *Let F be a finite field of characteristic p. Then F has order p^n for some $n \geqslant 1$.*

Proof. Since F is an extension of $\mathbb{F}_p$, it is a vector space over $\mathbb{F}_p$. As F is finite, it must be finite dimensional as a vector space. Let $\{b_1, \ldots, b_n\}$ be a basis for F as an $\mathbb{F}_p$-vector space. Then $F = \{\sum_{i=1}^n \beta_i b_i \mid \beta_i \in \mathbb{F}_p\}$ which has cardinality p^n since there are p choices for each of the n β_i. □

26.2 The group of units of a finite field is cyclic

In any commutative unital ring the set of units forms an abelian group under multiplication. In the case of a finite field we will show that this group is actually cyclic. We denote by $F^\times$ the group of units of the field F.

Proposition 26.3. *Let F be a finite field. Then $F^\times$ is cyclic.*

Proof. Since $F^\times$ is a finite abelian group,we know from the structure theorem, that

$$F^\times \cong C_{d_1} \times \cdots \times C_{d_m}$$

for some $d_i \in \mathbb{Z}$, $d_i \geqslant 2$, $d_1 | \cdots | d_m$. It follows that $q-1 = |F^\times| = d_1 d_2 \ldots d_m$. Since every element of $C_{d_1} \times \cdots \times C_{d_m}$ has order that divides d_m, every element of $F^\times$ is a root of the polynomial $X^{d_m} - 1 \in F[X]$. The polynomial $X^{d_m} - 1 \in F[X]$ has at most d_m roots in F. Therefore

$$q - 1 \leqslant d_m \quad \text{and} \quad q - 1 = d_1 \cdots d_m \geqslant d_m$$

It follows that $d_1 \cdots d_m = d_m$ and it must be the case that $m = 1$ and $F^\times \cong C_{d_1}$. □

26.3 Existence

Lemma 26.4. *Let F be a finite field of size $q = p^n$. Every element of F is a root of the polynomial $X^q - X \in F[X]$.*

Proof. By Lagrange's theorem, since $|F^\times| = q - 1$, each element a of $F^\times$ satisfies $a^{(q-1)} = 1$. It follows that every element of $F^\times$ is a root of the polynomial $X^q - X$. It is clear that zero is also a root of this polynomial. □

Remark. It follows that the polynomial $X^q - X$ can be written as a product of linear polynomials from $F[X]$.

Proposition 26.5. *Let $p \in \mathbb{N}$ be prime and $n \in \mathbb{N}$ with $n \geqslant 1$. There exists a field of size p^n.*

Proof. Let $q = p^n$ and let $f \in \mathbb{F}_p[X]$ be the polynomial $f = X^q - X$. By Proposition 23.3 (and induction) there is a field $E \supseteq \mathbb{F}_p$ such that the polynomial f factors as a product of q linear polynomials from $E[X]$. Let $K \subseteq E$ be given by

$$K = \{a \in E \mid f(a) = 0\}$$

We will show that K is a subfield of E and has exactly q elements.

Since f is a degree q polynomial, it has at most q roots. We need to show that it has no repeated roots. Suppose, for a contradiction, that $(X - a)^2$ divides f in $E[X]$. Let $g \in E[X]$ be such that $f = (X - a)g$. Notice that $(X - a)$ divides g. Applying the differentiation map $D : E[X] \to E[X]$ we get

$$\begin{aligned} D(f) &= D(X - a)g + (X - a)D(g) = g + (X - a)D(g) \\ \implies qX^{q-1} - 1 &= g + (X - a)D(g) \\ \implies qa^{q-1} - 1 &= 0 \qquad \text{(since } g(a) = 0\text{)} \\ \implies -1 &= 0 \qquad \text{(since } E \text{ has characteristic } p \text{ and } q = p^n\text{)} \end{aligned}$$

As this can not be the case in the field E, we conclude that f has no repeated roots, and therefore K has exactly q elements (and not fewer).

It remains to show that K is a subfield of E. Let $a, b \in K$ with $a \neq 0$. Then $a^q = a$ and $b^q = b$, and we have

$$\begin{aligned}(ab)^q &= a^q b^q = ab \\ (a^{-1})^q &= (a^q)^{-1} = a^{-1} \\ (-a)^q &= (-1)^q a^q = -1a = -a \\ (a+b)^q &= a^q + b^q \qquad \text{(see Exercise 29)}\end{aligned}$$

It follows that K is a subfield of E. □

Exercise 153. Use the above results to show that for all $p \in \mathbb{N}$ prime and all $n \in \mathbb{N}$ $n \geqslant 1$, there exists an irreducible polynomial in $\mathbb{F}_p[X]$ of degree n. (Hint: let F be a field of size p^n and $a \in F$ a generator for the group of units. Consider $\operatorname{irr}(a, \mathbb{F}_p)$.)

26.4 Uniqueness

Proposition 26.6. *If two finite fields have the same cardinality, then they are isomorphic,*

Proof. Let F and F' be two fields of cardinality $q = p^n$. We know from Proposition 26.3 that the group $F^\times$ is cyclic. Let $a \in F$ be a generator for $F^\times$. The evaluation map $\varphi_a : \mathbb{F}_p[X] \to F$ is surjective since $\operatorname{Im}(\varphi_a)$ contains 0 and contains $F^\times$. We therefore have

$$F \cong \mathbb{F}_p[X]/\langle \operatorname{irr}(a, \mathbb{F}_p)\rangle$$

Also, $\operatorname{irr}(a, \mathbb{F}_p)$ divides $X^q - X$ in $\mathbb{F}_p[X]$ since it divides any polynomial having a as a root. In $F'[X]$ the polynomial $X^q - X$ factors as a product of linear polynomials. It follows that, considered as an element of $F'[X]$, $\operatorname{irr}(a, \mathbb{F}_p)$ factors into linear polynomials. Therefore, $\operatorname{irr}(a, \mathbb{F}_p)$ has a root a' in F'. Then $\operatorname{irr}(a', \mathbb{F}_p) = \operatorname{irr}(a, \mathbb{F}_p)$ and

$$F \cong \mathbb{F}_p[X]/\langle \operatorname{irr}(a, \mathbb{F}_p)\rangle = \mathbb{F}_p[X]/\langle \operatorname{irr}(a', \mathbb{F}_p)\rangle \cong \mathbb{F}_p(a') \subseteq F'$$

But as K and K' have the same (finite) cardinality it must be the case that $F \cong F'$. □

26.5 Exercises

154. Give an example of two infinite fields that have the same cardinality, but are not isomorphic.

155. Let $\mathbb{F}_4$ be the field containing 4 elements. Write out the addition and multiplication tables for $\mathbb{F}_4$.

156. Show that if $\psi : E \to F$ is a (ring) homomorphism from one field to another and $\ker(\psi) \neq E$, then $\psi(1_E) = 1_F$.

