MAST30005 Algebra

Lawrence Reeves
School of Mathematics and Statistics
University of Melbourne

2025 semester 1

DO NOT POST ON ANY INTERNET SITE

© University of Melbourne 2025

Lectures

© University of Melbourne 2025

Rings

We investigate the properties of rings — general algebraic structures in which there are two operations. Good examples to keep in mind are the integers $\mathbb{Z}$ and 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 not to be true in every ring, but is true of $\mathbb{R}[X]$ for example. We will define and consider various classes of rings: integral domains, unique factorisation domains, principal ideal domains and Euclidean domains.

1.1 Definition

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:	$\mathbb{R}[X] = \{a_0 + a_1X + \cdots + a_nX^n \mid n \in \mathbb{N}, a_i \in \mathbb{R}\}$
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 set R together with two binary operations $+$ and $\times$, called addition and multiplication respectively, that satisfy the following conditions:

1)	$(R, +)$ forms an abelian group (with the identity element being denoted by 0)		
2)	multiplication is associative:	$x \times (y \times z) = (x \times y) \times z$	for all $x, y, z \in R$
3)	there is an element $1 \in R$ that satisfies:	$x \times 1 = 1 \times x = x$	for all $x \in R$
4)	distributive laws:	$x \times (y + z) = (x \times y) + (x \times z)$	for all $x, y, z \in R$
		$(x + y) \times z = (x \times z) + (y \times z)$	for all $x, y, z \in R$

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

Remark.

- Multiplication will often be represented by concatenation, that is we write ab in place of $a \times b$.
- The additive inverse of an element a is denoted $-a$.
- If we were to drop the condition that there is a multiplicative identity, the resulting structure is called a 'pseudo-ring'. An example is the even integers.

Exercise 1. Suppose that R is a ring and $e \in R$ satisfies $\forall x \in R,\ ex = xe = x$. Show that $e = 1$.
(The point is that there is a unique multiplicative identity, and it is uniquely determined by the property in axiom 3 in the definition of a ring.)

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

(a) $0x = x0 = 0$ (b) $x(-y) = (-x)(y) = -(xy)$ (c) $(-x)(-y) = xy$

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

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

Remark. From now on, all rings will be assumed to be non-trivial in the sense of having at least two elements.

Definition 1.3

A ring $(R, +, \times)$ is said to be **commutative** if multiplication is commutative:

$$x \times y = y \times x \quad \text{for all } x, y, \in R$$

Examples 1.4 (Some rings)**.**

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

$$A + B = (A \cup B) \setminus (A \cap B)$$
$$A \times B = A \cap B$$

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

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

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

   makes R into a commutative ring.

3. Consider the following subset 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) 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 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 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 ring. What is the multiplicative identity ?

© University of Melbourne 2025

1.2 Units and zero-divisors

Definition 1.5

Let R be a ring. An element $x \in R$ is called a **unit** (of an **invertible element**) if there exists $y \in R$ such that $xy = yx = 1$. The element y is called the **multiplicative inverse** of x and is denoted x^{-1}. The set of units in R, together with the operation of multiplication, forms a group called the **group of units**. We denote it $R^\times$.

Remark.

- The multiplicative identity is always a unit.
- The zero element is never a unit (see Exercise 2(a)).

Examples 1.6.

1. The units in $\mathbb{Z}$ are $1, -1$.
2. The units in $\mathbb{Z}/6\mathbb{Z}$ are 1 and 5.
3. The units in $\mathbb{R}[X]$ are the non-zero constant polynomials.

Definition 1.7

A ring R is called a **division ring** if every non-zero element is a unit. A **field** is a commutative division ring.

Examples 1.8.

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 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

   We will see later that this field is 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. $\mathbb{Z}/6\mathbb{Z}$ is not a field.
5. The ring of quaternions $\mathcal{H}$ is a division ring, but is not a field.

© University of Melbourne 2025

Definition 1.9

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 1.10. 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. Let R be a ring, and $x \in R$. Show that x cannot be both a unit and a zero-divisor.

Lemma 1.11: Cancellation Law

Let R be a ring. Then R has no zero-divisors if and only if the following condition holds for all $x, y, z \in R$ with $x \neq 0$

$$xy = xz \implies y = z$$
$$yx = zx \implies y = z$$

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

Proof (of Lemma 1.11). First note that $xy = xz \iff xy - xz = 0 \iff x(y - z) = 0$.

Suppose there are no zero-divisors. Then $xy = xz \implies x(y - z) = 0 \implies x = 0$ or $y - z = 0$. If $x \neq 0$, we therefore have $xy = xz \implies y = z$. The second condition follows in exactly the same way.

Now suppose that both conditions hold. If $x \neq 0$ and $xy = 0$, then we have $xy = x0 \implies y = 0$. Similarly if $yx = 0$. □

1.3 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 (the operations are those inherited from $\mathbb{C}$).

6. 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) $\mathbb{R}[X]$

7. 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.

8. 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?

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

© University of Melbourne 2025

10. 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 $\mathbb{Z}[\sqrt{2}]$.

© University of Melbourne 2025

© University of Melbourne 2025

Integral domains, subrings and ideals, homomorphisms

2.1 Integral domains

Definition 2.1

An **integral domain** is a non-zero (i.e., $0 \neq 1$) commutative ring in which there are no zero-divisors.[a]

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

Examples 2.2. The rings $\mathbb{Z}, \mathbb{Z}/3\mathbb{Z}, \mathbb{R}, \mathbb{R}[X]$ are integral domains. The rings $\mathbb{Z}/6\mathbb{Z}, M_2(\mathbb{R})$ are not integral domains.

Proposition 2.3

Every field is an integral domain.

Proof. Every field is commutative. 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 (Theorem 2.4). Of course, although $\mathbb{Z}$ is not 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**.

Theorem 2.4

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 1.11. 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 later).

Example 2.5. Let

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

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 $n \geqslant 1$ 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.

2.2 Subrings and ideals

Definition 2.6

A **subring** of a ring R is a non-empty subset $S \subseteq R$ which contains the multiplicative identity of R and, when equipped with the operations from R, forms a ring.

We will denote this by $S \leqslant R$ (meaning S is a subring of R).

Lemma 2.7

A non-empty subset $S \subseteq R$ of a ring $(R, +, \times)$ is a subring if and only if it satisfies the following:

(a) $1 \in S$ (b) $a - b \in S \quad \forall a, b \in S$ (c) $a \times b \in S \quad \forall a, b \in S$

Remark. The second condition is equivalent to requiring that $(S, +)$ is a subgroup of $(R, +)$. In particular, $0 \in S$.

Exercise 11. Prove Lemma 2.7.

Definition 2.8

An **ideal** in a ring R is a non-empty subset $I \subseteq R$ that satisfies the following:

(a) $a - b \in I \quad \forall a, b \in I$ (b) $r \times a \in I$ and $a \times r \in I \quad \forall a \in I \quad \forall r \in R$

We will 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.

Examples 2.9.

1. The set $I = \{\sum_{i=1}^{n} \alpha_i X^i \mid n \geqslant 1, \alpha_i \in \mathbb{R}\} \subseteq \mathbb{R}[X]$ of polynomials having constant term equal to zero is an ideal in $\mathbb{R}[X]$.

2. $J = \{\sum_{i=0}^{n} \alpha_i X^i \mid n \geqslant 0, \alpha_i \in \mathbb{Z}, \alpha_0 \text{ is even}\} \subseteq \mathbb{Z}[X]$ is an ideal in $\mathbb{Z}[X]$.

Example 2.10. $\mathbb{Z} \subseteq \mathbb{R}$ 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 Homomorphisms

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.

© University of Melbourne 2025

Definition 2.11

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 $a, b \in R$:

(a) $\varphi(a+b) = \varphi(a) + \varphi(b)$ (b) $\varphi(ab) = \varphi(a)\varphi(b)$ (c) $f(1) = 1$

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 $(S, +)$.

Lemma 2.12

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

1. The **kernel** of φ, $\ker(\varphi) = \{r \in R \mid \varphi(r) = 0\}$, 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

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

□

Example 2.13. 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 I of Example 2.9.

© University of Melbourne 2025

Lemma 2.14

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 $a - b = 0$. □

2.4 Exercises

12. Show that if R is an integral domain, then $R[X]$ is an integral domain.
13. Let R be an integral domain such that $x^2 = x$ for all $x \in R$. Show that R has exactly two elements.
14. Let R be a ring and I an ideal in R. Show that if I contains a unit from R, then $I = R$.
15. Show that a field F has only two ideals, namely F and $\{0\}$. Conversely, show that if a commutative ring has exactly two ideals, then it is a field.
16. Find an example of a ring R and a subset $I \subseteq R$ such that I is left ideal but not a right ideal.
17. Find an example of a homomorphism whose image is not an ideal in the codomain.
18. 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 $R \times R$?

19. (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)?

20.⋆ The **characteristic** of a (non-zero) 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 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 $p \in \mathbb{N}$.)

21.⋆ 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).

© University of Melbourne 2025

22.⋆ Let R be a commutative (non-zero) 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**).

23.⋆ 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:

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

(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.

© University of Melbourne 2025

© University of Melbourne 2025

Quotient rings and the isomorphism theorems

3.1 Quotient rings

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

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

Define operations on this set by

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

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 because I is an ideal $xb, ay, xy \in I$ and hence $xb + ay + xy \in I$.

Exercise 24. Check that the first operation above is also well-defined, and that with these operations R/I is a ring. What are the additive and multiplicative identities in R/I?

Definition 3.1

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

Examples 3.2.

1. For any $m \in \mathbb{Z}$ we can form the quotient $\mathbb{Z}/m\mathbb{Z}$.
2. $\mathbb{R}[X]/\langle X^2 + 1\rangle \cong \mathbb{C}$ (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 3.3

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

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

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

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

□

3.2 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 3.4: 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

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

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 3.5: 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.

Exercise 25. Write out a proof of second and third isomorphism theorems.

© University of Melbourne 2025

3.3 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 2.12 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 3.6

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 $\mathrm{im}(\varphi)$.
2. If S' is a subring (or ideal) in $\mathrm{im}(\varphi)$, 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 $r' \in \mathrm{im}(\varphi)$, 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 $\mathrm{im}(\varphi)$.

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 $\mathrm{im}(\varphi)$ 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 3.7: Correspondence Theorem

Let $\varphi : R \to R'$ be a ring homomorphism. The maps

$$\Phi : \{S \leqslant R \mid \ker(\varphi) \subseteq S\} \to \{S' \leqslant R' \mid S' \subseteq \mathrm{im}(\varphi)\}, \qquad \Phi(S) = \varphi(S)$$

$$\Psi : \{I \lhd R \mid \ker(\varphi) \subseteq I\} \to \{I' \subseteq R' \mid I' \lhd \mathrm{im}(\varphi)\}, \qquad \Psi(I) = \varphi(I)$$

are inclusion-preserving bijections.

Proof. We give the argument for Ψ and leave the other case as an exercise. Given $I' \lhd \mathrm{im}(\varphi)$, 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

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

© University of Melbourne 2025

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 $I' \subseteq J' \subseteq \mathrm{im}(\varphi) \implies \varphi^{-1}(I') \subseteq \varphi^{-1}(J')$. □

3.4 Exercises

26. 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 $R/(I \cap J) \cong R/I \times R/J$.

27. Using the above exercise 26 show that $\mathbb{Z}_{mn} \cong \mathbb{Z}_m \times \mathbb{Z}_n$ if and only if $\gcd(m, n) = 1$.

28. Complete the prove the Correspondence Theorem.

© University of Melbourne 2025

Constructions and generating sets

4.1 Constructions

We record here some standard ways of combining rings to produce new ring. Some have been mentioned already and will be used extensively in the sequel.

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

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

equipped with the operations

$$(r, s) + (r', s') = (r + r', s + s') \qquad (r, s)(r', s') = (rr', ss')$$

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**.

Polynomial rings. Let R be a commutative ring. Elements of the ring $R[X]$ 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. The units in $R[X]$ are precisely the degree zero polynomials that are units in R. Since $R[X]$ is itself a ring, this construction can be iterated to give $R[X, Y] = (R[X])[Y]$ and $R[X_1, \ldots, X_n]$.

Matrix rings. Let R be a commutative 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 of matrices 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.

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, that is, for $f, g : R \to R$ define

$$(f + g)(a) = f(a) + g(a) \qquad (fg)(a) = (f \circ g)(a)$$

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

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

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

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

4.2 Generating sets

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

Definition 4.1

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

In the case that $A = \{a_1, \ldots, a_k\}$ we denote the ideal generated by A by $\langle a_1, \ldots, a_k \rangle$ or $(a_1, \ldots, a_k)$.

Definition 4.2

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

Notice that if $u \in R$ is a unit, then $\langle u \rangle = R$. The following lemma states that the ideal generated by $A \subseteq R$ is the set of all R-linear combinations of elements from A.

Lemma 4.3

Let R be a commutative ring and $A \subseteq R$. Then

$$\langle A \rangle = \{r_1 a_1 + \cdots + r_n a_n \mid n \in \mathbb{N}, r_i \in R, a_i \in A\}$$

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

Examples 4.4.

1. We have already seen 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 4.3, 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 $2 + X \notin \langle 2 \rangle$. □

4.3 Exercises

29. 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.
30. Let $\varphi : R \to S$ be a homomorphism, and define a map $\Phi : R[X] \to S[X]$ by

    $$\Phi(a_0 + a_1 X + \cdots + a_n X^n) = \varphi(a_0) + \varphi(a_1) X + \cdots + \varphi(a_n) X^n$$

    Show that Φ is a homomorphism.
31. Show that the units in $F[X]$, where F is a field, are the elements of $F \setminus \{0\}$.
32. 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

    $$a_0 + a_1 X + \cdots + a_n X^n \mapsto a_0 + a_1 (X - a) + \cdots + a_n (X - a)^n$$

© University of Melbourne 2025

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 $b_i \in R$.

33. Let R b a commutative 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 $\varphi(X) = r$.

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

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

© University of Melbourne 2025

© University of Melbourne 2025

PIDs and divisors in IDs

5.1 Principal ideal domains

We come now to 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 5.1

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

Examples 5.2.

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.

Exercise 36. Show that every ideal in $\mathbb{Z}/12\mathbb{Z}$ is principal. Is $\mathbb{Z}/12\mathbb{Z}$ a PID?

5.2 Divisors in integral domains

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

Definition 5.3

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 5.4.

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 37.

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 38. 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 39. Show that

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

5.3 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 5.5

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 5.6.

1. The irreducibles in $\mathbb{Z}$ are exactly the prime integers (where we allow negative primes, eg -5).
2. Any degree 1 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 $X^2 + X + 1 = (X - \frac{1}{2}(1 + i\sqrt{3}))(X - \frac{1}{2}(1 - i\sqrt{3}))$.

5.4 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 5.7

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

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

© University of Melbourne 2025

Example 5.8.

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 5.9

Let R be an integral domain and $x \in R$. If x is prime, then x is 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 (noting that $p \neq 0$). It follows that c is a unit. Thus p is irreducible.

□

The next example demonstrates that the converse to the above result does not hold.

Example 5.10 (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 $N(x) = 1$, 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 first that it is not a unit. Now suppose that $2 = uv$ for some $u, v \in \mathbb{Z}[\sqrt{-5}]$. We then have that $N(u)N(v) = 4$, from which it follows that $N(u) = 1$ or $N(v) = 1$. Therefore one of u or v must be a unit.

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*.

5.5 Exercises

40. Let D be an integral domain, and $p, q \in D$ with $q \mid p$. 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.

41.* Let $d \in \mathbb{Z}$ be square-free, and $R = \mathbb{Z}[\sqrt{d}] \subset \mathbb{C}$.
    (a) Show that $x + y\sqrt{d} = a + b\sqrt{d}$ only if $x = a$ and $y = b$.

    Define $N : R \to \mathbb{N}$ by $N(x + y\sqrt{d}) = |x^2 - y^2 d|$
    (b) Show that $N(r_1 r_2) = N(r_1)N(r_2)$.
    (c) Show that $r \in R$ is a unit if and only if $N(r) = 1$.
    (d) Use induction on $N(r)$ to show that all non-unit elements $r \in R \setminus \{0\}$ can be written as a product of irreducibles.

© University of Melbourne 2025

© University of Melbourne 2025

Unique factorisation domains and prime and maximal ideals

6.1 Definition of a UFD

In the integers the 'fundamental theorem of arithmetic' states that every integer can be written as a product of irreducibles and that this factorisation is essentially unique. Not all integral domains have this property.

Definition 6.1

An integral domain R is called a **unique factorisation domain** (UFD for short) if the following hold:

1. *Existence of factorisation:* 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 factorisation:* If $a = b_1 \dots b_m$ is another factorisation 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 factorisations differ only by re-ordering and replacing each factor by an associate.

Examples 6.2.
1. $\mathbb{Z}$ is a UFD. This is the Fundamental Theorem of Arithmetic.
2. $\mathbb{Q}$, $\mathbb{R}$ are UFDs since there are no elements that are nonzero and non-unit.
3. $\mathbb{F}_5[X]$, $\mathbb{R}[X]$ are UFDs (we will show later that they are PIDs)
4. $\mathbb{Z}[X]$, $\mathbb{R}[X, Y]$ are UFDs (even though they are not PIDs)

Exercise 42. Show that in a UFD irreducible elements are prime.

It follows from this and Example 5.10, 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 in $\mathbb{Z}[\sqrt{-5}]$. The ring $\mathbb{Z}[\sqrt{-5}]$ is therefore not a UFD because it fails the second part of the definition (uniqueness), although it does satisfy the first part (existence).

