Algebra III*

Thomas Walker

Autumn 2023

Contents

*Text coloured blue is material from John Nicholson's notes for the same course taught in the autumn of 2022. It is not examinable for the course taught in the autumn of 2023, however they supplement the existing material.

1 Rings

1.1 Basic Theory

Definition 1.1.1. *A monoid $(M, \cdot)$ is a set M equipped with a binary operation $\cdot : M \times M \to M$ and an element $1_M \in M$, called the multiplicative identity, such that the following hold.*

- *$\cdot$ is associative, that is $x \cdot (y \cdot z) = (x \cdot y) \cdot z$ for all $x, y, z \in M$.*
- *$1_M \cdot m = m \cdot 1_M = m$ for all $m \in M$.*

A monoid M is commutative if $x \cdot y = y \cdot x$ for all $x, y \in M$.

Remark 1.1.2. *Any group is a monoid as associativity and the identity element are group axioms. A group is stronger than a monoid as it requires elements to have multiplicative inverses, whereas a monoid does not.*

Example 1.1.3.

- *The natural numbers $\mathbb{N} = \{1, 2, 3, \dots\}$ form a commutative monoid under multiplication, with the multiplicative identity being $1_{\mathbb{N}} = 1$.*
- *The non-negative integers $\mathbb{Z}_{\geq 0}$ form a commutative monoid under multiplication.*
- *The set of $n \times n$ real matrices $M_n(\mathbb{R})$ form a monoid under matrix multiplication, with the multiplicative identity being the identity matrix, I. For $n > 1$ the monoid $M_n(\mathbb{R})$ is not commutative.*

Definition 1.1.4. *A ring is a set R together with operations $+ : R \times R \to R$, $\cdot : R \times R \to R$, and elements $0_R, 1_R \in R$, such that the following hold.*

- *$(R, +)$ is an abelian group with identity 0_R.*
- *$(R, \cdot)$ is a monoid with multiplicative identity 1_R.*
- *The distributive properties $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$ hold for all $a, b, c \in R$.*

A ring R is commutative if $(R, \cdot)$ is a commutative monoid.

Remark 1.1.5.

- *$+$ is a function $R \times R \to R$ so it would make sense to write $+(x, y)$ for $x, y \in R$, but for sanity we will write $x + y$ for $+(x, y)$. Similarly, we will write $x \cdot y$ for $\cdot(x, y)$.*
- *We refer to $+$ as addition and $\cdot$ as multiplication.*
- *For $r \in R$, we write $-r$ for the additive inverse of r in the group $(R, +)$.*
- *We will often just write $1 = 1_R$ and $0 = 0_R$.*

Definition 1.1.6. *A subset $S \subseteq R$ is a subring if the following conditions are satisfied.*

- *$0_R, 1_R \in S$.*
- *For all $r, s \in S$ we have $-r, r + s, rs \in S$.*

Namely, S is a ring with operations $+$ and $\cdot$. We write $S \leq R$ to denote that S is a subring of R.

Example 1.1.7.

- *For the usual sets of numbers as rings with standard addition and multiplication, we have*

$$\mathbb{Z} \leq \mathbb{Q} \leq \mathbb{R} \leq \mathbb{C}.$$

  However, $\mathbb{N} \not\leq \mathbb{Z}$ *as* $-1 \notin \mathbb{N}$.

- *The set* $\mathbb{Z}/2\mathbb{Z}$ *with standard addition and multiplication is a ring.*
- *The Gaussian integers* $\mathbb{Z}[i] := \{a + ib : a, b \in \mathbb{Z}\}$ *is a subring of* $\mathbb{C}$.
- *The set* $\mathbb{Q}\left[\sqrt{2}\right] := \left\{a + b\sqrt{2} : a, b \in \mathbb{Z}\right\}$ *is a subring of* $\mathbb{R}$.

Proposition 1.1.8. *Let* R *be a ring and* $r \in R$. *Then* $r \cdot 0_R = 0_R \cdot r = 0_R$.

Proof. As 0_R is the identity element in the group $(R, +)$ we have that $0_R + 0_R = 0_R$. Hence,

$$\begin{aligned} r \cdot 0_R &= r \cdot (0_R + 0_R) \\ &= r \cdot 0_R + r \cdot 0_R. \end{aligned}$$

Adding $-(r \cdot 0_R)$ to both sides implies that $r \cdot 0_R = 0_R$. Similarly,

$$\begin{aligned} 0_R \cdot r &= (0_R + 0_R) \cdot r \\ &= 0_R \cdot r + 0_R \cdot r \end{aligned}$$

and so $0_R \cdot r = 0_R$. □

Example 1.1.9. *The trivial ring is the ring over the set* $R = \{0\}$ *where* $0 \cdot 0 = 0$ *and* $0 + 0 = 0$. *It is the only ring with a single element.*

Proposition 1.1.10. *Let* R *be a ring. Then* $1_R = 0_R$ *if and only if* $R = \{0\}$ *is the trivial ring.*

Proof. ($\Leftarrow$). We must have $1_R = 0_R$.
($\Rightarrow$). Let $r \in R$. Then

$$r = r \cdot 1_R = r \cdot 0_R = 0.$$

□

Throughout we will assume rings are non-trivial.

Definition 1.1.11. *An element* $u \in R$ *is a unit if there is another element* $v \in R$ *such that* $u \cdot v = v \cdot u = 1_R$. *We denote the set of units of* R *as* $R^\times \subseteq R$.

Remark 1.1.12. *With the notation of Definition 1.1.11 note that the element* v *is also a unit.*

Definition 1.1.13. *A ring* R *is a division ring if every non-zero element is a unit. That is,* $R^\times = R \backslash \{0\}$.

Definition 1.1.14. *A field is a commutative division ring.*

Example 1.1.15.

- $\mathbb{Z}$ *is not a field, as* 2 *is not invertible. In fact,* $\mathbb{Z}^\times = \{\pm 1\}$.
- $\mathbb{Q}$, $\mathbb{R}$ *and* $\mathbb{C}$ *are all fields.*
- $\mathbb{Z}[i]$ *is not a field as* 2 *is not invertible in* $\mathbb{Z}[i]$. *Indeed, suppose* $1 = 2(a+bi)$, *then by equating real and imaginary parts we have* $1 = 2a$ *and* $b = 0$. *No* $a, b \in \mathbb{Z}$ *satisfies these conditions, so* 2 *is not invertible.*
- $\mathbb{Q}\left[\sqrt{2}\right]$ *is a field.*
- *The quaternions* $\mathbb{H}$ *is the abelian group* $\mathbb{R}^4$ *with standard basis vectors labelled* $\{1, i, j, k\}$. *More specifically,* $\mathbb{H}$ *is a ring with* 1 *being the unit, and multiplication determined by* $ij = -ji = k$, $i^2 = j^2 = -1$, *and* $(r \cdot 1) \cdot s = rs$ *for any* $r \in \mathbb{R}$ *and* $s \in \mathbb{H}$. *These conditions are sufficient to define multiplication in* $\mathbb{H}$, *for instance*

$$\begin{aligned} k^2 &= (ij)(-ji) \\ &= -ijji \\ &= i^2 \\ &= -1. \end{aligned}$$

We similarly deduce that $jk = i$. *Note that* $\mathbb{H}$ *is not commutative since* $ij \neq ji$. *However, it is a division ring since*

$$(a1 + bi + cj + dk)(a - bi - cj - dk) = a^2 + b^2 + c^2 + d^2$$

and so if $a1 + bi + cj + dk \neq 0$ *it has an inverse*

$$\frac{1}{a^2 + b^2 + c^2 + d^2} \cdot (a - bi - cj - dk) \in \mathbb{H}.$$

Proposition 1.1.16. *Multiplicative inverses are unique. That is, if* $r \in R^\times$ *has inverses* u *and* v, *then* $u = v$.

Proof. By assumption, we have that $1 = ru = rv$ and so $r(u - v) = 0$. Since $ur = 1$ we get that

$$0 = ur(u - v) = u - v.$$

Therefore, $u = v$. □

Proposition 1.1.17. *For any ring* R, *the set of units,* $R^\times$, *is a group under multiplication.*

Proof. If $a, b \in R^\times$, then there are some $c, d \in R^\times$ such that $ac = ca = bd = db = 1$. In particular,

$$(ab)(dc) = a(bd)c = ac = 1,$$

and similarly $(dc)(ab) = 1$. Hence, $ab \in R^\times$ which means $R^\times$ is closed under multiplication. Moreover, the element $1 \in R^\times$ and $R^\times$ is closed under inverses by Remark 1.1.12. Therefore, $R^\times$ is a group under multiplication. □

1.2 Constructing Rings

Let R and S be rings. Then their product $R \times S$ is a ring with addition and multiplication given by

$$(r, s) + (r', s') := (r + r', s + s')$$

and

$$(r, s) \cdot (r', s') := (r \cdot r', s \cdot s').$$

Moreover, $0_{R\times S} = (0_R, 0_S)$ and $1_{R\times S} = (1_R, 1_S)$. Note that $R \times S$ is commutative if and only if both R and S are.

Definition 1.2.1. *Let R be a ring. Then a polynomial f with coefficients in R is a sequence $f = (a_0, a_1, a_2, \dots)$ in R which is eventually zero. That is, $a_i \neq 0$ for finitely many i.*

Remark 1.2.2. *If $a_i = 0$ for $i > N$ then we can write*

$$f = a_0 + a_1 X + a_2 X^2 + \ldots + a_N X^N.$$

The representation of f in this notation is not unique as

$$f = a_0 + a_1 X + a_2 X^2 + \ldots + a_N X^N + 0X^{N+1}.$$

Definition 1.2.3. *For f a non-zero polynomial with coefficients in R the degree of f is*

$$\deg(f) = \max\left(\left\{i : a^i \neq 0\right\}\right).$$

If $a_i = 0$ for each $i \in \mathbb{N}$, such that $f = 0$, then $\deg(f) := -\infty$.

Definition 1.2.4. *Let R be a ring. Then its polynomial ring, $R[X]$, is the set of polynomials with coefficients in R with the following operations. For $f = a_0 + \ldots a_n + X^n$ and $g = b_0 + \ldots b_m X^m$, where we can assume $n = m$ by adding copies of $0X^i$ if necessary, let*

$$f + g := (a_0 + b_0) + (a_1 + b_1) X + \ldots + (a_n + b_n) X^n$$

and

$$\begin{aligned} f \cdot g &:= (a_0 b_0) + (a_1 b_0 + a_0 b_1) X + (a_2 b_0 + a_1 b_1 + a_0 b_2) X^2 + \cdots + a_n b_n X^{2n} \\ &= \sum_{i=0}^{n+m} \left(\sum_{j=0}^{i} a_j b_{i-j} \right) X^i. \end{aligned}$$

The subset of constant polynomials, that is those with $a_i = 0$ for $i > 0$, form a subring which we can identify with R.

Exercise 1.2.5. *For a ring R, show that $R[X]$ is a ring which is commutative if and only if R is commutative.*

Remark 1.2.6. *We can take the polynomial ring of a polynomial ring, $(R[X])[Y]$. We write $R[X, Y] := (R[X])[Y]$ to conveniently iterate this notion to get*

$$R\left[X_1, \ldots, X_n\right] := \left(\ldots\left(\left(R\left[X_1\right]\right)\left[X_2\right]\right)\ldots\right)\left[X_n\right].$$

Definition 1.2.7. *If $f = a_0 + \ldots + a_n X^n$ is a polynomial of degree n, we say f is monic if $a_n = 1$.*

Example 1.2.8. *The polynomial $f = 2 + X^2$ is monic whereas $f = 2$ is not.*

A polynomial $f = a_0 + a_1 X + \cdots + a_n X^n \in R[X]$ can be thought of as determining a function $R \to R$, where

$$r \mapsto f(r) := a_0 + a_1 r + \cdots + a_n r^n.$$

However, the polynomial f itself is just a sequence of elements of R, and X is some formal symbol. Thus, we cannot identify the polynomial with the function. Different polynomials can have the same function. For

example, let $R = \mathbb{Z}/2\mathbb{Z}$, and consider $f = X \in R[X]$, $g = X^2 \in R[X]$. Then f and g are distinct polynomials, but $f(r) = g(r)$ for all $r \in R$.

Definition 1.2.9. *A Laurent polynomial f in R is a sequence $f = (\dots, a_{-1}, a_0, a_1, \dots)$, with only finitely many a_i non-zero. A Laurent polynomial can be written in the form*

$$f = \sum_{i \in \mathbb{Z}} a_i X^i.$$

We let $R\left[X, X^{-1}\right]$ denote the set of Laurent polynomials on R and impose the same polynomial ring structure as that on $R[X]$.

Definition 1.2.10. *A power series f in R is a sequence $f = (a_0, a_1, \dots)$ in R, where infinitely many a_i can be non-zero. We use $R[\![X]\!]$ to denote the ring of the power series in R.*

The ring structure on $R[\![X]\!]$ is the same as that on $R[X]$ and so $R[X] \leq R[\![X]\!]$.

Definition 1.2.11. *Let M be a monoid, and R a ring. The monoid ring of M over R denoted $R[M]$, is the set of tuples $f = (a_m)_{m \in M}$, where each $a_m \in R$, and only finitely many are non-zero. We write such a tuple f in the form*

$$f = \sum_{m \in M} a_m m.$$

For $f = \sum_{m \in M} a_m m$ and $g = \sum_{m \in M} b_m m$ in $R[M]$ we let

$$f + g := \sum_{m \in M} (a_m + b_m)\, m$$

and

$$f \cdot g := \sum_{m \in M} \left(\sum_{\substack{k,l \in M \\ k \cdot l = m}} a_k b_l \right) m.$$

Exercise 1.2.12. *Check that $R[M]$ is a ring, which is commutative if and only if both R and M are.*

If M is a group, then $R[M]$ is the group ring of M over R.

Example 1.2.13. *Let M be the monoid $\mathbb{Z}_{\geq 0}$ of non-negative integers, with addition as the binary operation. Then we can identify $R[X]$ with $R[M]$, by identifying aX^n with $a \cdot n$, where $a \in R$ and $n \in M$. If we take $M = \mathbb{Z}$, with addition, we can identify $R\left[X, X^{-1}\right]$ with $R[M]$ in a similar way.*

If R is a ring and $n \geq 1$, the set of $n \times n$ matrices $M_n(R)$ forms a ring under the usual rules for matrix addition and multiplication. When $n = 2$ the ring $M_n(R)$ is not commutative as

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

whereas

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

More generally, $M_n(R)$ is not commutative when $n \geq 2$. However, we may identify $M_1(R)$ with R and so $M_1(R)$ is commutative if and only if R is commutative.

1.3 Homomorphisms and Ideals

Definition 1.3.1. *Let R and S be rings. A function $\phi : R \to S$ is a ring homomorphism if for all $a, b \in R$ we have*

1. $\phi(a + b) = \phi(a) + \phi(b)$,
2. $\phi(ab) = \phi(a)\phi(b)$,
3. $\phi(0_R) = 0_S$, *and*
4. $\phi(1_R) = 1_S$.

A ring isomorphism is a ring homomorphism that is bijective.

Exercise 1.3.2. *Show that the inverse of an isomorphism is also an isomorphism.*

Definition 1.3.3. *Let $\phi : R \to S$ be a homomorphism.*

- *The kernel of ϕ is*
$$\ker(\phi) := \{r \in R : \phi(r) = 0\} \subseteq R.$$
- *The image of ϕ is*
$$\mathrm{im}(\phi) := \{s \in S : s = \phi(r) \text{ for some } r \in R\} \subseteq S.$$

Proposition 1.3.4. *Let $\phi : R \to S$ be a homomorphism. Then $\mathrm{im}(\phi) \subseteq S$ is a subring.*

Proof. As a ring homomorphism is a group homomorphism it follows that $(\mathrm{im}(\phi), +)$ is an abelian subgroup of S. For $s, s' \in \mathrm{im}(\phi)$ we have that $s = \phi(r)$ and $s' = \phi(r')$ for some $r, r' \in R$. Therefore,

$$s \cdot s' = \phi(r) \cdot \phi(r') = \phi(r \cdot r') \in \mathrm{im}(\phi).$$

Meaning, $(\mathrm{im}(\phi), \cdot)$ forms a monoid. Therefore, $\mathrm{im}(\phi) \leq S$. □

Remark 1.3.5. *If R is non-trivial then since $\phi(1) = 1$ we have that $1 \notin \ker(\phi)$ meaning $\ker(\phi)$ is not a subring of R.*

Proposition 1.3.6. *A homomorphism $\phi : R \to S$ is injective if and only if $\ker(\phi) = \{0\}$.*

Proof. ($\Rightarrow$). As $\phi(0) = 0$ it follows that $\phi(r) \neq 0$ for $r \in R \setminus \{0\}$ and so $\ker(\phi) = \{0\}$.
($\Leftarrow$). If $r, r' \in R$ are such that $\phi(r) = \phi(r')$, then $\phi(r - r') = 0$. Meaning $r - r' \in \ker(\phi)$ which implies that $r - r' = 0$ and so $r = r'$. Hence, ϕ is injective. □

Definition 1.3.7. *Let $I \subseteq R$.*

- *I is a left ideal if it is an additive subgroup of R such that for $i \in I$ and $r \in R$ we have $ri \in I$.*
- *I is a right ideal if it is an additive subgroup of R such that for $i \in I$ and $r \in R$ we have $ir \in I$.*
- *I is a two-sided ideal if it is an additive subgroup of R such that for $i \in I$ and $r \in R$, we have $ri \in I$ and $ir \in I$.*

We will normally use ideal to refer to a left ideal. Note that if R is commutative the different types of ideals coincide.

Example 1.3.8. *For any ring R the subsets $\{0\}$ and R are both ideals.*

Suppose $I \subseteq R$ is an ideal and $1 \in R$. Then for any $r \in R$ we have that $r \cdot 1 = r \in I$ which implies that $I = R$. Hence, the only subring which is also an ideal is the whole of R. We call an ideal $I \subseteq R$ such that $I \neq R$ a proper ideal. More generally, if $I \subseteq R$ is an ideal containing some unit u then $I = R$.

Lemma 1.3.9. *Let $\phi : R \to S$ be a homomorphism. Then $\ker(\phi) \subseteq R$ is a two-sided ideal.*

Proof. Since ϕ is a homomorphism of abelian groups, $\ker(\phi) \subseteq R$ is a subgroup of R. Now suppose $i \in \ker(\phi)$ and $r \in R$. Then

$$\phi(ri) = \phi(r)\phi(i) = \phi(r) \cdot 0 = 0,$$

so $ri \in \ker(\phi)$. Similarly, $ir \in \ker(\phi)$. Therefore, $\ker(\phi)$ is a two-sided ideal. □

Example 1.3.10. *Let $R = \mathbb{Z}$ and $n \in \mathbb{Z}$. Then $n\mathbb{Z} \subseteq \mathbb{Z}$ is an ideal. Conversely, suppose that I is an ideal. If $I = \{0\}$ then $I = 0\mathbb{Z}$. If I is not zero let $n \in I$ be its smallest positive element. Then $rn \in I$ for all $r \in \mathbb{Z}$ and so $n\mathbb{Z} \subseteq I$. For $i \in I$, using the Euclidean algorithm we can write*

$$i = an + b$$

for $0 \leq b < n$. As $n \in I$ we have that $an \in I$, and as $i \in I$ it follows that $b = i - an \in I$. Since n was the smallest positive element of I, we must have $b = 0$, and so $i \in n\mathbb{Z}$. We conclude that $I = n\mathbb{Z}$. Therefore, all ideals of $\mathbb{Z}$ are of the form $n\mathbb{Z}$ for some $n \in \mathbb{N}$.

Lemma 1.3.11. *Let $\phi : R \to S$ and $\psi : S \to T$ be homomorphisms of rings. Then their composition $\psi \circ \phi$ is also a homomorphism of rings.*

Definition 1.3.12.

1. *The ideal generated by an element $a \in R$ is*

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

2. *An ideal $I \subseteq R$ is principal if $I = (a)$ for some $a \in R$.*
3. *If $S \subseteq R$ is any subset, the ideal generated by S is*

$$(S) := R \cdot S = \left\{ \sum_{s \in S} r_s s : r_s \in R, \text{ with only finitely many } r_s \text{ are non-zero} \right\}.$$

Using Example 1.3.10 we can say that all the ideals of $\mathbb{Z}$ are principal.