157. Let F be a field of size $q = p^n$. Show that every irreducible polynomial in $\mathbb{F}_p[X]$ of degree n is a factor of $X^q - X \in \mathbb{F}_p[X]$.

158. If $f \in \mathbb{F}_p[X]$ and if u is a root of f in some extension of $\mathbb{F}_p$, show that u^p is also a root of f in that extension.

159. If the finite field F has p^n elements, write down a polynomial in $F[X]$ that has no roots in F. Conclude that no finite field is algebraically closed.

160. If E is a finite field of order p^n, show that E has exactly one subfield of order p^d for any $d|n$.
(Hint: If $n = qd + r$, then $(X^n - 1) = (X^d - 1)(X^{n-d} + X^{n-2d} + \cdots + X^{n-qd}) + (X^r - 1)$. It follows that $(p^d - 1)|(p^n - 1)$ if and only if $d|n$.)

27 Introduction to Galois Theory

Galois theory gives a connection between certain field extensions and the subgroups of an associated group.

Note. In this section we will be assuming that the field F under consideration (and therefore any extension of it) is of characteristic zero.

27.1 Galois group of an extension

The set of all automorphisms of a field E forms a group (the operation being composition) which will be denoted $\mathrm{Aut}(E)$. Let H be a subgroup of $\mathrm{Aut}(E)$. The **fixed field** of H is defined by

$$E^H = \{a \in E \mid \varphi(a) = a \text{ for all } \varphi \in H\}$$

Exercise 161. Show that E^H is a subfield of E.

Now suppose that $F \leqslant E$ is a subfield. An automorphism $\varphi \in \mathrm{Aut}(E)$ is called an ***F*-automorphism** if it fixes F pointwise, that is, $\varphi(a) = a$ for all $a \in F$.

Remark. By definition, all elements of $H \leqslant \mathrm{Aut}(E)$ are E^H-automorphisms.

Example 27.1. Complex conjugation is an $\mathbb{R}$-automorphism of $\mathbb{C}$.

Exercise 162. Let $E \supseteq F$, $f \in F[X]$ and $\varphi \in \mathrm{Aut}(E)$ an F-automorphism. Show that if $a \in E$ is a root of f, then $\varphi(a)$ is also a root of f. (It follows that φ permutes of the roots of f.)

Definition 27.2. For a given subfield $F \leqslant E$, the set of all F-automorphisms forms a subgroup of $\mathrm{Aut}(E)$, called the **Galois group** of the extension. It is denoted $\mathrm{Gal}(E/F)$. That is,

$$\mathrm{Gal}(E/F) = \{\varphi \in \mathrm{Aut}(E) \mid \varphi(a) = a \text{ for all } a \in F\}$$

Example 27.3. $\mathrm{Gal}(\mathbb{Q}(\sqrt{3}, \sqrt{5})/\mathbb{Q})$ is isomorphic to the Klein four group (i.e., it has four elements and is not cyclic).

We shall see that if E is a finite extension of F, then $|\mathrm{Gal}(E/F)|$ divides $[E : F]$.

Definition 27.4. An extension E of F is called a **Galois extension** if it is a finite extension and $|\mathrm{Gal}(E/F)| = [E : F]$.

For Galois extensions, there is a correspondence between subgroups of $\mathrm{Gal}(E/F)$ and intermediate fields L, $F \leqslant L \leqslant E$. We now state the main theorem of this section. The proof will be developed later.

Theorem 27.5 (The Fundamental Theorem of Galois Theory)**.**

Let E be a Galois extension of F.

1. *The map*
$$\Phi : \{H \mid H \text{ is a subgroup of } \mathrm{Gal}(E/F)\ \} \to \{L \mid L \text{ is a subfield of } E \text{ with } F \subseteq L \subseteq E\}$$
given by
$$\Phi(H) = E^H$$
is a bijection. It has inverse given by $L \mapsto \mathrm{Gal}(E/L)$.
2. *$[E : L] = |H|$, where $L = E^H$.*

3. *$L = E^H$ is a Galois extension of F if and only if H is normal in* $\mathrm{Gal}(E/F)$*. If it is the case that L is a Galois extension of F, then* $\mathrm{Gal}(L/F) \cong \mathrm{Gal}(E/F)/H$*.*

□

27.2 Splitting fields

We give an alternative characterisation of Galois extensions. Given a polynomial in $F[X]$ we want to extend F just enough so that f has $\deg(f)$ roots.

Definition 27.6. Let $f \in F[X]$ be a non-constant polynomial. An extension field E of F is a **splitting field** for f if:

1. In $E[X]$, f factors into a product of linear polynomials, $f = (X - a_1)\dots(X - a_m)$
2. $E = F(a_1, \dots, a_m)$

It follows from Proposition 23.3 that every polynomial has a splitting field. Before giving the next somewhat technical lemma, we illustrate some of the ideas with two examples.

Example 27.7. Let $f \in \mathbb{Q}[X]$ be an irreducible quadratic. Let a, $b \in \mathbb{C}$ be its (necessarily) distinct roots. Let $E = \mathbb{Q}(a, b)$. Then $E = \mathbb{Q}(a) = \mathbb{Q}(b) = \mathbb{Q}(\sqrt{d})$ where $d \in \mathbb{Q}$ is the descriminant of f. Therefore $[E : \mathbb{Q}] = 2$. Also, we know from exercise 151 that there is an isomorphism $\mathbb{Q}(a) \to \mathbb{Q}(b)$ that sends a to b and fixes $\mathbb{Q}$. Together with the identity map, we therefore have two distinct $\mathbb{Q}$-automorphisms of E. But there can be no others, since such an automorphism permutes the roots of f. Therefore E is a Galois extension.

Exercise 163. Let $F \subseteq \mathbb{C}$ be a field and suppose that $f \in F[X]$ is an irreducible quadratic. Let the roots of f be $a, b \in \mathbb{C}$. Show that

1. $F(a) = F(a, b)$
2. $|\mathrm{Gal}(F(a)/F)| = 2$
3. The non-trivial element in $\mathrm{Gal}(F(a)/F)$ interchanges a and b.

Example 27.8. We show that if $E \subset \mathbb{C}$ is a splitting field of the polynomial $f = X^3 + 3X + 1 \in \mathbb{Q}[X]$, then E is a Galois extension of $\mathbb{Q}$.

Note that the polynomial f is irreducible, since its image in $\mathbb{F}_2[X]$ is irreducible. Therefore f has no repeated roots (see exercise 142). Let $\alpha, \beta, \gamma \in \mathbb{C}$ be the three roots in $\mathbb{C}$ of this polynomial. Let $E = \mathbb{Q}(\alpha, \beta, \gamma) \subseteq \mathbb{C}$. We will show that E is a Galois extension of $\mathbb{Q}$ and find the Galois group $\mathrm{Gal}(E/\mathbb{Q})$.