Example 6.3. Here is an example in which the first part (in the definition of UFD) 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 factorisation as a product of irreducibles (and is not a unit):

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

6.2 Prime and maximal ideals

Definition 6.4

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 6.5

Let R be a commutative 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. Let I be any ideal in R. Note that since R is commutative, so too is R/I. If $I \neq R$, then R/I is non-zero. 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,

$$\begin{aligned} xy = 0 &\implies \varphi(a)\varphi(b) = 0 \implies \varphi(ab) = 0 \implies ab \in I \implies b \in I \quad \text{(since } I \text{ is prime and } a \notin I\text{)} \\ &\implies \varphi(b) = 0 \implies y = 0 \end{aligned}$$

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

$$\begin{aligned} ab \in I &\implies \varphi(ab) = 0 \implies \varphi(a)\varphi(b) = 0 \implies \varphi(b) = 0 \quad \text{(since } R/I \text{ is an ID and } \varphi(a) \neq 0\text{)} \\ &\implies b \in I \end{aligned}$$

2) Since R/I is commutative, it is a field if and only if its only ideals are itself and $\{0\}$ (Exercise 15). We have

$$\begin{aligned} I \text{ is maximal} &\iff R/I \text{ contains only two ideals} && \text{(Correspondence Theorem 3.7)} \\ &\iff R/I \text{ is a field} && \text{(Exercise 15)} \end{aligned}$$

□

Since every field is an integral domain, we have the following as an immediate consequence.

Corollary 6.6

Every maximal ideal is prime. □

© University of Melbourne 2025

Lemma 6.7

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

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

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

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

□

Exercise 43.

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 6.8

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

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

© University of Melbourne 2025

6.3 Exercises

44. Show that the following is equivalent to the definition of Unique factorisation 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.

45. 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.

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

47. Determine the maximal ideals in the following rings:

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

48. (a) Show that $\langle 2 \rangle \lhd \mathbb{Z}[\sqrt{-5}]$ is not prime.
    (b) Show that $\langle 11 \rangle \lhd \mathbb{Z}[\sqrt{-5}]$ is prime.

49. 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.

© University of Melbourne 2025

$F[X]$ 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 lecture, F denotes a field.

7.1 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 $\mathbb{Z}$.

Proposition 7.1

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 7.2

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 7.3: Vandermonde's Theorem

A polynomial equation of degree n over a field has at most n roots. □

7.2 $F[X]$ is a PID

Theorem 7.4

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. We know that $F[X]$ is an ID since F, being a field, is an ID (see Exercise 12). We need to show that all ideals in $F[X]$ are principal.

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 Proposition 7.1, 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 7.5. 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.

7.3 Exercises

50. Let $f, g \in \mathbb{F}_5[X]$ be given by $f = X^4 - 3X^3 + 2X^2 + 4X - 1$ and $g = X^2 - 2X + 3$. Find $q, r \in \mathbb{F}_5[X]$ such that $\deg(r) < \deg(g)$ and $f = qg + r$.

51. Show that $X^2 - 2$ is irreducible in $\mathbb{Q}[X]$. Show that $X^2 - 2$ *is not* irreducible in $\mathbb{R}[X]$.

52. Show that $X^3 + 3X + 2$ is irreducible in $\mathbb{F}_5[X]$.

53. Let R be an integral domain. Show that $R[X]/\langle X - a\rangle$ is isomorphic to R for any $a \in R$.

54. 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 $\mathbb{R}[X]/\langle X^2 + 1\rangle \cong \mathbb{C}$.

55. 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 $\deg(r) < \deg(g)$.

56. Show that if $R[X]$ is a PID, then R is a field. (This is the converse of Theorem 7.4)

© University of Melbourne 2025

57. Which are fields? (a) $\mathbb{Q}[X]/\langle X^2 - 5X + 6\rangle$ (b) $\mathbb{Q}[X]/\langle X^2 - 6X + 6\rangle$

58. (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 a root that is an integer that divides a_0.

59. 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) $\mathbb{C}[X]/\langle X^2 + 1\rangle$

© University of Melbourne 2025

© University of Melbourne 2025

Every PID is a UFD

To show that an integral domain is a unique factorisation 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 factorisation 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).

8.1 Ascending chain condition

Definition 8.1

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

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

there exists $N \in \mathbb{N}$ 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, Artin p.469.

Examples 8.2.

1. $\mathbb{Z}$ satisfies the ACC. Every ideal is of the form $\langle m \rangle$ for some $m \in \mathbb{Z}$ and $\langle m \rangle \subseteq \langle n \rangle$ iff $n \mid m$.
2. $\mathbb{R}[X]$ satisfies the ACC, as we shall see shortly.
3. $\mathbb{R}[X_1, X_2, \dots]$, the polynomial ring on infinitely many variables, is not Noetherian. The chain of ideals
$$\langle X_1 \rangle \subseteq \langle X_1, X_2 \rangle \subseteq \langle X_1, X_2, X_3 \rangle \subseteq \cdots$$
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, we know 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 8.3

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 factorisations. Then we would have elements $a_0 = a, a_1, a_2 \dots$ 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 8.4

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

8.2 Prime versus irreducible

We saw in Proposition 5.9 that, in an integral domain, prime elements are irreducible. The converse does not hold in general (see Example 5.10). We saw in Exercise 42 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 factorisations are essentially unique.

Lemma 8.5

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, $a_1 \mid a_1 a_2 \ldots a_m$, so we must have that $a_1 \mid b_1 b_2 \ldots b_n$. 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 8.6

In a PID irreducible elements are prime.

Proof. Applying Lemma 6.7, Corollary 6.6 and Lemma 6.8 gives

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

□

© University of Melbourne 2025

Remark. Once we have established that all PIDs are UFDs the above lemma follows from Exercise 42. However, we need the lemma in order to prove that PIDs are UFDs.

8.3 PID implies UFD

Assembling the results of the previous sections we have the following

Theorem 8.7

Every principal ideal domain is a unique factorisation domain.

Proof. Let R be a PID and $a \in R$ a non-zero non-unit element. By Proposition 8.4, R satisfies the ascending chain condition, and therefore a can be written as a product of irreducibles by Proposition 8.3. That the second part of the definition of UFD is satisfied, is precisely the statement of Lemma 8.5, which applies by Lemma 8.6. □

Corollary 8.8

For any field F, the polynomial ring $F[X]$ is a unique factorisation domain. □

8.4 Exercises

61. 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.

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

© University of Melbourne 2025

© University of Melbourne 2025

If R is a UFD, then $R[X]$ is a UFD

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.

9.1 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 exactly the same definition.

Definition 9.1

Let R be an integral domain. A **greatest common divisor** (or gcd) of a finite number of elements $a_1, \ldots, a_n \in R$ is an element $d \in R$ satisfying:

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

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 always exist a gcd.

Exercise 63. Let $a_1, a_2 \in \mathbb{Z}[\sqrt{-5}]$ be $a_1 = 6$ and $a_2 = 2 + 2\sqrt{-5}$.

(a) Use the function N from Example 5.10 to list all common divisors of a_1 and a_2.

(b) Show that a_1 and a_2 do not have a greatest common divisor.

However, in a UFD any collection of elements does have a gcd.

Lemma 9.2

Let R be a UFD and $a_1, \ldots, a_n \in R$ (not all zero). There exists a gcd of the elements $a_1, \ldots, a_n$.

Proof. We first show that for any two elements $a, b \in R$ (that aren't both zero), 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 factorisations of a and b as products of irreducibles. Rearranging, we can write these factorisations as

$$a = p_1^{m_1} \ldots p_k^{m_k}$$
$$b = p_1^{n_1} \ldots p_k^{n_k} u$$

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 $d = p_i^{\min\{m_1,n_1\}} \ldots p_k^{\min\{m_k,n_k\}}$. 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 factorisation $d' = c_1 \ldots c_l$,

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

Definition 9.4

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

Example 9.5. The polynomials $2X - 2$ and $2X^2 - 2X - 4$ are relatively prime in $\mathbb{C}[X]$.

9.2 Primitive polynomials and the Gauss Lemma

Definition 9.6

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

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

Exercise 64. Prove the following lemma.

Lemma 9.7

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 factorisation in $\mathbb{Q}[X]$ and factorisation in $\mathbb{Z}[X]$.

Lemma 9.8: 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 30). Since R is a UFD and p is irreducible, p is prime (Exercise 42), which in turn implies that $R/\langle p \rangle$ is an integral domain (Lemma 6.8 and Proposition 6.5). 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 9.9

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 $g_1' = g_1$.

© University of Melbourne 2025

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 $h_i = d_i g_i$ 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

$$\begin{aligned}
f = g_1g_2 &\implies d_1d_2f = h_1h_2 \\
&\implies d_1d_2c\hat{f} = c_1c_2\hat{h_1}\hat{h_2} \\
&\implies \hat{f} \sim \hat{h_1}\hat{h_2} && \text{by Lemmas 9.7 and 9.8} \\
&\implies \hat{f} = u\hat{h_1}\hat{h_2} && \text{for some unit } u \in R \\
&\implies f = cu\hat{h_1}\hat{h_2} \\
&\implies f = g_1'g_2' && \text{where } g_1' = \hat{h_1} \text{ and } g_2' = cu\hat{h_2}
\end{aligned}$$

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 9.10

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 9.11

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

9.3 If R a UFD, then $R[X]$ 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 the previous lecture, 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 factorisation as a product of irreducibles in $F[X]$. In order to obtain irreducibles in $R[X]$ we use the following technical lemma.

Lemma 9.12

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 9.13

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

© University of Melbourne 2025

Proof. Let $f \in R[X]$ be non-zero, non-unit. If $\deg(f) = 0$ then, since R is a UFD, we can factorise 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

$$\begin{aligned} & \hat{f} = f_1 \cdots f_k && \text{(where each } f_i \in F[X] \text{ is irreducible in } F[X]) \\ \implies & \hat{f} = f_1' \cdots f_k' && (f_i' \in R[X], \text{ Lemma 9.9, irreducible in } F[X]) \\ \implies & f_i' \text{ is primitive (since } \hat{f} \text{ is) and irreducible in } R[X] \text{ for all } i \text{ (by Lemma 9.12)} \end{aligned}$$

□

Proposition 9.14

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 9.10 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 9.11. □

Theorem 9.15

If R is a unique factorisation domain, then so too is $R[X]$.

Proof. Follows from Lemma 8.5 and Propositions 9.13 and 9.14. □

Examples 9.16. It follows from this theorem that $\mathbb{Z}[X], \mathbb{Z}[X, Y], \mathbb{R}[X, Y], \mathbb{R}[X_1, \ldots, X_n]$ are all UFDs. (None of them are PIDs.)

9.4 Exercises

65. 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 (an associate of) p appears in every factorisation of a.
66. 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 $\mathbb{Z}/7\mathbb{Z}[X]$

© University of Melbourne 2025

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

    (a) $\mathbb{Z}[X]$ (b) $\mathbb{Q}[X]$ (c) $\mathbb{F}_{11}[X]$

68. 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. That is, show that if d is a gcd of a and b, then $d = \alpha a + \beta b$ for some $\alpha, \beta \in R$. (Hint: consider the ideal $I = \langle a, b \rangle$ generated by a and b.)

69. 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.

70. Show that if $p, q \in \mathbb{Z}$ are relatively prime in $\mathbb{Z}$, then they are relatively prime in $\mathbb{Z}[i]$.

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

72. 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.

73. 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.

74. Consider the homomorphism $\varphi : \mathbb{Z}[X] \to \mathbb{R}$ 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.

75. 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 $1/2$. Note that $\mathbb{Z}[1/2] = \{m/2^k \mid m \in \mathbb{Z}, k \in \mathbb{N}\}$.

© University of Melbourne 2025

© University of Melbourne 2025

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 7-1 7.2. 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 76. 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.

10.1 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 10.1: Eisenstein's irreducibility criterion

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:

1) p divides a_i for all $i \in \{0, \ldots, n-1\}$
2) p does not divide a_n
3) 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 9.9 that $f = gh$ for some $g, h \in \mathbb{Z}[X]$ with g and h of degree at least 1. Consider the homomorphism $\varphi : \mathbb{Z}[X] \to \mathbb{F}_p[X]$ induced by the projection $\mathbb{Z} \to \mathbb{Z}/\langle p\rangle$, $a \mapsto \overline{a} = a + \langle p\rangle$ (see Exercise 30). The conditions of the theorem ensure that $\overline{a}_n \neq 0$ and $\varphi(f) = \overline{a}_nX^n$. Since φ is a homomorphism, $\varphi(g)\varphi(h) = \overline{a}_nX^n$. 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. But then the constant term of f would be divisible by p^2. □

Example 10.2. Using Eisenstein's criterion we conclude that the polynomial $X^4 + 50X^2 + 30X + 20$ is irreducible in $\mathbb{Q}[X]$. (Since it's primitive, it is also irreducible in $\mathbb{Z}[X]$.)

Corollary 10.3

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 $X = Y + 1$ gives

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

The last polynomial is irreducible by Theorem 10.1. 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 32 would give a contradiction. □

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

10.2 Computation modulo p

Proposition 10.4

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 $\overline{a}_0 + \overline{a}_1X + \cdots + \overline{a}_nX^n \in \mathbb{F}_p[X]$ 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 $\mathbb{F}_p[X]$. 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. □

Example 10.5. $f = X^4 + 9X^3 + 2X^2 + 6X + 1 \in \mathbb{Z}[X]$ is irreducible in $\mathbb{Q}[X]$ since $\overline{f} = X^4 + X^3 + 1 \in \mathbb{F}_2[X]$ is irreducible.

10.3 A factorisation algorithm for $\mathbb{Z}[X]$

There is a systematic, though possibly very long, method to factorise 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 factorise 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 $a \in \{0, 1, \dots, m\}$, let $D = \{(d_0, d_1, \dots, d_m) \in \mathbb{Z}^{m+1} \mid d_i \text{ is a divisor of } f(i)\}$. This is a finite set. For each $d = (d_0, d_1, \dots, d_m) \in D$ let $g_d \in \mathbb{Q}[X]$ be the unique polynomial with $\deg(g_d) \leqslant m$ and $g(i) = d_i$ for all $i \in \{0, 1, \dots, m\}$.
      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.

© University of Melbourne 2025

10.4 Exercises

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

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

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

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

78$\frac{1}{2}$. Test for irreducibility the following polynomials $\mathbb{Q}[X]$:

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

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

    (a) $X^5 - 4X + 22$
    (b) $2X^5 + 12X^4 - 15X^3 + 18X^2 - 45X + 3$
    (c) $X^4 + 1$

80. 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 $\mathbb{Q}[X]$.

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

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

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

84. Determine all irreducible polynomials of degree at most 4 in $\mathbb{F}_2[X]$.

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

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

© University of Melbourne 2025

© University of Melbourne 2025

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 7.1). We make a definition of this property.

11.1 Definition of Euclidean domain

Definition 11.1

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**.

Note.

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 11.2. It follows from Proposition 7.1 that, for any field F, $F[X]$ is a ED, with a suitable function being $\sigma(f) = \deg(f)$.

Example 11.3 (cf. Example 5.10)**.** 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 11.4

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 7.4).

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.

11.2 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 Algorithm

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 73 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 11.5. 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

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

© University of Melbourne 2025

11.3 Exercises

86. Show that every field is a ED (you should give an explicit norm function).

87. 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.

88. Find a gcd of $X^3 - 6X^2 + X + 4$ and $X^5 - 6X + 1$ in $\mathbb{Q}[X]$.

89. 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 $d = af + bg$.

90. 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.

91.* (This is a long question! Feel free to skip it. It is here mainly so that we have an example to show that not every PID is a ED.)
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.

   (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.

© University of Melbourne 2025

© University of Melbourne 2025

Modules

12.1 Definition

A module is a generalisation of a vector space in which the scalars do not necessarily form a field, but may be any commutative 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.

Definition 12.1

Let R be a commutative ring. An ***R*-module** M is an abelian group (whose operation will be denoted by addition) together with a map $R \times M \to M$ (the image of (ρ, u) being denoted ρu) that satisfies the following for all $\rho, \sigma \in R$ and all $u, v \in M$:

(1) $1u = u$

(2) $(\rho\sigma)u = \rho(\sigma u)$

(3) $(\rho + \sigma)u = \rho u + \sigma u$

(4) $\rho(u + v) = \rho u + \rho v$

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. In this section on modules the ring R is always assumed to be commutative.

Examples 12.2.

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. $R^n = \{(r_1, \ldots, r_n) \mid r_i \in R\}$ is an R-module. The operations are the usual coordinatewise addition and scalar multiplication:
$$(r_1, \ldots, r_n) + (s_1, \ldots, s_n) = (r_1 + s_1, \ldots, r_n + s_n)$$
$$r(r_1, \ldots, r_n) = (rr_1, \ldots, rr_n)$$
4. Any abelian group can be regarded as a $\mathbb{Z}$-module, and vice-versa.
5. $R[X]$ forms a module over R.

12.2 Submodules, homomorphisms, quotients and products

Definition 12.3

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

Exercise 92. Show that a subset $N \subseteq M$ is a submodule if and only if the following hold

(a) N is non-empty

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

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

Example 12.4.

1. If $I \lhd R$ is an ideal, then ${}_RI$ is a submodule of ${}_RR$.
2. If S is a commutative ring and R is a subring of S, then S is an R-module.

Definition 12.5

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

(1) $\varphi(u + v) = \varphi(u) + \varphi(v)$ (2) $\varphi(\rho u) = \rho\varphi(u)$

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

Exercise 93. Show that $\ker(\varphi)$ is a submodule of V and that $\mathrm{im}(\varphi)$ is a submodule of W.