Exercise 1.3.13. *Show that*

$$I := \{f \in R[X] : \text{the constant coefficient of } f \text{ is zero}\}$$

is an ideal of $R[X]$. Moreover, show that I is a principal ideal generated by the polynomial X.

Lemma 1.3.14. *Let R be a commutative ring, and consider $r \in R \setminus \{0\}$. Then $(r) = R$ if and only if $r \in R^\times$.*

Proof. ($\Rightarrow$). As $(r) = R$, there exists an $s \in R$ such that $sr = 1$. As R is commutative we also have that $rs = 1$, which implies that s is the inverse of r and so $r \in R^\times$.

($\Leftarrow$). Let $r \in R^\times$ with inverse s. Then $1 = sr \in (r)$. So for any $\tilde{r} \in R$ we have $\tilde{r} \cdot 1 = \tilde{r} \in (r)$. Therefore, $R \subseteq (r)$, implying that $R = (r)$. □

Definition 1.3.15. *Let $I \subseteq R$ be a two-sided ideal. The quotient ring R/I consists of additive cosets of the form $r + I$, with $0_{R/I} = 0_R + I$ and $1_{R/I} = 1_R + I$. Moreover, the operations are given by*

$$(r + I) + (s + I) := (r + s) + I$$

and

$$(r + I) \cdot (s + I) := (rs) + I.$$

Proposition 1.3.16. *The quotient ring R/I is a ring and the map $R \to R/I$ given by $r \mapsto r+I$ is a surjective ring homomorphism.*

Proof. Addition is well-defined as $(R/I, +)$ is the quotient of $(R, +)$ by a normal subgroup. Let $r, s \in R$, and let $i \in I$. Then $r + I = r + i + I$, so that

$$\begin{aligned}(r + i + I) \cdot (s + I) &= rs + is + I \\ &= rs + I.\end{aligned}$$

Similarly, $(r + I)(s + i + I) = (rs + I)$ meaning multiplication is well-defined. Associativity and distributivity follow from R. Moreover, $0 + I$ and $1 + I$ satisfy the requirements to be the additive and multiplicative identity respectively. □

Remark 1.3.17. *Proposition 1.3.16 motivates the definition of a two-sided, as a two-sided ideal provides the necessary conditions for R/I to be a well-defined ring.*

Example 1.3.18. *Consider the ideal $n\mathbb{Z} \subseteq \mathbb{Z}$. Elements of the quotient ring $\mathbb{Z}/n\mathbb{Z}$ are cosets of the form $r + n\mathbb{Z}$ for $r = 0, \ldots, n - 1$. Operations in $\mathbb{Z}/n\mathbb{Z}$ are addition and multiplication and addition modulo n.*

Exercise 1.3.19. *Let R be a ring and $\phi : G \to H$ a group homomorphism. Then ϕ induces a ring homomorphism $\phi_* : R[G] \to R[H]$, given by*

$$\phi_* \left(\sum_{g \in G} a_g g \right) := \sum_{g \in G} a_g \phi(g).$$

Note that $N = \ker(\phi) \trianglelefteq G$ is a normal subgroup of G. Consider the ideal $(N - 1) \subseteq R[G]$ generated by elements of $R[G]$ of the form $g - 1$ for $g \in N$. Then $\phi_*(g - 1) = 0$ for all $g \in N$ and so $(N - 1) \subseteq \ker(\phi_*)$. Show that $(N - 1)$ is a two-sided ideal, and $\ker(\phi_*) = (N - 1)$.

Example 1.3.20. *Let R be a commutative ring. Consider the ideal $(X) \subseteq R[X]$. Elements of this quotient ring $R[X]/(X)$ are of the form*

$$a_0 + a_1 X + \ldots + a_n X^n + (X).$$

Note that all terms apart from the constant term are in (X). More specifically, elements of $R[X]/(X)$ are uniquely represented as $a_0 + (X)$. Thus, there is an isomorphism $R \to R[X]/(X)$ given by $a \mapsto a + (X)$.

To understand more elaborate quotient rings, such as $\mathbb{R}[X]/\left(X^2 + 1\right)$, we can utilise Proposition 1.3.21.

Proposition 1.3.21. *Let F be a field and $f, g \in F[X]$, with $g \neq 0$. Then there exist $r, q \in F[X]$ such that*

$$f = gq + r$$

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

Proof. Let $\deg(f) = n$ and $\deg(g) = m$. Then if $m > n$ then we can choose $q = 0$ and $r = f$. Therefore, suppose that $m \leq n$ and $f = \sum_{i=0}^{n} a_i X^i$ with $a_n \neq 0$. We proceed by induction on n. Suppose that $g = \sum_{i=0}^{m} b_i X^i$ and let $q = a_n b_m^{-1} X^{n-m}$. It follows that

$$f = qg + f_1 \tag{1.3.1}$$

where $\deg(f_1) \leq n - 1$. If $n = m$ then $\deg(f_1) < m = \deg(g)$. Otherwise we can apply the induction hypothesis to f_1 to get that

$$f_1 = gq_1 + r_1$$

where $\deg(r_1) < \deg(g) = m$. Substituting this into (1.3.1) we see that

$$f = (q + q_1)\, g + r_1$$

completing the proof. □

Remark 1.3.22.

1. *Proposition 1.3.21 says that the Euclidean algorithm applies to polynomials over a field.*
2. *It is essential that F is a field. Indeed, if $F = \mathbb{Z}$ then with $f = 3X + 1$ and $g = 2X + 1$, one notes that q must be zero so that $r = 3X + 1$. As $\deg(f) = \deg(r)$ we see that the conclusion of Proposition 1.3.21 does not hold.*

Example 1.3.23. *Let $R = \mathbb{R}[X]/\left(X^2 + 1\right)$. Elements of R are of the form*

$$a_0 + a_1 X + \ldots + a_n X^n + \left(X^2 + 1\right).$$

Letting $f = a_0 + \ldots + a_n X^n$, we can apply the Euclidean algorithm to find $q, r \in R$ such that

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

with $\deg(r) < 2$. Which implies $r = b_0 + b_1 X$ for some $b_0, b_1 \in \mathbb{R}$. So any element of R is of the form $a + bX + \left(X^2 + 1\right)$ for some $a, b \in \mathbb{R}$. In particular, this representation is unique as if

$$a + bX + \left(X^2 + 1\right) = a' + b'X + \left(X^2 + 1\right),$$

then

$$(a - a') + (b - b')\, X = \left(X^2 + 1\right) g$$

for some $g \in R$. But if $g \neq 0$ then $\deg\left(\left(X^2 + 1\right) g\right) > 1$. So we must have $g = 0$ meaning $a = a'$ and $b = b'$. Therefore, every element of R is of the form $a + bX + \left(X^2 + 1\right)$. Note that $X + \left(X^2 + 1\right)$ squares to -1 and so the structure of the elements of R starts to resemble the complex numbers. Let $\phi : R \to \mathbb{C}$ be given by

$$\phi\left(a + bX + \left(X^2 + 1\right)\right) = a + bi \in \mathbb{C}.$$

The map ϕ is a well-defined, bijective ring homomorphism. Consequently, $R = \mathbb{R}[X]/\left(X^2 + 1\right) \cong \mathbb{C}$.

Theorem 1.3.24 (First Isomorphism Theorem for Rings)**.** *Let $\phi : R \to S$ be a ring homomorphism. Then*

$$R/\ker(\phi) \cong \mathrm{im}(\phi).$$

Proof. Let $\psi : R/\ker(\phi) \to \mathrm{im}(\phi)$ be given by

$$\psi(r + \ker(\phi)) = \phi(r).$$

This is well-defined, since if $r + \ker(\phi) = r' + \ker(\phi)$ then $r - r' \in \ker(\phi)$ and so $\phi(r) = \phi(r')$. Moreover, it is a ring homomorphism as ϕ is a ring homomorphism. As it is bijective we conclude that

$$R/\ker(\phi) \cong \mathrm{im}(\phi).$$

□

Example 1.3.25. *Let $\phi : \mathbb{R}[X] \to \mathbb{C}$ be the homomorphism given by $\phi(f) = f(i)$. One can check that ϕ is surjective and that $\ker(\phi) = \left(X^2 + 1\right)$. Therefore, by Theorem 1.3.24 we have that $\mathbb{C} \cong \mathbb{R}(X)/\left(X^2 + 1\right)$. Note how it is more efficient to establish this result using Theorem 1.3.24, than the approach used in Example 1.3.23.*

Exercise 1.3.26. *Let R be a ring with two-sided ideals $I, J \subseteq R$. Show that $I \cap J$ is a two-sided ideal. Moreover, show that $R/(I \cap J)$ is isomorphic to a subring of $R/I \times R/J$.*

Theorem 1.3.27 (Second Isomorphism Theorem for Rings)**.** *Let R and S be rings with $R \leq S$. Let $I \subseteq S$ be a two-sided ideal. Then the following hold.*

1. *$R + I := \{r + i : r \in R, i \in I\} \leq S$.*
2. *$I \subseteq R + I$ and $R \cap I \subseteq R$ are two-sided ideals.*
3. *$(R + I)/I = \{r + I : r \in R\} \leq S/I$, and*

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

Proof.

1. Since $R \subseteq R + I$, it follows that $0, 1 \in R + I$. Let $r, s \in R$ and $i, j \in I$, so that $r + i, s + j \in R + I$. Then
$$(r + i) + (s + j) = (r + s) + (i + j) \in R + I,$$
and
$$(r + i) \cdot (s + j) = (rs) + (is + rj + ij)$$
which is in $R + I$ as the second term is in I since I is a two-sided ideal. Therefore, $R + I \leq S$.
2. Note that, $I \subseteq R + I$ is a two-sided ideal since $I \subseteq S$ is a two-sided ideal. Let $\phi : R \to S/I$ be the homomorphism given by $\phi(r) = r + I$. Then $\ker(\phi)$ consists of elements $r \in R$ such that $r + I = I$, that is, $r \in I$. So $\ker(\phi) = R \cap I$ which means that $R \cap I$ is a two-sided ideal in R.
3. With ϕ as given above, we have that
$$\mathrm{im}(\phi) = \{r + I : r \in R\} = (R + I)/I \leq S/I.$$
On the other hand, using Theorem 1.3.24 we have
$$\mathrm{im}(\phi) \cong R/(R \cap I).$$
Therefore,
$$R/(R \cap I) \cong (R + I)/I.$$

□

Theorem 1.3.28 (Third Isomorphism Theorem for Rings)**.** *Let R be a ring, and $I, J \subseteq R$ two-sided ideals such that $I \subseteq J$. Then $J/I \subseteq R/I$ is a two-sided ideal and*

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

Proof. Let $\phi : R/I \to R/J$ be the homomorphism given by

$$\phi(r + I) := r + J.$$

This is a well-defined and surjective ring homomorphism. Note that $\ker(\phi)$ consists of elements $r+I \in R+I$ such that $r + J = 0$, that is, $r \in J$. Therefore, $\ker(\phi) = J/I$ and so we conclude by applying Theorem 1.3.24. □

Example 1.3.29. *Applying Theorem 1.3.28 to $R = \mathbb{Z}[X]$, $I = \left(X^2 + 1\right)$, and $J = \left(n, X^2 + 1\right)$ gives*

$$\left(\mathbb{Z}[X]/\left(X^2 + 1\right)\right)/(n) \cong \mathbb{Z}[X]/\left(n, X^2 + 1\right).$$

Therefore,

$$\mathbb{Z}[i]/(n) \cong \mathbb{Z}[X]/\left(n, X^2 + 1\right),$$

where we have used that $\mathbb{Z}[X]/\left(X^2 + 1\right) \cong \mathbb{Z}[i]$. On the other hand, applying Theorem 1.3.28 to $R = \mathbb{Z}[X]$, $I = (n)$, and $J = \left(n, X^2 + 1\right)$ we deduce that

$$(\mathbb{Z}[X]/(n))/\left(X^2 + 1\right) \cong \mathbb{Z}[X]/\left(n, X^2 + 1\right).$$

Therefore,

$$\mathbb{Z}[i]/(n) \cong (\mathbb{Z}/n\mathbb{Z})[X]/\left(X^2 + 1\right).$$

Proposition 1.3.30. *Let R be a ring, and $I \subseteq R$ a two-sided ideal. Then there is a bijection between the two-sided ideals of R/I and the two-sided ideals of R containing I.*

Proof. Let

$$\alpha : \{\text{two-sided ideals of } R/I\} \to \{\text{two-sided ideals of } R \text{ containing } I\}$$

be given by

$$\alpha(J) = \{r \in R : r + I \in J\}$$

and let

$$\beta : \{\text{two-sided ideals of } R \text{ containing } I\} \to \{\text{two-sided ideals of } R/I\}$$

be given by

$$\beta(K) = K/I.$$

These are well-defined maps. Moreover,

$$\begin{aligned}\alpha(\beta(K)) &= \alpha(K/I) \\ &= \{r \in R : r + I \in K/I\} \\ &= K,\end{aligned}$$

where the last equality follows as $I \subseteq K$. Similarly,

$$\begin{aligned}\beta(\alpha(J)) &= \beta(\{r \in R : r + I \in J\}) \\ &= \{r + I : r + I \in J\} \\ &= J.\end{aligned}$$

Therefore, α and β are bijections. □

Let R be any ring. Then there is unique ring homomorphism $\iota : \mathbb{Z} \to R$ given by

$$\iota(n) = \begin{cases} \underbrace{1_R + \ldots + 1_R}_{n} & n \geq 0 \\ -\underbrace{(1_R + \ldots + 1_R)}_{|n|} & n < 0. \end{cases}$$

This is a homomorphism by the distributivity property of R. It is unique, as any homomorphism $\mathbb{Z} \to R$ must send 1 to 1_R and thus from the structure of $\mathbb{Z}$ one finds that any such homomorphism must be equal to ι. As $\ker(\iota) \subseteq \mathbb{Z}$ is an ideal we have $\ker(\iota) = n\mathbb{Z}$ for some $n \in \mathbb{Z}$.

Definition 1.3.31. *The characteristic of R is the unique $n \geq 0$ such that $\ker(\iota) = n\mathbb{Z}$, where ι is the unique ring homomorphism $\mathbb{Z} \to R$.*

Example 1.3.32.

- *The ring $\mathbb{Q}$ has characteristic zero since $\iota : \mathbb{Z} \to \mathbb{Q}$ is injective which means that $\ker(\iota) = \{0\}$. Similarly, $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{Z}[i]$ have characteristic zero.*
- *For any $n \geq 1$, the homomorphism $\iota : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ sends r to $r + n\mathbb{Z}$ and has kernel $n\mathbb{Z}$. Meaning $\mathbb{Z}/n\mathbb{Z}$ has characteristic n.*

Consequently, we see that any $n \geq 0$ is the characteristic of some commutative ring.

1.4 Solution to Exercises

Exercise 1.2.5

Solution. ($\Rightarrow$). For $f = a_0 + \cdots + a_n X^n$ and $g = b_0 + \cdots + b_m X^m$ in $R[X]$ we have

$$\begin{aligned} f \cdot g &= \sum_{i=0}^{n+m} \left(\sum_{j=0}^{i} a_j b_{i-j} \right) X^i \\ &= \sum_{i=0}^{n+m} \left(\sum_{j=0}^{i} b_{i-j} a_j \right) X^i \\ &= \sum_{i=0}^{n+m} \left(\sum_{j=0}^{i} b_j a_{i-j} \right) X^i \\ &= g \cdot f. \end{aligned}$$

($\Leftarrow$). As any $r, s \in R$ can be viewed as elements in $R[X]$ we have that $r \cdot s = s \cdot r$. □

Exercise 1.2.12

Solution. This follows similar arguments to the proof of Exercise 1.2.5. □

Exercise 1.3.2

Solution. Let $\phi : R \to S$ be an isomorphism. In particular, ϕ is bijective and so $\phi^{-1} : S \to R$ is well-defined and bijective. It is clear that $\phi^{-1}(0_S) = 0_R$ and $\phi^{-1}(1_S) = 1_R$. Let $s_1, s_2 \in S$, then there exists $r_1, r_2 \in R$ such that $\phi(r_1) = s_1$ and $\phi(r_2) = s_2$. Consequently,

$$\begin{aligned} \phi^{-1}(s_1 + s_2) &= \phi^{-1}(\phi(r_1) + \phi(r_2)) \\ &= \phi^{-1}(\phi(r_1 + r_2)) \\ &= r_1 + r_2 \\ &= \phi^{-1}(s_1) + \phi^{-1}(s_2). \end{aligned}$$

Similarly,

$$\begin{aligned}\phi^{-1}(s_1 s_2) &= \phi^{-1}(\phi(r_1)\phi(r_2)) \\ &= \phi^{-1}(\phi(r_1 r_2)) \\ &= r_1 r_2 \\ &= \phi^{-1}(s_1)\phi^{-1}(s_2).\end{aligned}$$

Therefore, ϕ^{-1} is a homomorphism and thus an isomorphism. □

Exercise 1.3.13

Solution. Clearly $(X) \subseteq I$. For $f \in I$ we can write

$$f = a_1 X + \cdots + a_n X^n = \left(a_1 + \cdots + a_n X^{n-1}\right) X,$$

where $a_1 + \cdots + a_n X^{n-1} \in R[X]$ and so $f \in (X)$. Thus, we conclude that $I = (X)$. □

Exercise 1.3.19

Solution. Let $f \in (N-1)$ with $f = \sum_{i \in I} g_i(n_i - 1)$ where $g_i \in R[G]$ and $n_i \in N$ for each $i \in I$. Then

$$\phi_*(f) = \sum_{i \in I} \phi_*(g_i)\phi(n_i - 1) = \sum_{i \in I} \phi_*(g_i)(1-1) = 0.$$

Therefore, $f \in \ker(\phi_*)$ and thus $(N-1) \subseteq \ker(\phi_*)$. Now let $f \in \ker(\phi_*)$ and be of the form $f = \sum_{g \in G} a_g g$. Note that if $\phi(g) = \phi(g')$ then $\phi\left(g^{-1}g'\right) = e$ so that $g \in g'N$. Suppose that $G/N = (g_i N)_{i \in I}$, then we can write

$$\phi_*(f) = \sum_{i \in I} \left(\sum_{g \in g_i N} a_g \right) g_i.$$

As $\phi_*(f) = 0$ by assumption it must be the case that

$$\sum_{g \in g_i N} a_g = 0 \tag{1.4.1}$$

for each $i \in I$. For each $h \in g_i N$ we can write $h = g_i n_h$ for some $n_h \in N$. Therefore,

$$\begin{aligned} f &= \sum_{g \in G} a_g g \\ &= \sum_{i \in I} \left(\sum_{h \in g_i N} a_h g_i n_h \right) \\ &= \sum_{i \in I} \left(\sum_{h \in g_i N} a_h n_h \right) g_i \\ &\overset{(1.4.1)}{=} \sum_{i \in I} \left(\sum_{h \in g_i N} a_h (n_h - 1) \right) g_i \in (N-1). \end{aligned}$$

Therefore, $\ker(\phi_*) = (N-1)$. As ϕ_* is a homomorphism, it follows that $(N-1)$ is a two-sided ideal. □

Exercise 1.3.26

Solution. The intersection of additive groups is also an additive group and so $I \cap J$ is an additive group. Next, let $r \in R$ and $i \in I \cap J$, then $ri, ir \in I$ as I is a two-sided ideal and $ri, ir \in J$ as J is a two-sided ideal. Hence, $ri, ir \in I \cap J$ meaning $I \cap J$ is a two-sided ideal of R. Now consider the map $\varphi : R \to R/I \times R/J$ given by $\varphi(r) = (r + I, r + J)$. Recall, that $I = 0_{R/I}$ and $1_R + I = 1_{R/I}$ and similarly for J. In particular, we have that $0_{R/I \times R/J} = (I, J)$ and $1_{R/I \times R/J} = (1_R + I, 1_R + J)$. Using this we make the following observations.