Exactly one of the roots, α say, is real (and therefore $\gamma = \overline{\beta}$). Let $L = \mathbb{Q}(\alpha)$. Notice that $L \neq E$ since $L \subseteq \mathbb{R}$. Also, $[L : \mathbb{Q}] = \deg(\alpha, \mathbb{Q}) = 3$. In $L[X]$ we have the factorisation $f = (X-\alpha)h$ for some quadratic $h \in L[X]$ with the roots of h being exactly β and γ. Since $\beta \notin L$, h is irreducible and $\deg(\beta, L) = 2$. Since $E = \mathbb{Q}(\alpha, \beta, \gamma) = L(\beta, \gamma) = L(\beta)$, we have $[E : L] = 2$. From Lemma 24.11 we get

$$[E : \mathbb{Q}] = [E : L][L : \mathbb{Q}] = 2 \times 3 = 6$$

Since the elements of $\mathrm{Gal}(E/\mathbb{Q})$ permute the roots of f (exercise 162) we know that $\mathrm{Gal}(E/\mathbb{Q})$ is isomorphic to a subgroup of S_3 (the symmetric group on three objects). Since $|S_3| = 6$ we conclude that $|\mathrm{Gal}(E/\mathbb{Q})|$ divides 6. We'll show that $\mathrm{Gal}(E/\mathbb{Q})$ has at least 4 distinct elements, from which it follows that $\mathrm{Gal}(E/\mathbb{Q}) \cong S_3$ and $|\mathrm{Gal}(E/\mathbb{Q})| = 6 = [E : \mathbb{Q}]$.

The identity and complex conjugation are $\mathbb{Q}$-automorphisms that permute the roots of f and are therefore in $\mathrm{Gal}(E/\mathbb{Q})$. We demonstrate two other elements in $\mathrm{Gal}(E/F)$.

Let $F = \mathbb{Q}(\gamma)$ and let $g \in F[X]$ be such that $f = (X - \gamma)g$. Note that $E = F(\alpha) = F(\beta)$. Applying exercise 163 to F and g, we see that there is an element of $\mathrm{Gal}(E/F)$ (which is a subset of $\mathrm{Gal}(E/\mathbb{Q})$)) that interchanges α and β. The same argument, with the roles of β and γ interchaged, shows that there is an element in $\mathrm{Gal}(E/\mathbb{Q})$ that fixes β and swaps α and γ.

In fact all splitting fields are Galois extensions. To prove this we will use the following technical result.

Lemma 27.9. *Let $\varphi : F \to F'$ be an isomorphism of fields. Let $f \in F[X]$ be a polynomial and let $f' \in F'[X]$ be the image of f by (the extension to $F[X]$ of) φ. Let $E \supseteq F$ and $E' \supseteq F'$ be splitting fields for f and f' respectively. Then, φ extends to an isomorphism from E to E'. Moreover, the number of such isomorphisms is $[E : F]$.*

Proof. Denote by $\tilde{\varphi} : F[X] \to F'[X]$ the map defined by extending φ, that is

$$\tilde{\varphi}(a_0 + \cdots + a_m X^m) = \varphi(a_0) + \cdots + \varphi(a_m)X^m$$

By hypothesis $\tilde{\varphi}(f) = f'$. We proceed by induction on $[E : F]$.

If $[E : F] = 1$, then $E = F$ and f factors into linear polynomials in $F[X]$. It follows that f' factors into linear polynomials in $F'[X]$, and therefore $E' = F'$. Then φ itself is an isomorphism from E to E', and it is obviously the only such.

Now suppose that $[E : F] > 1$ and that the result holds for all cases with lower degree. Let $a \in E$ be a root of f, with $a \notin F$. Let $g = \mathrm{irr}(a, F) \in F[X]$ and $g' = \tilde{\varphi}(g)$. Then $g' \in F'[X]$ is irreducible and $\deg(g') = \deg(g) = [F(a) : F]$. Since g' is irreducible and F' has characteristic zero, g' has no repeated roots (see exercise 142). For each of the $[F(a) : F]$ (distinct) roots b of g' there is exactly one injective homomorphism $\xi : F(a) \to E'$ such that $\xi|_F = \varphi$ and $\xi(a) = b$ (c.f. exercise 151). Moreover, any injective homomorphism from $F(a)$ to E' that restricts to φ, must send a to one of the roots of g'. It follows that there are exactly $[F(a) : F]$ homomorphisms from $F(a)$ to E' that restrict to φ. Since $[E : F(a)] < [E : F]$ we can apply the induction hypothesis, to conclude that there are $[E : F(a)]$ isomorphisms from E to E' that extend ξ. Combining, we see that the total number of isomorphisms from $E \to E'$ that extend φ is $[E : F(a)][F(a) : F] = [E : F]$. Note that any isomorphism $\psi : E \to E'$ is an extension of $\psi|_{F(a)}$ and $\psi(a)$ is necessarily a root of g'.

$$\begin{tikzcd}
E \arrow[r, "\psi"] & E' \\
F(a) \arrow[u] \arrow[r, "\xi"] & F'(b) \arrow[u] \\
F \arrow[u] \arrow[r, "\varphi"] & F' \arrow[u]
\end{tikzcd}$$

This diagram illustrates the above argument. The vertical arrows are just inclusions, and each horizontal map is a restriction of the one above it. The inductive hypothesis applies to the extension of ξ to ψ. Note that E is a splitting field for f over $F(a)$ and E' is a splitting field for f' over $F'(b)$.

□

Corollary 27.10. *Any two splitting fields of $f \in F[X]$ are isomorphic.*

Proof. Apply the lemma with $F' = F$, E and E' the two splitting fields and $\varphi = \mathrm{Id}_F$. □

Corollary 27.11. *A splitting field of $f \in F[X]$ is a Galois extension of F.*

Proof. Let E be a splitting field for F. Apply the lemma with $F' = F$, $E' = E$ and $\varphi = \mathrm{Id}_F$. The extensions of φ are precisely the F-automorphisms of E. Therefore $|\,\mathrm{Gal}(E/F)| = [E : F]$ by the lemma. □

Exercise 164. Use the above lemma to show the following: Let F be a field, $f \in F[X]$ and $E \supseteq F$ a splitting field for f. Let $g \in F[X]$ be irreducible and such that g divides f. Let $a, b \in E$ be two roots of g. Then there is an F-automorphism of E sending a to b.

27.3 Primitive elements, the orbit lemma and Artin's theorem

In order to prove Artin's Theorem and to see that all Galois extensions are splitting fields, we will use the following

Proposition 27.12. *Let E be a finite extension of F. There exists an element $a \in E$ such that $E = F(a)$.*

Remark. Such an element is called a **primitive element** of the extension.