Definition 12.6

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:

(1) $(u + W) + (v + W) = (u + v) + W$ (2) $\rho(v + W) = \rho v + W$

Definition 12.7

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.

© University of Melbourne 2025

12.3 Exercises

94. Let M be an R-module. Show that for all $\rho \in R$ and $u \in M$ we have:

(a) $0_R\, u = 0_M$ (b) $\rho\, 0_M = 0_M$ (c) $(-\rho)u = -(\rho u) = \rho(-u)$

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

96. (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) $U' \cong U$, $V' \cong V$

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

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

98. 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 $R \to M$, $a \mapsto a + \langle p^k \rangle$.)

© University of Melbourne 2025

© University of Melbourne 2025

Free modules and bases

The notion of a basis is extremely useful when studying vector spaces,. We now consider the corresponding notion in a module.

Definition 13.1

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 99. 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 13.2

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 13.3

A subset of 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 (i.e., all vector spaces) 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 $\mathbb{Z}/2\mathbb{Z}$. 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 ${}_R I$ must contain at least two elements. But if $u, v \in {}_R I$ are two distinct elements, the identity $vu + (-u)(v) = 0$ implies that no linearly independent subset of ${}_R I$ 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 13.4

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 100. Prove Lemma 13.4.

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

Lemma 13.5

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 13.6

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.

© University of Melbourne 2025

Proposition 13.7

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

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

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.

Definition 13.8

The number of elements in a basis is called the **rank** of the free R-module.

Lemma 13.9

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 13.5). Since $\mathrm{im}(\varphi)$ contains a generating set, φ is surjective. □

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

© University of Melbourne 2025

13.1 Exercises

101. Let R be a ring (commutative) 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.

102. 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.

103. 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 $m + n$.

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

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

106. 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?

107. 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 $(6, 9)$.

108. 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 $f = X^2 + 2X - 3 \in \mathbb{Q}[X]$. 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.

© University of Melbourne 2025

Torsion and submodules of free modules

14.1 Torsion

Definition 14.1

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

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

An element $u \in M$ is said to be **torsion** if $\operatorname{ann}_R(u) \neq \{0\}$. 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 109.

a) Show that $\operatorname{ann}_R(u)$ 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 finitely generated torsion-free module over an integral domain that is not free. (Hint: The ring should not be a PID.)

Proposition 14.2

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.

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

□

14.2 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.

Lemma 14.3: The splitting lemma

Let R be a commutative 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 $F' \cap \ker(\varphi) = \{0\}$. □

Theorem 14.4: Submodules of free modules over a PID are free

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 13.6), and 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 $\{x_1, \ldots, x_r\}$ be a basis for F, and let $F' \subseteq F$ be the submodule generated by $\{x_1, \ldots, x_{r-1}\}$. Then F' is free and $\{x_1, \ldots, x_{r-1}\}$ 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 14.3, $N = L \oplus (N \cap F')$ 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. □

© University of Melbourne 2025

14.3 Exercises

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

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

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

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

114. 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 $p \in R$.

115. 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 14.4?

116. 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.

© University of Melbourne 2025

© University of Melbourne 2025

Smith normal form

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 the structure of such matrices, we will be able to analyse the structure of F/N.

15.1 The matrix of a homomorphism

Let R be an integral domain, and F and G two finitely generated free R-modules. Fix bases for $\mathcal{B} = \{f_1, \ldots, f_m\}$ and $\mathcal{C} = \{g_1, \ldots, g_n\}$ 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 117. 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 15.1

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.

15.2 Smith normal form

Theorem 15.2

Let R be a PID and $A \in M_{m\times n}(R)$. The matrix A is equivalent to a diagonal matrix $D \in M_{m\times n}(R)$ satisfying $D = \operatorname{diag}(d_1, d_2, \ldots, d_{\min\{m,n\}})$ and $d_1 \mid d_2 \mid \cdots \mid d_{\min\{m,n\}}$.

Definition 15.3

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 required 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 68) that in a PID we have $d = \gcd(a, b) = xa + yb$ for some $x, y \in R$. We have

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

© University of Melbourne 2025

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 15.4. 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 Theorem 15.2.

$$A = \begin{bmatrix} 6 & 4 & 4 \\ 4 & 8 & 0 \end{bmatrix} \xrightarrow{C1 \mapsto C1-C2} \begin{bmatrix} 2 & 4 & 4 \\ -4 & 8 & 0 \end{bmatrix} \xrightarrow[C3 \mapsto C3-2C1]{C2 \mapsto C2-2C1} \begin{bmatrix} 2 & 0 & 0 \\ -4 & 16 & 8 \end{bmatrix}$$
$$\xrightarrow{R2 \mapsto R2+2R1} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 16 & 8 \end{bmatrix} \xrightarrow{C2 \mapsto C2-2C3} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 8 \end{bmatrix} \xrightarrow{C2 \leftrightarrow C3} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 8 & 0 \end{bmatrix} = D$$

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 15.5. 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 Theorem 15.2.

$$A = \begin{bmatrix} 1-X & 1+X & X \\ X & 1-X & 1 \\ 1+X & 2X & 1 \end{bmatrix} \xrightarrow{C1 \leftrightarrow C3} \begin{bmatrix} X & 1+X & 1-X \\ 1 & 1-X & X \\ 1 & 2X & 1+X \end{bmatrix} \xrightarrow{R1 \leftrightarrow R3} \begin{bmatrix} 1 & 2X & 1+X \\ 1 & 1-X & X \\ X & 1+X & 1-X \end{bmatrix}$$
$$\xrightarrow[R3 \mapsto R3 - XR1]{R2 \mapsto R2 - R1} \begin{bmatrix} 1 & 2X & 1+X \\ 0 & 1-3X & -1 \\ 0 & 1+X-2X^2 & 1-2X-X^2 \end{bmatrix}$$
$$\xrightarrow[C3 \mapsto C3 - (1+X)C1]{C2 \mapsto C2 - 2XC1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1-3X & -1 \\ 0 & 1+X-2X^2 & 1-2X-X^2 \end{bmatrix} \xrightarrow{C2 \leftrightarrow C3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 1-3X \\ 0 & 1-2X-X^2 & 1+X-2X^2 \end{bmatrix}$$
$$\xrightarrow{C3 \mapsto C3 + (1-3X)C2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1-2X-X^2 & 2-4X+3X^2+3X^3 \end{bmatrix}$$
$$\xrightarrow{R3 \mapsto R3 + (1-2X-X^2)R2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2-4X+3X^2+3X^3 \end{bmatrix} = D$$

© University of Melbourne 2025

15.3 Exercises

118. 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

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

119. For the matrices $A, D \in M_{3\times 3}(\mathbb{Q}[X])$ from Example 15.5 find invertible matrices $L, R \in M_{3\times 3}(\mathbb{Q}[X])$ satisfying $LAR = D$.

120. 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 $d_1 | d_2 | d_3$.

(a) $A = \begin{bmatrix} 7 & 8 & 9 \\ 4 & 5 & 6 \\ 1 & 2 & 3 \end{bmatrix} \in M_{3\times 3}(\mathbb{Z})$

(b) $A = \begin{bmatrix} 1-X & 1+X & X \\ X & 1-X & 1 \\ 1+X & 2X & 1 \end{bmatrix} \in M_{3\times 3}(\mathbb{Q}[X])$

121. 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:

(a) $\begin{bmatrix} 3 & 1 \\ -1 & 2 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$

(c) $\begin{bmatrix} -4 & -6 & 7 \\ 2 & 2 & 4 \\ 6 & 6 & 15 \end{bmatrix}$

(d) $\begin{bmatrix} X & 1 & -2 \\ -3 & X+4 & -6 \\ -2 & 2 & X-3 \end{bmatrix}$

(e) $\begin{bmatrix} X & 0 & 0 \\ 0 & 1-X & 0 \\ 0 & 0 & 1-X^2 \end{bmatrix}$

122. 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

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

(b) φ is an isomorphism;

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

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

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

126.* 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 $|\det A|$.

© University of Melbourne 2025

The structure theorem

Theorem 16.1: Structire theorem for finitly generated modules over a PID

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

$$M \cong R/\langle d_1\rangle \oplus R/\langle d_2\rangle \oplus \cdots \oplus R/\langle d_k\rangle \qquad (*)$$

Proof. Since M if finitely generated (by k elements say), there is a surjective homomorphism $\varphi : R^k \to M$ (Lema 13.9). Let $N = \ker(\varphi)$. By Theorem 14.4, 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 Theorem 15.2, A is equivalent to a matrix of the form

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

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 $\{f_1, f_2, \ldots, f_k\}$ of R^k such that $\{d_1 f_1, d_2 f_2, \ldots, d_s f_s\}$ 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 $\psi(\sum_{i=1}^k r_i f_i) = (r_1 + \langle d_1\rangle, \ldots, r_k + \langle d_k\rangle)$ is a homomorphism of R-modules. The result follows from the first isomorphism theorem, since ψ is surjective, and $\ker(\psi) = N$. □

Remark.

1. Some of the d_i might be zero and some might be units. 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 $j \leqslant i$.
2. A matrix $A \in M_{k\times s}(R)$ as above is sometimes called a **presentation matrix** for the module $R^k/(AR^s) \cong M$.

Corollary 16.2

Let M be a finitely generated module over a PID.

1. If M is torsion-free, then M is free.
2. $M = F \oplus T_M$, 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 14.2) 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 14.3, $M = \ker(\varphi) \oplus F$ 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 16.3

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 16.4

For a given M as in the Structure Theorem 16.1, the invariant factors are all uniquely determined by M (up to associates). The torsion-free rank is uniquely determined by M.

We postpone the proof until later.

Example 16.5. 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 15.4 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 97) that

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

© University of Melbourne 2025

Example 16.6. 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 $N = \mathrm{im}(\varphi)$ and $M = \mathbb{Q}[X]^3/N$. From Example 15.5, 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 $\mathrm{ann}_R(M) = (2 - 4X + 3X^2 + 3X^3) \lhd \mathbb{Q}[X]$.

16.1 Exercises

127. 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.

© University of Melbourne 2025

© University of Melbourne 2025

Primary decomposition

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

Lemma 17.1

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

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

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 17.2

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

$$M \cong \quad R/\langle p_1^{n_1}\rangle \oplus R/\langle p_2^{n_2}\rangle \oplus \cdots \oplus R/\langle p_s^{n_s}\rangle \oplus R^r \qquad (\dagger)$$

Proof. From Theorem 16.1 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 17.1 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 17.3

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

Example 17.4. Suppose that M is a $\mathbb{Q}[X]$-module such that

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

From the irreducible factorisations

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

We obtain the primary decomposition

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

17.1 Exercises

128. 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.

129. Let $R = \mathbb{Q}[X]$ and suppose that the torsion R-module M is a direct sum of four cyclic modules whose annihilators are $\langle (X - 1)^3\rangle$, $\langle (X^2 + 1)^2\rangle$, $\langle (X - 1)(X^2 + 1)^4\rangle$, and $\langle (X + 2)(X^2 + 1)^2\rangle$. 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 $\mathbb{Q}$.

© University of Melbourne 2025

Applications of the structure theorem

18.1 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 18.1: Structure Theorem for Finitely Generated Abelian Groups

Let G be a finitely generated abelian group. Then

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

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

Exercise 130. Find a direct sum of cyclic groups which is isomorphic to the abelian group $\mathbb{Z}^3/N$, where $N \leqslant \mathbb{Z}^3$ is generated by $\{(2,2,2),(2,2,0),(2,0,2)\}$.

18.2 Application to linear transformations

Suppose we have a finite dimensional vector space V over a field $\mathbb{F}$. Given a linear transformation $T : V \to V$, we'd like to find a matrix representation of T that is as simple as possible (while using the same basis for domain and codomain).

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

$$(\sum_{i=0}^{n} a_i X^i)v = \sum_{i=0}^{n} a_i T^i(v)$$

where $a_i \in \mathbb{F}$ and $v \in V$.

Since V is finite dimensional as an $\mathbb{F}$-module, it is finitely generated as a $\mathbb{F}[X]$-module. Indeed, any generating set for ${}_{\mathbb{F}}V$ will be a generating set for ${}_{\mathbb{F}[X]}V$. Since ${}_{\mathbb{F}[X]}V$ is finitely generated and $\mathbb{F}[X]$ is a PID, we can apply the structure theorem to obtain

$${}_{\mathbb{F}[X]}V \cong \frac{\mathbb{F}[X]}{\langle d_1 \rangle} \oplus \frac{\mathbb{F}[X]}{\langle d_2 \rangle} \oplus \cdots \oplus \frac{\mathbb{F}[X]}{\langle d_k \rangle} \oplus (\mathbb{F}[X])^r$$

where $r \geqslant 0$, each $d_i \in \mathbb{F}[X]$ is non-zero and non-unit, and $d_1 \mid d_2 \mid \cdots \mid d_k$.

In fact it must be the case that $r = 0$, that is, that the torsion-free rank of ${}_{\mathbb{F}[X]}V$ is zero. To see this, note that the set $\{1, X, X^2, \dots\} \subseteq \mathbb{F}[X]$ is linearly independent over $\mathbb{F}$. 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 vector space.

We thus have

$${}_{\mathbb{F}[X]}V \cong \frac{\mathbb{F}[X]}{\langle d_1 \rangle} \oplus \frac{\mathbb{F}[X]}{\langle d_2 \rangle} \oplus \cdots \oplus \frac{\mathbb{F}[X]}{\langle d_k \rangle} \tag{$\circledast$}$$

Remark.

1. It follows from this decomposition that $d_k = \operatorname{ann}_{\mathbb{F}[X]}(V)$ and is therefore the **minimal polynomial** of T. That is, $m_T(X) = d_k(X)$.

2. The **characteristic polynomial** is given by $c_T(X) = d_1 d_2 \dots d_k$. (The justification for this assertion will be clear shortly.)

Each summand in (⁕) is a submodule of ${}_{\mathbb{F}[X]}V$ and is therefore a subspace of ${}_{\mathbb{F}}V$ that is preserved by the linear transformation T. We therefore want to understand the structure of each summand as a vector space over $\mathbb{F}$ together with the linear transformation obtained by restricting T to W.

Let W denote one of the summands in (⁕). That is, $W = \mathbb{F}[X]/\langle d \rangle$ for some $d = X^m + a_{m-1}X^{m-1} + \dots + a_1 X + a_0 \in \mathbb{F}[X]$ with $m \geqslant 1$.

We will analyse the restriction of T to W. For $f \in \mathbb{F}[X]$, denote the element $f + \langle d \rangle \in W$ by $\bar{f}$.

Lemma 18.2

The set $\mathcal{B}_W = \{\bar{1}, \bar{X}, \bar{X}^2, \dots, \bar{X}^{m-1}\}$ is a basis for ${}_{\mathbb{F}}W$.

Proof. Let $\xi_i \in \mathbb{F}$, $1 \leqslant i \leqslant m-1$. Then

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

Hence the set is linearly independent.

From the division algorithm for $\mathbb{F}[X]$, we know that for any $f \in \mathbb{F}[X]$, there is a $g \in \mathbb{F}[X]$ such that $\deg(g) < \deg(d) = m$ and $\bar{g} = \bar{f}$. It follows that $\bar{f} \in \operatorname{span}\{\bar{1}, \bar{X}, \dots, \bar{X}^{m-1}\}$. □

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 $T(\bar{f}) = X\bar{f}$.

$$\begin{aligned}
T(\bar{X}^i) &= \bar{X}^{i+1} \qquad (\text{for } 0 \leqslant i < m-1) \\
T(\bar{X}^{m-1}) &= -a_0 - a_1\bar{X} - \dots - a_{m-1}\bar{X}^{m-1}
\end{aligned}$$

The matrix of $T|_W$ (with respect to $\mathcal{B}_W$) is therefore

$$\begin{bmatrix}
0 & 0 & 0 & & 0 & -a_0 \\
1 & 0 & 0 & \cdots & 0 & -a_1 \\
0 & 1 & 0 & & 0 & -a_2 \\
 & & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & & 0 & -a_{m-2} \\
0 & 0 & 0 & \cdots & 1 & -a_{m-1}
\end{bmatrix} \in M_{m\times m}(\mathbb{F})$$

Definition 18.3

A matrix in the above form is called the **companion matrix** of $d = X^m + a_{m-1}X^{m-1} + \dots + a_1 X + a_0 \in \mathbb{F}[X]$ ($m \geqslant 1$) and will be denoted $\mathcal{C}_d$.

Exercise 131. Show that the characteristic polynomial of $\mathcal{C}_d$ is d.

Applying this to each summand in (⁕) we obtain the following.

© University of Melbourne 2025

Theorem 18.4: Rational canonical form

Let V be a finite dimensional vector space over a field $\mathbb{F}$ and let $T : V \to V$ be a linear transformation. There exists of basis $\mathcal{B}$ for V such that

$$[T]_{\mathcal{B}} = \begin{bmatrix} \mathcal{C}_{d_1} & & & \\ & \mathcal{C}_{d_2} & 0 & \\ & 0 & \ddots & \\ & & & \mathcal{C}_{d_k} \end{bmatrix}$$

with $d_i \in \mathbb{F}[X]$, $\deg(d_i) \geqslant 1$, and $d_1 \mid d_2 \mid \cdots \mid d_k$.

This matrix is called the **rational canonical form** of T.

Proof. Discussed in the lecture. □

18.3 Exercises

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

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

134. 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) $2x + y = 0$, $x - y + 3z = 0$.

135. 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.