- $\varphi(0_R) = (0_R + I, 0_R + J) = (I, J) = 0_{R/I \times R/J}$.
- $\varphi(1_R) = (1_R + I, 1_R + J) = 1_{R/I \times R/J}$.
- For $r, s \in R$ we have

$$\begin{aligned}\varphi(r+s) &= (r+s+I, r+s+J)\\ &= (r+I, r+J) + (s+I, s+J)\\ &= \varphi(r) + \varphi(s).\end{aligned}$$

- For $r, s \in R$ we have

$$\begin{aligned}\varphi(rs) &= (rs+I, rs+J)\\ &= \left(rs + rI + Is + I^2, rs + rJ + Js + J^2\right)\\ &= (r+I, r+J)(s+I, s+J)\\ &= \varphi(r)\varphi(s).\end{aligned}$$

Therefore, φ is a ring homomorphism. With $\varphi(r) = 0_{R/I \times R/J}$ if and only if $r + I = I$ and $r + J = J$ which happens if and only if $r \in I \cap J$. Therefore, $\ker(\varphi) = I \cap J$. Hence, using the Theorem 1.3.24 we conclude that

$$R/(I \cap J) \cong \operatorname{im}(\varphi),$$

where $\operatorname{im}(\varphi)$ is a subring of $R/I \times R/J$ by Proposition 1.3.4. □

2 Integral Domains

The integers as a ring have properties, such as unique factorisation, that we would like to study in a more general context. In this section, we will assume all rings are commutative and non-trivial.

2.1 Integral Domains and Ideals

Definition 2.1.1. *Let R be a commutative ring. An element $r \in R$ is a zero divisor if $r \neq 0$ and there is some $s \neq 0$ such that $rs = 0$.*

Definition 2.1.2. *A ring R is an integral domain if it contains no zero divisors. That is, if $rs = 0$ then either $r = 0$ or $s = 0$.*

Example 2.1.3.

- *The ring of integers $\mathbb{Z}$ is an integral domain.*
- *Any field is an integral domain. Indeed, if $rs = 0$ with $r \neq 0$, then letting r^{-1} be the multiplicative inverse of r we see that*
$$0 = r^{-1}rs = s.$$
Meaning no element of a field is a zero divisor.
- *The ring $\mathbb{Z}/6\mathbb{Z}$ is not an integral domain, since $2 + 6\mathbb{Z}$ and $3 + 6\mathbb{Z}$ are non-zero but their product is*
$$(2 + 6\mathbb{Z})(3 + 6\mathbb{Z}) = \mathbb{Z}.$$

Lemma 2.1.4. *Let $n \geq 1$. Then $\mathbb{Z}/n\mathbb{Z}$ is an integral domain if and only if n is prime.*

Proof. ($\Rightarrow$). Suppose $n \mid rs$ for some $r, s \in \mathbb{Z}$. Then
$$(r + n\mathbb{Z})(s + n\mathbb{Z}) = n\mathbb{Z}.$$
Since $\mathbb{Z}/n\mathbb{Z}$ is an integral domain, either $r + n\mathbb{Z} = n\mathbb{Z}$ or $s + n\mathbb{Z} = n\mathbb{Z}$. Hence, $n \mid r$ or $n \mid s$ meaning n is prime. ($\Leftarrow$). Suppose that $(r + n\mathbb{Z})(s + n\mathbb{Z}) = 0$. Then $rs + n\mathbb{Z} = 0$ meaning $n \mid rs$. Which implies that $n \mid r$ or $n \mid s$ and so $r + n\mathbb{Z} = 0$ or $s + n\mathbb{Z} = 0$. □

Example 2.1.5. *If R is an integral domain and $S \leq R$, then S is also an integral domain since a zero divisor in S would be a zero divisor in R. Therefore, as $\mathbb{C}$ is an integral domain we deduce that $\mathbb{Z}[i] \leq \mathbb{C}$ is an integral domain.*

Lemma 2.1.6. *Let R be an integral domain. Then $R[X]$ is also an integral domain.*

Proof. Let $f, g \in R[X]$ be non-zero with
$$f = a_0 + \ldots + a_n X^n$$
and
$$g = b_0 + \ldots + b_m X^m$$
for $a_n, b_m \in R \setminus \{0\}$. Then $n = \deg(f)$ and $m = \deg(g)$. Moreover, the coefficient of X^{n+m} in fg is $a_n b_m \neq 0$ since R is an integral domain. Therefore, $\deg(fg) > 0$ meaning $fg \neq 0$. □

Remark 2.1.7. *Iterating Lemma 2.1.6 we see that if R is an integral domain, then $R[X_1, \ldots, X_n]$ is an integral domain for all $n \in \mathbb{N}$.*

Lemma 2.1.8. *A non-trivial, commutative ring R is a field if and only if its only ideals are $\{0\}$ and R.*

Proof. ($\Rightarrow$). Suppose R is a field and $I \subseteq R$ is a non-zero ideal, such that it contains some non-zero element r. Then since R is a field, r is a unit, which implies $I = R$.
($\Leftarrow$). Let $r \in R$ be non-zero. Then $(r) \subseteq R$ is a non-zero ideal, and so must be R by assumption. Consequently, there is some s such that $rs = 1$, and so r is a unit. Therefore, every non-zero $r \in R$ is a unit, or in other words, R is a field. □

Remark 2.1.9. *Since $\{0\}$ and R are always ideals in R, the statement of Lemma 2.1.8 is equivalent to saying that R only has two ideals.*

Exercise 2.1.10. *Suppose R is a finite and commutative integral domain. Show that R is a field.*

Definition 2.1.11. *A proper ideal $I \subseteq R$ is maximal if any proper ideal $J \subseteq R$ containing I is equal to I.*

Lemma 2.1.12. *A proper ideal $I \subseteq R$ is maximal if and only if R/I is a field.*

Proof. ($\Rightarrow$). Suppose I is maximal. Then the only ideals in R containing I are I and R. As we have a bijection between such ideals and ideals in R/I, we see that the only ideals in R/I are $\{0\}$ and R/I. Therefore, R/I is a field by Lemma 2.1.8.
($\Leftarrow$). Suppose R/I is a field, and suppose $J \subseteq R$ is a proper ideal containing I. Then $J/I \subseteq R/I$ is a proper ideal of a field, and so $J = \{0\}$. Therefore, $J = I$ meaning I is a maximal ideal. □

Definition 2.1.13. *A proper ideal $I \subseteq R$ is prime if whenever there are $r, s \in R$ such that $rs \in I$, we either have $r \in I$ or $s \in I$.*

Example 2.1.14. *Let $n \in \mathbb{Z}$ be non-zero. Then the ideal $n\mathbb{Z} \subseteq \mathbb{Z}$ is prime if and only if n is prime. Indeed, suppose n is prime. If $rs \in n\mathbb{Z}$, then $n \mid rs$. So $n \mid r$ or $n \mid s$, that is, $r \in n\mathbb{Z}$ or $s \in n\mathbb{Z}$. Conversely, suppose $n\mathbb{Z}$ is prime, and for contradiction that $n = uv$ where $u, v \notin \{0, \pm 1\}$. Then $uv \in n\mathbb{Z}$ but $u, v \notin n\mathbb{Z}$, since $0 < |u|, |v| < |n|$. This contradicts $n\mathbb{Z}$ being prime and so n itself must be prime.*

Lemma 2.1.15. *A proper ideal $I \subseteq R$ is prime if and only if R/I is an integral domain.*

Proof. ($\Rightarrow$). Suppose that $I \subseteq R$ is prime and $r + I, s + I \in R/I$ are such that $(r + I)(s + I) = 0 + I$. This means $rs \in I$. Since I is prime, one of r and s is in I meaning one of $r + I$ and $s + I$ is zero.
($\Leftarrow$). Suppose R/I is an integral domain, and $r, s \in R$ are such that $rs \in I$. Then $(r + I)(s + I) = 0 + I$, so $r + I$ or $s + I$ is I. Meaning r or s is in I. □

Corollary 2.1.16. *If $I \subseteq R$ is a maximal ideal then it is a prime ideal.*

Proof. Since I is maximal we have that R/I is a field thus an integral domain. Therefore, I is a prime ideal. □

Remark 2.1.17. *The converse of Corollary 2.1.16 is not true in general. Take $(0) \subseteq \mathbb{Z}$, which is a prime ideal that is not maximal. However, if R is a finite, commutative ring, then for an ideal I we have that R/I is a finite, commutative ring. Therefore, if I is prime we additionally have that R/I is an integral domain. Thus, using Exercise 2.1.10 we deduce that R/I is a field and hence I is also a maximal ideal.*

Example 2.1.18. *Let R be a ring. Then $R[X]/(X) \cong R$ and so the ideal $(X) \subseteq R[X]$ is prime if and only if R is an integral domain, and it is maximal if and only if R is a field.*

Lemma 2.1.19. *Let R be an integral domain. Then its characteristic is either zero or prime.*

Proof. Let $\iota : \mathbb{Z} \to R$ be the unique ring homomorphism, with $\ker(\iota) = n\mathbb{Z}$, where $n \geq 0$ is the characteristic of R. Then by Theorem 1.3.24 we have that

$$\operatorname{im}(\iota) \cong \mathbb{Z}/n\mathbb{Z}.$$

Since R is an integral domain and $\operatorname{im}(\iota) \leq R$ it follows that $\mathbb{Z}/n\mathbb{Z}$ is an integral domain. Therefore, using Lemma 2.1.4, if $n > 0$ then n is prime. As $\mathbb{Z}$ is an integral domain we can also have $n = 0$. □

2.2 Factorisation

Throughout this subsection, rings will be integral domains.

Definition 2.2.1. *Let R be an integral domain with $r, s \in R$.*

- *We say that r divides s, written $r \mid s$, if there is some $u \in R$ such that $s = ru$. Equivalently, $(s) \subseteq (r)$.*
- *We say r and s are associates if there is some unit $u \in R^\times$ such that $s = ru$. Equivalently, $(r) = (s)$ or $r \mid s$ and $s \mid r$.*

Example 2.2.2. *Elements $r, s \in \mathbb{Z}$ are associates if and only if $r = \pm s$. However, this is not true in general as $2i, 2 \in \mathbb{Z}[i]$ are associates.*

Definition 2.2.3. *Let R be an integral domain. An element $r \in R$ is irreducible if the following statements hold.*

1. *$r \neq 0$.*
2. *r is not a unit.*
3. *If $r = uv$ then u or v is a unit.*

Definition 2.2.4. *Let R be an integral domain. An element $r \in R$ is prime if the following statements hold.*

1. *$r \neq 0$.*
2. *r is not a unit.*
3. *Whenever $r \mid uv$, either $r \mid u$, $r \mid v$, or both.*

Example 2.2.5. *An element being prime or irreducible is dependent on the ring and not just the element.*

- *Observe that 2 is prime in $\mathbb{Z}$, but not in $\mathbb{Q}$. In $\mathbb{Q}$ it is a unit.*

- *The polynomial $2X$ is irreducible in $\mathbb{Q}[X]$, but not in $\mathbb{Z}[X]$.*

In $\mathbb{Z}$ primality and irreducibility coincide, but this is not the case for general rings. Moreover, any $n \in \mathbb{Z}$ has an essentially unique prime factorisation, up to reordering and signs, but again this is not true for general rings.

Example 2.2.6. *Consider the ring*

$$R = \mathbb{Z}\left[\sqrt{-5}\right] = \left\{a + b\sqrt{-5} : a, b \in \mathbb{Z}\right\}.$$

This is a subring of $\mathbb{C}$ so is an integral domain. Consider $3 \in R$. Note that 2 and 3 both divide 6. However, $6 = (1 - \sqrt{-5})(1 + \sqrt{-5})$ with neither 2 nor 3 dividing either of the factors on the right-hand side. Therefore, 2 and 3 cannot be prime. Now suppose that $3 = \left(a + b\sqrt{-5}\right)\left(c + d\sqrt{-5}\right)$ for $a, b, u, v \in \mathbb{Z}$. Applying $|\cdot|^2$ to both sides, we see that

$$9 = \left(a^2 + 5b^2\right)\left(c^2 + 5d^2\right).$$

The only solutions to which are $a + b\sqrt{-5} = \pm 3$ and $c + d\sqrt{-5} = \pm 1$, or the other way around. Hence, 3 is irreducible. A similar argument shows that 2 and $1 \pm \sqrt{-5}$ are irreducible. So 6 can be factored into irreducibles in distinct ways.

Lemma 2.2.7. *For R an integral domain, the principal ideal (r) is a prime ideal if and only if $r = 0$ or r is prime.*

Proof. ($\Rightarrow$). Suppose (r) is a prime ideal. If $r = 0$, then we are done. So assume $r \neq 0$. Since prime ideals are proper ideals, r cannot be a unit. If $r \mid uv$ then $uv \in (r)$. Since (r) is prime we have $u \in (r)$ or $v \in (r)$. That is, $r \mid u$ or $r \mid v$. Hence, r is prime.
($\Leftarrow$). If $r = 0$, then the ideal $(r) = \{0\}$ is prime since R is an integral domain. Suppose $r \neq 0$ is prime. If $uv \in (r)$, this means $r \mid uv$ and so $r \mid u$ or $r \mid v$. Meaning $u \in (r)$ or $v \in (r)$. Hence, (r) is prime. □

Lemma 2.2.8. *For an integral domain R, if $r \in R$ is prime then r is irreducible.*

Proof. Let $r \in R$ be prime, and suppose $r = uv$. As $r \mid uv$ and r is prime, we have $r \mid u$ or $r \mid v$. Without loss of generality assume $r \mid u$. Then there is some $s \in R$ such that $rs = u$. Hence, $r = uv = r(sv)$. Since R is an integral domain, this implies $sv = 1$ meaning v is a unit and thus r is irreducible. □

In Example 2.2.6 we saw that $3 \in \mathbb{Z}\left[\sqrt{-5}\right]$ was irreducible but not prime, hence, the converse of Lemma 2.2.8 is not true.

Definition 2.2.9. *An integral domain R is a Euclidean domain if there is a function $\theta : R \backslash \{0\} \to \mathbb{Z}_{\geq 0}$, called a Euclidean function, that satisfies the following.*

- *$\theta(rs) \geq \theta(r)$ for all $r, s \in R \setminus \{0\}$.*
- *For all $a, b \in R$ with $b \neq 0$, there are $q, r \in R$ such that*

$$a = qb + r$$

with $r = 0$ or $\theta(r) < \theta(b)$.

Example 2.2.10.

1. *$\mathbb{Z}$ is a Euclidean domain with $\theta(n) = |n|$.*
2. *If F is a field, then $F[X]$ with $\theta(f) = \deg(f)$ is a Euclidean domain.*
3. *If F is a field, then F with $\theta(r) = 0$ for all $r \in F$ is a Euclidean domain.*

Exercise 2.2.11. *Show that the converse to statement 3 of Example 2.2.10 is true. Namely, an integral domain R that is a Euclidean domain with Euclidean function $r \mapsto 0$ is a field.*

Proposition 2.2.12. *The Gaussian integers $\mathbb{Z}[i]$ is a Euclidean domain with $\theta(r) = |r|^2$.*

Proof. Note that $\theta(rs) = \theta(r)\theta(s)$. Moreover, if $r \in \mathbb{Z}[i] \setminus \{0\}$ then $\theta(r) \geq 1$. So if $r, s \neq 0$ we have that $\theta(rs) \geq \theta(r)$. Now let $a, b \in \mathbb{Z}[i]$ with $b \neq 0$. Choose $q \in \mathbb{Z}[i]$ such that $\left|\frac{a}{b} - q\right| < 1$. We can do this as every complex number has a distance of at most one from a Gaussian integer. In particular, we can write

$$\frac{a}{b} = q + c$$

where $c \in \mathbb{Q}[i]$ is such that $|c| < 1$. Multiplying by b and setting $r = bc = a - bq \in \mathbb{Z}[i]$, we have

$$a = qb + r.$$

Since $|c| < 1$ and $r = bc$ we either have $r = 0$ or $\theta(r) = |r|^2 < \theta(b)$. □

This strategy of proof for Proposition 2.2.12 works for any $R \leq \mathbb{C}$ where for any $r \in \mathbb{C}$, there is a point in R of distance less than one from r. This does not hold for $\mathbb{Z}\left[\sqrt{-5}\right]$, and in fact, $\mathbb{Z}\left[-\sqrt{5}\right]$ is not a Euclidean domain.

Exercise 2.2.13. *Show that $\mathbb{Z}\left[\sqrt{2}\right]$ is a Euclidean domain.*

Definition 2.2.14. *Let R be an integral domain. Then R is a principal ideal domain if every ideal is principal. That is, for any ideal $I \subseteq R$ we have $I = (r)$ for some $r \in R$.*

We have already seen in Example 1.3.10 that $\mathbb{Z}$ is a principal ideal domain.

Theorem 2.2.15. *Let R be a Euclidean domain. Then R is a principal ideal domain.*

Proof. Let θ be the Euclidean function for R. Let $I \subseteq R$ be an ideal, which we can assume to be non-zero. Choose $b \in I \setminus \{0\}$ such that $\theta(b)$ is minimised, and consider $a \in I$. Then there are $q, r \in R$ such that $a = qb + r$ with $r = 0$ or $\theta(r) < \theta(b)$. Since I is an ideal and $a, b \in I$ we observe that $r = a - qb \in I$. We cannot have $\theta(r) < \theta(b)$ and $r \neq 0$, by construction of b, so $r = 0$. Hence, $a = qb$ meaning $a \in (b)$. Since $a \in I$ was arbitrary we have that $I = (b)$. □

Example 2.2.16. *Recall that for a field F, the polynomial ring $F[X]$ is a Euclidean domain. Therefore, by Theorem 2.2.15 we have that $F[X]$ is a principal ideal domain. Similarly, we deduce that $\mathbb{Z}[i]$ is a principal ideal domain.*

Proposition 2.2.17. *The ideal $(2, X) \subseteq \mathbb{Z}[X]$ is not generated by a single element.*

Proof. Suppose there is some $f \in \mathbb{Z}[X]$ such that $(2, X) = (f)$. Then as $2 \in (f)$, there is some $g \in \mathbb{Z}[X]$ such that $fg = 2$. So f must have degree zero, which implies f is ± 1 or ± 2. We know $f \neq \pm 1$ as $1 \notin (2, X)$. Similarly $f \neq \pm 2$ since 2 does not divide X. Therefore, $(2, X)$ cannot be a principal ideal. □

Remark 2.2.18. *Proposition 2.2.17 tells us that $\mathbb{Z}[X]$ is not a principal ideal domain, and hence not a Euclidean domain.*

Example 2.2.19. *Let F be a field and $A \in M_n(F)$. Consider the set*

$$I := \{f \in F[X] : f(A) = 0\} \subseteq F[X].$$

If $f, g \in I$ then $(f+g)(A) = f(A) + g(A) = 0$. If $f \in I$ and $h \in F[X]$ then $(fh)(A) = f(A)h(A) = 0$. Therefore, I is an ideal. Since $F[X]$ is a principal ideal domain we have that $I = (m)$ for some $m \in F[X]$. In particular, $m(A) = 0$ and for any $f \in F[X]$ such that $f(A) = 0$ we have $m \mid f$. The polynomial m is called the minimal polynomial of A.

Definition 2.2.20. *An integral domain R is a unique factorisation domain if the following statements hold.*

- *Every non-unit $r \in R \setminus \{0\}$ is a product of irreducibles.*
- *If $p_1 \dots p_n = q_1 \dots q_m$ with each $p_i, q_j \in R$ irreducible, then $n = m$. In particular, they can be reordered such that each p_i is an associate of q_i.*

A unique factorisation domain ensures the factorisation of elements into irreducibles exists and is unique up to associates.