Proof. Since E is a finite extension, there are elements $b_i \in E$ such that $E = F(b_1, \ldots, b_k)$. By induction it is enough to consider the case in which $E = F(b, c)$ with $b, c \in E \setminus F$. Let $f = \operatorname{irr}(b, F)$, $g = \operatorname{irr}(c, F)$ and let $L \subseteq E$ be the splitting field for the polynomial fg. Let $b = b_1, b_2, \ldots, b_m \in L$ be the roots of f and $c = c_1, c_2, \ldots, c_n$ be the roots of g. Since f and g are irreducible and F is of characteristic zero, both f and g have no repeated roots. Since F is of characteristic zero, it is infinite, and we can therefore choose $d \in F$ such that

$$d \notin \{(b_j - b)(c - c_i)^{-1} \mid 2 \leqslant i \leqslant m,\ 1 \leqslant j \leqslant n\} \subseteq L$$

Let $a = b + dc \in F(b, c) \subseteq L$. We will show that $F(b, c) \subseteq F(a)$. Let $h \in F(a)[X]$ be given by $h = f(a - dX)$. Then $h(c) = f(b) = 0$ and by the choice of d we have $h(c_i) = f(a - dc_i) \neq 0$ if $i \geqslant 2$.

Therefore c is the only common root of h and g. Since g factors as a product of linear terms in $L[X]$ we have that the gcd of g and h in $L[X]$ is $(X - c)$. On the other hand, any gcd of g and h in $F(a)[X]$ is also a gcd of g and h in $L[X]$ (see exercise I.10.4.5). Therefore $(X - c)$ is a gcd of g and h in $F(a)[X]$. It follows that $c \in F(a)$ and therefore also $b \in F(a)$. Thus $F(b, c) \subseteq F(a)$. The reverse inclusion is clear from the choice of a. □

As well as being used in our proof of Artin's Theorem, the next lemma is often useful in determining the irreducible polynomial of an element.

Lemma 27.13 (Orbit Lemma)**.** *Let E be a field and let G be a finite subgroup of $\operatorname{Aut}(E)$. Let $a \in E$ and let $\{a = a_1, a_2, \ldots, a_m\} \subseteq E$ be the orbit of a under the action of G. Then in $E[X]$ we have $\operatorname{irr}(a, E^G) = (X - a_1)(X - a_2) \cdots (X - a_m)$.*

Proof. Let $f = (X - a_1) \cdots (X - a_m)$ and let $F = E^G$. First note that $f \in F[X]$ since for all $g \in G$

$$\tilde{g}(f) = \tilde{g}((X - a_1) \cdots (X - a_m)) = (X - g(a_1)) \cdots (X - g(a_m)) = (X - a_1) \cdots (X - a_m) = f$$

where $\tilde{g} : E[X] \to E[X]$ is the homomorphism induced by g.

Then note that

$$\begin{aligned} \operatorname{irr}(a, F)(a_i) &= \operatorname{irr}(a, F)(ga) && \text{(for some } g \in G\text{)} \\ &= g(\operatorname{irr}(a, F)(a)) = g(0) = 0 \end{aligned}$$

So all the a_i are roots of $\operatorname{irr}(a, F)$. Therefore f divides $\operatorname{irr}(a, F)$. But since $f(a) = 0$, we also have that $\operatorname{irr}(a, F)$ divides f.

□

Remark. Since the order of an orbit divides the order of the group acting, we know that m divides $|G|$. It need not be equal to $|G|$.

Theorem 27.14 (Artin's Theorem)**.** *Let E be a field and let G be a finite subgroup of $\operatorname{Aut}(E)$. Then*

$$[E : E^G] = |G|$$

Corollary 27.15. *If E is a finite extension of F, then $|\operatorname{Gal}(E/F)|$ divides $[E:F]$.*

Proof. We have that $F \subseteq E^{\operatorname{Gal}(E/F)} \subseteq E$, which implies that

$$[E:F] = [E:E^{\operatorname{Gal}(E/F)}][E^{\operatorname{Gal}(E/F)}:F] = |\operatorname{Gal}(E/F)|[E^{\operatorname{Gal}(E/F)}:F]$$

□

Corollary 27.16. *Let E be a field and let G be a finite subgroup of $\operatorname{Aut}(E)$. Then E is a Galois extension of E^G and $\operatorname{Gal}(E/E^G) = G$.*

Proof. Clearly $G \subseteq \operatorname{Gal}(E/E^G)$ since all elements of G fix E^G pointwise. Corollary 27.15 implies that $|\operatorname{Gal}(E/E^G)| \leqslant [E:E^G]$. Then from Theorem 27.14 we have

$$[E:E^G] = |G| \leqslant |\operatorname{Gal}(E/E^G)| \leqslant [E:E^G]$$

It follows that $G = \operatorname{Gal}(E/E^G)$ and $|\operatorname{Gal}(E/E^G)| = [E:E^G]$. □

Corollary 27.17. *If E be a Galois extension of F, then $E^{\operatorname{Gal}(E/F)} = F$.*

Proof. Let $G = \operatorname{Gal}(E/F)$. We have $F \subseteq E^G \subseteq E$ and therefore $\operatorname{Gal}(E/E^G) \subseteq G$. It is also the case that $G \subseteq \operatorname{Gal}(E/E^G)$ since all elements of G fix its own fixed field. Therefore $G = \operatorname{Gal}(E/E^G)$. Also,

$$\begin{aligned} & |\operatorname{Gal}(E/E^G)| \text{ divides } [E:E^G] && \text{(by Corollary 27.15)} \\ \Longrightarrow\ & |G| \text{ divides } [E:E^G] && \\ \Longrightarrow\ & [E:F] \text{ divides } [E:E^G] && \text{(since } E \text{ is a Galois extension of } F) \end{aligned}$$

But is is also the case that $[E:E^G]$ divides $[E:F]$ since

$$[E:F] = [E:E^G][E^G:F]$$

Therefore $[E:F] = [E:E^G]$ and $[E^G:F] = 1$. □

We now give the proof of Artin's Theorem.

Proof of Theorem 27.14. Let $F = E^G$. We first show that E is a finite extension of F. By Lemma 27.13 every element $a \in E$ is algebraic over F and $\deg(a, F)$ divides $|G|$. Starting with $F_0 = F$ we define a sequence of extensions F_i of F. If $F_i \neq E$, let $a_i \in E \setminus F_i$ and define $F_{i+1} = F_i(a_i)$.

Suppose, for a contradiction, that this process continues indefinitely to give an infinite chain of subfields

$$F \subsetneqq F_1 \subsetneqq F_2 \subsetneqq \cdots$$

Noting that F_{i+1} is a finite extension of F_i, we have that for all i, F_i is a finite extension of F and also that $[F_i : F] \geqslant 2^i$. By Proposition 27.12, for all i, there exists an element $b_i \in E$ such that $F_i = F(b_i)$ and therefore $[F_i : F] = \deg(b_i, F)$. As noted at the beginning of the proof, $\deg(b_i, F)$ divides $|G|$. A contradiction.