136. Let $T : \mathbb{R}^7 \to \mathbb{R}^7$ be a linear transformation. Suppose that the corresponding $\mathbb{R}[X]$-module can be written as

$$\frac{\mathbb{R}[X]}{\langle X - 2 \rangle} \oplus \frac{\mathbb{R}[X]}{\langle X^3 - X^2 - X - 2 \rangle} \oplus \frac{\mathbb{R}[X]}{\langle X^3 - X^2 - X - 2 \rangle}$$

    (a) Write down the rational canonical form of T.

    (b) What are the minimal and characteristic polynomials of T?

© University of Melbourne 2025

© University of Melbourne 2025

Rational Canonical Form

We saw last lecture that every linear transformation $T : V \to V$ from a finite dimensional vector space over $\mathbb{F}$ to itself has a matrix representation in *rational canonical form*. That is, there exists a basis $\mathcal{B}$ of V such that the matrix of T with respect to $\mathcal{B}$ is in block diagonal form

$$[T]_{\mathcal{B}} = \begin{bmatrix} \mathcal{C}_{d_1} & & & \\ & \mathcal{C}_{d_2} & 0 & \\ & 0 & \ddots & \\ & & & \mathcal{C}_{d_k} \end{bmatrix}$$

with monic $d_i \in \mathbb{F}[X]$, $\deg(d_i) \geqslant 1$, and $d_1 \mid d_2 \mid \cdots \mid d_k$. The square matrices $\mathcal{C}_{d_i}$ are *companion matrices*.

$$\mathcal{C}_d = \begin{bmatrix} 0 & 0 & 0 & & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & & 0 & -a_2 \\ & & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & & 0 & -a_{m-2} \\ 0 & 0 & 0 & \cdots & 1 & -a_{m-1} \end{bmatrix} \in M_{m\times m}(\mathbb{F})$$

where $d = a_0 + a_1X + \cdots + a_{m-1}X^{m-1} + X^m$

Example 19.1.

$$\mathcal{C}_{X-2} = \begin{bmatrix} 2 \end{bmatrix} \qquad \mathcal{C}_{X^2-2X+1} = \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} \qquad \mathcal{C}_{X^3+2X^2-3X-4} = \begin{bmatrix} 0 & 0 & 4 \\ 1 & 0 & 3 \\ 0 & 1 & -2 \end{bmatrix}$$

Example 19.2. The following matrices are in rational canonical form

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

The following matrices are *not* in rational canonical form

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

19.1 Minimal and characteristic polynomials

The polynomials $d_1, \ldots, d_k$ determine the minimal and characteristic polynomials of the linear transformation (or matrix). To see this, recall that given $T : V \to V$ we can consider V as a $\mathbb{F}[X]$-module

by defining scalar multiplication such that $Xu = T(u)$. Applying the structure thereom for finitely generated modules over a PID (Theorem 16.1) we obtained

$$_{\mathbb{F}[X]}V \cong \frac{\mathbb{F}[X]}{\langle d_1\rangle} \oplus \frac{\mathbb{F}[X]}{\langle d_2\rangle} \oplus \cdots \oplus \frac{\mathbb{F}[X]}{\langle d_k\rangle}$$

with monic $d_i \in \mathbb{F}[X]$, $\deg(d_i) \geqslant 1$, and $d_1 \mid d_2 \mid \cdots \mid d_k$.

Proposition 19.3

The minimal and characteristic polynomials of T are given by

$$m_T(X) = d_k(X) \qquad\qquad c_T(X) = d_1 d_2 \cdots d_k$$

Proof. Both follow from the existence of a rational canonical form matrix for T and that for any (monic, non-constant) $d \in \mathbb{F}[X]$ the companion matrix $\mathcal{C}_d$ has characteristic polynomial equal to d and minimal polynomial to d. □

As a corollary of the existence of rational canonical form (together with the above proposition) we get the following well-known result.

Cayley Hamilton Theorem

A square matrix $A \in M_n(\mathbb{F})$ satisfies its own characteristic equation. □

19.2 Calculating the rational canonical form

Given a matrix $A \in M_n(\mathbb{F})$ we can calculate its rational canonical form by considering the matrix $XI_n - A \in M_n(\mathbb{F}[X])$. Let V be a vector space over $\mathbb{F}$ with basis $\{v_1, \ldots, v_n\}$. Let $T : V \to V$ be the linear transformation whose matrix (with respect to the the basis $\{v_1, \ldots, v_n\}$) is A, and define $_{\mathbb{F}[X]}V$ as above. Then $_{\mathbb{F}[X]}V \cong \mathbb{F}[X]^n/N$, where N is generated by the columns of the matrix $XI - A$. We then find $d_1, \ldots, d_n \in \mathbb{F}[X]$ such that $XI - A \sim \operatorname{diag}(d_1, \ldots, d_n)$ (Theorem 15.2) and obtain that $_{\mathbb{F}[X]}V \cong \mathbb{F}[X]/\langle d_1\rangle \oplus \cdots \oplus \mathbb{F}[X]/\langle d_n\rangle$. From this we then get the rational canonical form of A as explained in the previous section.

Example 19.4.

An an example, let's calculate the rational canonical form of the matrix $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \in \mathbb{Q}[X]$.

$$XI - A = \begin{bmatrix} X-1 & -1 & 0 \\ 0 & X-1 & 0 \\ 0 & -1 & X-1 \end{bmatrix} \xrightarrow{C_1 \leftrightarrow C_2} \begin{bmatrix} -1 & X-1 & 0 \\ X-1 & 0 & 0 \\ -1 & 0 & X-1 \end{bmatrix}$$

$$\xrightarrow{C_2+(X-1)C_1} \begin{bmatrix} -1 & 0 & 0 \\ X-1 & (X-1)^2 & 0 \\ -1 & -(X-1) & X-1 \end{bmatrix} \xrightarrow[R_3-R_1]{R_2+(X-1)R_1} \begin{bmatrix} -1 & 0 & 0 \\ 0 & (X-1)^2 & 0 \\ 0 & -(X-1) & X-1 \end{bmatrix}$$

$$\xrightarrow{R_2 \leftrightarrow R_3} \begin{bmatrix} -1 & 0 & 0 \\ 0 & -(X-1) & X-1 \\ 0 & (X-1)^2 & 0 \end{bmatrix} \xrightarrow{C_3+C_2} \begin{bmatrix} -1 & 0 & 0 \\ 0 & -(X-1) & 0 \\ 0 & (X-1)^2 & (X-1)^2 \end{bmatrix}$$

$$\xrightarrow{R_3+(X-1)R_2} \begin{bmatrix} -1 & 0 & 0 \\ 0 & -(X-1) & 0 \\ 0 & 0 & (X-1)^2 \end{bmatrix}$$

© University of Melbourne 2025

From which we get that ${}_{\mathbb{F}[X]}V \cong \frac{\mathbb{F}[X]}{\langle X-1\rangle} \oplus \frac{\mathbb{F}[X]}{\langle (X-1)^2\rangle}$. The rational canonical form of A is therefore

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

19.3 Exercises

137. Write down the minimal and characteristic polynomials of the following matrices. (Hint: they are in rational canonical form)

(a) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \in M_4(\mathbb{F}_2)$

(b) $\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix} \in M_6(\mathbb{F}_3)$

138. Find the rational canonical forms of the following matrices. State their minimal and characteristic polynomials.

(a) $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \in M_2(\mathbb{R})$

(b) $\begin{bmatrix} 7 & 6 & 9 \\ 0 & 1 & 0 \\ -4 & -4 & -5 \end{bmatrix} \in M_3(\mathbb{R})$

(c) $\begin{bmatrix} 0 & -3 & 3 & 1 \\ 0 & 2 & 0 & 0 \\ -2 & -3 & 5 & 1 \\ 2 & 3 & -3 & 1 \end{bmatrix} \in M_4(\mathbb{R})$

139. Show that if a square matrix $A \in M_n(\mathbb{F})$ has minimal polynomial equal to its characteristic polynomial (both equal to $d \in \mathbb{F}[X]$), then A is similar to the companion matrix $\mathcal{C}_d$.

© University of Melbourne 2025

© University of Melbourne 2025

Jorndan normal form

Another useful standard matrix representation for a linear transformation is Jordan normal form, which is based on the primary decomposition.

Suppose that V is a finite-dimensional *complex* vector space, and $T : V \to V$ a linear transformation. As discussed above we can equip V with the structure of a $\mathbb{C}[X]$-module. Noting that the prime elements in $\mathbb{C}[X]$ are exactly the linear polynomials, from Theorem 17.2, the module ${}_{\mathbb{C}[X]}V$ has a primary decomposition of the form

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

for some $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$ and $m_1, \ldots, m_k \in \mathbb{Z}_{\geqslant 0}$. That there are no summands of the form $\mathbb{C}[X]$ follows, as above, from the fact that V is finite dimensional as a vector space.

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 $\bar{f}$.

Exercise 140. Show that the set

$$ \mathcal{B}_W = \{(\bar{X}-\bar{\lambda})^{m-1}, (\bar{X}-\bar{\lambda})^{m-2}, \ldots, (\bar{X}-\bar{\lambda})^2, \bar{X}-\bar{\lambda}, \bar{1}\} $$

is a basis for ${}_{\mathbb{C}}W$.

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 $T(\bar{f}) = X\bar{f}$, we have

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

The matrix of $T|_W$ is therefore

$$ [T|_W]_{\mathcal{B}_W} = \begin{bmatrix} \lambda & 1 & 0 & & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda & & 0 & 0 \\ & & \vdots & \ddots & 1 & 0 \\ 0 & 0 & 0 & & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{bmatrix} \in M_m(\mathbb{C}) $$

Definition 20.1

A matrix in the above form is called an **elementary Jordan matrix** and will be denoted $J_{\lambda,m}$.

Exercise 141. (a) Show that the characteristic polynomial of $J_{\lambda,m}$ is $(X-\lambda)^m$.

(b) Show that the minimal polynomial of $J_{\lambda,m}$ is $(X-\lambda)^m$.

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

Theorem 20.2: Jordan normal form

Let V be a finite-dimensional complex vector space and $T : V \to V$ a linear transformation. There exists a basis $\mathcal{B}$ of V and $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$ and $m_1, \ldots, m_k \in \mathbb{Z}_{\geqslant 0}$ such that the matrix of T is in block diagonal form

$$[T]_{\mathcal{B}} = \begin{bmatrix} J_{\lambda_1, m_1} & & & \\ & J_{\lambda_2, m_2} & & \\ & & \ddots & \\ & & & J_{\lambda_k, m_k} \end{bmatrix}$$

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

Definition 20.3

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

Example 20.4. The following are examples of matrices in Jordan Normal Form:

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

20.1 Calculating the Jordan normal form of a matrix

We can proceed as with rational canonical form.

Given a matrix $A \in M_n(\mathbb{C})$ we can calculate its Jordan normal form by considering the matrix $XI_n - A \in M_n(\mathbb{C}[X])$. Let V be a complex vector space with basis $\{v_1, \ldots, v_n\}$. Let $T : V \to V$ be the linear transformation whose matrix (with respect to the the basis $\{v_1, \ldots, v_n\}$) 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 $XI_n - A$. We find $d_1, \ldots, d_n \in \mathbb{C}[X]$ such that $XI - A \sim \operatorname{diag}(d_1, \ldots, d_n)$ (Theorem 15.2) and obtain 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.

In summary, calculating the Smith normal form of $XI - A \in M_n(\mathbb{C}[X])$ enables us to write down the primary decomposition of ${}_{\mathbb{C}[X]}V$ and hence the Jordan Normal Form of the matrix A.

To justify the above process for calculating the Jordan normal form (and that for calculating the rational canonical form) we note the following lemma.

© University of Melbourne 2025

Lemma 20.5

Let $\mathbb{F}$ be a field and $A \in M_n(\mathbb{F})$. Let $\mathcal{F} = \{f_1, \ldots, f_n\}$ be the standard basis for $(\mathbb{F}[X])^n$. Let $\{v_1, \ldots, v_n\}$ be a basis the vector space $V = \mathbb{F}^n$. Let π be the surjective $\mathbb{C}[X]$-module homomorphism $\pi : (\mathbb{F}[X])^n \to {}_{\mathbb{C}[X]}V$ determined by $\pi(f_i) = v_i$.

Let $\varphi : F \to F$ be the homomorphism whose matrix with respect to $\mathcal{F}$ is $XI - A$. Then $\ker(\pi) = \mathrm{im}(\varphi)$. In particular

$$ {}_{\mathbb{F}[X]}V \cong \frac{(\mathbb{F}[X])^n}{\mathrm{im}(\varphi)} $$

Proof. We first show that $\mathrm{im}(\varphi) \subseteq \ker(\pi)$. Let $\mathcal{V} = \{v_1, \ldots, v_n\}$. 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 $(XI - A)_{ij} = \delta_{ij}X - a_{ij}$ and

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

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 $\alpha_i \in \mathbb{F}$. (Note that the α_i are in $\mathbb{F}$ not $\mathbb{F}[X]$.)

Then

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

□

© University of Melbourne 2025

Example 20.6. 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

$$XI - A = \begin{bmatrix} X-2 & 0 & 0 & 0 \\ 1 & X-1 & 0 & 0 \\ 0 & 1 & X & 1 \\ -1 & -1 & -1 & X-2 \end{bmatrix} \xrightarrow{R1\leftrightarrow R2} \begin{bmatrix} 1 & X-1 & 0 & 0 \\ X-2 & 0 & 0 & 0 \\ 0 & 1 & X & 1 \\ -1 & -1 & -1 & X-2 \end{bmatrix}$$

$$\xrightarrow{C2-(X-1)C1} \begin{bmatrix} 1 & 0 & 0 & 0 \\ X-2 & -(X-1)(X-2) & 0 & 0 \\ 0 & 1 & X & 1 \\ -1 & X-2 & -1 & X-2 \end{bmatrix}$$

$$\xrightarrow[R4+R1]{R2-(X-2)R1} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -(X-1)(X-2) & 0 & 0 \\ 0 & 1 & X & 1 \\ 0 & X-2 & -1 & X-2 \end{bmatrix}$$

$$\xrightarrow{R2\leftrightarrow R3} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & X & 1 \\ 0 & -(X-1)(X-2) & 0 & 0 \\ 0 & X-2 & -1 & X-2 \end{bmatrix}$$

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

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

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

$$\xrightarrow{C4-XC3} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & (X-1)(X-2) & 0 \\ 0 & 0 & 0 & -(X-1)^2 \end{bmatrix}$$

This matrix is not in Smith normal form, but it is sufficient to conclude that (in the notation from the explanation above)

$${}_{\mathbb{C}[X]}V \cong \frac{\mathbb{C}[X]}{\langle 1\rangle} \oplus \frac{\mathbb{C}[X]}{\langle 1\rangle} \oplus \frac{\mathbb{C}[X]}{\langle (X-1)(X-2)\rangle} \oplus \frac{\mathbb{C}[X]}{\langle -(X-1)^2\rangle}$$

From which we get that the primary decomposition is

$${}_{\mathbb{C}[X]}V \cong \frac{\mathbb{C}[X]}{\langle X-1\rangle} \oplus \frac{\mathbb{C}[X]}{\langle (X-1)^2\rangle} \oplus \frac{\mathbb{C}[X]}{\langle X-2\rangle}$$

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

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

© University of Melbourne 2025

20.2 Exercises

142. Find the Smith normal of the matrix $XI - A$ from Example 20.6. (The point is to continue with the given sequence of row and column operations.)

143. 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?

144. 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 $XI - A \in M_{3\times 3}(\mathbb{C})$.

145. Find all possible Jordan normal forms for a matrix (over $\mathbb{C}$) whose characteristic polynomial is $(X+2)^2(X-5)^3$

© University of Melbourne 2025

© University of Melbourne 2025

More on calculating normal forms

We look in more detail at the process for calculating the normal forms of a matrix $A \in M_n(\mathbb{F})$ and give examples of finding an invertible matrix $P \in M_n(\mathbb{F})$ such that $P^{-1}AP$ is in normal form.

21.1 Recap

We recall the process we've discussed for finding the normal form of a matrix.

Let ${}_{\mathbb{F}}V$ be the $\mathbb{F}$-vector space $\mathbb{F}^n$ and let $\mathcal{S} = \{e_1, e_2, \ldots, e_n\} \subseteq V$ be the standard basis for $\mathbb{F}^n$. Given $A \in M_n(\mathbb{F})$ define $T : {}_{\mathbb{F}}V \to {}_{\mathbb{F}}V$ to be the linear transformation given by $[T]_{\mathcal{S}} = A$.

We equip V with the structure of an $\mathbb{F}[X]$-module by defining $Xu = T(u)$. That is, the $\mathbb{F}[X]$-module ${}_{\mathbb{F}[X]}V$ has the same set of vectors and vector addition as ${}_{\mathbb{F}}V$, but with scalar multiplication given by

$$(\sum_{i=0}^{N} a_i X^i)u = \sum_{i=0}^{N} a_i T^i(u)$$

Note that $\mathcal{S}$ is a generating set for ${}_{\mathbb{F}[X]}V$. Let $\mathcal{F} = \{u_1, \ldots, u_n\}$ by the standard basis for $\mathbb{F}[X]^n$ and define a module homomorphism $\pi : \mathbb{F}[X]^n \to {}_{\mathbb{F}[X]}V$ by $\pi(u_i) = e_i$. This homomorphism is surjective because $\mathcal{S}$ is a generating set for ${}_{\mathbb{F}[X]}V$. By the first isomorphism theorem, we have that ${}_{\mathbb{F}[X]}V \cong \frac{\mathbb{F}[X]^n}{\ker(\pi)}$.