Lemma 2.2.21. *Let R be a principal ideal domain. Then a non-zero principal ideal $(r) \subseteq R$ is maximal if and only if r is irreducible, or $r = 0$ when R is a field.*

Proof. ($\Rightarrow$). Suppose (r) is a maximal ideal with r not being irreducible. If $r = 0$, then as $R/(r)$ is a field we have that R is a field. Otherwise, we can assume that $r = xy$ for $x, y \in R$ non-units such that $(r) \subseteq (x) \subseteq R$. Since x is not a unit we have that $(x) \neq R$. Since (r) is maximal it follows that $(r) = (x)$. Meaning $r = xz$ for some unit $z \in R^\times$. Hence, $r = xz = xy$ which implies that $x(z - y) = 0$ and so $z = y$, by using the fact that $x \neq 0$ and R is an integral domain. However, y is not a unit whereas z is and thus we arrive at a contradiction. ($\Leftarrow$). If $r = 0$ and R is a field, then (r) is maximal. So we instead assume r is irreducible and that (r) is not maximal. Meaning there exists some proper ideal $(r) \subseteq I \subseteq R$ such that $I \neq (r)$. Since R is a principal ideal domain we have $I = (s)$ for some non-unit $s \in R$. Since $(r) \subseteq (s)$ we have that $r = sz$ for some non-unit $z \in R$, which contradicts the irreducibility of r. □

Lemma 2.2.22. *Let R be a principal ideal domain. If $r \in R$ is irreducible then it is prime.*

Proof. If $r \in R$ is irreducible, then $(r) \subseteq R$ is a maximal ideal, by Lemma 2.2.21. Therefore, it is also a prime ideal. Since the ideal (r) is prime and $r \neq 0$, Lemma 2.2.7 implies that r is prime. □

Remark 2.2.23. *Primes are irreducible in any integral domain. In Example 2.2.6 we showed that 2 was irreducible but not prime. So Lemma 2.2.22 shows $\mathbb{Z}\left[\sqrt{-5}\right]$ cannot be a principal ideal domain.*

Corollary 2.2.24. *Let R be a principal ideal domain. Then every non-zero prime ideal is maximal.*

Proof. Let $I \subseteq R$ be a non-zero prime ideal. Since R is a principal ideal domain, we know $I = (r)$ for some $r \in R$. Since $I \neq 0$ is a prime ideal, we know r is prime by Lemma 2.2.7. Therefore, r is irreducible as all primes are irreducible in integral domains. Therefore, by Lemma 2.2.21 (r) is maximal. □

In general rings, maximal ideals are prime. Therefore, Corollary 2.2.24 says that in a principal ideal domain, non-zero prime ideals and maximal ideals are equivalent. Except for the zero ideal which is always prime if R is a principal ideal domain, but maximal if and only if R is a field.

Exercise 2.2.25. *Let R be an integral domain. Suppose that $(I_n)_{n \in \mathbb{N}}$ is a family of increasing ideals of R.*

Show that

$$I = \bigcup_{n\in\mathbb{N}} I_n \subseteq R$$

is an ideal of R.

Proposition 2.2.26. *Let R be a principal ideal domain, and $I_1 \subseteq I_2 \subseteq \cdots \subseteq R$ an increasing sequence of ideals in R. Then $(I_n)_{n\in\mathbb{N}}$ is eventually constant. That is, there is some integer N such that $I_n = I_{n+1}$ for all $n \geq N$.*

Proof. Let $I = \bigcup_{n\in\mathbb{N}} I_n \subseteq R$. Using Exercise 2.2.25 we have that I is an ideal of R. Since R is a principal ideal domain it follows that $I = (r)$ for some $r \in I = \bigcup_{n\in\mathbb{N}} I_n$. Hence, there is some $N \in \mathbb{N}$ such that $r \in I_N$. Then,

$$(r) \subseteq I_N \subseteq I_{N+1} \subseteq \cdots \subseteq I = (r).$$

which implies that for $n \geq N$ we have $I_n = (r)$. □

Example 2.2.27. *Let $I_n = \left(2^n, 2^{n-1}X, \ldots, 2X^{n-1}, X^n\right)$. Note that I_n contains polynomials such that*

- *the X^0 coefficient is divisible, by 2^n,*
- *the X^1 coefficient is divisible by 2^{n-1},*
- *$\vdots$*
- *the X^{n-1} coefficient is divisible by 2, and*
- *higher-order coefficients being arbitrary.*

Note that $I_{n+1} \subseteq I_n$ as ideals. By the third isomorphism theorem $I_n/I_{n+1} \subseteq \mathbb{Z}[X]/I_{n+1}$. Let $f \in I_n/I_{n+1}$ where $f = a_0 + a_1X + \cdots + a_nX^n + \ldots$, such that $2^n|a_0 \ldots, 2|a_{n-1}$. Let $\tilde{a}_k$ be the representation of a_k in I_n/I_{n+1}. Observe that for $k = 0, \ldots, n$ if $2^{n-k+1}|a_k$ then $\tilde{a}_k = 0$, otherwise $\tilde{a}_k = 2^{n-k}$. Consider the map $\varphi : I_n \to (\mathbb{Z}/2\mathbb{Z})^{n+1}$ where

$$f \mapsto \left(1 - \mathbf{1}_{2^{n+1}|a_0}, \ldots, 1 - \mathbf{1}_{2|a_n}\right).$$

Note $\varphi(f) = 0$ if and only if $2^{n-k+1}|a_k$ for each $k = 0, \ldots, n$, which happens if and only if $f \in I_{n+1}$. Clearly, the map is surjective. Therefore,

$$I_n/I_{n+1} \cong (\mathbb{Z}/2\mathbb{Z})^{n+1}.$$

Observe that $(\mathbb{Z}/2\mathbb{Z})^{n+1}$ has $n+1$ generators. Therefore, I_n/I_{n+1} also has $n+1$ generators which implies that I_n has $n+1$ generators. In particular, this means that the increasing sequence of ideals $(I_n)_{n\in\mathbb{N}} \subseteq \mathbb{Z}[X]$ is not eventually constant. Indeed, from Proposition 2.2.17 we know that $\mathbb{Z}[X]$ is not a principal ideal domain, and so this example does not contradict Proposition 2.2.26.

Remark 2.2.28. *A ring R with the property that any increasing sequence of ideals is eventually constant is referred to as a Noetherian ring. Thus, Proposition 2.2.26 says that every principal ideal domain is a Noetherian ring.*

Theorem 2.2.29. *Suppose R is a principal ideal domain. Then R is a unique factorisation domain.*

Proof. Consider statement 2 of Definition 2.2.20. Suppose $p_1 \ldots p_n = q_1 \ldots q_m$, where each $p_i, q_j \in R$ is irreducible. Without loss of generality suppose that $n \geq m$. In particular, $p_1 \mid q_1 \ldots q_m$. Since p_1 is irreducible, it is prime and so $p_1 \mid q_j$ for some $j \in \{1, \ldots, m\}$. Reordering if necessary, we can assume that $p_1 \mid q_1$, so $ap_1 = q_1$

for some a. Since q_1 is irreducible, a must be a unit meaning p_1 and q_1 associates. Since R is an integral domain, we can cancel p_1 to give

$$p_2 \dots p_n = (aq_2)\, q_3 \dots q_m,$$

and replacing q_2 with aq_2 we get

$$p_2 \dots p_n = q_2 \dots q_m.$$

Repeating the process m times, we find that, possibly after reordering, p_i and q_i are associates for each $i = 1, \dots, m$ and

$$p_{m+1} \dots p_n = 1.$$

If $n > m$ then this would imply that p_n is a unit giving a contradiction. Therefore, $n = m$. Now consider statement 1 of Definition 2.2.20. Suppose $r \in R$ is a non-unit that cannot be factored as a product of irreducibles. In particular, r cannot be irreducible. So we can write $r = r_1 s_1$, with $r_1, s_1 \in R$ non-units. Since r is not a product of irreducibles, at least one of r_1 and s_1 cannot be a product of irreducibles either. Without loss of generality suppose that r_1 cannot be factored into irreducibles. Then we can write $r_1 = r_2 s_2$ with $r_2, s_2 \in R$ non-units. Again, suppose that r_2 is not a product of irreducibles so that we can write $r_2 = r_3 s_3$ with $r_3, s_3 \in R$ non-units. By assumption, we can continue this process indefinitely. Consequently, we can construct an increasing sequence of ideals,

$$(r) \subseteq (r_1) \subseteq (r_2) \subseteq \dots \subseteq R.$$

Since R is a principal ideal domain we can apply Proposition 2.2.26 to find an $N \in \mathbb{N}$ such that

$$(r_N) = (r_{N+1}) = (r_{N+2}) = \dots.$$

Since, $(r_N) = (r_{N+1})$ and $r_N = r_{N+1} s_{N+1}$ we have that s_{N+1} must be a unit, which is a contradiction. So r must be a product of irreducibles. Therefore, we can conclude that R is a unique factorisation domain. □

Example 2.2.30. *Recall, that $\mathbb{Z}[i]$ is a Euclidean domain meaning it is a principal ideal domain and thus a unique factorisation domain.*

Definition 2.2.31. *Let R be a ring. An element $d \in R$ is a greatest common divisor of a finite sequence $a_1, \dots, a_n \in R$ if $d \mid a_i$ for each $i = 1, \dots, n$, and if for any other $d' \in R$ such that $d' \mid a_i$ for each $i = 1, \dots, n$ it follows that $d' \mid d$.*

In general, the greatest common divisor may not exist, however, if it does exist then it is unique up to multiplication by units.

Proposition 2.2.32. *Let R be a unique factorisation domain. Then any $a_1, \dots, a_n \in R$ that are not all zero have a greatest common divisor d, and any other greatest common divisor d' is an associate of d.*

Proof. Let $p_1, \dots, p_m \in R$ be a collection of irreducibles in R such that any irreducible factor of any a_i is an associate of some p_j. Moreover, for $i \neq j$ the irreducibles p_i and p_j are not associates. Consequently, for each $i = 1, \dots, n$ we can write

$$a_i = u_i \prod_{j=1}^{m} p_j^{k_{ij}}$$

where each $k_{ij} \in \mathbb{Z}_{\geq 0}$ and each u_i is a unit. For each $j = 1, \dots, m$ let $l_j = \min_{i \in \{1,\dots,n\}} (k_{ij})$ and let

$$d := \prod_{j=1}^{m} p_j^{l_j}.$$

Since $l_j \leq k_{ij}$ for each $i = 1, \dots, n$, it follows that $d \mid a_i$ for each $i = 1, \dots, n$. Suppose $d' \in R$ also satisfies $d' \mid a_i$ for each $i = 1, \dots, n$. Then we can write

$$d' = v \prod_{j=1}^{m} p_j^{c_j}$$

for some unit v, and some $c_j \in \mathbb{Z}_{\geq 0}$. Since $d' \mid a_i$ for each $i = 1, \ldots, n$ we must have $c_j \leq l_j$ for each $i = 1, \ldots, m$. So $d' \mid d$, and hence d is a greatest common divisor. Suppose d and d' are both greatest common divisors of $a_1, \ldots, a_n$. Then they must divide each other and hence are associates. □

We can summarise the main relationships we have established so far with the following hierarchy. Where implications go down the order.

1. $\mathbb{Z}$.
2. Euclidean domains.
3. Principal ideal domains.
4. Unique factorisation domains.
5. Integral domains.
6. Commutative rings.
7. Rings.

However, we have no implications going up the hierarchy.

1. $\mathbb{Q}$ and $\mathbb{Z}[i]$ are Euclidean domains but are not isomorphic to $\mathbb{Z}$ as rings.
2. $\mathbb{Z}\left[\frac{1+\sqrt{-19}}{2}\right]$. One can check this is a principal ideal domain but not a Euclidean domain.
3. $\mathbb{Z}[X]$ is not a principal ideal domain, but we will see it is a unique factorisation domain.
4. $\mathbb{Z}\left[\sqrt{-5}\right]$, is an integral domain but not a unique factorisation domain.
5. $\mathbb{Z}/6\mathbb{Z}$, is a commutative ring but not an integral domain.
6. $M_2(\mathbb{R})$, is a ring that is not commutative.

2.3 Localisation

Definition 2.3.1. *Let R be a commutative ring and $(S, \cdot) \subseteq (R, \cdot)$ a submonoid. The localisation $S^{-1}R$ is the set of equivalence classes of pairs (r, s) with $r \in R$ and $s \in S$ where $(r, s) \sim (r', s')$ if and only if there exists a $t \in S$ such that $t(rs' - r's) = 0$.*

Remark 2.3.2. *A pair (r, s), as in Definition 2.3.1, is often denoted as $\frac{r}{s}$.*

If R is an integral domain we add the condition that $0 \notin S$ to Definition 2.3.1. So that $t \neq 0$ which implies that $(r, s) \sim (r', s')$ if and only if $rs' - r's = 0$ which happens if and only if $rs' = r's$. If we did not add the condition that $0 \notin S$ then $(r, s) \sim (0, 0)$ for all $r \in R$ and $s \in S$ so that $S^{-1}R = 0$. Henceforth, we assume that R is an integral domain with S a multiplicative submonoid with $0 \notin S$.

Lemma 2.3.3. *Show that the relation $\sim$ given in Definition 2.3.1 is an equivalence relation.*

Proof. Symmetry and reflexivity are straightforward. For transitivity, suppose $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$. This means $ad = bc$ and $cd = de$. Multiplying these equations by f and b respectively, tells us that $adf = bcf$ and $bcf = bde$. So $adf = bde$. Since $d \neq 0$ and R is an integral domain, we can cancel d to find that $af = be$ which means $(a, b) \sim (e, f)$. □

On the localisation $S^{-1}R$, we define the following operations.

- $(r, s) + (r', s') := (rs' + r's, ss')$.

- $(r,s)\cdot(r',s') := (rr',ss')$.

Notice how the notation $(r,s) = \frac{r}{s}$ is justified.

- $\frac{r}{s} + \frac{r'}{s'} = (r,s) + (r',s') = (rs' + r's, ss') = \frac{rs'+r's}{ss'}$.
- $\frac{r}{s} \cdot \frac{r'}{s'} = (r,s)\cdot(r',s') = (rr',ss') = \frac{rr'}{ss'}$.

Let $(a,b),(c,d),(e,f) \in S^{-1}R$. Then

$$\begin{aligned}\frac{0}{1} + \frac{a}{b} &= \frac{0\cdot b + 1\cdot a}{1\cdot b}\\ &= \frac{a}{b}\end{aligned}$$

and so $0_{S^{-1}R} = \frac{0}{1}$. Similarly,

$$\begin{aligned}\frac{1}{1}\cdot\frac{a}{b} &= \frac{1a}{1b}\\ &= \frac{a}{b},\end{aligned}$$

means $1_{S^{-1}R} = \frac{1}{1}$. Furthermore,

$$\begin{aligned}\left(\frac{a}{b}\cdot\frac{c}{d}\right)\cdot\frac{e}{f} &= \left(\frac{ac}{bd}\right)\cdot\frac{e}{f}\\ &= \frac{ace}{bdf}\\ &= \frac{a}{b}\cdot\left(\frac{c}{d}\cdot\frac{e}{f}\right)\end{aligned}$$

and so $\cdot$ on $S^{-1}R$ is associative.

Exercise 2.3.4. *Show that the operations $+$ and $\cdot$ on $S^{-1}R$ are well-defined. That is, they respect the equivalence relation $\sim$.*

Proposition 2.3.5. *The map $\iota : R \to S^{-1}R$ given by $\iota(r) = (r,1)$ is injective.*

Proof. If $\iota(r') = \iota(r)$ then $(r',1) = (r,1)$ which happens if and only if $r' = r$. Therefore, ι is injective. □

Corollary 2.3.6. *If R is an integral domain R then $R \leq S^{-1}R$.*

Proof. Let $\iota : R \to S^{-1}R$ be the map of Proposition 2.3.5. For $r_1, r_2 \in R$ we have

$$\iota(r_1 + r_2) = (r_1 + r_2, 1) = (r_1,1) + (r_2,1) = \iota(r_1) + \iota(r_2),$$

and

$$\iota(r_1 r_2) = (r_1 r_2, 1) = (r_1,1)(r_2,1) = \iota(r_1)\iota(r_2).$$

Therefore, ι is a homomorphism with $\ker(\iota) = \{0\}$ by Proposition 2.3.5. Therefore, by Theorem 1.3.24 we have

$$R \cong \mathrm{im}(\iota) \leq S^{-1}R.$$

□

Definition 2.3.7. *If R is an integral domain and $S = R \setminus \{0\}$, then $S^{-1}R$ is a field, and is referred to as the field of fractions of R denoted $\mathrm{Frac}(R)$.*

Remark 2.3.8. *Equivalently, for R an integral domain, its field of fractions is*

$$\mathrm{Frac}(R) = \{(a, b) \in R \times R : b \neq 0\}/\sim$$

where $(a, b) \sim (c, d)$ if and only if $ad = bc$.

Remark 2.3.9. *Referring to $\mathrm{Frac}(R)$ as a field can be justified.*

- *The inclusion $\mathrm{Frac}(R)^\times \subseteq \mathrm{Frac}(R) \setminus \{0\}$ is clear as $0 \notin \mathrm{Frac}(R)^\times$.*
- *Let $(r, s) \in \mathrm{Frac}(R) \setminus \{0\}$, then by construction $r, s \in R \setminus \{0\}$ so that (s, r) is also in $\mathrm{Frac}(R) \setminus \{0\}$. Therefore,*
$$(r, s) \cdot (s, r) = (rs, sr) = (1, 1)$$
which implies that $\mathrm{Frac}(R) \setminus \{0\} \subseteq \mathrm{Frac}(R)^\times$.

Example 2.3.10. *The field of fractions of $\mathbb{Z}$ is $\mathbb{Q}$. As expected from Corollary 2.3.6 we have that $\mathbb{Z}$ is a subring of $\mathbb{Q}$. The field of fractions of $\mathbb{C}[X]$ is*

$$\left\{\frac{p}{q} : p,\, q \in \mathbb{C}[X],\, q \neq 0\right\}.$$

In general, if F is a field, we write $F(X)$ for the field of fractions of the polynomial ring $F[X]$.

Proposition 2.3.11. *If A is a commutative ring and $\varphi : R \to A$ is a ring homomorphism such that $\varphi(S) \subseteq A^\times$, then there exits a unique $\tilde{\varphi} : S^{-1}R \to A$ such that $\varphi = \tilde{\varphi} \circ \iota$. Where ι is the map given in Proposition 2.3.5.*

Proof. Let $\tilde{\varphi} : S^{-1}R \to A$ be given by $\tilde{\varphi}((a, b)) = \varphi(a)\varphi(b)^{-1}$. Suppose $(a, b) = (c, d)$, then $ad = bc$. Applying φ it follows that $\varphi(a)\varphi(d) = \varphi(b)\varphi(c)$, as φ is a homomorphism. Therefore, $\varphi(a)\varphi(b)^{-1} = \varphi(c)\varphi(d)^{-1}$, which implies that $\tilde{\varphi}((a, b)) = \tilde{\varphi}((c, d))$. That is, $\tilde{\varphi}$ respect conjugacy classes and so is well-defined. Let $(a, b), (c, d) \in S^{-1}R$. Then

$$\begin{aligned}
\tilde{\varphi}((a, b) + (c, d)) &= \tilde{\varphi}((ad + bc, bd)) \\
&= \varphi(ad + bc)\varphi(bd)^{-1} \\
&= (\varphi(a)\varphi(d) + \varphi(b)\varphi(c))\varphi(b)^{-1}\varphi(d)^{-1} \\
&= \varphi(a)\varphi(b)^{-1} + \varphi(c)\varphi(d)^{-1} \\
&= \tilde{\varphi}((a, b)) + \tilde{\varphi}((c, d)).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\tilde{\varphi}((a, b)(c, d)) &= \tilde{\varphi}((ac, bd)) \\
&= \varphi(ac)\varphi(bd)^{-1} \\
&= \varphi(a)\varphi(b)^{-1}\varphi(c)\varphi(d)^{-1} \\
&= \tilde{\varphi}((a, b))\tilde{\varphi}((c, d)).
\end{aligned}$$

Therefore, $\tilde{\varphi}$ is a homomorphism. Suppose that $\tilde{\psi}$ is a homomorphism with the property that $\varphi = \tilde{\psi} \circ \iota$. Then

$$\varphi(b)\tilde{\psi}((1, b)) = \tilde{\psi}((b, 1))\tilde{\psi}((1, b)) = \tilde{\psi}((1, 1)) = 1.$$

Therefore, $\tilde{\psi}((1,b))$ is the inverse of $\varphi(b)$, which is unique. Thus using the homomorphism properties of $\tilde{\psi}$ it follows that $\tilde{\psi}((a,b)) = \varphi(a)\varphi(b)^{-1} = \tilde{\varphi}((a,b))$. Hence, $\tilde{\varphi}$ is the unique homomorphism such that $\varphi = \tilde{\varphi} \circ \iota$. □

Definition 2.3.12 provides an alternative, but equivalent, definition of localisation.