We have shown that there exists an element $b \in E$ such that $E = F(b)$. Notice that b must have trivial stabiliser in G since if $g \in G$ fixes b it fixes the whole of E (pointwise) and is therefore the identity homomorphism. Since b has trivial stabiliser, the size of its orbit is exactly $|G|$. Lemma 27.13 tells us that the size of the orbit of b is equal to $\deg(b, F) = [F(b) : F]$.

□

And finally, we show that all Galois extensions are splitting fields.

Proposition 27.18. *Let E be a Galois extension of F. Then there exists a polynomial $f \in F[X]$ such that E is a splitting field for f.*

Proof. Let $a \in E$ be such that $E = F(a)$, and let $f = \operatorname{irr}(a, F)$. Let $\{a = a_1, a_2, \ldots, a_m\}$ be the orbit of a under $\operatorname{Gal}(E/F)$. Then $F = E^{\operatorname{Gal}(E/F)}$ by Corollary 27.17 and Lemma 27.13 tells us that in $E[X]$ $f = (X - a_1) \cdots (X - a_m)$. Therefore E is a splitting field for f. □

27.4 Proof of the fundamental theorem

Recall the statement of the theorem.

Theorem 27.19 (The main theorem of Galois theory)**.** *Let E be a Galois extension of F.*

1. *The map*
$$\Phi : \{H \mid H \textit{ is a subgroup of } \operatorname{Gal}(E/F)\ \} \to \{L \mid L \textit{ is a subfield of } E \textit{ with } F \subseteq L \subseteq E\}$$
given by
$$\Phi(H) = E^H$$
is a bijection. It has inverse given by $L \mapsto \operatorname{Gal}(E/F)$.
2. *$[E : L] = |H|$, where $L = E^H$.*
3. *$L = E^H$ is a Galois extension of F if and only if H is normal in $\operatorname{Gal}(E/F)$. If it is the case that L is a Galois extension of F, then $G(L/F) \cong \operatorname{Gal}(E/F)/H$.*

□

Proof. Let $G = \operatorname{Gal}(E/F)$ and let

$$\Psi : \{L \mid L \text{ is a subfield of } E \text{ with } F \subseteq L \subseteq E\} \to \{H \mid H \text{ is a subgroup of } G\ \}$$

be the map $\Psi(L) = \operatorname{Gal}(E/L)$. By Corollary 27.16, $\Psi \circ \Phi(H) = \Psi(E^H) = \operatorname{Gal}(E/E^H) = H$. Applying Corollary 27.17 gives $\Phi \circ \Psi(L) = \Phi(\operatorname{Gal}(E/L)) = E^{\operatorname{Gal}(E/L)} = L$. Hence Ψ and Φ are mutually inverse bijections. This proves the first part of the statement.

The second part is a direct consequence of Artin's Theorem (Theorem 27.14).

For the third part note that given any $g \in G$ and any subgroup $H \leqslant G$ we have $E^{gHg^{-1}} = gE^H$. It follows that H is normal in G if and only if $gE^H = E^H$ for all $g \in G$.

Suppose that H is a normal subgroup of G. Then for all $g \in G$, $gL = \Phi(gHg^{-1}) = \Phi(H) = L$. We therefore have, by restriction, a map $G \to G(L/F)$. The kernel of this homomorphism is equal to H (all the elements of G that fix L pointwise). Then G/H is isomorphic to a subgroup of $G(L/F)$ and noting that $|G| = [E : F] = [E : L][L : F] = |H|[L : F]$ we get

$$\begin{aligned} & |G/H| \leqslant |G(L/F)| && \text{(since it is isomorphic to a subgroup)} \\ \Longrightarrow\ & |G|/|H| \leqslant |G(L/F)| && \\ \Longrightarrow\ & [L : F] \leqslant |G(L/F)| && \text{(since } |G| = |H|[L : F]\text{)} \end{aligned}$$

It is also the case that $|G(L/F)|$ divides $[L : F]$ by Corollary 27.15. Therefore $|G(L/F)| = [L : F]$ and so L is a Galois extension of F.

Conversely, suppose that L is a Galois extension of F. Then L is a splitting field for some $f \in F[X]$ and every element of G permutes the roots of f. This implies that $gL = L$. □

27.5 Examples

Having proved the main theorem we now give some examples in which we calculate the Galois group and list the subgroups together with corresponding subfileds.

Example 27.20 (Quadratic extension)**.** Let E be an extension of $\mathbb{Q}$ with $[E : \mathbb{Q}] = 2$. Then $E = \mathbb{Q}(a)$ for some a with $\deg(a, \mathbb{Q}) = 2$. Let b be the other root of the polynomial $\operatorname{irr}(a, \mathbb{Q})$. Note that, being irreducible over $\mathbb{Q}$, $\operatorname{irr}(a, \mathbb{Q})$

has no repeated roots and so $b \neq a$. Consider the group $G = \text{Gal}(E/\mathbb{Q})$. For any element $g \in G$, we have either $g(a) = a$ (and therefore $g = \text{Id}$) or $g(a) = b$. If $g(a) = b$, then we must similarly have $g(b) = a$. Therefore G has exactly two elements, and E is a Galois extension of $\mathbb{Q}$. Since C_2 has no proper subgroups, there are no fields lying between $\mathbb{Q}$ and E. Note that we know from Exercise 2.3.6 that there exists an element of $\text{Gal}(E/\mathbb{Q})$ that sends a to b.

Example 27.21 (Non Galois extension). Let $E = \mathbb{Q}(2^{\frac{1}{3}})$. Then $E \subseteq \mathbb{R}$. Any element of $\text{Gal}(E/\mathbb{Q})$ must permute the roots of the polynomial $\text{irr}(2^{\frac{1}{3}}, \mathbb{Q}) = X^3 - 2$. Since only one of these roots lies in E (since the others are not in $\mathbb{R}$), any element of $\text{Gal}(E/\mathbb{Q})$ must send $2^{\frac{1}{3}}$ to $2^{\frac{1}{3}}$. Such an automorphism fixes E pointwise. Therefore $|\text{Gal}(E/\mathbb{Q})| = 1 \neq 3 = \deg(2^{\frac{1}{3}}, \mathbb{Q}) = [E : \mathbb{Q}]$ and E is not a Galois extension.