To analyse $\frac{\mathbb{F}[X]^n}{\ker(\pi)}$ we define a homomorphism of free modules $\varphi : \mathbb{F}[X]^n \to \mathbb{F}[X]^n$ by $[\varphi]_{\mathcal{F}} = XI_n - A$. Then $\mathrm{im}(\varphi) = \ker(\pi)$ (see Lemma 20.5).

Consider the Smith normal form of $XI_n - A$, $D = \mathrm{diag}(d_1, \ldots, d_n)$. Then $D = Z[\varphi]_{\mathcal{F}}Y$ for some invertible matrices $Z, Y \in M_n(\mathbb{F}[X])$. Therefore, $D = [\varphi]_{\mathcal{D},\mathcal{C}}$ for some bases $\mathcal{C}$ for the domain and $\mathcal{D}$ for the codomain of φ. Let $\mathcal{D} = \{v_1, \ldots, v_n\}$. Then $\{d_1v_1, \ldots, d_nv_n\}$ is a generating set for $\mathrm{im}(\varphi)$ and

$${}_{\mathbb{F}[X]}V \cong \frac{\mathbb{F}[X]^n}{\mathrm{im}(\varphi)} = \frac{\langle v_1\rangle \oplus \cdots \oplus \langle v_n\rangle}{\langle d_1v_1\rangle \oplus \cdots \oplus \langle d_nv_n\rangle} \cong \frac{\langle v_1\rangle}{\langle d_1v_1\rangle} \oplus \cdots \oplus \frac{\langle v_n\rangle}{\langle d_nv_n\rangle} \cong \frac{\mathbb{F}[X]}{\langle d_1\rangle} \oplus \cdots \oplus \frac{\mathbb{F}[X]}{\langle d_n\rangle}$$

Calculating the d_i enables us to deduce the normal form (either rational or Jordan). In addition, knowing the v_i enables us to calculate P such that $P^{-1}AP$ is in normal form. Since the matrix Z is the transition matrix $P_{\mathcal{D},\mathcal{F}}$, its inverse is $P_{\mathcal{F},\mathcal{D}}$. The columns of Z^{-1} are therefore $[v_1]_{\mathcal{F}}, \ldots, [v_n]_{\mathcal{F}}$.

21.2 Examples

Example 21.1. In Example 19.4 we determined the rational canonical form of the following matrix $A \in M_3(\mathbb{Q})$ by putting $XI - A \in M_3(\mathbb{Q}[X])$ into Smith normal form:

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \qquad XI - A \sim \begin{bmatrix} -1 & 0 & 0 \\ 0 & -(X-1) & 0 \\ 0 & 0 & (X-1)^2 \end{bmatrix} = D = ZAY$$

To find the matrix Z^{-1} we consider the row operations that were applied in obtaining D from $XI - A$. They were (in the order applied):

1. $R_2 + (X-1)R_1$ 2. $R_3 - R_1$ 3. $R_2 \leftrightarrow R_3$ 4. $R_3 + (X-1)R_2$

Denoting the corresponding elementary matrices by Z_1, Z_2, Z_3, Z_4, we have

$$\begin{aligned} Z &= Z_4 Z_3 Z_2 Z_1 \\ Z^{-1} &= Z_1^{-1} Z_2^{-1} Z_3^{-1} Z_4^{-1} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ -(X-1) & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -(X-1) & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ -(X-1) & -(X-1) & 1 \\ 1 & 1 & 0 \end{bmatrix} \end{aligned}$$

As in the explanation above, let $\mathcal{F}$ denote the standard basis for $\mathbb{F}[X]^3$ and $\mathcal{S} = \{e_1, e_2, e_3\}$ the standard basis for $\mathbb{F}^3$. Let $v_1, v_2, v_3 \in \mathbb{F}[X]^3$ be given by the columns of Z^{-1}. That is,

$$[v_1]_{\mathcal{F}} = \begin{bmatrix} 1 \\ -(X-1) \\ 1 \end{bmatrix} \qquad [v_2]_{\mathcal{F}} = \begin{bmatrix} 0 \\ -(X-1) \\ 1 \end{bmatrix} \qquad [v_3]_{\mathcal{F}} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

The images of these elements $\pi(v_1), \pi(v_2), \pi(v_3) \in {}_{\mathbb{F}[X]}V$ are given by

$$\begin{aligned} \pi(v_1) &= e_1 - (X-1)e_2 + e_3 = e_1 + e_2 + e_3 - Xe_2 \\ &= e_1 + e_2 + e_3 - T(e_2) && (Xv = T(v) \text{ in } {}_{\mathbb{F}[X]}V) \\ &= e_1 + e_2 + e_3 - (e_1 + e_2 + e_3) && ([T(e_2)]_{\mathcal{S}} = A[e_2]_{\mathcal{S}}) \\ &= \vec{0} && \text{(as expected since } d_1 \text{ is a unit)} \end{aligned}$$

Similarly,

$$\pi(v_2) = -(X-1)e_2 + e_3 = e_2 + e_3 - Xe_2 = e_2 + e_3 - T(e_2) = e_2 + e_3 - (e_1 + e_2 + e_3) = -e_1$$
$$\pi(v_3) = e_2$$

Now define $\mathcal{B} = \{b_1, b_2, b_3\} \subseteq V$ by

$$\begin{aligned} b_1 &= \pi(v_2) = -e_1 \\ b_2 &= \pi(v_3) = e_2 \\ b_3 &= T(b_2) = T(\pi(v_3)) = T(e_2) = e_1 + e_2 + e_3 \end{aligned}$$

Noting that $T(b_1) = b_1$, $T(b_2) = b_3$, and $T(b_3) = -b_2 + 2b_3$, we have

$$[T]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix}$$

which is in rational canonical form.

Letting $P = P_{\mathcal{S},\mathcal{B}} = \begin{bmatrix} -1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$, the change of basis for the matrix representation of the linear transformation $\varphi : V \to V$ gives

$$\begin{aligned} [T]_{\mathcal{B}} &= P_{\mathcal{B},\mathcal{S}} [T]_{\mathcal{S}} P_{\mathcal{S},\mathcal{B}} \\ &= P^{-1} A P \end{aligned}$$

To obtain a Q such that $Q^{-1}AQ$ is in Jordan normal form we make a different choice of basis for $\mathbb{F}^3$. Define $\mathcal{C} = \{c_1, c_2, c_3\} \subseteq V$ by

$$\begin{aligned} c_1 &= \pi(v_2) = -e_1 \\ c_3 &= \pi(v_3) = e_2 \\ c_2 &= (T - I)(c_3) = e_1 + e_3 \end{aligned}$$

© University of Melbourne 2025

Noting that $T(c_1) = c_1$, $T(c_2) = c_2$, and $T(c_3) = c_2 + c_3$, we have

$$[T]_{\mathcal{C}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

which is in Jordan normal form.

With $Q = \begin{bmatrix} -1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, we have $Q^{-1}AQ$ is in Jordan normal form.

Example 21.2. In Example 20.6 we determined the Jordan normal form of the following matrix $A \in M_4(\mathbb{C})$ by using row and column operations to show

$$A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & -1 \\ 1 & 1 & 1 & 2 \end{bmatrix} \qquad XI - A \sim \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & (X-1)(X-2) & 0 \\ 0 & 0 & 0 & -(X-1)^2 \end{bmatrix}$$

We will find a matrix Q such that $Q^{-1}AQ$ is in Jordan normal form. We use the same technique and notation as in the previous example. The row operations applied to $XI - A$ were, in order:

1. $R_1 \leftrightarrow R_2$
2. $R_2 - (X-2)R_1$
3. $R_4 + R_1$
4. $R_2 \leftrightarrow R_3$
5. $R_3 + (X-1)(X-2)R_2$
6. $R_4 - (X-2)R_2$

Labelling the corresponding elementary matrices as $Z_1, Z_2, Z_3, Z_4, Z_5, Z_6$, we have

$$Z_1^{-1}Z_2^{-1}Z_3^{-1}Z_4^{-1}Z_5^{-1}Z_6^{-1} =$$
$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ (X-1) & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -(X-1)(X-2) & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & (X-2) & 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} (X-2) & -(X-1)(X-2) & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & (X-2) & 0 & 1 \end{bmatrix}$$

Let $v_1, v_2, v_3, v_4 \in \mathbb{C}[X]^4$ be given by the columns of the above matrix.

$$\begin{aligned} \pi(v_1) &= -2e_1 + e_2 - e_4 + T(e_1) = -2e_1 + e_2 - e_4 + (2e_1 - e_2 + e_4) = \vec{0} \\ \pi(v_2) &= -2e_1 + e_3 - 2e_4 + T(3e_1 + e_4) - T^2(e_1) \\ &= -2e_1 + e_3 - 2e_4 + (6e_1 - 3e_2 - e_3 + 5e_4) - (4e_1 - 3e_2 + 3e_4) = \vec{0} \\ \pi(v_3) &= e_1 \\ \pi(v_4) &= e_4 \end{aligned}$$

Let $\mathcal{C} = \{c_1, c_2, c_3, c_4\}$ be the basis for $\mathbb{C}^4$ given by

$$\begin{aligned} c_1 &= (T - 2I)e_1 = -e_2 + e_4 \\ c_3 &= e_4 \\ c_2 &= (T - I)e_4 = -e_3 + e_4 \\ c_4 &= (T - I)e_1 = e_1 - e_2 + e_4 \end{aligned}$$

Letting $Q = \begin{bmatrix} 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & -1 \\ 0 & -1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$ we have that $Q^{-1}AQ = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$

© University of Melbourne 2025

21.3 Exercises

146. For each of the matrices A from Exercise 138 (repeated below), find P such that $P^{-1}AP$ is the rational canonical form of A.

(a) $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \in M_2(\mathbb{R})$

(b) $\begin{bmatrix} 7 & 6 & 9 \\ 0 & 1 & 0 \\ -4 & -4 & -5 \end{bmatrix} \in M_3(\mathbb{R})$

(c) $\begin{bmatrix} 0 & -3 & 3 & 1 \\ 0 & 2 & 0 & 0 \\ -2 & -3 & 5 & 1 \\ 2 & 3 & -3 & 1 \end{bmatrix} \in M_4(\mathbb{R})$

147. Let $\mathbb{F}$ be a field and let $p, q \in \mathbb{F}[X]$ be relatively prime.

(a) Let M be a cyclic $\mathbb{F}[X]$-module, and let $u \in M$ be such that $M = \langle u \rangle$. Suppose that $\operatorname{ann}_{\mathbb{F}[X]}(u) = \langle pq \rangle$. Show that $M = \langle pu \rangle \oplus \langle qu \rangle$ (internal direct sum).

(b) Let M be an $\mathbb{F}[X]$-module and suppose that there exist $u, v \in M$ such that $\operatorname{ann}_{\mathbb{F}[X]}(u) = \langle p \rangle$, $\operatorname{ann}_{\mathbb{F}[X]}(u) = \langle p \rangle$, and $M = \langle u \rangle \oplus \langle v \rangle$. Show that $M = \langle u + v \rangle$ and $\operatorname{ann}(u + v) = \langle pq \rangle$.

© University of Melbourne 2025

Uniqueness of the decompositions

To finish our discussion of finitely generated modules over a PID we will establish the uniqueness of the invariant factor decomposition (Theorem 16.1) and the primary decomposition (Theorem 17.2).

22.1 p-primary submodules

We first consider the following special case.

Proposition 22.1

Let $p \in R$ be prime. Suppose $k, l \in \mathbb{Z}_{\geqslant 1}$ and $m_1, \dots, m_k, n_1, \dots n_l \in \mathbb{Z}_{\geqslant 1}$ are such that

$$\frac{R}{\langle p^{m_1} \rangle} \oplus \cdots \oplus \frac{R}{\langle p^{m_k} \rangle} \cong \frac{R}{\langle p^{n_1} \rangle} \oplus \cdots \oplus \frac{R}{\langle p^{n_l} \rangle}$$

with $m_1 \leqslant m_2 \leqslant \cdots \leqslant m_k$ and $n_1 \leqslant n_2 \leqslant \cdots \leqslant n_l$.

Then $k = l$ and $m_i = n_i$ for all i.

Proof. Let $M = \frac{R}{\langle p^{m_1} \rangle} \oplus \cdots \oplus \frac{R}{\langle p^{m_k} \rangle}$ and $N = \frac{R}{\langle p^{n_1} \rangle} \oplus \cdots \oplus \frac{R}{\langle p^{n_l} \rangle}$. Note that $m_k = n_l$ since

$$\langle p^{m_k} \rangle = \operatorname{ann}_R(M) = \operatorname{ann}_R(N) = \langle p^{n_l} \rangle$$

We will use induction on m_k, the highest power of p appearing.

If $m_k = n_l = 1$, we have

$$\left(\frac{R}{\langle p \rangle}\right)^k \cong \left(\frac{R}{\langle p \rangle}\right)^l$$

and therefore $k = l$ since $\frac{R}{\langle p \rangle}$ is a field.

For the induction step consider the submodules $pM \leqslant M$ and $pN \leqslant N$. We can apply the induction hypothesis since $pM \cong pN$ and $\operatorname{ann}_R(pM) = \langle p^{m_k - 1} \rangle$. Let $\alpha \in \{0, 1, \dots, k\}$ be such that exactly α of the m_i are equal to 1, and let $\beta \in \{0, 1, \dots, l\}$ be such that exactly β of the n_i are equal to 1. We have

$$\begin{aligned} pM &\cong \frac{R}{\langle p^{m_{\alpha+1}-1} \rangle} \oplus \cdots \oplus \frac{R}{\langle p^{m_k - 1} \rangle} && (k - \alpha \text{ summands}) \\ \cong pN &\cong \frac{R}{\langle p^{n_{\beta+1}-1} \rangle} \oplus \cdots \oplus \frac{R}{\langle p^{n_l - 1} \rangle} && (l - \beta \text{ summands}) \end{aligned}$$

By induction we have $k - \alpha = l - \beta$ and $(m_{\alpha+1}, \dots, m_k) = (n_{\beta+1}, \dots, n_l)$. We also have (see Exercise 150) that

$$\left(\frac{R}{\langle p \rangle}\right)^k \cong \frac{M}{pM} \cong \frac{N}{pN} \cong \left(\frac{R}{\langle p \rangle}\right)^l$$

Therefore $k = l$ and $\alpha = \beta$ and we conclude that $m_i = n_i$ for all i. □

22.2 Primary and invariant factor decompositions are unique

Let M and N be finitely generated modules over a PID R. Suppose we have two decompositions as described in Theorem 17.2

$$\begin{aligned} M = &\quad R/\langle p_1^{m_1}\rangle \oplus R/\langle p_2^{m_2}\rangle \oplus \cdots \oplus R/\langle p_k^{m_k}\rangle \oplus R^r \\ N = &\quad R/\langle q_1^{n_1}\rangle \oplus R/\langle q_2^{n_2}\rangle \oplus \cdots \oplus R/\langle q_l^{n_l}\rangle \oplus R^t \end{aligned}$$

where all $p_i, q_i \in R$ are prime, $m_i, n_i \in \mathbb{Z}_{\geqslant 1}$, and $r, t \in \mathbb{Z}_{\geqslant 0}$.

Theorem 22.2

The primary decomposition is unique. That is, with the setup above, if $M \cong N$ then $r = t$, $k = l$, and (after permuting the indices if necessary) $m_i = n_i$ and $p_i \sim q_i$ for all i.

Proof outline. Fix an isomorphism $\varphi : M \to N$.

First note that M and N have isomorphic torsion submodules: $\varphi(T_M) = T_N$. It follows that $M/T_M \cong N/T_N$. An explicit isomorphism is given by $u + T_M \mapsto \varphi(u) + T_N$. Since $M/T_M \cong R^r$ and $N/T_N \cong R^t$ we have $R^r \cong R^t$ and hence $r = t$ (Proposition 13.7).

Since $\operatorname{ann}_R(T_M) = \operatorname{ann}_R(T_N)$ we know that every prime that appears in the expressions for M also appears in the expression for N, and vice versa.

For a prime $p \in R$ consider the submodules

$$\begin{aligned} M_p = &\{u \in M \mid p^e u = 0 \quad \text{for some } e \in \mathbb{Z}_{\geqslant 1}\} \\ N_p = &\{u \in N \mid p^e u = 0 \quad \text{for some } e \in \mathbb{Z}_{\geqslant 1}\} \end{aligned}$$

Note that M_p is given by taking those summands in the expression for M that involve a power of p. Similarly for N_p. The result then follows by applying Proposition 22.1. □

Since the invariant factor decomposition is determined by the primary decomposition, and vice versa, we have the following.

Corollary 22.3

The invariant factor decomposition is unique.