Definition 2.3.12. *For R a commutative ring and $S \subseteq R$ a multiplicative submonoid. The localisation, $S^{-1}R$, is the unique ring R' such that there exists a map $\iota : R \to R'$ with the following properties.*

1. *$\iota(S) \subseteq (R')^\times$.*
2. *For all commutative rings A, and maps $\varphi : R \to a$ with $\varphi(S) \subseteq A^\times$, there exists a unique map $\tilde{\varphi} : R' \to A$ such that $\varphi = \tilde{\varphi} \circ \iota$.*

Corollary 2.3.13. *Let R be an integral domain, F a field and $\varphi : R \to F$ an injective homomorphism. Then $\varphi = \tilde{\varphi} \circ \iota$, with $\iota : R \to \mathrm{Frac}(R)$, the canonical map and $\tilde{\varphi} : \mathrm{Frac}(R) \to F$ injective. That is φ factors through the field of fractions of R.*

Proof. Let $S = R \setminus \{0\}$ and $A = F$ in the context of Proposition 2.3.11. Then as φ is an injective homomorphism we have that $\varphi(S) \subseteq F \setminus \{0\} = F^\times$. So we can apply the statement of Proposition 2.3.11 to obtain a homomorphism $\tilde{\varphi} : \mathrm{Frac}(R) \to F$ such that $\varphi = \tilde{\varphi} \circ \iota$. As $\mathrm{Frac}(R)$ and F are both fields and $\tilde{\varphi}$ is non-zero, it must be the case that $\tilde{\varphi}$ is injective. □

Corollary 2.3.14. *If F is a field of characteristic zero, it contains a subfield isomorphic to $\mathbb{Q}$. If F is a field of characteristic p, it contains a subfield isomorphic to $\mathbb{F}_p$.*

Proof. The characteristic of F is zero if and only if the kernel of the unique ring homomorphism $\mathbb{Z} \to F$ is trivial. Using Corollary 2.3.13 it follows that the unique ring homomorphism factors through $\mathbb{Q}$, which implies that $\mathbb{Q} \leq F$. If the characteristic of F is a prime p, then the kernel of the unique homomorphism $\mathbb{Z} \to F$ is $p\mathbb{Z}$. Therefore, by the first isomorphism theorem we conclude that $\mathbb{F}_p \cong \mathbb{Z}/p\mathbb{Z} \cong F$. □

Lemma 2.3.15. *Let R be a ring and F a field such that $F \leq R$. Then R is a vector space over F.*

Proof. Observe that R is an abelian group with a notion of scalar multiplication by F coming from the multiplication of R. Using the axioms of the ring one can check that the scalar multiplication by F satisfies the vector space axioms. □

Corollary 2.3.16. *If F is a field of characteristic zero then it is a vector space over $\mathbb{Q}$. If F is a field of characteristic p then it is a vector space over $\mathbb{F}_p$.*

Proof. This follows directly from Lemma 2.3.15 and Corollary 2.3.14. □

Note that for a commutative ring with a prime ideal I the set $S = R \setminus I$ is a multiplicative submonoid. The localisation $S^{-1}R$ is denoted R_I.

Example 2.3.17. *Recall, that for a prime p the ideal $(p) \subseteq \mathbb{Z}$ is a prime ideal, and so we can consider $\mathbb{Z}_{(p)}$. We can think of this as a subset of $\mathbb{Q}$ generated by numbers of the form qp, where q is either*

1. *a prime number, or*
2. *the reciprocal of a prime number not equal to p.*

Definition 2.3.18. *A local ring is a ring with a unique maximal ideal.*

Lemma 2.3.19. *A commutative ring R is local if and only if there is an ideal I such that all $r \in R \setminus I$ are units.*

Proof. $(\Rightarrow)$. Let I be the unique maximal ideal of R. Consider $r \in R \setminus I$ and suppose for contradiction that $(r) \neq R$. As I is a maximal ideal it follows that $(r) \subseteq I$ which implies that $r \in I$, which is a contradiction. Therefore, $(r) = R$, which in particular means that r is a unit by Lemma 1.3.14.
$(\Leftarrow)$. Let $J \subseteq R$ be a proper ideal. Therefore, by Lemma 1.3.14, all $j \in J$ are not units and so $j \in I$. Consequently, $J \subseteq I$ which implies that I is maximal and moreover unique. □

Proposition 2.3.20. *Let R be a commutative ring and let $I \subseteq R$ be a prime ideal. Then R_I has a unique maximal ideal given by*

$$\bar{I} = \{(r, s) : r \in I,\, s \in R \setminus I\}.$$

Proof. If $a \notin \bar{I}$, then $a = (r, s)$ for $s \notin I$. Therefore, a is invertible in R_I, and so we can conclude using Lemma 2.3.19. □

Definition 2.3.21. *For a commutative ring R, an ideal $I \subseteq R$, and a multiplicative submonoid $S \subseteq R$, the set*

$$S^{-1}I = \left\{ \frac{i}{s} : s \in S,\, i \in I \right\} \subseteq S^{-1}R$$

is an ideal referred to as the image of I under the localisation.

Proposition 2.3.22. *Every ideal $I \subseteq S^{-1}R$ is of the form $S^{-1}J$ for some ideal $J \subseteq R$.*

Proof. For $r \in R$ and $s \in S$ note that $(r, s) \in I$ if and only if $(r, 1) \in I$. Let

$$J := \{r \in R : (r, 1) \in I\}.$$

Then $J \subseteq R$ is an ideal and $S^{-1}J = I$. □

2.4 Polynomial Rings

Recall that if R is an integral domain then so is $R[X]$. An important case of this is $F[X]$ when F is a field as then $F[X]$ is also a Euclidean domain and therefore a principal ideal domain and thus a unique factorisation domain. Consequently, we deduce the following.

1. If $I \subseteq F[X]$ is a non-zero ideal, then $I = (f)$ for some non-zero $f \in F[X]$, and I is maximal if and only if I is prime.
2. An element $f \in F[X]$ is irreducible if and only if it is prime.

Moreover, for $f \in F[X]$ the following are equivalent.

1. f is irreducible.
2. f is prime.
3. $F[X]/(f)$ is an integral domain.
4. $F[X]/(f)$ is a field.

Example 2.4.1.

1. *$\mathbb{F}_2[X]/\left(X^2+1\right)$ is not an integral domain as $(X+1)^2 = X^2+1 = 0$, but $X+1 \neq 0$.*
2. *Since X^2+X+1 has no roots in $\mathbb{F}_2$ it is irreducible, and thus $\mathbb{F}_2[X]/\left(X^2+X+1\right)$ is a field.*

Definition 2.4.2. *If R is a ring, and $f \in R[X]$, then f is monic if $f = a_0 + a_1X + \ldots a_{n-1}X^{n-1} + 1 \cdot X^n$. That is, $f \in R[X]$ is monic if the coefficient of the highest order term is one.*

Definition 2.4.3. *Let R be a unique factorisation domain, and $f = a_0 + \ldots + a_nX^n \in R[X]$. Then the content of f, written $c(f)$, is the greatest common divisor of the set $\{a_0, \ldots, a_n\}$.*

The content of a polynomial is only well-defined up to associates, though the ideal it generates, $(c(f))$, is well-defined.

Definition 2.4.4. *Let R be a unique factorisation domain. Then $f \in R[X]$ is primitive if $c(f)$ is a unit. In other words, the coefficients of f are all coprime.*

Example 2.4.5.

- *Let $f = a_0 \in R[X]$ be a constant polynomial. Then $c(f) = a_0$, up to associates.*
- *Let $f = 2 + 3X + 4X^2 \in \mathbb{Z}[X]$. Then $c(f)$ is a unit meaning f is primitive.*
- *Let $g = 2 + 4X + 6X^2 \in \mathbb{Z}[X]$. Then $c(f) = 2$, up to associates, so g is not primitive in $\mathbb{Z}[X]$. However, 2 is a unit in $\mathbb{Q}$ and so g is primitive in $\mathbb{Q}[X]$.*

Lemma 2.4.6. *Let R be a unique factorisation domain, and $f \in R[X]$. Then $f = c(f) \cdot f'$ for some primitive $f' \in R[X]$.*

Proof. If $f = a_0 + \cdots + a_nX^n$ then $c(f) \mid a_i$ for each $i = 0, \ldots, n$. Therefore, $a_i = c(f)b_i$ for some $b_i \in R$. In particular, the greatest common divisor of $\{b_1, \ldots, b_n\}$ is one. Let $f' = b_0 + \ldots b_nX^n$. Then f' is primitive and $f = c(f) \cdot f'$. □

Lemma 2.4.7. *Let R be a unique factorisation domain. If $f, g \in R[X]$ are primitive, then $f \cdot g$ is also primitive.*

Proof. Let $f = a_0 + \ldots a_nX^n$ and $g = b_0 + \ldots b_mX^m$, with $a_n, b_m \neq 0$, be primitive in $R[X]$. Suppose fg is not primitive so that $c(f \cdot g)$ is not a unit. Since R is a unique factorisation domain, there is some irreducible $p \in R$ with $p \mid c(f \cdot g)$. As $c(f)$ and $c(g)$ are units we know that p does not divide $c(f)$ or $c(g)$. Therefore, p does not divide all a_i or all b_j, meaning there exists $k, l \geq 0$ with the following properties.

1. p divides $a_0, \ldots, a_{k-1}$ but not a_k.
2. p divides $b_0, \ldots, b_{l-1}$ but not b_l.

Consider the X^{k+l} coefficient in $f \cdot g$ which is given by

$$\sum_{i+j=k+l} a_ib_j = \underbrace{(a_{k+l}b_0 + \cdots + a_{k+1}b_{l-1})}_{(1)} + a_kb_l + \underbrace{(a_{k-1}b_{l+1} + \cdots + a_0b_{l+k})}_{(2)}.$$

We know p divides (1) as p divides each b_j. Similarly, we know p divides (2) as p divides each a_i. Hence, as p divides the X^{k+l} coefficient, it must be that p divides a_kb_l. Since p is irreducible and hence prime, $p \mid a_k$ or $p \mid b_l$, either of which provides a contradiction. So $c(f \cdot g)$ is a unit which implies that $f \cdot g$ is primitive. □

Corollary 2.4.8. *Let R be a unique factorisation domain. Then for $f, g \in R[X]$ the content $c(f \cdot g)$ is an associate of $c(f) \cdot c(g)$.*

Proof. Using Lemma 2.4.6, we can write $f = c(f) \cdot f'$ and $g = c(g) \cdot g'$ for $f', g' \in R[X]$ primitive. Then $f \cdot g = c(f)c(g)\,(f' \cdot g')$. As f' and g' are primitive it follows by Lemma 2.4.7 that $f' \cdot g'$ is primitive. Therefore, $c(f)c(g)$ is a greatest common divisor of the coefficients of $f \cdot g$ meaning it is an associate of $c(f \cdot g)$. □

Remark 2.4.9.

- *Note that Corollary 2.4.8 implies Lemma 2.4.7. However, Corollary 2.4.8 is derived from Lemma 2.4.7.*
- *We cannot say $c(f \cdot g) = c(f) \cdot c(g)$, since both are only well-defined up to associates.*

Lemma 2.4.10 (Gauss' Lemma)**.** *Let R be a unique factorisation domain and $f \in R[X]$ a primitive polynomial. Let F be the field of fractions of R. Then f is irreducible in $R[X]$ if and only if it is irreducible in $F[X]$.*

Proof. Let R be a unique factorisation domain, with F its field of fractions, and $f \in R[X]$ a primitive polynomial. ($\Leftarrow$) Suppose f is reducible in $R[X]$. Then $f = gh$ where $g, h \in R[X]$ are both non-units. As f is primitive both g and h are also primitive. Therefore, if g had degree zero it would be a unit and so under our assumptions, we have $\deg(g) > 0$. Similarly, $\deg(h) > 0$. Consequently, g and h cannot be units in $F[X]$. Writing $f = gh$ and viewing f, g and h as elements of $F[X]$, we see that f is reducible in $F[X]$. It follows that if f is irreducible in $F[X]$ then it ought to be irreducible in $R[X]$ too.
($\Rightarrow$) Suppose f is reducible in $F[X]$ with $f = gh$, for $g, h \in F[X]$ non-units. As before we must have $\deg(g), \deg(h) > 0$. We can clear denominators by choosing $a, b \in R\backslash\{0\}$ such that ag and bh lie in $R[X]$. For example, a can be taken to be the product of all denominators of coefficients of g and similarly for b. Let $g' = ag$ and $h' = bh$ and note these lies in $R[X]$. However, h may not be in $R[X]$ and so $h' = bh$ is not necessarily a factorisation in $R[X]$. Similarly, $g' = ag$ is not necessarily a factorisation in $R[X]$. Using $f = gh$ we see that

$$abf = g'h' \in R[X] \tag{2.4.1}$$

where each factor lie in $R[X]$. Hence, we can write $g' = c\,(g') \cdot g''$ and $h' = c\,(h') \cdot h''$ for $g'', h'' \in R[X]$ primitive. Since f is primitive we note that $c(abf) = ab$ which, by (2.4.1), is an associate of $c\,(g'h')$ and thus also an associate of $c\,(g')\,c\,(h')$ by Corollary 2.4.8. That is, $uab = c\,(g')\,c\,(h')$ where $u \in R^\times$ is some unit. Therefore

$$\begin{aligned} abf &= g'h' \\ &= c\,(g')\,c\,(h')\,g''h'' \\ &= uabg''h''. \end{aligned}$$

Since $R[X]$ is an integral domain and $ab \neq 0$, we get

$$f = u \cdot g'' \cdot h'',$$

where $u \in R^\times$ is a unit and $g''h'' \in R[X]$. Moreover, as g'' and h'' have positive degrees they are non-units which means that f is reducible in $R[X]$. Therefore, we conclude that if f is irreducible in $R[X]$ then it ought to be irreducible in $F[X]$ too. □

Corollary 2.4.11. *Let R be a unique factorisation domain, and let F be the field of fractions R. If $\frac{p}{q} \in F$, with p and q coprime, is a root for $f = a_0 + a_1X + \cdots + a_nX^n \in R[X]$, then $p|a_0$ and $q|a_n$.*

Proof. Let $f = c(f_1)f_1$, where f_1 is primitive. Then $\frac{p}{q}$ is a root of f, hence, $f_1 = (qX - p)g$ for some $g \in F[X]$ as $F[X]$ is a Euclidean domain. Therefore, by Lemma 2.4.10, f_1 is reducible over $R[X]$. In particular, $f_1 = (qX - p)g$ for $g \in R[X]$. Writing $g = b_0 + \cdots + b_nX^n$ it follows that f_1 has X^0 coefficient $-pb_0$ and X^n coefficient qb_{n-1}. Thus, $-b_0|a_0$ and $qb_{n-1}|a_n$, which implies that $p|a_0$ and $q|a_n$. □

Lemma 2.4.10 is useful as checking irreducibility in $F[X]$ is generally harder than checking it in $R[X]$.

Example 2.4.12.

1. *Let $f = 1+X+X^3 \in \mathbb{Z}[X]$. Note that f is primitive as $c(f) = 1$. Suppose f is reducible in $\mathbb{Q}[X]$. Then by Lemma 2.4.10 we have that f is reducible in $\mathbb{Z}[X]$. Let $f = g \cdot h$ for $g, h \in \mathbb{Z}[X]$ non-units. As the coefficients of f are ones, if g and h were constants then they would have to be ± 1 or 0. However, this would mean that g and h were either units or zero, which is not the case and so $\deg(g), \deg(h) > 0$. As $\deg(g)+\deg(h) = \deg(f) = 3$ we can suppose without loss of generality that $\deg(g) = 1$ and $\deg(h) = 2$ and write $g = b_0 + b_1 X$ and $h = c_0 + c_1 X + c_2 X^2$, where $b_i, c_j \in \mathbb{Z}$. Since $f = g \cdot h$, the X^0 coefficient of f is*

$$1 = b_0 c_0$$

and the X^3 coefficient is

$$1 = b_1 c_2.$$

As $b_i, c_j \in \{\pm 1\}$ for each i and j we have $g = \pm 1 \pm X$ meaning ± 1 is a root of f. However, computing $f(1)$ and $f(-1)$ we see that they are both non-zero, giving a contradiction. Therefore, f is irreducible in $\mathbb{Z}[X]$ and thus irreducible in $\mathbb{Q}[X]$. Consequently, $\mathbb{Q}[X]/(f)$ is a field. The utility of Lemma 2.4.10 for this problem was in restricting the possible coefficients of g and h.

2. *The polynomial f in Lemma 2.4.10 needs to be primitive. Indeed, take $f = 2X + 2 \in \mathbb{Z}[X]$. Then $f = 2(X + 1)$ is reducible, however, as $2 \in \mathbb{Q}[X]$ is a unit the polynomial f is not reducible in $\mathbb{Q}[X]$.*

Theorem 2.4.13. *Let R be a unique factorisation domain. Then $R[X]$ is a unique factorisation domain.*

Proof. Consider statement 1 of Definition 2.2.20. Let $f \in R[X]$ be non-zero and not a unit. Write $f = c(f) \cdot f_1$, with $f_1 \in R[X]$ a primitive polynomial. Since R is a unique factorisation domain we can factor the content as

$$c(f) = p_1 \dots p_n$$

where each $p_i \in R$ is irreducible in R. As $R \subseteq R[X]$ each p_i is also irreducible in $R[X]$. Next, suppose f_1 is not the product of irreducibles. In particular, f_1 itself is not irreducible and so we can write $f_1 = f_2 g_2$ with $f_2, g_2 \in R[X]$ non-units. Since f_1 is primitive we must have that f_2 and g_2 are primitive, and so they cannot be constants. Hence, $\deg(f_2), \deg(g_2) > 0$. Since $\deg(f_1) = \deg(f_2) + \deg(g_2)$, we must also have $\deg(f_2), \deg(g_2) < \deg(f_1)$. If f_2 and g_2 were products of irreducibles then f_1 would be too. Assume without loss of generality that f_2 is not a product of irreducibles. Apply the same argument to f_2 to write $f_2 = f_3 g_3$ as a product of non-units. Continuing gives a sequence $(f_n)_{n \in \mathbb{N}} \subseteq R[X]$ of non-zero elements, with $\deg(f_1) > \deg(f_2) > \ldots > 0$. However, we cannot have an infinite sequence of positive integers which is strictly decreasing, and so we arrive at a contradiction. Therefore we can write $f_1 = q_1 \dots q_m$, with each $q_i \in R[X]$ irreducible. Thus,

$$f = p_1 \dots p_n \cdot q_1 \dots q_m,$$

where all the p_i are irreducible constant polynomials, and all the q_j are irreducible non-constant polynomials. We conclude that factorisations into irreducibles exist. Next, consider statement 2 of Definition 2.2.20. Let $f \in R[X]$ be non-zero and not a unit. Let $c(f) = p_1 \dots p_n$ be a factorisation of $c(f)$ into irreducibles of R. This is unique up to reordering and associates since R is a unique factorisation domain. Consequently, it suffices to consider primitive polynomials. Suppose $f' \in R[X]$ is primitive and that we have factorisations

$$f' = q_1 \dots q_m = r_1 \dots r_l$$

where each $q_i, r_j \in R[X]$ is irreducible. Each $c(q_i)$ and $c(r_j)$ is a factor of the unit $c(f')$ and so each q_i and r_j must be primitive. Let F be the field of fractions of R, and consider $q_i, r_j \in F[X]$. By Lemma 2.4.10 each q_i and r_j is irreducible in $F[X]$. As F is a field we have that $F[X]$ is a Euclidean domain, meaning it is also a principal ideal domain and hence a unique factorisation domain. So by the uniqueness of factorisation in $F[X]$, we find that $m = l$, and possibly after reordering we find that r_i and q_i are associates in $F[X]$ for each $i = 1, \dots, m$.