Example 27.22 (Biquadratic extension). Let $E = \mathbb{Q}(i, \sqrt{2})$. There are $\mathbb{Q}$-automorphisms $\sigma, \tau \in \text{Gal}(E/\mathbb{Q})$ determined by

$$\begin{aligned} \sigma(i) &= -i & \sigma(\sqrt{2}) &= \sqrt{2} \\ \tau(i) &= i & \tau(\sqrt{2}) &= -\sqrt{2} \end{aligned}$$

Since any element of $\text{Gal}(E/\mathbb{Q})$ must permute the roots of X^2+1 and the roots of X^2-2, $\text{Gal}(E/\mathbb{Q}) = \{id, \sigma, \tau, \sigma\tau\}$ and we have $\text{Gal}(E\mathbb{Q}) \cong C_2 \oplus C_2$ and $|\text{Gal}(E/\mathbb{Q}) = 4|$. Also, $\mathbb{Q} \subsetneq \mathbb{Q}(\sqrt{2}) \subsetneq \mathbb{Q}(\sqrt{2}, i)$ which implies that $[E : \mathbb{Q}] = [E : \mathbb{Q}(\sqrt{2})][\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2 \times 2 = 4$. Therefore $|\text{Gal}(E/\mathbb{Q}) = 4| = [E : \mathbb{Q}]$ and E is a Galois extension of $\mathbb{Q}$. It is a splitting field of the polynomial $(X^2 - 2)(X^2 + 1)$. The correspondence between subgroups and intermediate fields is given in the following table:

$E = \mathbb{Q}(i, \sqrt{2})$	
subgroup	**subfield**
$\text{Gal}(E/\mathbb{Q})$	$\mathbb{Q}$
$\{id, \tau\}$	$\mathbb{Q}(i)$
$\{id, \sigma\}$	$\mathbb{Q}(\sqrt{2})$
$\{id, \sigma\tau\}$	$\mathbb{Q}(i\sqrt{2})$
$\{id\}$	K

Since G is abelian, all subgroups are normal, and therefore all the intermediate fields are Galois extensions of $\mathbb{Q}$ (which also follows from the fact that all the (proper) intermediate fields are quadratic extensions of $\mathbb{Q}$).

Example 27.23 (Example 5.2.2 continued). Let $E \subseteq \mathbb{C}$ be the splitting field of $X^3 + 3X + 1$. We have already seen that $\text{Gal}(E/\mathbb{Q}) \cong S_3$. Let $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{C}$ be the three roots in $\mathbb{C}$ of this polynomial, with $\alpha_1 \in \mathbb{R}$ and $\alpha_2 = \overline{\alpha_3} \notin \mathbb{R}$. Since any element of $\text{Gal}(E/\mathbb{Q})$ must permute the elements of $\{\alpha_1, \alpha_2, \alpha_3\}$ and $\text{Gal}(E/\mathbb{Q}) = 6$, we know that all permutations of the roots are achievable by an element of $\text{Gal}(E/\mathbb{Q})$. We label each element of G by the corresponding permutation, e.g. (12) represents the automorphism determined by swapping α_1 and α_2 but leaving α_3 fixed. Knowing the subgroups of S_3, we can list all intermediate fields.

$f = X^3 + 3X + 1$	
subgroup	**subfield**
$\text{Gal}(E/\mathbb{Q}) = S_3$	$\mathbb{Q}$
$H = \{id, (123), (132)\}$	$L = \mathbb{Q}(\delta)$
$\{id, (12)\}$	$\mathbb{Q}(\alpha_3)$
$\{id, (13)\}$	$\mathbb{Q}(\alpha_2)$
$\{id, (23)\}$	$\mathbb{Q}(\alpha_1)$
$\{id\}$	K

The determination of L needs some explanation. By the Main Theorem there is some subfield that corresponds to the subgroup H. Call it L. Note that $[L : \mathbb{Q}] = [G : H]$, so $[L : \mathbb{Q}] = 2$. Now let $\delta = (\alpha_1 - \alpha_2)(\alpha_2 - \alpha_3)(\alpha_3 - \alpha_1)$. Then $\delta \in L = K^H$ since it is fixed by the automorphism corresponding to the permutation (123). Therefore $\mathbb{Q}(\delta) \subseteq L$. Also, $\delta \notin \mathbb{Q}$ since it is not fixed by the automorphism corresponding to (12) (it sends δ to $-\delta$). Then $[\mathbb{Q}(\delta) : \mathbb{Q}] \geqslant 2$, and $2 = [L : \mathbb{Q}] = [L : \mathbb{Q}(\delta)][\mathbb{Q}(\delta) : \mathbb{Q}] \geqslant [L : \mathbb{Q}(\delta)] \times 2$ which implies that $[L : \mathbb{Q}(\delta)] = 1$ and $L = \mathbb{Q}(\delta)$.

The subfields $\mathbb{Q}(\alpha_1)$, $\mathbb{Q}(\alpha_2)$ and $\mathbb{Q}(\alpha_3)$ are not Galois extensions of $\mathbb{Q}$ because the order 2 subgroups of S_3 are not normal (since, for example, $(23)(12)(23)^{-1} = (13)$). The field L is a Galois extension of $\mathbb{Q}$ since the subgroup H is normal in S_3.

27.6 Exercises

165. (a) Show that $\mathrm{Aut}(\mathbb{Q})$ is trivial.
    (b) Let $\varphi : \mathbb{R} \to \mathbb{R}$ be an automorphism of $\mathbb{R}$.
        i) Show that if $x > 0$, then $\varphi(x) > 0$.
        ii) Show that if $x > y$, then $\varphi(x) > \varphi(y)$.
        iii) Use this to conclude that φ is the identity map, that is that $\mathrm{Aut}(\mathbb{R})$ is trivial.
    (c) Show that the only *continuous* automorphisms of $\mathbb{C}$ are the identity and complex conjugation.

166. Determine the degree of the splitting fields of the following elements of $\mathbb{Q}[X]$.

    (a) $X^4 - 1$ (b) $X^3 - 2$ (c) $X^4 + 1$

167. Show that $K = \mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5})$ is a Galois extension of $\mathbb{Q}$ and identify its Galois group.

168. Show that $X^2 - 3$ and $X^2 - 2X - 2$ are irreducible in $\mathbb{Q}[X]$ and have the same splitting field.

169. Find the dimensions of the splitting fields over $\mathbb{Q}$ of

    (a) $X^3 - 56$; and (b) $X^4 - 4X^2 - 5$.

170. Find the dimension of a splitting field of $X^3 + X + 1$ over $\mathbb{F}_2$.

171. For each of the following polynomials $f \in \mathbb{Q}[X]$ calculate:

    (i) The size of the Galois group $G = \mathrm{Gal}(E/\mathbb{Q})$, where E is the splitting field of f (over $\mathbb{Q}$);
    (ii) Identify the group G;
    (iii) List the correspondence between subgroups of G and intermediate fields L, $\mathbb{Q} \subseteq L \subseteq E$.

    (a) $X^2 - 5X + 6$
    (b) $X^2 - 2$
    (c) $X^4 - X^2 - 2$
    (d) $X^3 - 7$ (Which subfields of E are Galois extensions of $\mathbb{Q}$?)
    (e) $X^3 - 1$
    (f) $X^5 - 1$
    (g) $X^4 - 2$

Appendix A

A tiny bit of group theory revision

Here is some elementary group theory as a reference and to present some results used in the text.