22.3 Exercises

148. Determine whether the following pairs of modules are isomorphic.

(a) $M = \frac{\mathbb{Q}[X]}{\langle 2\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle X^2-2X+1\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle 1\rangle}$ $\quad N = \frac{\mathbb{Q}[X]}{\langle 1\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle X-1\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle X-1\rangle}$

(b) $M = \frac{\mathbb{Q}[X]}{\langle 2\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle X^2-3X+2\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle 1\rangle}$ $\quad N = \frac{\mathbb{Q}[X]}{\langle 1\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle X-2\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle 2X-2\rangle}$

(c) $M = \frac{\mathbb{Q}[X]}{\langle 2\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle X^2-3X+2\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle 0\rangle}$ $\quad N = \frac{\mathbb{Q}[X]}{\langle 0\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle X-2\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle X-1\rangle}$

(d) $M = \frac{\mathbb{Q}[X]}{\langle 2\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle X^2-3X+2\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle 0\rangle}$ $\quad N = \frac{\mathbb{Q}[X]}{\langle 1\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle X-2\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle X-1\rangle}$

(e) $M = \frac{\mathbb{Q}[X]}{\langle X^4+2X^3+6X^2-4X+6\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle 1\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle 0\rangle}$ $\quad N = \frac{\mathbb{Q}[X]}{\langle 0\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle X^2+2X-6\rangle} \oplus \frac{\mathbb{Q}[X]}{\langle X^2-4X-2\rangle}$

© University of Melbourne 2025

149. Let R be a PID and $p \in R$ a prime. Suppose that M and N are torsion R-modules with $\mathrm{ann}_R(M) = \mathrm{ann}_R(N) = \langle p \rangle$. Show that M and N are isomorphic as R-modules iff they are isomorphic as vector spaces over $R/\langle p \rangle$.

150. Let R be a PID, $p \in R$ a prime, and M an R-module given by $M = \frac{R}{\langle p^{m_1} \rangle} \oplus \cdots \oplus \frac{R}{\langle p^{m_k} \rangle}$ for some $m_i \geqslant 1$. Show that $\frac{M}{pM} \cong \left(\frac{R}{\langle p \rangle}\right)^k$.

© University of Melbourne 2025

© University of Melbourne 2025

Fields

23.1 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 terminology 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. For example we can regard $\mathbb{C}$ as a vector space over $\mathbb{R}$ and also as a vector space over $\mathbb{Q}$. The vector spaces ${}_{\mathbb{C}}\mathbb{C}$, ${}_{\mathbb{R}}\mathbb{C}$, ${}_{\mathbb{Q}}\mathbb{C}$ are *not* isomorphic.

We know that a polynomial in $f \in F[X]$ need not have any roots in F. However, it is always possible to extend to a field $E \supseteq F$ such that f has a root in E.

Example 23.3. The polynomial $X^2 + X + 1 \in \mathbb{F}_2[X]$ has no roots in $\mathbb{F}_2$. The polynomial $X^2 + X + 1$ does have a root in the field F_4 of Example 2.5. The field F_4 contains a copy of $\mathbb{F}_2$ and can therefore be regarded as an extension of $\mathbb{F}_2$.

Proposition 23.4

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 $\alpha \in E$ such that $f(\alpha) = 0$.

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. Let $p \in F[X]$ be prime. Because p is prime, $E = F[X]/\langle p \rangle$ is a field (Proposition 6.5 and Lemma 6.7). The map $F \to E$ given by $f \mapsto f + \langle p \rangle$ is a homomorphism. Since p has degree at least 1, this homomorphism is injective, and so we can 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

$$\begin{aligned} p(X + I) &= (a_0 + I)(1 + I) + (a_1 + I)(X + I) + \cdots + (a_m + I)(X + I)^m \\ &= (a_0 + I)(1 + I) + (a_1 + I)(X + I) + \cdots + (a_m + I)(X^m + I) \\ &= (a_0 + I) + (a_1 X + I) + \cdots + (a_m X^m + I) \\ &= (a_0 + \cdots + a_m X^m) + I \\ &= 0 + I \qquad \text{(since } a_0 + \cdots + a_m X^m = p \in I\text{)} \\ &= 0_E \end{aligned}$$

□

Example 23.5. 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.

23.2 Algebraic and transcendental elements

Definition 23.6

Let E be an extension of the field F. An element $\alpha \in E$ is called **algebraic** over F if there is a non-zero element in $F[X]$ having α as a root. An element is called **transcendental** over F if it is not algebraic.

Example 23.7.

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 $\mathbb{Q}$.*
4. $\pi \in \mathbb{R}$ is algebraic over $\mathbb{R}$.

Given an element $\alpha \in E \supseteq F$ that is algebraic over F, the set $I = \{f \in F[X] \mid f(\alpha) = 0\}$ 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 151. Let $\alpha \in E \supseteq F$ be algebraic over F.

(a) Show that $I = \{f \in F[X] \mid f(\alpha) = 0\}$ 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 23.8

Let $E \supseteq F$ be fields and $\alpha \in E$. If α is algebraic over F, the unique monic irreducible polynomial having α as a root is called the **irreducible polynomial for α over F**. It will be denoted $\mathrm{irr}(\alpha, F)$. It is also sometimes called the minimal polynomial for α. The degree of $\mathrm{irr}(\alpha, F)$ will be called the **degree of α over F** and will be denoted $\deg(\alpha, F)$. The irreducible polynomial is also called the **minimal polynomial**.

Example 23.9. Let $\alpha = \sqrt{1 + \sqrt{3}} \in \mathbb{R}$. Then

$$\alpha = \sqrt{1 + \sqrt{3}} \implies \alpha^2 = 1 + \sqrt{3} \implies (\alpha^2 - 1)^2 = 3 \implies \alpha^4 - 2\alpha^2 - 2 = 0$$

Since the polynomial $X^4 - 2X^2 - 2$ is irreducible (by Eisenstein's criterion) we conclude that $\mathrm{irr}(\alpha, \mathbb{Q}) = X^4 - 2X^2 - 2$ and $\deg(\alpha, \mathbb{Q}) = 4$.

Example 23.10. 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

$$\begin{bmatrix} 1 & 0 & 5 & 0 & 49 \\ 0 & 1 & 0 & 11 & 0 \\ 0 & 1 & 0 & 9 & 0 \\ 0 & 0 & 2 & 0 & 20 \end{bmatrix} \text{ is row-equivalent to } \begin{bmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 10 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

*This was first proved by the German mathematician Ferdinand von Lindemann in 1882.

© University of Melbourne 2025

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 $\deg(\sqrt{2}+\sqrt{3}, \mathbb{Q}) = 4$.

23.3 Exercises

152. Find $\mathrm{irr}(\alpha, \mathbb{Q})$ and $\deg(\alpha, \mathbb{Q})$ for the following polynomials. You should justify why your answer for $\mathrm{irr}(\alpha, \mathbb{Q})$ is irreducible.

(a) $\alpha = \sqrt{3-\sqrt{6}}$ (b) $\alpha = \sqrt{(\frac{1}{3})+\sqrt{7}}$ (c) $\alpha = \sqrt{2}+i$

153. Show that the following elements of $\mathbb{C}$ are algebraic over $\mathbb{Q}$ and find their irreducible polynomials.

(a) $2^{\frac{1}{3}}$ (b) $\sqrt{3}+\sqrt{2}$ (c) $\frac{(\sqrt{5}+1)}{2}$ (d) $\frac{(i\sqrt{3}-1)}{2}$

154. Let F be a field and $D : F[X] \to F[X]$ the map given by

$$D(a_0 + a_1X + \cdots + a_nX^n) = a_1 + 2a_2X + 3a_3X^2 \cdots + na_nX^{n-1}$$

The polynomial $D(f)$ is called the **derivative** of f. Note that the coefficients are in F and the notation '3', for example, means $1+1+1 \in F$.

(a) Verify that $D(fg) = D(f)g + fD(g)$.
(This is over any field and is purely combinatorial as defined above. There is no calculus involved!)

An element $\alpha \in E$ in an extension $E \supseteq F$ is called a **multiple root** of $f \in F[X]$ if $(X-\alpha)^2$ divides f (in $E[X]$).

(b) Show that if $\alpha \in E$ is a multiple root of $f \in F[X]$, then α 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.

© University of Melbourne 2025

© University of Melbourne 2025

Algebraic extensions and finite extensions

The difference between algebraic and transcendental elements is reflected in the corresponding evaluation maps.

Definition 24.1

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 $F[a] \subseteq F(a)$.

Exercise 155. Show that $F(a)$ is isomorphic to the field of quotients of $F[a]$.

Lemma 24.2

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
$$F[X]/\langle \mathrm{irr}(a, F) \rangle \cong F[a] \quad \text{and} \quad F[a] = F(a)$$
3. If a is transcendental over F, then φ_a is injective and φ_a gives an isomorphism
$$F[X] \cong F[a] \quad \text{and} \quad F[a] \subsetneq F(a)$$

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 Exercise 151 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, then $F[a] \neq F(a)$ as $F[a] \cong F[X]$ and $F[X]$ is not a field.

□

24.1 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 24.3

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 24.4

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 24.3 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 156. Show that every finite extension is algebraic.

Example 24.5. 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 159.

Lemma 24.6

Let E F and K be fields, $K \supseteq E \supseteq F$, 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.

© University of Melbourne 2025

Given any $k \in K$ we have

$$\begin{aligned} k &= \sum_{j=1}^{n} b_j \beta_j && \text{(for some } b_i \in E\text{)} \\ &= \sum_{j=1}^{n} (\sum_{i=1}^{m} a_{ij}\alpha_i)\beta_j && \text{(for some } a_{ij} \in F\text{)} \\ &= \sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij}\alpha_i\beta_j \\ &\in \operatorname{span}\{\alpha_i\beta_j \mid 1 \leqslant i \leqslant m,\ 1 \leqslant j \leqslant n\} \end{aligned}$$

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 24.7

Let $a \in E \supseteq F$, 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 24.8. 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.

24.2 Exercises

157. Find the degree and a basis for the following extensions:
   (a) $\mathbb{R}(\sqrt{2} + i) \supseteq \mathbb{R}$
   (b) $\mathbb{Q}(\sqrt{2} + i) \supseteq \mathbb{Q}$
   (c) $\mathbb{Q}(\sqrt{2} + \sqrt{3}) \supseteq \mathbb{Q}$
   (d) $\mathbb{Q}(\sqrt{3}, i) \supseteq \mathbb{Q}$

158.* Let F be a field and $k \in F$ an element which is not a square in F (i.e, there does not exist an element $x \in F$ with $x^2 = k$). Show that

$$K = \left\{ \begin{pmatrix} a & kb \\ b & a \end{pmatrix} \mid a, b \in F \right\} \leqslant M_{2\times 2}(F)$$

is a field and that it is isomorphic to $F(\sqrt{k})$.

© University of Melbourne 2025

159. Let $A = \{a \in \mathbb{R} \mid a \text{ is algebraic over } \mathbb{Q}\}$ be the set of algebraic real numbers.

    (a) Show that A forms a subfield of $\mathbb{R}$. (Use that $a \in \mathbb{R}$ is algebraic iff $[\mathbb{Q}(a) : \mathbb{Q}]$ is finite.)

    (b) Show that A is an algebraic extension of $\mathbb{Q}$, but A is not a finite extension of $\mathbb{Q}$.

160. 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 $\mathrm{irr}(a, F) = \mathrm{irr}(b, F)$ if and only if there exists an isomorphism $\varphi : F(a) \to F(b)$ such that $\varphi(a) = b$ and $\varphi|_F = \mathrm{Id}_F$.

© University of Melbourne 2025

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.

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 25.1: 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 of Artin* and Stillwell†.

Given that a is constructible, we can construct a right triangle with non hypotenuse side lengths 1 and a as shown in Figure 25.2. 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. □

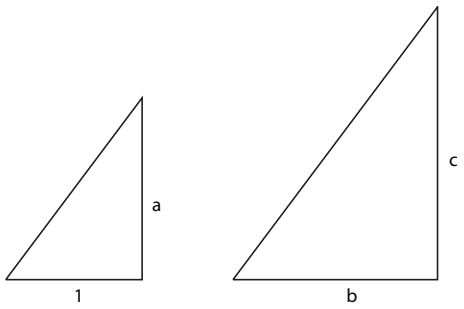


Figure 25.2: If a and b are contructible, then ab is constructible.

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 $n \in \mathbb{N}$.

Algebra, Michael Artin,1991
†*The four pillars of geometry*, John Stillwell, 20005

© University of Melbourne 2025

Proof. Let a_i and F_i be as in the previous proposition. Since $a_i \in F_i$, $\deg(\sqrt{a_i}, F_i)$ 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 24.3. Apply Lemma 24.6 repeatedly to conclude that $[F_n : \mathbb{Q}] = 2^m$ for some m. Then Corollary 24.7 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 $1 + 6\cos(\pi/9) - 8(\cos(\pi/9))^3 = 0$. 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 $\deg(2^{\frac{1}{3}}, \mathbb{Q}) = 3$.

25.4 Exercises

161. Let F be a field and let $E \supseteq F$ be an extension with $[E : F] = 2$.
   (a) Show that there exists $a \in E$ such that $\deg(a, F) = 2$ and $E = F[a]$.
   (b) Suppose in addition that $1 + 1 \neq 0$ in F. Show that there exists $a \in E$ such that $a^2 \in F$ and $E = F[a]$.
   (c) i) Let $K = \mathbb{F}_3[X]/\langle X^2 + X + 2\rangle$. Find an element $k \in K$ such that $k^2 \in \mathbb{F}_3$ and $K = \mathbb{F}_3[k]$.
   ii) Let $E = \mathbb{F}_2[X]/\langle X^2 + X + 1\rangle$. Show that no element $e \in E$ has the property that $e^2 \in \mathbb{F}_2$ and $E = \mathbb{F}_2[e]$.

162. Explain why a point $P(x, y)$ is constructible iff x and y are both constructible.

163. Let $C \subseteq \mathbb{R}$ be the field of constructible numbers. Show the C is an algebraic extension of $\mathbb{Q}$, but not a finite extension of $\mathbb{Q}$.

© University of Melbourne 2025

© University of Melbourne 2025

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.* 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 164. 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. □

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. □

*If $m < 0$, add 1 to itself $|m|$ times and then take the additive inverse.

26.2 The group of units of a finite field is cyclic

In any commutative 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. Let $q = |F|$. Note that since F is a field, $F^\times$ is the set of non-zero elements in F, and therefore $|F^\times| = q - 1$. 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 $q - 1 = |F^\times| = d_1 d_2 \dots d_m$. 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 Exercises

165. Prove the following. (It doesn't really need any result from this lecture!)

Fermat's Little Theorem

Let $p \in \mathbb{N}$ be prime. Then $\forall a \in \mathbb{Z}$, $a^p \equiv a \pmod{p}$.

166. Let F be a finite field of characteristic p. Show that the map $\varphi : F \to F$, $\varphi(a) = a^p$ is an isomorphism.

167. Let $f \in \mathbb{F}_p[X]$ and suppose that $\alpha \in E \supseteq \mathbb{F}_p$ is a root of f in some extension E of $\mathbb{F}_p$. Show that α^p is also a root of f in that extension.

168. Let F be a finite field.Write down a polynomial in $F[X]$ that has no roots in F. Conclude that no finite field is algebraically closed.

169. 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 $d|n$.)

© University of Melbourne 2025

Existence and uniqueness of a field of size p^n

27.1 Existence

To help motivate the construction we note the following.

Lemma 27.1

Let F be a 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 $a \in F^\times$ 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 q linear polynomials from $F[X]$.

Proposition 27.2

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 23.4 (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. It is clear that $0, 1 \in K$. Let $a, b \in K$ with $a \neq 0$. Then $a^q = a$ and $b^q = b$, and we have

$$\begin{aligned} (ab)^q &= a^q b^q = ab \\ (a^{-1})^q &= (a^q)^{-1} = a^{-1} \\ (-a)^q &= (-1)^q a^q = -1a = -a \\ (a+b)^q &= a^q + b^q = a + b \qquad \text{(see Exercise 22)} \end{aligned}$$

Therefore $ab, a^{-1}, -a, a + b \in K$. It follows that K is a subfield of E. □

Corollary 27.3

For all $p \in \mathbb{N}$ prime and all $n \geqslant 1$, there exists an irreducible polynomial in $\mathbb{F}_p[X]$ of degree n.

Proof. Let $p \in \mathbb{N}$ be prime and let $n \in \mathbb{Z}_{\geqslant 1}$. Let K be a field of size p^n. Note that $[K : \mathbb{F}_p] = n$. By Proposition 26.3, $K^\times$ is cyclic. Let $a \in K^\times$ be a generator for $K^\times$. Then $K = \{0, 1, a, a^2, \ldots, a^{|K|-2}\}$ and $\mathbb{F}_p(a) \subseteq K$ is in fact equal to K. Therefore $\deg(a, \mathbb{F}_p) = [K : \mathbb{F}_p] = n$. Therefore $\operatorname{irr}(a, \mathbb{F}_p)$ is of degree n (and is irreducible by definition). □

27.2 Uniqueness

Proposition 27.4

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'. Therefore $\operatorname{irr}(a', \mathbb{F}_p) = \operatorname{irr}(a, \mathbb{F}_p)$ 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 F and F' have the same (finite) cardinality it must be the case that $F \cong F'$. □

27.3 Exercises

170.⋆ Find an example of two infinite fields that have the same cardinality, but are not isomorphic.

171. Let $\mathbb{F}_8$ be the field containing 8 elements. Write out the addition and multiplication tables for $\mathbb{F}_8$.

172. 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 $X^q - X \in \mathbb{F}_p[X]$.

© University of Melbourne 2025

The Galois group of an extension

Galois theory gives a connection between certain field extensions and the subgroups of an associated group. Note that in this section we will be assuming that the field F under consideration (and therefore any extension of it) is of characteristic zero.

Definition 28.1

The set of all automorphisms of a field E forms a group (the operation being composition) which will be denoted $\mathrm{Aut}(E)$. For a subgroup H 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 173. Show that E^H is a subfield of E.

Definition 28.2

Now suppose that F is a subfield of E. An element $\varphi \in \mathrm{Aut}(E)$ 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 28.3. Complex conjugation is an $\mathbb{R}$-automorphism of $\mathbb{C}$.