That is, $r_i = u_i q_i$ for some unit $u_i \in F[X]$. In particular, each u_i must be a constant polynomial and thus $u_i \in F$, that is, $u_i = \frac{a_i}{b_i}$, for some elements $a_i, b_i \in R \setminus \{0\}$. Meaning $a_i r_i = b_i q_i$, where all factors are in $R[X]$. Recall that r_i and q_i are primitive, and so by taking the content of both sides we find that a_i, b_i are associates. Hence we can write $b_i = v_i a_i$, for some unit $v_i \in R^\times$. Therefore,

$$a_i r_i = v_i a_i q_i.$$

Cancelling the a_i factor, which we can do as it is non-zero, we find that $r_i = v_i \cdot q_i$, where v_i is a unit in R and hence a unit in $R[X]$. Consequently, r_i and q_i are associates in $R[X]$ for each $i = 1, \ldots, m$, implying that the factorisation into irreducibles is unique up to associates and reordering. □

Remark 2.4.14. *There are some comments to be made regarding the proof of Theorem 2.4.13.*

1. *Showing the existence of factorisations was similar to showing that a principal ideal domain is a unique factorisation domain.*
   (a) *Find factorisations $f_1 = f_2 g_2, f_2 = f_3 g_3, \ldots$ into non-units.*
   (b) *Show this sequence terminates.*
2. *Showing uniqueness involved taking factors out of the contents to reduce everything to the primitive case, and then using Lemma 2.4.10 to argue in $F[X]$ which is known to be a unique factorisation domain.*
3. *Iterating Theorem 2.4.13 we deduce that if R is a unique factorisation domain then $R[X_1, \ldots, X_n]$ is a unique factorisation domain, for $n \in \mathbb{N}$.*
4. *Theorem 2.4.13 shows that $\mathbb{Z}[X]$ is a unique factorisation domain which we know is not a principal ideal domain.*

Proposition 2.4.15 (Eisenstein's criterion)**.** *Let R be a unique factorisation domain, and let $f = a_0 + a_1 X + \ldots + a_n X^n \in R[X]$ be a primitive polynomial with $a_n \neq 0$. Let $p \in R$ be irreducible such that*

- *p does not divide a_n,*
- *p divides a_i for each $0 \leq i < n$, and*
- *p^2 does not divide a_0.*

Then f is irreducible in $R[X]$.

Proof. Suppose that f is reducible in $R[X]$, that is $f = gh$ for $g, h \in R[X]$ non-units. Let

- $g = \sum_{i=1}^{m} b_i X^i$, and
- $h = \sum_{i=1}^{l} c_i X^i$.

As f is primitive, g and h are primitive. Note that $m + l = n$. Moreover, $m, l \geq 1$ as if $m = 0$ then g would be a constant. Thus as g is primitive it follows that $g \in R^\times \subseteq R[X]^\times$, which is a contradiction. Let $\tilde{f}$ be the transportation of $f \in R[X]$ to $R[X]/(p)$. In other words, $\tilde{f}$ is f with each coefficient reduced modulo p, similarly let $\tilde{a} = a + (p)$ for $a \in R$. By assumption we have that $\tilde{f} = \tilde{a}_n X^n$. Moreover, $\tilde{f} = \tilde{g}\tilde{h}$. Looking at the X^0 coefficient we see that $p | b_0 c_0$. As p^2 does not divide a_0 it must be the case that p divides exactly one of b_0 or c_0. Without loss of generality suppose that $p | c_0$ and p does not divide b_0, that is $\tilde{c}_0 = 0$ and $\tilde{b}_0 \neq 0$. As $\tilde{h}$ is not zero, there is some $0 < j < l$ such that $p | c_0, \ldots, p | c_{j-1}$ but $p \nmid c_j$. Consequently, the X^j coefficient of $\tilde{f}$ is given by $b_0 c_j \bmod p$. We know that $j \neq n$ as $l < n$, and so $0 = b_0 c_j \bmod p$. However, this implies that $p | b_0$ or $p | c_j$, which is a contradiction. □

Example 2.4.16.

- *Consider $f = X^n - p \in \mathbb{Z}[X]$, where $p \in \mathbb{N}$ is prime. The conditions for Eisenstein's criterion hold and so f is irreducible in $\mathbb{Z}[X]$. Therefore, by Lemma 2.4.10, it is also irreducible in $\mathbb{Q}[X]$. Consequently, we deduce that $X^n - p$ has no rational zeroes. That is, $\sqrt[n]{p}$ is irrational for $n > 1$.*
- *Let $f = X^{p-1} + X^{p-2} + \cdots + 1 \in \mathbb{Z}[X]$, for $p \in \mathbb{N}$ prime. Note that*

$$f = \frac{X^p - 1}{X - 1}.$$

Since the coefficients in f are ones, Eisenstein's criterion cannot currently be applied. Instead, we let $Y = X - 1$ so that $f(X) = \hat{f}(Y)$ where $\hat{f} \in \mathbb{Z}[Y]$ is given by

$$\begin{aligned}\hat{f} &= \frac{(Y+1)^p - 1}{Y} \\ &= Y^{p-1} + \binom{p}{1} Y^{p-2} + \binom{p}{2} Y^{p-3} + \ldots + \binom{p}{p-1}.\end{aligned}$$

Now Eisenstein's criterion can be applied with p as $\hat{f}$ is primitive and $p \mid \binom{p}{i}$ for $1 \leq i \leq p-1$ but $p^2 \nmid \binom{p}{p-1}$. Therefore, $\hat{f} \in \mathbb{Z}[Y]$ is irreducible. Thus, $f = gh$ in $\mathbb{Z}[X]$ would mean $\hat{f}(Y) = g(Y+1)h(Y+1)$ which contradicts $\hat{f}(Y)$ being irreducible. Hence $f \in \mathbb{Z}[X]$ is irreducible. This is to be expected as we know the roots of f are $e^{\frac{2\pi i k}{p}}$ for $1 \leq k \leq p-1$, none of which are rational, or even real.

Proposition 2.4.17 (Generalised Eisenstein's Criterion)**.** *Let R be a commutative ring, and let $f = a_0 + a_1 X + \cdots + a_n X^n \in R[X]$ be a primitive polynomial with $a_n \neq 0$. Let $P \subseteq R$ be a prime ideal such that*

- *$a_n \notin P$,*
- *$a_i \in P$ for each $0 \leq i < n$, and*
- *$a_0 \notin P^2$.*

Then f is irreducible in $R[X]$.

Remark 2.4.18.

1. *In Proposition 2.4.17, P^2 denotes the ideal generated by elements of the form $p_1 p_2$ where $p_1, p_2 \in P$.*
2. *In Proposition 2.4.17, since we are not assuming R is a unique factorisation domain, the greatest common divisor of the coefficient need not exist. Proposition 2.4.17 only applies to f where the greatest common divisor exists.*

2.5 Noetherian Rings

Recall that a commutative ring R is Noetherian if for a chain of ideals

$$I_1 \subseteq I_2 \subseteq \cdots \subseteq R$$

there is some N for which $I_N = I_{N+1} = \ldots$.

Definition 2.5.1. *An ideal $I \subseteq R$ is finitely generated if there exists a set $\{r_1, \ldots, r_n\} \subseteq R$ such that $I = (r_1, \ldots, r_n)$.*

Proposition 2.5.2. *A ring is Noetherian if and only if every ideal is finitely generated.*

Proof. ($\Rightarrow$). Let R be a Noetherian ring. Suppose that an ideal $I \subseteq R$ is not finitely generated. Let $i_1 \in I$, it follows that $(i_1) \subseteq I$, with $(i_1) \neq I$ as I is not finitely generated. Consequently, we can choose $i_2 \in I \setminus (i_1)$. Then the ideal (i_1, i_2) is such that $(i_1) \subseteq (i_1, i_2) \subseteq I$ where $(i_1) \neq (i_1, i_2)$ as $i_2 \notin (i_1)$ and $(i_1, i_2) \neq I$ as I is not finitely generated. Continuing in the was we generate a chain of ideals

$$(i_1) \subseteq (i_1, i_2) \subseteq (i_1, i_2, i_3) \subseteq \cdots \subseteq I$$

which is not eventually constant. This contradicts R being Noetherian.
($\Leftarrow$). Let R be a ring for which every ideal is finitely generated. Let

$$I_1 \subseteq I_2 \subseteq I_3 \cdots \subseteq R$$

be an ascending chain of ideals. By Exercise 2.2.25 we have that

$$I = \bigcup_{i=1}^{\infty} I_i \subseteq R$$

is an ideal and thus must be finitely generated. Suppose that

$$(i_1, \ldots, i_n) = I$$

and let M be the minimum number such that $i_k \in I_M$ for each $k \in \{1, \ldots, n\}$. Consequently, $i_k \in I_m$ for each $k \in \{1, \ldots, n\}$ and $m \geq M$. Therefore, as ideals are additive groups it follows that

$$I = (i_1, \ldots, i_n) \subseteq I_m$$

for every $m \geq M$. However, as $I_m \subseteq I$ for all $m \geq M$ we conclude that $I_M = I_{M+1} = \ldots$. Therefore, R is Noetherian. □

Remark 2.5.3. *Using Proposition 2.5.2 we deduce the following statements.*

- *Any principal ideal domain is Noetherian.*
- *Any field is Noetherian.*
- *Any finite ring is Noetherian.*
- *For $\mathbb{Z}[X_1, X_2, \ldots]$ the chain of ideals*

  $$(X_1) \subseteq (X_1, X_2) \subseteq (X_1, X_2, X_3) \subseteq \ldots$$

  is not eventually constant and so $\mathbb{Z}[X_1, X_2, \ldots]$ cannot be a Noetherian ring.

Proposition 2.5.4. *Let R be a commutative Noetherian ring and let $I \subseteq R$ be an ideal. Then R/I is Noetherian.*

Proof. Let $\pi : R \to R/I$ be given by $x \mapsto x + I$. Let $J \subseteq R/I$ be an ideal. Then $\pi^{-1}(J) \subseteq R$ is an ideal. Therefore, $\pi^{-1}(J)$ is finitely generated by Proposition 2.5.2, that is

$$\pi^{-1}(J) = (r_1, \ldots, r_n)$$

for some $r_1, \ldots, r_n \in R$. Consequently,

$$J = (\pi(r_1), \ldots, \pi(r_n))$$

and so R/I is Noetherian by Proposition 2.5.2. □

Exercise 2.5.5. *Let R be a ring and let $I \subseteq R[X]$ be an ideal of $R[X]$. For $n \in \mathbb{N}$ let*

$$I_n := \{r \in R : \text{ there exists } f \in I \text{ such that } f = rX^n + \dots\} \cup \{0\} \subseteq R.$$

Show that I_n is an ideal of R.

Theorem 2.5.6 (Hilbert Basis Theorem)**.** *Let R be a Noetherian ring. Then so is $R[X]$ is a Noetherian ring.*

Proof. Let $I \subseteq R[X]$ be an ideal. For $n \in \mathbb{N}$ let

$$I_n := \{r \in R : \text{ there exists } f \in I \text{ such that } f = rX^n + \dots\} \cup \{0\}.$$

By Exercise 2.5.5 we know that $I_n \subseteq R$ is an ideal of R. In particular, as $Xf \in I$ for every $f \in I$ it follows that $(I_n)_{n\in\mathbb{N}} \subseteq R$ is an increasing sequence of ideals. Therefore, as R is Noetherian there exists a $N \in \mathbb{N}$ such that $I_N = I_{N+1} = \dots$. Furthermore, as R is Noetherian, we can use Proposition 2.5.2 to write

$$I_n = \left(r_1^{(n)}, \dots, r_{k(n)}^{(n)}\right)$$

for $0 \leq n \leq N$. For each $r_i^{(n)}$ there exists some $f_i^{(n)} \in I$ with $f_i^{(n)} = r_i^{(n)} X^n + \dots$.
Claim 1: The ideal generated by the polynomials $\left(f_i^{(n)}\right)_{0\leq n\leq N, 1\leq i\leq k(n)}$, which we denote J, equals I.

Proof of Claim 1. Proceed by contradiction and let $g \in I$ with $g = rX^m + \dots$ be of minimal degree such that $g \notin J$.

1. If $m \leq N$, then $r \in I_m$ and so
$$r = \sum_{i=1}^{k(m)} \lambda_i r_i^{(m)}$$
for some $\lambda_i \in R$. Consequently,
$$\tilde{g} := \sum_{i=1}^{k(m)} \lambda_i f_i^{(m)} = rX^m + \dots \in I.$$
Therefore, if $g \notin J$ then $g - \tilde{g} \notin J$. However, $\deg(g - \tilde{g}) < m$ which is a contradiction.

2. If $m > N$, then we know that $r \in I_N$ by construction of N. Therefore, we can write
$$r = \sum_{i=1}^{k(N)} \lambda_i r_i^{(N)}.$$
Consequently,
$$\tilde{g} := X^{n-N} \sum_{i=1}^{k(N)} \lambda_i r_i^{(N)} = rX^m + \dots \in I.$$
Therefore, if $g \notin J$ then $g - \tilde{g} \notin J$. However, $\deg(g - \tilde{g}) < m$ which is a contradiction.

With Claim 1 we deduce that I is a finitely generated ideal, and so $R[X]$ is Noetherian. □

Let F be a field and consider a set of equations $\mathcal{E} \subseteq F[X_1, \dots, X_n]$. Let I be the ideal $(\mathcal{E})$. As F is Noetherian the polynomial ring $F[X_1, \dots, X_n]$ is also Noetherian by applying an induction argument to the Theorem 2.5.6. Hence, every ideal of $F[X_1, \dots, X_n]$ is finitely generated meaning

$$(f_1, \dots, f_k) = I$$

for some $f_1, \ldots, f_k \in F[X_1, \ldots, X_n]$. For $\alpha = (\alpha_1, \ldots, \alpha_n) \in F^n$ consider the homomorphism $\phi_\alpha : F[X_1, \ldots, X_n] \to F$ defined by

$$\phi_\alpha(X_i) = \alpha_i$$

for $1 \leq i \leq n$. Then α is a solution to the equations $\mathcal{E}$ if and only if $(\epsilon) \subseteq \ker(\phi_\alpha)$ which happens if and only if $(f_1, \ldots, f_k) \subseteq \ker(\phi_\alpha)$. Remarkably, what we observe is that we can solve the potentially infinite set of equations $\mathcal{E}$ by determining whether α solves the finite set of equations $f_1, \ldots, f_k$.

2.6 Algebraic Integers

Previously, we encountered rings such as $\mathbb{Z}[i]$ and $\mathbb{Z}\left[\sqrt{-5}\right]$. We want to generalise these ideas to define rings such as $\mathbb{Z}\left[e^{\frac{2\pi i}{n}}\right]$.

Definition 2.6.1. *A complex number $\alpha \in \mathbb{C}$ is an algebraic integer if it is the root of some monic polynomial $f \in \mathbb{Z}[X]$. That is, $f(\alpha) = 0$.*

Example 2.6.2.

1. *$\alpha = i$ is an algebraic integer as $f(i) = 0$ for $f(X) = X^2 + 1$.*
2. *$\alpha = \sqrt{2}$ is an algebraic integer as $f\left(\sqrt{2}\right) = 0$ for $f(X) = X^2 - 2$.*

Remark 2.6.3. *There are only countably many polynomials with integer coefficients, and as each of these can only have finitely many roots we deduce that the set of algebraic integers is countable. Therefore, not all complex numbers can be algebraic integers as $\mathbb{C}$ is uncountable.*

Definition 2.6.4. *For $\alpha \in \mathbb{C}$ an algebraic integer, we write $\mathbb{Z}[\alpha] \leq \mathbb{C}$ for the smallest subring containing α. More specifically,*

$$\mathbb{Z}[\alpha] := \bigcap_{S \leq \mathbb{C}, \alpha \in S} S$$

where the intersection is over subrings $S \leq \mathbb{C}$ containing α.

Equivalently, we can let $\phi_\alpha : \mathbb{Z}[X] \to \mathbb{C}$ be the homomorphism sending f to $f(\alpha)$, and let $\mathbb{Z}[\alpha] = \operatorname{im}(\phi_\alpha) \leq \mathbb{C}$. By Theorem 1.3.24 we have that

$$\mathbb{Z}[\alpha] \cong \mathbb{Z}[X] / \ker(\phi_\alpha)$$

where $\ker(\phi_\alpha)$ is non-empty by construction.

Proposition 2.6.5. *Let $\alpha \in \mathbb{C}$ be an algebraic integer, and let $\phi_\alpha : \mathbb{Z}[X] \to \mathbb{C}$ be the homomorphism sending f to $f(\alpha)$. Let $I = \ker(\phi_\alpha) \subseteq \mathbb{Z}[X]$. Then the ideal I is principal with $I = (f_\alpha)$, for some $f_\alpha \in \mathbb{Z}[X]$ which is irreducible and monic.*

Proof. As α is an algebraic integer, there is some monic $f \in \mathbb{Z}[X]$ such that $f(\alpha) = 0$ and so $I \neq 0$. Let $f_\alpha \in I$ be a non-zero polynomial of minimal degree in I. If f_α is not primitive, we can write $f_\alpha = c(f_\alpha) \cdot f'_\alpha$ for some primitive $f'_\alpha \in I$ and thus we can assume that f_α is primitive. Let $h \in I$. By the Euclidean algorithm in $\mathbb{Q}[X]$, we can write

$$h = f_\alpha \cdot q + r$$

for some $q, r \in \mathbb{Q}[X]$, with $r = 0$ or $\deg(r) < \deg(f_\alpha)$. By clearing denominators we know there is some $a \neq 0 \in \mathbb{Z}$ such that $aq, ar \in \mathbb{Z}[X]$ with

$$ah = f_\alpha \cdot (aq) + (ar).$$

Evaluating at α shows that $ar(\alpha) = 0$ and so $ar \in I$. Since f_α has a minimal degree among non-zero elements in I, we must have $ar = 0$ which implies that $r = 0$. Hence, $ah = (aq) \cdot f_\alpha$ is a factorisation in $\mathbb{Z}[X]$. Taking contents we get

$$\begin{aligned} ac(h) &= c(ah) \\ &= c\left(f_\alpha\right) \cdot c(aq) \\ &= c(aq), \end{aligned}$$

where equality is only up to associates. So $a \mid c(aq)$ which implies $aq = a\bar{q}$, for some $\bar{q} \in \mathbb{Z}[X]$. Since $a \neq 0$ we must have $q = \bar{q} \in \mathbb{Z}[X]$. Therefore, $h = f_\alpha \cdot q \in I$ and so it follows that $I = (f_\alpha)$. Since $\mathbb{Z}[\alpha] \leq \mathbb{C}$ and $\mathbb{Z}[X]/I$ is an integral domain we know that I is a prime ideal and hence f_α is prime which implies it is irreducible. Since I contains some monic polynomial f and $f_\alpha \mid f$, the leading coefficient of f_α must be a unit in $\mathbb{Z}$, that is ± 1. If it is -1, we take $-f_\alpha$. Either way, we may assume f_α is monic. □

Remark 2.6.6. *Requiring f_α to be monic in Proposition 2.6.5 ensures that f_α is determined uniquely.*

Definition 2.6.7. *For $\alpha \in \mathbb{C}$ an algebraic integer, we call the polynomial f_α of Proposition 2.6.5 the minimal polynomial of α.*

Example 2.6.8.

1. *$\alpha = (1+i)\sqrt{3}$ is an algebraic integer as $f\left((1+i)\sqrt{3}\right) = 0$ for $f(X) = X^4 + 36$. In particular, the linear factors of f over $\mathbb{C}[X]$ are $\left(X - (\pm 1 \pm i)\sqrt{3}\right)$. As $\mathbb{C}[X]$ is a unique factorisation domain, f_α is a product of some of these linear factors. The only combination leading to an integer polynomial gives rise to f. Therefore, f is the minimal polynomial of α.*
2. *$\alpha = i + \sqrt{3}$ is an algebraic integer as $f(\alpha) = 0$ for $f(X) = X^4 - 4X^2 + 16$. Using similar arguments as to those made in statement 1 we deduce that f is the minimal polynomial of α.*

Exercise 2.6.9. *Show that $2\cos\left(\frac{2\pi}{7}\right)$ is an algebraic integer, and find its minimal polynomial.*

Lemma 2.6.10. *Let $\alpha \in \mathbb{Q}$ be an algebraic integer. Then $\alpha \in \mathbb{Z}$.*

Proof. Let $f_\alpha \in \mathbb{Z}[X]$ be the minimal polynomial of α. Then f_α is irreducible and primitive. Hence, by Lemma 2.4.10 we have that f_α is irreducible in $\mathbb{Q}[X]$. Since $f_\alpha(\alpha) = 0$ and $\alpha \in \mathbb{Q}$ we know that $X - \alpha \mid f_\alpha$ in $\mathbb{Q}[X]$. Since f_α is irreducible and monic this implies that $f_\alpha = X - \alpha$. Since $f_\alpha \in \mathbb{Z}[X]$ we must have $\alpha \in \mathbb{Z}$. □

Example 2.6.11.