Definition A.1. A **group** consists of a nonempty set G together with a map $G \times G \to G$ (the image of $(g, h) \in G \times H$ being denoted $g \times h$ or $g + h$ or simply gh) satisfying the following axioms:

1. *associativity*: For all $g_1, g_2, g_3 \in G$, $(g_1 \times g_2) \times g_3 = g_1 \times (g_2 \times g_3)$
2. *identity element*: There exists an element $e \in G$ such that for all $g \in G$, $eg = g$ and $ge = g$
3. *inverses*: For all $g \in G$ there exists $h \in G$ such that $gh = e$ and $hg = e$

The group is denoted $(G, \times)$ or $(G, +)$ or simply G. The identity element is often denoted 1 or 0 depending on whether the operation is denoted $\times$ or $+$. The inverse of an element g is denoted g^{-1} or $-g$. A group is called **finite** if the underlying set G is finite. A group is an **abelian group** if it satisfies the further condition

commutativity: for all $g_1, g_2 \in G$, $\quad g_1 \times g_2 = g_2 \times g_1$

Example A.2. Let $n \in \mathbb{N}^+$. The **symmetric group** S_n is the finite group consisting of all bijections from the set $\{1, 2, \ldots, n\}$ to itself. The group operation is given by composition of functions. There are various notations used for the elements of S_n. An efficient way is to use **cycle notation**. For $m \leqslant n$ and distinct elements $a_1, a_2, \ldots a_m \in \{1, 2, \ldots n\}$, we write $(a_1 a_2 \ldots a_m)$ to denote the bijection $\sigma : \{1, 2, \ldots n\} \to \{1, 2, \ldots n\}$ given by $\sigma(a_1) = a_2$, $\sigma(a_1) = a_2$,..., $\sigma(a_m) = a_1$ and $\sigma(b) = b$ for all $b \notin \{a_1, a_2, \ldots a_m\}$. Every element of S_n can be written as a product of disjoint cycles. Cycles of length 1 are usually omitted. An element of S_n the is a single cycle of length 2 is called a **transposition**. Some examples of multiplying elements, which you can check by thinking about the corresponding bijections, are:

$$\begin{aligned}
(132)(58) &= (58)(132) \quad \text{(disjoint cycles commute)} \\
(132)(53) &= (1352) \\
(53)(132) &= (1532) \quad (S_n \text{ is not abelian if } n \geqslant 3) \\
(1532) &= (12)(13)(15) \quad \text{(Every element of } S_n \text{ can be written as a product of transpositions)}
\end{aligned}$$

Definition A.3. A **subgroup** of a group G is a subset $H \subset G$ such that, when endowed with the operation inherited from G (i.e., the map $H \times H \to H$ is the restriction of that for G), H forms a group. A subgroup H of G is a **normal subgroup** if for all $g \in G$ and for all $h \in H$ we have $ghg^{-1} \in H$.

Example A.4.

1. For any group G, $\{e\}$ (called the trivial subgroup) and G itself are normal subgroups.
2. $A_3 = \{(123), (132), e\}$ and $\{(13), e\}$ are subgroups of S_3. The first is normal; the second is not.

Definition A.5. Let H be a subgroup of a group G. A (left) **coset** of H in G is a subset of the form

$$gH = \{gh \mid h \in H\}$$

Theorem A.6 (Lagrange's Theorem). *Let G be a finite group. If H is a subgroup of G, then $|H|$ divides $|G|$.*

Sketch of proof. The (left) cosets of H in G partition G. All cosets have cardinality $|H|$. Therefore $|G| = |H| \times [G : H]$, where $[G : H]$ denotes the number of cosets of H in G. □

It follows that the order of any element of G divides the order of G. If $n \in \mathbb{N}$ divides the order of G, there may or may not be an element in G of order n. For prime divisors of $|G|$ such an element does always exist, as we will see below (Cauchy's Theorem).

Definition A.7. Let G be a group and X a set. An **action** of G on X is a homomorphism $G \to \mathrm{Aut}(X)$, where $\mathrm{Aut}(X)$ is the group of all bijections from X to itself. In other words, a (left) action is a map $G \times X \to X$ (with the image of $(g, x) \in G \times X$ being denoted gx) satisfying:

1. for all $x \in X$, $1x = x$
2. for all $g, h \in G$ and for all $x \in X$, $(gh)x = g(hx)$

Given an action of G on X, the **stabiliser** of $x \in X$ is $\mathrm{Stab}_G\, x = \{g \in G \mid gx = x\}$

Proposition A.8 (Orbit Stabiliser Relation). *Let G be a finite group acting on a set X. Then for any $x \in X$,*

$$|Gx| \times |\mathrm{Stab}_G(x)| = |G|$$

Sketch of proof. Let $H = \mathrm{Stab}_G(x)$. Note that $gx = hx$ if and only if $gH = hH$. Therefore $|Gx| = [G : H]$. Apply Lagrange's Theorem. □

Proposition A.9. *Let G be a finite group and let $S \subset G$ be a subset. Let $H = \{g \in G \mid \forall s \in S, gs \in S\}$ be the stabiliser of S under the action of G on itself by left-multiplication. Then $|H|$ divides $|S|$.*

Proof. Consider the action of H on G by left multiplication. The set S is a disjoint union of H-orbits. Each H-orbit, being a right coset of H in G, has size equal to $|H|$. □

Theorem A.10 (First Sylow Theorem). *Let G be a finite group, and let $p \in \mathbb{N}$ be prime. Suppose G has order $p^m k$ where k is not divisible by p. Then G contains a subgroup of order p^m.*

Proof. We use the orbit-stabilizer relation and A.9. Let U be the collection of all subsets of G of size p^m. Then $|U| = \binom{p^m k}{p^m}$, which is not divisible by p. The group G acts on U. Since U is a disjoint union of orbits, and $|U|$ is not divisible by p, it follows that there is an orbit that has size not divisible by p. Let $S \in U$ be such that the G-orbit of S has size not divisible by p. Let H be the stabilizer of S. Then $|H|$ divides $|S| = p^m$ by A.9. Therefore $|H| = p^i$ for some $i \leqslant m$. Also, by the orbit-stabiiser theorem applied to the action of G on U, $p^m k = |H| \times |O_S|$, where $O_S \subset U$ is the orbit of S. Since $|O_S|$ is not divisible by p, we conclude that $|H| = p^m$. □

Corollary A.11 (Cauchy's Thoerem). *Let G be a finite group. If p divides $|G|$, then there exists an element $g \in G$ with $|g| = p$.*

Proof. Write $|G| = p^m k$ as in the theorem, and let $H \leqslant G$ be a subgroup of size p^m. Let $h \in H \setminus \{1\}$. Then h has order p^i for some $i \in \{1, \ldots, m\}$. The element $h^{p^{i-1}}$ has order p. □

A **simple group** is a group that has no normal subgroups other than itself and the trivial subgroup. A classification of all finite simple groups was completed in 2004. All groups of prime order are easily seen to be simple. The smallest non-cyclic simple group is the alternating group A_5.

Definition A.12. The **alternating group** A_n is the subgroup of S_n consisting of all those elements that can be written as a product of an even number of transpositions.