Lemma 28.4

Let $E \supseteq F$, $f \in F[X]$ and $\varphi \in \mathrm{Aut}(E)$ an F-automorphism. If $a \in E$ is a root of f, then $\varphi(a)$ is also a root of f.

Proof. Let $f = \sum_{i=0}^{n} \alpha_i X^i$ with $\alpha_i \in F$. Then

$$f(a) = 0 \implies \varphi(f(a)) = 0 \implies \varphi(\sum_{i=0}^{n} \alpha_i a^i) = 0 \implies \sum_{i=0}^{n} \varphi(\alpha_i)\varphi(a)^i = \sum_{i=0}^{n} \alpha_i \varphi(a)^i = 0$$
$$\implies f(\varphi(a)) = 0$$

□

Definition 28.5

Suppose that $E \supseteq F$ is an extension of the field F. The set of all F-automorphisms of E 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 28.6. $\mathrm{Gal}(\mathbb{Q}(\sqrt{3}, \sqrt{5})/\mathbb{Q})$ is isomorphic to the Klein four group (i.e., it has four elements and is not cyclic).

Definition 28.7

An extension E of F is called a **Galois extension** if it is a finite extension and

$$|\operatorname{Gal}(E/F)| = [E : F]$$

For Galois extensions, there is a correspondence between subgroups of $\operatorname{Gal}(E/F)$ 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 28.8: The Fundamental Theorem of Galois Theory

Let E be a Galois extension of F.

1) The map

$$\Phi : \{H \mid H \text{ is a subgroup of } \operatorname{Gal}(E/F)\} \to \{L \mid L \text{ is a subfield of } E \text{ with } F \subseteq L \subseteq E\}$$

given by

$$\Phi(H) = E^H$$

is a bijection. It has inverse given by $L \mapsto \operatorname{Gal}(E/L)$.

Let H be a subgroup of $\operatorname{Gal}(E/F)$.

2) $[E : E^H] = |H|$

3) 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 E^H is a Galois extension of F, then $\operatorname{Gal}(E^H/F) \cong \operatorname{Gal}(E/F)/H$.

Example 28.9 (Quadratic extension)**.** Let $E = \mathbb{Q}(\sqrt{2})$ and let $G = \operatorname{Gal}(E/\mathbb{Q})$. By Lemma 28.4 for any element $g \in G$, we have either $g(\sqrt{2}) = \sqrt{2}$ (and therefore $g = id_E$) or $g(\sqrt{2}) = -\sqrt{2}$. If $g(\sqrt{2}) = -\sqrt{2}$, then $g(-\sqrt{2}) = \sqrt{2}$. Therefore G has exactly two elements and E is a Galois extension of $\mathbb{Q}$. Since the group of size two has no proper subgroups, there are no fields lying between $\mathbb{Q}$ and E.

Example 28.10 (Non-Galois extension)**.** Let $E = \mathbb{Q}(2^{\frac{1}{3}})$. Then $E \subseteq \mathbb{R}$. Any element of $\operatorname{Gal}(E/\mathbb{Q})$ must permute the roots of the polynomial $\operatorname{irr}(2^{\frac{1}{3}}, \mathbb{Q}) = X^3 - 2$. Since only one of these roots lies in E (since the others are not in $\mathbb{R}$), any element of $\operatorname{Gal}(E/\mathbb{Q})$ must send $2^{\frac{1}{3}}$ to $2^{\frac{1}{3}}$. Such an automorphism fixes E pointwise. Therefore $|\operatorname{Gal}(E/\mathbb{Q})| = 1 \neq 3 = \deg(2^{\frac{1}{3}}, \mathbb{Q}) = [E : \mathbb{Q}]$ and E is not a Galois extension.

Example 28.11 (Biquadratic extension)**.** Let $E = \mathbb{Q}(i, \sqrt{2})$. There are $\mathbb{Q}$-automorphisms $\sigma, \tau \in \operatorname{Gal}(E/\mathbb{Q})$ 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 $\operatorname{Gal}(E/\mathbb{Q})$ must permute the roots of X^2+1 and the roots of X^2-2, $\operatorname{Gal}(E/\mathbb{Q}) = \{id, \sigma, \tau, \sigma\tau\}$ and we have $\operatorname{Gal}(E\mathbb{Q}) \cong C_2 \times C_2$ and $|\operatorname{Gal}(E/\mathbb{Q}) = 4|$. 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 $|\operatorname{Gal}(E/\mathbb{Q}) = 4 = [E : \mathbb{Q}]$ and E is a Galois extension of $\mathbb{Q}$. The correspondence between subgroups and intermediate fields is given in the following table:

© University of Melbourne 2025

$E = \mathbb{Q}(i, \sqrt{2})$	
subgroup	**subfield**
$\mathrm{Gal}(E/\mathbb{Q})$	$\mathbb{Q}$
$\{id, \tau\}$	$\mathbb{Q}(i)$
$\{id, \sigma\}$	$\mathbb{Q}(\sqrt{2})$
$\{id, \sigma\tau\}$	$\mathbb{Q}(i\sqrt{2})$
$\{id\}$	E

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 $\mathbb{Q}$).

28.1 Exercises

174. Let $E \supseteq F$ with $[E : F] = 2$ and let $\alpha \in E \setminus F$ be such that $\alpha^2 \in F$. Show that there is an element $\varphi \in \mathrm{Gal}(E/F)$ such that $\varphi(\alpha) = -\alpha$.

175. Let $E \supseteq F$ be fields and $\alpha_1, \ldots, \alpha_n \in E$. Suppose that $E = F(\alpha_1, \ldots, \alpha_n)$ and $\varphi \in \mathrm{Gal}(E/F)$ is such that $\varphi(\alpha_i) = \alpha_i$ for all i. Show that $\varphi = id_E$.

© University of Melbourne 2025

© University of Melbourne 2025

Splitting fields

29.0.1 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 29.1

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. $E = F(a_1, \dots, a_m)$

Exercise 176. Use Proposition 23.4 to prove the following lemma.

Lemma 29.2

Every polynomial $f \in F[X]$ has a splitting field.

We prove a technical lemma below that establishes the following.

Proposition 29.3

If $E \subseteq F$ is a splitting field for $f \in F[X]$, then E is a Galois extension of F.

Before giving the technical lemma, we illustrate some of the ideas with two examples.

Example 29.4. Let $f \in \mathbb{Q}[X]$ be an irreducible quadratic. Let $\alpha, \beta \in \mathbb{C}$ be its (necessarily) distinct roots. Let $E = \mathbb{Q}(\alpha, \beta)$. Then $E = \mathbb{Q}(\alpha) = \mathbb{Q}(\beta)$ and $[E : \mathbb{Q}] = 2$. Also, we know from Exercise 160 that there is an isomorphism $\mathbb{Q}(\alpha) \to \mathbb{Q}(\beta)$ that sends α to β 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.

Example 29.5. We consider a splitting field $E \subseteq \mathbb{C}$ of the polynomial $f = X^3 + 3X + 1 \in \mathbb{Q}[X]$. The polynomial f is irreducible, since its image in $\mathbb{F}_2[X]$ is irreducible. Therefore f has no repeated roots (see Exercise 154). 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, γ say, is real (and therefore $\beta = \overline{\alpha} \in \mathbb{C} \setminus \mathbb{R}$). Let $L = \mathbb{Q}(\gamma)$. Notice that $L \neq E$ since $L \subseteq \mathbb{R}$. Also, $[L : \mathbb{Q}] = \deg(\gamma, \mathbb{Q}) = 3$. In $L[X]$ we have the factorisation $f = (X - \gamma)h$ for some quadratic $h \in L[X]$ with $h(\alpha) = h(\beta) = 0$. Since $\alpha \notin L$, h is irreducible and $\deg(\beta, L) = 2$. From $E = \mathbb{Q}(\alpha, \beta, \gamma) = L(\beta, \gamma) = L(\beta)$, we have $[E : L] = 2$. From Lemma 24.6 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 (Lemma 28.4) we know that $\mathrm{Gal}(E/\mathbb{Q})$ is isomorphic to a subgroup of S_3 (the symmetric group on three objects). Because $|S_3| = 6$ we have 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 $F = \mathbb{Q}(\alpha)$ and let $g \in F[X]$ be such that $f = (X - \alpha)g$. Note that $E = F(\gamma) = F(\beta)$. Because g is irreducile and has roots β and γ, there is an element of $\mathrm{Gal}(E/F)$ (which is a subset of $\mathrm{Gal}(E/\mathbb{Q})$) that interchanges γ and β. The same argument, with the roles of β and α interchanged, shows that there is an element in $\mathrm{Gal}(E/\mathbb{Q})$ that fixes β and swaps α and γ.

Now for the technical lemma that shows that all splitting fields are Galois extensions.

Lemma 29.6

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 $[E : F]$.

Before proving the lemma we note some consequences.

Corollary 29.7

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 29.8

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. □

Corollary 29.9

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.

© University of Melbourne 2025

29.1 Exercises

177. 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

    (a) $F(a) = F(a, b)$

    (b) $|\mathrm{Gal}(F(a)/F)| = 2$

    (c) The non-trivial element in $\mathrm{Gal}(F(a)/F)$ interchanges a and b.

178. Prove Corollary 29.9.

179. Let $f = X^3 - 2 \in \mathbb{Q}[X]$ and let $E \subseteq \mathbb{C}$ be the splitting field of f. Show that $\mathrm{Gal}(E/F) \cong S_3$.

180. Let $f = X^3 - 1 \in \mathbb{Q}[X]$ and let $E \subseteq \mathbb{C}$ be the splitting field of f. Calculate $\mathrm{Gal}(E/F)$.

© University of Melbourne 2025

© University of Melbourne 2025

Primitive elements

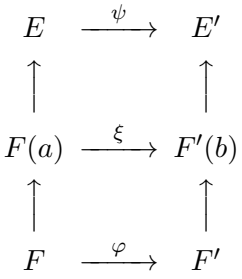
Proof of Lemma 29.6. Denote by $\tilde{\varphi} : F[X] \to F'[X]$ the map defined by extending φ, that is

$$\tilde{\varphi}(a_0 + \cdots + a_m X^m) = \varphi(a_0) + \cdots + \varphi(a_m) X^m$$

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 $g = \operatorname{irr}(a, F) \in F[X]$ 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 154). 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 $\xi(a) = b$ (cf. Exercise 160). 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 $\deg(g') = [F(a) : F]$ 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'.



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 $F'(b)$.

□

30.1 Primitive elements

We say last time (as a consequence of the lemma proved above) that all splitting fields are Galois extensions. In order to prove that all Galois extensions are splitting fields, we will use the following

Proposition 30.1

Let E be a finite extension of F. There exists an element $a \in E$ such that $E = F(a)$.

Definition 30.2

An element $a \in E$ such that $E = F(a)$ is called a **primitive element** of the extension $E \supseteq F$.

Proof of Proposition 30.1. 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 $f = \operatorname{irr}(b, F)$,

$g = \mathrm{irr}(c, F)$ 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 69). 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 (below), the next lemma is often useful in determining the irreducible polynomial of an element.

Lemma 30.3: Orbit Lemma

Let E be a field and let G be a finite subgroup of $\mathrm{Aut}(E)$. 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

$$\mathrm{irr}(a, E^G) = (X - a_1)(X - a_2) \cdots (X - a_m)$$

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

$$\begin{aligned} \mathrm{irr}(a, F)(a_i) &= \mathrm{irr}(a, F)(ga) && \text{(for some } g \in G\text{)} \\ &= g(\mathrm{irr}(a, F)(a)) = g(0) = 0 \end{aligned}$$

So all the a_i are roots of $\mathrm{irr}(a, F)$. Therefore f divides $\mathrm{irr}(a, F)$. But since $f(a) = 0$, we also have that $\mathrm{irr}(a, F)$ 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 $|G|$.

30.2 Exercises

181. Find an element $\alpha \in E = \mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5})$ such that $E = \mathbb{Q}(\alpha)$.

182. Let $\alpha = \sqrt{2} + i \in E = \mathbb{Q}(\sqrt{2}, i)$.

    (a) Find the orbit of α under the action of the group $\mathrm{Gal}(E/\mathbb{Q})$. That is, calculate the set $\{g(\alpha) \mid g \in \mathrm{Gal}(E/\mathbb{Q})\}$.

    (b) Use the Orbit Lemma to calculate $\mathrm{irr}(\alpha, \mathbb{Q})$.

© University of Melbourne 2025

Artin's Theorem

Theorem 31.1: Artin's Fixed Field Theorem

Let E be a field and let G be a finite subgroup of $\mathrm{Aut}(E)$. Then

$$[E : E^G] = |G|$$

Corollary 31.2

If E is a finite extension of F, then $|\operatorname{Gal}(E/F)|$ divides $[E : F]$.

Proof. We have that $F \subseteq E^{\mathrm{Gal}(E/F)} \subseteq E$, which implies that

$$[E : F] = [E : E^{\mathrm{Gal}(E/F)}][E^{\mathrm{Gal}(E/F)} : F] = |\operatorname{Gal}(E/F)|[E^{\mathrm{Gal}(E/F)} : F]$$

□

Corollary 31.3

Let E be a field and let G be a finite subgroup of $\mathrm{Aut}(E)$. Then E is a Galois extension of E^G and $\mathrm{Gal}(E/E^G) = G$.

Proof. Clearly $G \subseteq \mathrm{Gal}(E/E^G)$ since all elements of G fix E^G pointwise. Corollary 31.2 implies that $|\operatorname{Gal}(E/E^G)| \leqslant [E : E^G]$. Then from Artin's Theorem we have

$$[E : E^G] = |G| \leqslant |\operatorname{Gal}(E/E^G)| \leqslant [E : E^G]$$

It follows that $G = \mathrm{Gal}(E/E^G)$ and $|\operatorname{Gal}(E/E^G)| = [E : E^G]$. □

Corollary 31.4

If E be a Galois extension of F, then $E^{\mathrm{Gal}(E/F)} = F$.

Proof. Let $G = \mathrm{Gal}(E/F)$. We have $F \subseteq E^G \subseteq E$ and therefore $\mathrm{Gal}(E/E^G) \subseteq G$. It is also the case that $G \subseteq \mathrm{Gal}(E/E^G)$ since all elements of G fix its own fixed field. Therefore $G = \mathrm{Gal}(E/E^G)$. Also,

$$\begin{aligned}
&|\operatorname{Gal}(E/E^G)| \text{ divides } [E : E^G] && \text{(by 31-1 31.2)} \\
\implies &|G| \text{ divides } [E : E^G] \\
\implies &[E : F] \text{ divides } [E : E^G] && \text{(since } E \text{ is a Galois extension of } F\text{)}
\end{aligned}$$

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 Artin's Theorem (31.1). Let $F = E^G$. We first show that E is a finite extension of F. By the Orbit Lemma 30.3 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 30.1, 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 $|G|$. The Orbit Lemma 30.3 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 31.5

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 $\mathrm{Gal}(E/F)$. Then $F = E^{\mathrm{Gal}(E/F)}$ by Corollary 31.4 and Lemma 30.3 tells us that in $E[X]$ $f = (X - a_1) \cdots (X - a_m)$. Therefore E is a splitting field for f. □

31.1 Exercises

183. Determine the degree of the splitting fields of the following elements of $\mathbb{Q}[X]$.

(a) $X^4 - 1$ (b) $X^3 - 2$ (c) $X^4 + 1$

184. Show that $K = \mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5})$ is a Galois extension of $\mathbb{Q}$ and identify its Galois group.

185. Show that $X^2 - 3$ and $X^2 - 2X - 2$ are irreducible in $\mathbb{Q}[X]$ and have the same splitting field.

186. Find the dimensions of the splitting fields over $\mathbb{Q}$ of

(a) $X^3 - 56$ (b) $X^4 - 4X^2 - 5$

187. Find the dimension of a splitting field of $X^3 + X + 1$ over $\mathbb{F}_2$.

© University of Melbourne 2025

Proof of the fundamental theorem

32.1 Proof of the fundamental theorem

Recall the statement of the theorem.

Theorem 32.1: The main 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 $G(L/F) \cong \mathrm{Gal}(E/F)/H$.

□

Proof. Let $G = \mathrm{Gal}(E/F)$ 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 $\Psi(L) = \mathrm{Gal}(E/L)$. By Corollary 31.3, $\Psi \circ \Phi(H) = \Psi(E^H) = \mathrm{Gal}(E/E^H) = H$. Applying Corollary 31.4 gives $\Phi \circ \Psi(L) = \Phi(\mathrm{Gal}(E/L)) = E^{\mathrm{Gal}(E/L)} = L$. 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 31.1.

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

$$\begin{aligned} & |G/H| \leqslant |G(L/F)| && \text{(since it is isomorphic to a subgroup)} \\ \implies & |G|/|H| \leqslant |G(L/F)| \\ \implies & [L : F] \leqslant |G(L/F)| && \text{(since } |G| = |H|[L : F]\text{)} \end{aligned}$$

It is also the case that $|G(L/F)|$ divides $[L : F]$ by Corollary 31.2. 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 $gL = L$. □

32.2 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 32.2 (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 $G=\operatorname{Gal}(E/\mathbb{Q})$. For any element $g\in G$, we have either $g(a)=a$ (and therefore $g=\mathrm{Id}$) 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.