1. *$\alpha = \frac{1}{2}\left(1 + \sqrt{3}\right)$ is not an algebraic integer. Suppose it were an algebraic integer with $f(\alpha) = 0$ for $f \in \mathbb{Z}[X]$. As $f(X(1-X))$ is monic it follows that $\alpha(1-\alpha) = -\frac{1}{2}$ is an algebraic integer. However, this contradicts Lemma 2.6.10.*
2. *$\alpha = \frac{\sqrt{5}}{\sqrt{7}}$ is not an algebraic integer. Suppose it were an algebraic integer with $f(\alpha) = 0$ for $f \in \mathbb{Z}[X]$ monic. As $f\left(X^2\right) \in \mathbb{Z}[X]$ is monic it follows that $\alpha^2 = \frac{5}{7}$ is an algebraic integer. However, this contradicts Lemma 2.6.10.*

Remark 2.6.12. *The set of algebraic integers is a subring of $\mathbb{C}$.*

We now turn our attention to a specific subring of algebraic integers, namely $\mathbb{Z}[i]$. This is a unique factorisation domain for which we will characterise its prime, and hence irreducible, elements.

Proposition 2.6.13. *Let $p \in \mathbb{N}$ be a prime number. Then p is prime in $\mathbb{Z}[i]$ if and only if p cannot be written as $a^2 + b^2$ for $a, b \in \mathbb{Z} \setminus \{0\}$.*

Proof. ($\Rightarrow$). If $p = a^2 + b^2$, then $p = (a + ib)(a - ib)$. Meaning p is not irreducible and so it is not prime.
($\Leftarrow$). Suppose p is reducible in $\mathbb{Z}[i]$. That is, $p = uv$ with $u, v \in \mathbb{Z}[i]$ non-units. Taking the norm squared, we find that $p^2 = |u|^2|v|^2$. Since u, v are not units, we must have that $|u|^2, |v|^2 \neq 1$ and so $|u|^2 = p$ and $|v|^2 = p$. If $u = a + ib$, this says that $p = a^2 + b^2$. □

Lemma 2.6.14 is a technical lemma that will allow us to say more about the primes in $\mathbb{Z}[i]$.

Lemma 2.6.14. *Let $p \in \mathbb{N}$ be a prime number, and let $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ be the field with p elements. Then $\mathbb{F}_p^\times \cong C_{p-1}$ is cyclic of order $p - 1$.*

Proof. Note that $\mathbb{F}_p^\times$ has order $p - 1$, and is abelian. From the classification of finite abelian groups, if $\mathbb{F}_p^\times$ is not cyclic then it contains a subgroup isomorphic to $C_m \times C_m$ for some $m > 1$. Consider the polynomial $f = X^m - 1$. All elements of $C_m \times C_m$ are roots of f, so f has at least m^2 roots and hence it has at least m^2 distinct linear factors. However, $\mathbb{F}_p[X]$ is a unique factorisation domain so it cannot have more than m distinct linear factors. Therefore, $\mathbb{F}_p^\times$ is cyclic and of order $p - 1$ meaning $\mathbb{F}_p^\times \cong C_{p-1}$. □

Theorem 2.6.15. *The primes in $\mathbb{Z}[i]$ are one of the following, up to associates.*

1. *Prime numbers $p \in \mathbb{N} \subseteq \mathbb{Z}[i]$ which are equal to $3 \bmod 4$.*
2. *$z \in \mathbb{Z}[i]$ such that $|z|^2 = p$, with $p \in \mathbb{N}$ a prime number which is either 2 or $1 \bmod 4$.*

Proof. If $p \in \mathbb{N}$ is prime and equal to $3 \bmod 4$, then $p \neq a^2 + b^2$ for any $a, b \in \mathbb{Z}$ as any square number is always 0 or $1 \bmod 4$. Therefore, using Proposition 2.6.13 it follows that p is prime in $\mathbb{Z}[i]$. Similarly, if $|z|^2 = p$ is prime and $z = uv$, then $|u|^2|v|^2 = p$. Without loss of generality suppose that $|u|^2 = 1$ which implies it is a unit. Hence, z is irreducible and thus prime in $\mathbb{Z}[i]$. Now we show that irreducibles, and thus prime, elements of $\mathbb{Z}[i]$ satisfy statement 1 or statement 2. Let $z \in \mathbb{Z}[i]$ be irreducible. It cannot be a unit and so $|z|^2 > 1$. Moreover, $\bar{z}$ is also irreducible. So $|z|^2 = z\bar{z}$ is a factorisation of $|z|^2$ into irreducibles. Let $p \in \mathbb{N}$ be a prime factor of $|z|^2 > 1$. Then $p \mid z\bar{z}$ in $\mathbb{Z}[i]$.

1. If p is $3 \bmod 4$, then p is prime in $\mathbb{Z}[i]$. So $p \mid z$ or $p \mid \bar{z}$. Noting $p = \bar{p}$, either way, we must have $p \mid z$. Since both p and z are irreducible, they must be associates. So z satisfies statement 1.

2. If p is $1 \bmod 4$, then $p - 1 = 4k$ for some $k \in \mathbb{Z}$ and so $\mathbb{F}_p^\times \cong C_{4k}$ by Lemma 2.6.14. Let $a \in \mathbb{F}_p^\times$ be an element of order 4. Then a^2 is an element of order 2, of which there is only one, namely $a^2 = -1$. So there is some $a \in \mathbb{Z}$ with $p \mid a^2 + 1$. In other words, $p \mid (a + i)(a - i) = a^2 + 1$. If $p = 2$, we also note that $p \mid (a + i)(a - i)$ for $a = 1$. As, $p \nmid a + i$ for any prime p we note p is not prime in $\mathbb{Z}[i]$. Thus we can write $p = z_1 z_2$, with $z_1, z_2 \in \mathbb{Z}[i]$ non-units. Taking the norm squared we deduce that

$$p^2 = |z_1|^2 |z_2|^2 .$$

   Since z_1, z_2 are not units, we must have $|z_1|^2 = |z_2|^2 = p$. Therefore, $p = z_1\bar{z}_1 = z_2\bar{z}_2 = z_1 z_2$ which implies that $z_1 = \bar{z}_2$. So $p = z_1\bar{z}_1$ divides $|z|^2 = z\bar{z}$. As z is irreducible and $\mathbb{Z}[i]$ is a unique factorisation domain we must have that $z = z_1$ or $z = \bar{z}_1$. Either way $|z|^2 = p$. and z satisfies statement 2.

□

Remark 2.6.16. *The upshot of Theorem 2.6.15 is that the problem of finding primes in $\mathbb{Z}[i]$ is made equivalent*

to finding sums of two squares which are prime.

Corollary 2.6.17. *An integer $n \in \mathbb{N}$ is a sum of two squares, $n = x^2 + y^2$, if and only if when we write $n = p_1^{n_1} \dots p_k^{n_k}$ as a product of powers of distinct primes, n_i is even whenever p_i is $3 \bmod 4$.*

Proof. ($\Rightarrow$). Suppose $n = x^2 + y^2$ so that $n = |x + iy|^2$. Let $z = x + iy$. Write $z = \alpha_1 \dots \alpha_q$, where each $\alpha_i \in \mathbb{Z}[i]$ is irreducible. Then $n = z\bar{z} = |\alpha_1|^2 \dots |\alpha_q|^2$. By Theorem 2.6.15 each $|\alpha_i|^2$ is either p^2 for p equal to $3 \bmod 4$, or is a prime p which is 2 or equal to $1 \bmod 4$. Consequently, each prime which is $3 \bmod 4$ appears an even number of times in the prime factorisation of n.
($\Leftarrow$). Write $n = p_i^{n_1} \dots p_k^{n_k}$ as a product of powers of distinct primes. If p_1 is 2 or equal to $1 \bmod 4$, then $p_i = |\alpha_i|^2$ for some $\alpha_i \in \mathbb{Z}[i]$ which implies that $p_i^{n_i} = |\alpha_i^{n_i}|^2$. If p_i is $3 \bmod 4$, then $p_i^{n_i} = \left|p_i^{\frac{n_i}{2}}\right|^2$. Since $|\cdot|^2$ is multiplicative, we find that $n = |\beta|^2$ for some $\beta \in \mathbb{Z}[i]$. That is, n is the sum of two squares. □

Example 2.6.18. *Consider $65 = 5 \times 13$. As 5 and 13 are both congruent to $1 \bmod 4$, Corollary 2.6.17 tells us that 65 is a sum of two squares. Moreover, the proof of Corollary 2.6.17 gives us a way to write 65 as the sum of two squares. First, we factor 5 and 13 into irreducibles in $\mathbb{Z}[i]$ as*

$$5 = (2 + i)(2 - i)$$

and

$$13 = (2 + 3i)(2 - 3i).$$

Then we can write

$$65 = |2 + i|^2 \cdot |2 + 3i|^2 = |(2 + i)(2 + 3i)|^2 = |1 + 8i|^2 = 1^2 + 8^2$$

and

$$65 = |(2 + i)(2 - 3i)|^2 = |7 - 4i|^2 = 4^2 + 7^2.$$

As $\mathbb{Z}[i]$ is a unique factorisation domain we know these are the only ways of writing 65 as a sum of two squares.

2.7 Solution to Exercises

Exercise 2.1.10

Solution. Let $R = \{r_0, r_1, \dots, r_n\}$ be a ring where $r_0 = 0$ and $r_1 = 1$. For $r \in R \setminus \{0\}$ consider the map $\varphi_r : R \to R$ given by $\varphi_r(r_i) = rr_i$. If $\varphi_r(r_i) = \varphi_r(r_j)$ then $rr_i = rr_j$ which implies that $r(r_i - r_j) = 0$ meaning $r_i = r_j$ as R is an integral domain. Therefore, φ_r is injective and thus surjective as R is finite. In particular, φ_r is a bijection. Therefore, there exists an r_i such that $\varphi_r(r_i) = 1$ which implies that $rr_i = 1$. Moreover, as R is commutative we have that $rr_i = 1$, hence r_i is the unique multiplicative inverse of r. Repeating this for each $r \in R \setminus \{0\}$ we conclude that each non-zero element has a multiplicative inverse and hence R is a commutative division ring, namely R is a field. □

Exercise 2.2.11

Solution. Let $p \in R \setminus \{0\}$. Then $1 = qp + r$ for some $q, r \in R$ with $\theta(r) < \theta(1)$ or $r = 0$. Since $\theta(s) = 0$ for every $s \in R$ it follows that $r = 0$. Therefore, $1 = qp$ which means that $p \in R^\times$. It follows that $R = R^\times\{0\}$ which means that R is a field. □

Exercise 2.2.13

Solution. Let $\phi : \mathbb{Z}\left[\sqrt{2}\right] \to \mathbb{Z}_{\geq 0}$ be given by

$$a + b\sqrt{2} \mapsto \left|a^2 - 2b^2\right|.$$

For $a, b, c, d \in \mathbb{Z}$ we have

$$\begin{aligned}\phi\left(\left(a+b\sqrt{2}\right)\left(c+d\sqrt{2}\right)\right) &= \phi\left(ac+2bd+(ad+bc)\sqrt{2}\right)\\ &= \left|(ac)^2+4abcd+4(bd)^2-2\left((ad)^2+2abcd+(bc)^2\right)\right|\\ &= \left|(ac)^2+4(bd)^2-2(ad)^2-2(bc)^2\right|\\ &= \left|\left(a^2-2b^2\right)\left(c^2-2d^2\right)\right|\\ &= \phi\left(a+b\sqrt{2}\right)\phi\left(c+d\sqrt{2}\right).\end{aligned}$$

Moreover, if $\phi\left(a+b\sqrt{2}\right)=0$ then $a^2=2b^2$. We can assume that a, b are positive, as if a, b is a solution then so are $-a, b$ and $a, -b$ and $-a, -b$. Similarly, we can assume that a, b are non-zero, as if $a=0$ then $b=0$ and vice versa. So $2|a$ meaning $a=2k$, which implies that $4k^2=2b^2$ which implies $2k^2=b^2$. Therefore, $b=2p$ which implies that $k^2=2p^2$. As $k<a$ and $p<b$ we get ever smaller solutions which is a contradiction. Therefore, for $a+b\sqrt{2}\neq 0$ we have $\phi\left(a+b\sqrt{2}\right)>0$, and as it is an integer it must be greater than equal to one. Therefore, $\phi(zw)\geq\phi(w)$ for all $z, w\in\mathbb{Z}\left[\sqrt{2}\right]\setminus\{0\}$. Next, let $a+b\sqrt{2}, c+d\sqrt{2}\in\mathbb{Z}\left[\sqrt{2}\right]$. Then

$$\frac{a+b\sqrt{2}}{c+d\sqrt{2}}=\frac{(ac-2bd)+(bc-ad)\sqrt{2}}{c^2-2d^2}=:r_1+r_2\sqrt{2}\in\mathbb{Q}\left[\sqrt{2}\right].$$

We can find $q_1, q_2\in\mathbb{Z}$ such $|r_1-q_1|\leq\frac{1}{2}$ and $|r_2-q_2|\leq\frac{1}{2}$. Then let

$$\tilde{r}=(r_1-q_1)+(r_2-q_2)\sqrt{2}$$

so that

$$a+b\sqrt{2}=\left(q_1+q_2\sqrt{2}\right)\left(c+d\sqrt{2}\right)+\left(c+d\sqrt{2}\right)\tilde{r}.$$

Note that $\left(c+d\sqrt{2}\right)\tilde{r}\in\mathbb{Z}\left[\sqrt{2}\right]$ as the two other terms in the above expression are in $\mathbb{Z}\left[\sqrt{2}\right]$. Moreover,

$$\begin{aligned}\phi\left(\left(c+d\sqrt{2}\right)\tilde{r}\right) &= \phi\left(c+d\sqrt{2}\right)\phi\left(\tilde{r}\right)\\ &= \phi\left(c+d\sqrt{2}\right)\left|(r_1-q_1)^2-2(r_2-q_2)^2\right|\\ &\leq \phi\left(c+d\sqrt{2}\right)\left(\frac{1}{4}+2\frac{1}{4}\right)\\ &< \phi\left(c+d\sqrt{2}\right).\end{aligned}$$

Therefore, $\mathbb{Z}\left[\sqrt{2}\right]$ is a Euclidean domain. □

Exercise 2.2.25

Solution. As I_1 is an abelian group we have $0_R\in I_1$ which implies that $0_R\in I$. For $x, y\in I$, we have that $x\in I_i$ and $y\in I_j$ for some $i, j\in\mathbb{N}$. Suppose without loss of generality that $i\leq j$, then as $I_i\subseteq I_j$ we have $x\in I_j$. As I_j is an abelian group we have $x+y\in I_j$ which implies that $x+y\in I$. Therefore, I is an abelian group. Similarly, let $r\in R$, then for $x\in I$ we have that $x\in I_i$ for some $i\in\mathbb{N}$. As I_i is an ideal we have $rx\in I_j$, which implies that $rx\in I$. Therefore, I is an ideal. □

Exercise 2.3.4

Solution. Suppose that $(a_1, b_1)\sim(a_2, b_2)$ such that $a_1b_2=a_2b_1$. It follows that

$$\begin{aligned}(a_1, b_1)+(c, d) &= (a_1d+b_1c, b_1d)\\ &= (b_2a_1d+b_2b_1d, b_2b_1d)\\ &= (a_2b_1d+b_2b_1d, b_2b_1d)\\ &= (a_2d+b_2d, b_2d)\\ &= (a_2, b_2)+(c, d).\end{aligned}$$

Similarly,

$$\begin{aligned}(a_1, b_1)(c, d) &= (a_1 c, b_1 d) \\ &= (b_2 a_1 c, b_2 b_1 d) \\ &= (a_2 b_1 c, b_2 b_1 d) \\ &= (a_2 c, b_2 d) \\ &= (a_2, b_2)(c, d).\end{aligned}$$

□

Exercise 2.6.9

Solution. Let $\gamma = \cos\left(\frac{2\pi}{7}\right) + i \sin\left(\frac{2\pi}{7}\right)$ and $\alpha = 2\cos\left(\frac{2\pi}{7}\right)$. Note that

$$\gamma^3 + \gamma^2 + \gamma + 1 + \frac{1}{\gamma} + \frac{1}{\gamma^2} + \frac{1}{\gamma^3} = 0.$$

As

- $\alpha = \gamma + \frac{1}{\gamma}$,
- $\alpha^2 = \gamma^2 + \frac{1}{\gamma^2} + 2$, and
- $\alpha^3 = \gamma^3 + \frac{1}{\gamma^3} + 3\alpha$

we deduce that

$$\alpha^3 + \alpha^2 - 2\alpha - 1 = 0,$$

which implies that α is an algebraic integer. Moreover, this tells us that $f_\alpha | X^3 + X^2 - 2X - 1$. As $X^3 + X^2 - 2X - 1$ is irreducible in $\mathbb{Z}[X]$ it follows that

$$f_\alpha = X^3 + X^2 - 2X - 1.$$

□

3 Modules

3.1 Definition and Examples

A module to a ring is similar to what a vector space is to a field. Various parts of the theory of vector spaces will carry over, but many will not. In this section, we will consider a ring R that is not necessarily commutative.

Definition 3.1.1. *An R-module is a set M with operations $+ : M \times M \to M$, $\cdot : R \times M \to M$ and an element $0_M \in M$ such that $(M, +)$ is an abelian group with additive identity $0_M \in M$. Moreover, for all $r, r' \in R$ and $m, m' \in M$ the following statements are satisfied.*

1. $(r + r') \cdot m = r \cdot m + r' \cdot m$.
2. $r \cdot (m + m') = r \cdot m + r \cdot m'$.
3. $r \cdot (r' \cdot m) = (r \cdot r') \cdot m$.
4. $1_R \cdot m = m \cdot 1_R = m$.

Remark 3.1.2. *In Definition 3.1.1 we multiply elements of M by elements of R on the left. Consequently, we refer to such modules as left modules. There is an analogous notion of a right module, whose theory is essentially the same.*

An alternative characterisation of modules is given by homomorphisms. Recall that for an abelian group A, the set $\mathrm{End}(A)$ is the set of group homomorphisms $A \to A$. In particular, $\mathrm{End}(A)$ is a ring.

Definition 3.1.3. *An R-module is an abelian group M equipped with a ring homomorphism $\phi : R \to \mathrm{End}(M)$.*

Given such a ϕ, for $r \in R$ and $m \in M$ we write $r \cdot m$ for $(\phi(r))(m)$.

Proposition 3.1.4. *Definition 3.1.1 and Definition 3.1.3 are equivalent.*

Proof. Suppose that $R \times M \mapsto R$ satisfies the structure of a left R-module as given in Definition 3.1.1 and consider $\varphi : R \to \mathrm{End}(M)$ where $\varphi(r)(m) = r \cdot m$.

- By the structure of a R-module we know that

  $$\varphi(r)(m_1 + m_2) = r \cdot (m_1 + m_2) = r \cdot m_1 + r \cdot m_2 = \varphi(r)(m_1) + \varphi(r)(m_2).$$

  Moreover, $\varphi(r)(0) = r \cdot 0 = 0$. Therefore, $\varphi(r) : M \to M$ is a homomorphism, and thus φ is well-defined.

- By the structure of a R-module it follows that

  $$\varphi(r_1 + r_2)(m) = (r_1 + r_2) \cdot m = r_1 \cdot m + r_2 \cdot m = \varphi(r_1)(m) + \varphi(r_2)(m).$$

  Moreover,

  $$(\varphi(r_1) \circ \varphi(r_2))(m) = r_1 \cdot (r_2 \cdot m) = (r_1 \cdot r_2) \cdot m = \varphi(r_1 \cdot r_2)(m).$$

  So, φ is a homomorphism.

Conversely, it is clear that if we have a homomorphism $\varphi : R \to \mathrm{End}(M)$ then M possess the structure of an R-module due to the properties of φ. Moreover, constructing the homomorphism on the R-module that φ represents is the same as φ which shows there is a one-to-one correspondence between Definition 3.1.1 and Definition 3.1.3. □

One can think of an R-module as an R-action on an abelian group.