Example A.13.

1. $A_3 = \{e, (123), (132)\}$
2. $A_4 = \{e, (123), (132), (124), (142), (134), (143), (234), (243), (12)(34), (13)(24), (14)(23)\}$
3. A_5 consists of 60 elements: the identity, 20 3-cycles, 24 5-cycles and 15 double transpositions (product of two disjoint transpositions).

Proposition A.14. *The alternating group A_5 is simple.*

Proof. Suppose that $H \lhd A_5$ is a normal subgroup and $H \neq \{e\}$. We will show that $H = A_5$.

First we show that H contains a 3-cycle. Let $\sigma \in A_5 \setminus \{e\}$. Then σ is either a 5-cycle, a 3-cycle or a product of two 2-cycles. In the first case, say $\sigma = (abcde)$. Then

$$\sigma(ace)\sigma(ace)^{-1} = (adb) \in H$$

If $\sigma = (ab)(cd)$, then

$$\sigma(ae)(cd)\sigma(ae)(cd) = (abe) \in H$$

It follows that H contains a 3-cycle.

Now we show that H contains *every* 3-cycle. Suppose $(abc) \in H$ and let $(xyz) \in A_5$ be any other 3-cycle. Let $\tau, \rho \in S_5$ be given by

$$\begin{array}{lllll} \tau(a) = x & \tau(b) = y & \tau(c) = z & \tau(d) = u & \tau(e) = v \\ \rho(a) = x & \rho(b) = y & \rho(c) = z & \rho(d) = v & \rho(e) = u \end{array}$$

(Where $\{a, b, c, d, e\} = \{x, y, z, u, v\} = \{1, 2, 3, 4, 5\}$.) Since $\rho = (uv)\tau$, exactly one of ρ or τ is in A_5. Noting that

$$\rho(abc)\rho^{-1} = (xyz) \quad \text{and} \quad \tau(abc)\tau^{-1} = (xyz)$$

we conclude that $(xyz) \in H$. Therefore H contains all 3-cycles.

Finally, we note that all even permutations can be written as a product of 3-cycles. To see this, it's enough to note that

$$(xy)(uv) = (vxu)(xyv) \quad \text{and} \quad (xy)(yz) = (xyz)$$

It follows that H contains all even permutations, that is, $H = A_5$. □

Appendix B

Some other algebraic structures

This is currently not much more than a very short list of some other standard algebraic structures of which it's good to be aware.

Algebras

Definition B.1. Let k be a commutative unital ring. A k-**algebra** is a ring R that is also a k-module and satisfies the property that $a(rs) = (ar)s = r(as)$ for all $a \in k$ and $r, s \in R$.

Examples B.2.

1. Every ring is a $\mathbb{Z}$-algebra.

2. $\mathbb{C}[X]$ is a $\mathbb{C}$-algebra.

3. For any commutative unital ring k, $M_n(k)$ is a k-algebra.

C^*-algebras

Given a compact space X we can consider the $\mathbb{C}$-algebra $C(X)$ of continuous complex valued functions on X. This algebra comes equipped with an involution determined by complex conjugation: $f \to f^*$, where $f^*(x) = \overline{f(x)}$. There is also a norm on $C(X)$ making it into a Banach space. This is an example of a C^*-algebra.

Definition B.3. A C^*-**algebra** is a Banach algebra over $\mathbb{C}$ with a conjugate-linear involution $a \mapsto a^*$ such that

$$(ab)^* = b^*a^* \qquad \text{and} \qquad \|a^*a\| = \|a\|^2$$

Coalgebras

The definition of an algebra can be phrased in terms of diagrams as follows. (This exposition follows that in [Str07].) Let R be any ring (commutative unital). An algebra over R is an R-module M together with module morphisms $\mu : M \otimes_R M \to M$ and $\eta : R \to M$ such that the following diagrams commute:

$$\begin{tikzcd}
M \otimes_R M \otimes_R M \arrow[r, shift left, "\eta \otimes 1_M"] \arrow[r, shift right, "1_M \otimes \eta"'] & M \otimes_R M \arrow[r, "\mu"] & M
\end{tikzcd} \qquad \text{(associativity)}$$

$$\begin{tikzcd}
M \arrow[r, shift left, "\eta \otimes 1_M"] \arrow[r, shift right, "1_M \otimes \eta"'] \arrow[rr, bend right, "1_M"'] & M \otimes_R M \arrow[r, "\mu"] & M
\end{tikzcd}$$

(identity)

Note (for the second diagram) that there are canonical R-module isomorphisms $M \cong R \otimes M \cong M \otimes R$. A ring structure on M can then be defined by setting $ab = \mu(a \otimes b)$. The multiplicative identity is then $\eta(1_r)$.

The definition of a coalgebra is obtained by reversing the directions of the arrows in the above diagrams.

Definition B.4. A **coalgebra** over R is an R-module C together with module homomorphisms $\delta : C \to C \otimes_R C$ and $\epsilon : C \to R$ (called comulitiplication and co-unit respectively) such that the following commute:

$$\begin{tikzcd}
C \arrow[r, "\delta"] & C \otimes_R C \arrow[r, shift left, "\delta \otimes 1_C"] \arrow[r, shift right, "1_C \otimes \delta"'] & C \otimes_R C \otimes_R C
\end{tikzcd}$$

$$\begin{tikzcd}
C \arrow[r, "\delta"] \arrow[rr, bend right, "1_C"'] & C \otimes_R C \arrow[r, shift left, "\epsilon \otimes 1_C"] \arrow[r, shift right, "1_C \otimes \epsilon"'] & C
\end{tikzcd}$$

Example B.5. Let C be the free R-module on the set $\mathbb{N}$. Defining

$$\delta(n) = \sum_{a+b=n} a \otimes b \qquad \text{and} \qquad \epsilon(n) = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{if } n > 0 \end{cases}$$

makes C into a coalgebra over R.

Bibliography

[Art91] Michael Artin. *Algebra*. Prentice Hall Inc., Englewood Cliffs, NJ, 1991.

[Fra67] John B. Fraleigh. *A first course in abstract algebra*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1967.

[HH70] B. Hartley and T. O. Hawkes. *Rings, modules and linear algebra. A further course in algebra describing the structure of Abelian groups and canonical forms of matrices through the study of rings and modules*. Chapman and Hall Ltd., London, 1970.

[Hun96] Thomas W. Hungerford. *Abstract algebra : an introduction*. Cengage Learning, UK, second edition, 1996.

[Rot02] Joseph J. Rotman. *Advanced modern algebra*. Prentice Hall Inc., Upper Saddle River, NJ, 2002.

[Sti05] John Stillwell. *The four pillars of geometry*. Undergraduate Texts in Mathematics. Springer, New York, 2005.

[Str07] Ross Street. *Quantum groups*, volume 19 of *Australian Mathematical Society Lecture Series*. Cambridge University Press, Cambridge, 2007. A path to current algebra.