Example 32.3 (Non-Galois extension)**.** Let $E=\mathbb{Q}(2^{\frac{1}{3}})$. Then $E\subseteq\mathbb{R}$. Any element of $\operatorname{Gal}(E/\mathbb{Q})$ must permute the roots of the polynomial $\operatorname{irr}(2^{\frac{1}{3}},\mathbb{Q})=X^3-2$. Since only one of these roots lies in E (since the others are not in $\mathbb{R}$), any element of $\operatorname{Gal}(E/\mathbb{Q})$ must send $2^{\frac{1}{3}}$ to $2^{\frac{1}{3}}$. Such an automorphism fixes E pointwise. Therefore $|\operatorname{Gal}(E/\mathbb{Q})|=1\neq 3=\deg(2^{\frac{1}{3}},\mathbb{Q})=[E:\mathbb{Q}]$ and E is not a Galois extension.

Example 32.4 (Biquadratic extension)**.** Let $E=\mathbb{Q}(i,\sqrt{2})$. There are $\mathbb{Q}$-automorphisms $\sigma,\tau\in\operatorname{Gal}(E/\mathbb{Q})$ 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 $\operatorname{Gal}(E/\mathbb{Q})$ must permute the roots of X^2+1 and the roots of X^2-2, $\operatorname{Gal}(E/\mathbb{Q})=\{id,\sigma,\tau,\sigma\tau\}$ and we have $\operatorname{Gal}(E\mathbb{Q})\cong C_2\oplus C_2$ and $|\operatorname{Gal}(E/\mathbb{Q})=4|$. 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 $|\operatorname{Gal}(E/\mathbb{Q})=4|=[E:\mathbb{Q}]$ 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**
$\operatorname{Gal}(E/\mathbb{Q})$	$\mathbb{Q}$
$\{id,\tau\}$	$\mathbb{Q}(i)$
$\{id,\sigma\}$	$\mathbb{Q}(\sqrt{2})$
$\{id,\sigma\tau\}$	$\mathbb{Q}(i\sqrt{2})$
$\{id\}$	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 32.5. Let $E\subseteq\mathbb{C}$ be the splitting field of X^3+3X+1. We have already seen that $\operatorname{Gal}(E/\mathbb{Q})\cong S_3$. 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 $\operatorname{Gal}(E/\mathbb{Q})$ must permute the elements of $\{\alpha_1,\alpha_2,\alpha_3\}$ and $\operatorname{Gal}(E/\mathbb{Q})=6$, we know that all permutations of the roots are achievable by an element of $\operatorname{Gal}(E/\mathbb{Q})$. 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.

© University of Melbourne 2025

$f = X^3 + 3X + 1$	
subgroup	**subfield**
$\mathrm{Gal}(E/\mathbb{Q}) = S_3$	$\mathbb{Q}$
$H = \{id, (123), (132)\}$	$L = \mathbb{Q}(\delta)$
$\{id, (12)\}$	$\mathbb{Q}(\alpha_3)$
$\{id, (13)\}$	$\mathbb{Q}(\alpha_2)$
$\{id, (23)\}$	$\mathbb{Q}(\alpha_1)$
$\{id\}$	E

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 $-\delta$). On the other hand $\delta^2 \in \mathbb{Q}$ since it remains unchanged after any permutation of the roots. Therefore $[\mathbb{Q}(\delta) : \mathbb{Q}] = 2$.

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.

Example 32.6. Consider the splitting field $E \subseteq \mathbb{C}$ of the irreducible polynomial $f = X^3 - 3X + 1 \in \mathbb{Q}[X]$. The roots α, β, γ of this polynomial are all real. Let $\xi = e^{\frac{2\pi i}{9}}$. The roots of f are $\alpha = \xi + \xi^{-1}$, $\beta = \xi^2 + \xi^{-2}$, and $\gamma = \xi^4 + \xi^{-4}$. Noting that $\alpha^2 = \xi^2 + \xi^{-2} + 2 = \beta + 2$, and therefore $\beta \in \mathbb{Q}(\alpha)$ we have

$$\mathbb{Q} \subseteq \mathbb{Q}(a) = \mathbb{Q}(\alpha, \beta) = \mathbb{Q}(\alpha, \beta, \gamma) = E$$

Therefore $|\mathrm{Gal}(E/\mathbb{Q})| = 3$. It follows that $\mathrm{Gal}(E/\mathbb{Q}) = A_3$. Recall that A_3 is the index 2 subgroup of S_3 given by $A_3 = \{id, (123), (132)\}$. The only subgroups are $\{id\}$ and A_3. The corresponding fields are E and $\mathbb{Q}$ respectively.

Remark. Note that in this case the element $\delta = (\alpha - \beta)(\beta - \gamma)(\gamma - \alpha)$ turns out to be an element of $\mathbb{Q}$. We could have concluded that $\mathrm{Gal}(E/\mathbb{Q}) = A_3$ from that fact.

The discriminant $D = \delta^2$ can be calculated without knowing the roots. For cubic $X^3 + pX + q$ it is given by $D = -4p^3 - 27q^2$. The Galois group of an irreducible cubic is A_3 if $\delta \in \mathbb{Q}$ and is S_3 if $\delta \notin \mathbb{Q}$.

32.3 Exercises

188. 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) $X^4 - 2$

© University of Melbourne 2025

© University of Melbourne 2025

Solubility by radicals

Definition 33.1

Let F be a subfield of $\mathbb{C}$. An extension $E \supseteq F$ is called a **radical extension** if there are subfields

$$F = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_n = E \subseteq \mathbb{C}$$

such that for all $k \in \{1, \ldots, n\}$, $F_k = F_{k-1}(\alpha_k)$ for some $\alpha_k \in F_k$ and $n_k \in \mathbb{N}$ with $\alpha_k^{n_k} \in F_{k-1}$

A polynomial $f \in F[X]$ is **soluble by radicals** if there is a splitting field for f that is contained in a radical extension of F.

From the quadratic formula we know that all quadratic polynomials are soluble. In fact, this is also true for polynomials of degree 3 and 4.

Theorem 33.2

If $f \in \mathbb{Q}[X]$ has degree at most 4, then f is soluble by radicals. □

Famously, this does not extend to degree 5 or higher. To show this we will establish a few preliminary results.

Lemma 33.3

Let $p \in \mathbb{N}$ be prime, let F be a subfield of $\mathbb{C}$ with $\zeta = e^{\frac{2\pi i}{p}} \in F$. If $E \supseteq F$ is a Galois extension with $[E : F] = p$, then $E = F(\alpha)$ for some $\alpha \in E$ that satisfies $\alpha^p \in F$.

Proof. Since $|\operatorname{Gal}(E/F)| = [E : F] = p$, $\operatorname{Gal}(E/F)$ is a cyclic group. Let $\varphi \in \operatorname{Gal}(E/F) \setminus \{id\}$. Note that $|\varphi| = p$. The map φ gives a linear transformation from $T_\varphi : {}_FE \to {}_FE$, $T_\varphi(u) = \varphi(u)$. (We don't really need a new name for it!) Since $\varphi^p = 1$ and $\zeta \in F$, T_φ is diagonalisable. This is because the minimal polynomial of T_φ divides $X^p - 1$ and $X^p - 1$ factors as a product of linear terms in $F[X]$. The eigenvalues of T_φ can not all be equal to 1 because T_φ is diagonalisable and not the identity. Let $\lambda \in \mathbb{F} \setminus \{1\}$ be an eigenvalue of T_φ. Then $\lambda^p = 1$ because $T^p = id$. Let $\alpha \in E$ be an eigenvector with eigenvalue λ. Then

$$\varphi(\alpha^p) = \varphi(\alpha)^p = (\lambda\alpha)^p = \lambda^p\alpha^p = \alpha^p$$

Since φ generates $\operatorname{Gal}(E/F)$, it follows that $\alpha^p \in E^{\operatorname{Gal}(E/F)} = F$. Also, $\alpha \notin F$ because $\varphi(\alpha) = \lambda\alpha \neq \alpha$. Therefore $\deg(\alpha, E) = p$ and $E = E(\alpha)$. □

Lemma 33.4

Let $p \in \mathbb{N}$ be prime and let $\zeta = e^{\frac{2\pi i}{p}} \in \mathbb{C}$.

1) $\operatorname{Gal}(\mathbb{Q}(\zeta)/\mathbb{Q}) \cong \mathbb{F}_p^\times$ (therefore cyclic of order $p - 1$)
2) For any subfield $F \subseteq \mathbb{C}$ we have that $\operatorname{Gal}(F(\zeta)/F)$ is cyclic.

Proof. Let $G = \mathrm{Gal}(F(\zeta)/F)$ and let $\varphi \in G$. Then $\varphi(\zeta) \in \{\zeta, \zeta^2, \ldots, \zeta^{p-1}\}$ since those are the roots on the irreducible polynomial $X^{p-1} + X^{p-2} + \cdots + X + 1 \in \mathbb{Q}[X]$. We define a map $f : G \to \mathbb{F}_p^\times$ by $f(\varphi) = \bar{i}$ where $\varphi(\zeta) = \zeta^i$. Note that $f(id) = \bar{1}$ and that if $\varphi, \psi \in G$ with $\varphi(\zeta) = \zeta^i$ and $\psi(\zeta) = \zeta^j$, then $f(\varphi\psi) = f(\varphi)f(\psi)$ because $\varphi\psi(\zeta) = \varphi(\zeta^j) = \varphi(\zeta)^j = \zeta^{ij}$. That is, f is a group homomorphism. Moreover, f is injective because $f(\varphi) = \bar{1}$ means that $\varphi(\zeta) = \zeta$ and therefore φ fixes the whole of $F(\zeta)$. Hence G is isomorphic to a subgroup of a cyclic group and is therefore cyclic. For the first part note that $|G| = [\mathbb{Q}(\zeta) : \mathbb{Q}] = \deg(\zeta, \mathbb{Q}) = p - 1$ and $|\mathbb{F}_p^\times| = p - 1$. □

Lemma 33.5

Let F be a subfield of $\mathbb{C}$ and $E \supseteq F$ a radical extension. Then there are subfields

$$F = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_n$$

such that $F_n \supseteq E$ and for all $k \in \{1, \ldots, n\}$

1) $F_k = F_{k-1}(\alpha_k)$ for some $\alpha_k \in F_k$ and $n_k \in \mathbb{N}$ with $\alpha_k^{n_k} \in F_{k-1}$
2) F_k is a Galois extension of F_{k-1} and $\mathrm{Gal}(F_k/F_{k-1})$ is cyclic

Proof. Since E is a radical extension, there exist F_k, α_k, n_k satisfying the first condition (it's exactly the definition). There is no lose in generality in assuming that the n_k are prime (by increasing the number of subfields if needed). Let's rename them as p_k. Let $\zeta_k = e^{\frac{2\pi i}{p_k}}$. Consider the chain of fields

$$F \subseteq F(\zeta_1) \subseteq F(\zeta_1, \zeta_2) \subseteq \cdots \subseteq F(\zeta_1, \ldots, \zeta_n) \subseteq F(\zeta_1, \ldots, \zeta_n, \alpha_1) \subseteq F(\zeta_1, \ldots, \zeta_n, \alpha_1, \alpha_2) \subseteq \cdots$$
$$\cdots \subseteq F(\zeta_1, \ldots, \zeta_n, \alpha_1, \ldots, \alpha_n) = E \subseteq \mathbb{C}$$

By Lemmas 33.3 and 33.4 each of these extensions is Galois with a cyclic Galois group. □

Although we won't prove the general result below, it's worth stating here.

Theorem 33.6

Let $f \in F[X]$ and let $E \supseteq F$ be a splitting field. Then f is soluble by radicals if and only if $\mathrm{Gal}(E/F)$ is a soluble group. □

Definition 33.7

A finite group G is called a **soluble group** if there are subgroups

$$\{1\} = G_0 \subseteq G_1 \subseteq \cdots \subseteq G_n = G$$

such that $G_{i-1} \lhd G_i$ and G_i/G_{i-1} is cyclic.

© University of Melbourne 2025

Qunitics

Rather than establishing the general result of Theorem 33.6, we'll merely aim to show that there are quintics that are not soluble by radicals. It relies on the following technical looking result. Recall that a group is called **simple** if it has no proper normal subgroups (and hence no proper quotients).

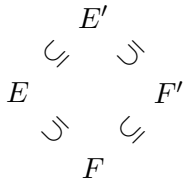
Proposition 34.1

Let $f \in F[X]$ and let $E \supseteq F$ be a splitting field for f. Suppose that $\mathrm{Gal}(E/F)$ is simple and non-abelian. Let $F' \supseteq F$ be a Galois extension of F with $\mathrm{Gal}(F'/F)$ an abelian group and let $E' \supseteq F'$ be a splitting field for $f \in F'[X]$. Then $\mathrm{Gal}(E'/F') \cong \mathrm{Gal}(E/F)$.

Remark. The crucial point is that extending from F to F' has not gotten us any closer to a splitting field for f.

Proof. [M. Artin] Consider first the case in which $[F' : F] = p$ is prime and (therefore) $\mathrm{Gal}(F'/F)$ is cyclic of size p. The splitting field E' contains a copy of E. From the Main Theorem 32.1 we have that

$$\mathrm{Gal}(E/F) \cong \frac{\mathrm{Gal}(E'/F)}{\mathrm{Gal}(E'/E)} \qquad \mathrm{Gal}(F'/F) \cong \frac{\mathrm{Gal}(E'/F)}{\mathrm{Gal}(E'/F')}$$



The natural projection maps from the above quotients give a map $\mathrm{Gal}(E'/F) \to \mathrm{Gal}(E/F) \times \mathrm{Gal}(F'/F)$. Moreover this map is injective since anything in the kernel fixes all the roots of f and all elements of F'. Therefore $\mathrm{Gal}(E'/F)$ is isomorphic to a subgroup of $\mathrm{Gal}(E/F) \times \mathrm{Gal}(F'/F)$. We have that $|\mathrm{Gal}(E/F)|$ divides $|\mathrm{Gal}(E'/F)|$ which divides $|\mathrm{Gal}(E/F) \times \mathrm{Gal}(F'/F)| = p|\mathrm{Gal}(E/F)|$. In fact we must have $|\mathrm{Gal}(E/F) \times \mathrm{Gal}(F'/F)| = p|\mathrm{Gal}(E/F)|$ because

$$|\mathrm{Gal}(E/F)| = |\mathrm{Gal}(E'/F)| \implies |\mathrm{Gal}(E'/E)| = 1 \implies E' = E$$

which would imply that $\mathrm{Gal}(F'/F)$ is a quotient of $\mathrm{Gal}(E/F)$ contradicting the hypothesis that $\mathrm{Gal}(E/F)$ is simple and non-abelian. We have then that $\mathrm{Gal}(E'/F) = \mathrm{Gal}(E/F) \times \mathrm{Gal}(F'/F)$. Applying the Main Theorem to the extensions $E' \supseteq F' \supseteq F$, we get that $\mathrm{Gal}(E'/F') \cong \mathrm{Gal}(E/F)$.

For the general case, in which $\mathrm{Gal}(F'/F)$ is abelian, we can proceed by induction on $[F' : F]$. Being abelian, $\mathrm{Gal}(F'/F)$ has a quotient H that is of prime order. This quotient determines an intermediate field $F_1 \supseteq F$ that is a Galois extension of F and $\mathrm{Gal}(F_1/F) = H$. Let E_1 be the splitting field of f over F_1. Since $[F_1 : F]$ is prime, we know (from above) that $\mathrm{Gal}(E_1/F_1) = \mathrm{Gal}(E/F)$. By induction we have that $\mathrm{Gal}(E_1/F_1) = \mathrm{Gal}(E'/F')$. □

Theorem 34.2

If an irreducible $f \in F[X]$ has degree 5 and has Galois group isomorphic to S_5 or A_5, then f is not soluble by radicals.

Proof. Let the roots be $\alpha_1, \ldots, \alpha_5$ and let E be a splitting field for f. Consider $\delta = \prod_{i<j}(\alpha_i - \alpha_j)$. That is, $D = \delta^2 \in F$ is the discriminant of the polynomial. If the Galois group $G = S_5$, then $\delta \notin F$. The

Galois group $\mathrm{Gal}(E/F(\delta))$ is then A_5. It is enough, therefore, to consider the case in which the Galois group of the polynomial is A_5.

We will use the fact that A_5 is a simple group. Suppose that f were soluble by radicals. Then we have

$$F = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_n \ni \alpha_1$$

with each extension being Galois and with a cyclic Galois group. From the above Proposition 34.1 we have that the Galois group of f over F_k is A_5 for all k. In particular, the Galois group of f over F_n is A_5. But this contradicts the assumption that $\alpha_1 \in F_n$. □

To get an explicit example of such a quintic we note the following.

Lemma 34.3

Let $f \in \mathbb{Q}[X]$ be an irreducible quintic with exactly 3 roots in $\mathbb{R}$. Then the Galois group of f is S_5.

Proof. Let the roots be $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ and $\alpha_4, \alpha_5 \in \mathbb{C} \setminus \mathbb{R}$. Let G be the Galois group of f over $\mathbb{Q}$. Since G acts transitively on the set of 5 roots, we have that $|G|$ is divisible by 5 (Orbit-Stabiliser relation). Therefore G contains an element of order 5 (Cauchy's Theorem). The only elements of order 5 in S_5 are the cycles of length 5. We have $\mathbb{Q}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5) \supseteq \mathbb{Q}(\alpha_1, \alpha_2, \alpha_3$ is a quadratic extension. Therefore there is an element in the Galois group of that extension that interchanges α_4 and α_5. Therefore G contains a transposition. Since G is a subgroup of S_5 that contains a 5-cycle and a transposition, we have $G = S_5$. □

Example 34.4. The polynomial $f = X^5 - 16X + 2 \in \mathbb{Q}[X]$ is irreducible and has exactly three roots in $\mathbb{R}$. Therefore, it is not soluble by radicals.

© University of Melbourne 2025