Example 3.1.5.

- *Let F be a field. Then an F-vector space is a F-module.*
- *Let $I \subseteq R$ be an ideal. Then I is an R-module, through the R-action $r \cdot a := r \cdot_R a$, for $r \in R$ and $a \in I$.*
- *R is an R-module. More generally, for $n \geq 1$ we have that R^n is an R-module through the R-action*

$$r \cdot (r_1, \ldots, r_n) := (r \cdot r_1, \ldots, r \cdot r_n)\,.$$

- *If $I \subseteq R$ is a two-sided ideal, then R/I is an R-module through the R-action $r \cdot (a + I) := (r \cdot a) + I$.*
- *A $\mathbb{Z}$-module is equivalent to an abelian group. Let A be an abelian group. For $n \in \mathbb{Z}$ and $a \in A$ we let*

$$n \cdot a = \begin{cases} \underbrace{a + \cdots + a}_{n} & n \geq 0 \\ \underbrace{(-a) + \cdots + (-a)}_{|n|} & n < 0. \end{cases}$$

Alternatively, there is a unique ring homomorphism $\mathbb{Z} \to \mathrm{End}(A)$ and so we can use this to endow A with the structure of a $\mathbb{Z}$-module.

3.2 Constructions

Definition 3.2.1. *Let $M_1, \ldots, M_k$ be R-modules. Their direct sum written $M_1 \oplus \ldots \oplus M_k$, is the abelian group $M_1 \times \cdots \times M_k$ with*

$$r \cdot (m_1, \ldots, m_k) := (r \cdot m_1, \ldots, r \cdot m_k)\,.$$

Example 3.2.2. *As R is an R-module we can let $R^n = \underbrace{R \oplus \cdots \oplus R}_{n}$.*

Definition 3.2.3. *A subset $N \subseteq M$ is an R-submodule if it is a subgroup of M such that for $r \in R$ and $n \in N$ we have $r \cdot n \in N$. In such a case, we write $N \leq M$.*

Example 3.2.4.

1. *For an R-module M, the sets $\{0\}$ and M are submodules M.*
2. *A subset $I \subseteq R$ is a submodule if and only if I is an ideal.*
3. *If F is a field, submodules and sub-vector spaces coincide.*

Definition 3.2.5. *Let $N \leq M$ be a submodule. The quotient module M/N is the abelian group M/N, with R-action $r \cdot (m + N) := (r \cdot m) + N$.*

Remark 3.2.6. *Groups have both subgroups and normal subgroups. Normal subgroups are a special type of the former that ensures quotient groups are well-defined. Similarly, rings have subrings and ideals, neither of which is a special type of the other but ideals ensure quotient rings are well-defined. Modules have submodules, which are sufficient to ensure that quotient modules are well-defined.*

Definition 3.2.7. *Let R, R' be rings, with M an R-module and M' an R'-module. Then $M \times M'$ is an $R \times R'$-module, with action*

$$(r, r') \cdot (m, m') := (r \cdot m, r' \cdot m') .$$

Definition 3.2.8. *Let R be a commutative ring with $S \subseteq R$ a multiplicative submonoid and M an R-module. The localisation of M by S denoted $S^{-1}M$, is the set of equivalence classes of pairs (m, s) for $m \in M$ and $s \in S$ where $(m, s) \sim (m', s')$ if and only if there is a $t \in S$ such that $t(ms' - m's) = 0$.*

Remark 3.2.9. *With the natural structure of the abelian group and the $S^{-1}R$-action $(r, t) \cdot (m, s) = (rm, ts)$ for $(r, t) \in S^{-1}R$ and $(m, s) \in S^{-1}M$ the localisation of M by S is a $S^{-1}R$-module.*

3.3 Homomorphisms

Definition 3.3.1. *A map $\phi : M \to N$ between R-modules is an R-module homomorphism if it is a homomorphism of abelian groups, and $\phi(r \cdot m) = r \cdot \phi(m)$ for all $r \in R$ and $m \in M$.*

An isomorphism is a bijective homomorphism.

Proposition 3.3.2. *The composition of R-module homomorphisms is an R-module homomorphism.*

Example 3.3.3. *The ideals $(2), (3) \subseteq \mathbb{Z}$ are distinct as ideals of $\mathbb{Z}$. However, through the map $\phi : (2) \to (3)$ given by $2k \mapsto 3k$ we see that (2) and (3) are isomorphic as R-modules. More generally, if R is an integral domain and $r \in R \setminus \{0\}$, then R and (r) are isomorphic as R-modules.*

Many proofs for this section are omitted as they follow analogous arguments as those made for rings.

Theorem 3.3.4 (First Isomorphism Theorem for Modules)**.** *Let $\phi : M \to N$ be an R-module homomorphism. Then the following statements hold.*

1. $\ker(\phi) \leq M$.
2. $\mathrm{im}(\phi) \leq N$.
3. $M/\ker(\phi) \cong \mathrm{im}(\phi)$.

Definition 3.3.5. *The cokernel of an R-module homomorphism $\phi : M \to N$, written $\mathrm{coker}(\phi)$, is the quotient module $N/\mathrm{im}(\phi)$.*

Remark 3.3.6. *The image of a ring homomorphism is not necessarily an ideal, similarly, the image of a group homomorphism is not necessarily a normal subgroup. Therefore, the cokernel for these objects is not well-defined.*

Theorem 3.3.7 (Second Isomorphism Theorem for Modules)**.** *Let M be an R-module and consider $A, B \leq M$. Then the following statements hold.*

1. $A + B := \{a + b : a \in A, b \in B\} \leq M$.
2. $A \cap B \leq M$.

3. $(A+B)/A \cong B/(A \cap B)$.

Remark 3.3.8. *More generally, for $S \subseteq \mathbb{N}$, if $(A_i)_{i\in S}$ is a collection of submodules, then*

$$\sum_{i\in S} A_i \leq M,$$

where the left-hand side is the set of finite sums of elements of the A_i. Similarly

$$\bigcap_{i\in S} A_i \leq M.$$

Theorem 3.3.9 (Third Isomorphism Theorem for Modules)**.** *For M and R-module consider $N \leq L \leq M$. Then the following statements hold.*

1. $L/N \leq M/N$.
2. $M/L \cong (M/N)/(L/N)$.

Proposition 3.3.10. *For M an R-module let $N \leq M$. Then there is a bijection between submodules of M/N and submodules of M that contain N.*

3.4 Generating Modules

Definition 3.4.1. *Let M be an R-module and consider $m \in M$. The submodule generated by m is*

$$Rm := \{r \cdot m : r \in R\} \leq M.$$

Alternatively, one can consider the homomorphism $\phi_m : R \to M$ given by $r \mapsto r \cdot m$. Then

$$Rm := \mathrm{im}(\phi_m).$$

Using this we have $\ker(\phi_m) \leq R$ which is referred to as the annihilator of m, in particular,

$$\mathrm{Ann}(m) := \{r \in R : r \cdot m = 0\} = \ker(\phi_m).$$

As $\mathrm{Ann}(m) \leq R$, it follows that it is a two-sided ideal of R. Moreover, by the first isomorphism theorem for modules we have that

$$Rm \cong R/\mathrm{Ann}(m).$$

Definition 3.4.2. *An R-module M is finitely generate if there are elements $\{m_1, \ldots, m_n\} \subseteq M$, such that*

$$\begin{aligned} M &= Rm_1 + \cdots + Rm_n \\ &= \{r_1m_1 + \cdots + r_nm_n : r_1, \ldots, r_n \in R\}. \end{aligned}$$

Example 3.4.3. *If F is a field and M is an F-module, then M is finitely generated as an F-module if and only if it is finite-dimensional as an F-vector space.*

Lemma 3.4.4. *An R-module M is finitely generated if and only if there is a surjective R-module homomorphism $R^n \to M$.*

Proof. ($\Rightarrow$). Suppose $M = Rm_1 + \cdots + Rm_n$ and let $\phi : R^n \to M$ be the R-module homomorphism given by

$$(r_1, \ldots, r_n) \mapsto r_1 m_1 + \cdots + r_n m_n.$$

Note that ϕ is surjective by assumption.
($\Leftarrow$). Suppose $\phi : R^n \to M$ is a surjective R-module homomorphism. Let

$$m_i := \phi(\underbrace{(0, \ldots, 0, 1_R}_{i}, 0, \ldots, 0)).$$

Then $M = Rm_1 + \cdots + Rm_n$ as ϕ is surjective. □

Corollary 3.4.5. *Let M be a finitely generated R-module. If $N \leq M$ then M/N is finitely generated.*

Proof. By Lemma 3.4.4 there exists $\phi : R^n \to M$ a surjective R-module homomorphism. As the quotient map $\psi : M \to M/N$ is surjective, the composition $\phi \circ \psi : R^n \to M/N$ is also surjective. Therefore, by Lemma 3.4.4 it follows that M/N is finitely generated. □

Proposition 3.4.6. *Let $R = \mathbb{Z}[X_1, X_2, \ldots]$ and let I be the ideal $I = (X_1, X_2, \ldots) \subseteq R$. Then $I \leq R$ is not a finitely generated R-module.*

Proof. We note that $I \leq R$ as $I \subseteq R$ is an ideal of R. Suppose $I = (f_1, \ldots, f_k)$ is finitely generated. Let $p \in \mathbb{N}$ be the largest number such that X_p appears in any of the f_i. Then $X_{p+1} \in I$ but $X_{p+1} \notin (f_1, \ldots, f_k)$ which contradicts $(f_1, \ldots, f_k)$ generating I. □

Remark 3.4.7. *The converse of Corollary 3.4.5 does not hold. That is if M is a finitely generated R-module and $N \leq M$, then it is not necessarily the case that N is finitely generated. An explicit example is given by Proposition 3.4.6 as $R = R1$ is a finitely generated R-module.*

3.5 Free modules

Definition 3.5.1. *Let S be a set. The free module over S, written $R^{(S)}$, is*

$$R^{(S)} = \bigoplus_{i \in S} R := \left\{(x_i)_{i \in S} : x_i \in R,\ x_i = 0 \text{ for all but finitely many } i\right\}$$

with coordinate-wise addition and R-action.

Proposition 3.5.2. *For a ring R and a set S, the free module $R^{(S)}$ is finitely generated if and only if S is finite.*

Proof. ($\Leftarrow$). If $|S| = n$, then $R^{(S)} \cong R^n$, so there is a surjective homomorphism $R^n \to R^{(S)}$. Therefore, by Lemma 3.4.4 it follows that $R^{(S)}$ is finitely generated.
($\Rightarrow$). Suppose S is infinite and $R^{(S)} = Rm_1 + \cdots + Rm_n$. Write each $m_k = \left(x_i^{(k)}\right)_{i \in S} \in R^{(S)}$ and consider

$$T := \left\{i \in S : x_i^{(k)} \neq 0 \text{ for some } 1 \leq k \leq n\right\} \subseteq S.$$

Note that T is a finite set, however, S is infinite and so we can find some $s \in S \backslash T$. Let $a = (a_i)_{i \in S} \in R^{(S)}$ be the element where

$$a_i := \begin{cases} 1 & i = s \\ 0 & i \neq s. \end{cases}$$

Then $a \notin Rm_1 + \cdots + Rm_n$, which contradicts $\{m_1, \ldots, m_n\}$ generating $R^{(S)}$. Therefore, if $R^{(S)}$ is finitely generated it must be the case that S is finite. □

Theorem 3.5.3. *Let R be a non-trivial commutative ring. Then R is a field if and only if every finitely generated R-module is free.*

Proof. $(\Rightarrow)$. Let R be a field and let M be a finitely generated R-module. Let n be the minimum size of a generating set for M.
Claim 1: If $S \subseteq M$ is linearly independent then $|S| \leq n$.
Proof of Claim 1: If $|S| > n$ then $\{v_1, \ldots, v_{n+1}\} \subseteq S$ is linearly independent. Suppose M is generated by $\{m_1, \ldots, m_n\}$. The for $1 \leq i \leq n+1$ we can write

$$v_i = \sum_{j=1}^{n} a_{ij} m_j.$$

Let $A = (a_{ij}) \in M_{(n+1)\times n}(R)$, where recall R is a field. For $x \in R^n$ problem $A^\top x = 0$ is a system of $n+1$ equations in n variables and so there exists a non-zero solution $x \in R^n \setminus \{0\}$. Hence,

$$\sum_{i=1}^{n+1} x_i v_i = \sum_{j=1}^{n} \left(A^\top x\right)_j m_j = 0,$$

which contradicts the linear independence assumption.
Claim 2: If $S \subseteq M$ is linearly independent and $v \notin \operatorname{span}(S)$, then $S \cup \{v\}$ is linearly independent.
Proof of Claim 2: Suppose that $S \cup \{v\}$ is not linearly independent such that exists $a, a_i \in R$, with at least one being non-zero, such that

$$av + \sum_{i=1}^{n} a_i v_i = 0.$$

Since S is linearly independent it must be the case that $a \neq 0$. Since R is a field we have $a \in R^\times$ and so

$$v = -a^{-1} \sum_{i=1}^{n} a_i v_i$$

which contradicts v not being in the span of S.
If $M = \{0\}$ then M is free. If $M \neq \{0\}$ let $S_1 = \{v_1\}$ for some $v_1 \neq 0$. Then S is linearly independent as R is a field meaning it has no non-zero zero divisors. If S_1 spans M then S_1 is a basis for M and we are done. If not, using Claim 2, we can extend S_1 to and linearly independent set $S_2 = \{v_1, v_2\}$ where $v_2 \notin \operatorname{span}(S_1)$. We can continue this process, which must terminate by Claim 1. Therefore, M has a basis and is thus free.
$(\Leftarrow)$. Suppose that R is not a field. Then there exists an $a \in R \setminus \{0\}$ that is not a unit. Consider the R-module $M = R/(a)$. Let $m \in M$, then $a \cdot m = 0$. Thus if $f : M \to R^{(S)}$ is an isomorphism it follows that $a \cdot m = 0$ for all $m \in R^{(S)}$.
Claim 3: If $r \cdot m = 0$ for all $m \in R^{(S)}$ then $r = 0$.
Proof of Claim 3: If $S = \emptyset$ then the claim is true. Suppose $S \neq \emptyset$. Let $s \in S$ and let $m = (x_i)_{i \in S}$ be such that $x_s = 1$. Restricting $r \cdot m = 0$ to the s^{th} coordinate gives $r = r \cdot 1 = 0 \in R$.
Using Claim 3 it follows that $a = 0$ which is a contradiction. Therefore, R is a field. □

Remark 3.5.4. *Theorem 3.5.3 is equivalent to the axiom of choice.*

Definition 3.5.5. *A subset $S \subseteq M$ generates M freely if the following statements hold.*

1. *S generates M as an R-module, namely*

$$R \cdot S := \left\{ \sum_{s \in S} r_s s : r_s \in R \text{ with only finitely many non-zero} \right\} = M$$

2. *For any R-module N, a map $\psi : S \to N$ can be extended to an R-module homomorphism $\phi : M \to N$.*

Remark 3.5.6. *Suppose that $S \subseteq M$ generates M freely as formulated in Definition 3.5.5 and suppose that $\phi, \phi' : M \to N$ are extensions of the map $\psi : S \to N$. Then $\phi - \phi' : M \to N$ is a homomorphism sending all of S to zero, meaning $S \subseteq \ker(\phi - \phi')$. Since S generates M, this implies $\ker(\phi - \phi') = M$. In other words, $\phi = \phi'$ and so the extension of ψ to an R-module homomorphism is unique.*

Definition 3.5.7 provides an alternative, but equivalent, definition for a free module.

Definition 3.5.7. *An R-module M is free if it is freely generated by some subset $S \subseteq M$. Such a subset $S \subseteq M$ is called a basis for M.*

Proposition 3.5.8. *Definition 3.5.1 and Definition 3.5.7 are equivalent.*

Proof. Suppose that $M \cong R^{(S)}$. Note that we can identify S as a subset of M by considering the elements $(x_i^s)_{i \in S}$ where

$$x_i^s = \begin{cases} 1 & i = s \\ 0 & \text{otherwise.} \end{cases}$$

Then $(x_i^s)_{i \in S}$ generates M as an R-module. Let N be an R-module and let $\psi : S \to N$ be a map. Then the map $\phi : M \to N$ given by

$$(x_i)_{i \in S} \mapsto \sum_{i \in S} x_i \cdot \psi(i)$$

is a homomorphism extending ψ. Therefore, S generates M freely. Conversely, suppose that $S \subseteq M$ freely generates M. Let $\psi : S \to R^{(S)}$ be given by $s \mapsto (x_i^s)_{i \in S}$. This extends to a homomorphism $\phi : M \to R^{(S)}$ which is an isomorphism as the map $(x_i)_{i \in S} \mapsto \sum_{i \in S} x_i \cdot i$ is its inverse. Therefore, $M \cong R^{(S)}$. □

Lemma 3.5.9. *Suppose M and N are R-modules where M is free with a basis $S \subseteq M$ and N is free with a basis T. If there is a bijection between S and T then $M \cong N$ as R-modules.*

Proof. Consider the injective functions $i_M : S \to M$, given by the inclusion map, and $i_N : S \to N$, given by composing the bijection from S to T with the inclusion map of T into N. Since S generates M freely it follows that i_N extends to a R-module homomorphism $\theta_N : M \to N$. Similarly, i_M extends to a R-module homomorphism $\theta_M : N \to M$. Note that $\theta_M \circ \theta_N \circ i_M = i_M$ and $\mathrm{id}_M \circ i_M = i_M$. Hence, $i_M : S \to M$ extends to $\theta_M \circ \theta_N : M \to M$ and $\mathrm{id}_M : M \to M$. As the extension is unique, by the arguments of Remark 3.5.6, it follows that $\theta_M \circ \theta_N = \mathrm{id}_M$. Similarly, we have $\theta_N \circ \theta_M = \mathrm{id}_N$ and so $M \cong N$. □

Definition 3.5.10. *A set of elements $\{m_1, \ldots, m_n\} \subseteq M$ is linearly independent, if whenever*

$$r_1 \cdot m_1 + \cdots + r_n \cdot m_n = 0$$

for $r_i \in R$ we have $r_1 = \cdots = r_n = 0$.

Proposition 3.5.11. *For a subset $S = \{m_1, \ldots, m_n\} \subseteq M$ the following statements are equivalent.*

1. *S generates M freely.*
2. *S generates M and S is linearly independent.*
3. *For every $m \in M$ we can write*
$$m = r_1 \cdot m_1 + \ldots + r_n \cdot m_n$$

uniquely for $r_i \in R$.

Example 3.5.12.

1. *Consider $\mathbb{Z}/2\mathbb{Z}$ as a $\mathbb{Z}$-module. Suppose $\mathbb{Z}/2\mathbb{Z}$ were freely generated by $S \subseteq \mathbb{Z}/2\mathbb{Z}$, then $S = \{1\}$ since any set containing zero cannot be linearly independent. But S is not linearly independent as $2 \cdot 1 = 0$ with $2 \neq 0$. Therefore, $\mathbb{Z}/2\mathbb{Z}$ is not freely generated as a $\mathbb{Z}$-module.*
2. *The subset $\{2, 3\} \subseteq \mathbb{Z}$ generates $\mathbb{Z}$, however, it is not linearly independent as $3 \cdot 2 + (-2) \cdot 3 = 0$. Therefore, $\{2, 3\}$ does not generate $\mathbb{Z}$ freely. For an F-vector space V recall that if S spans V but is not linearly independent, then one can discard elements of S to arrive at $T \subseteq S$ that is linearly independent whilst still spanning V. However, no subset of $\{2, 3\}$ freely generates $\mathbb{Z}$.*

Definition 3.5.13. *For a finitely generated R-module M, there exists a surjective R-module homomorphism $\varphi : R^n \to M$ for some $n \in \mathbb{N}$ by Lemma 3.4.4. The relation module for those generators is $\ker(\varphi)$.*

Definition 3.5.14. *A finitely generated R-module M is finitely presented if there exists a surjective homomorphism $R^n \to M$ that has a finitely generated kernel.*

Remark 3.5.15.

1. *Equivalently, we can say that an R-module M is finitely presented if M is isomorphic to the cokernel of some homomorphism $\phi : R^m \to R^n$. To see this we note that a surjective homomorphism $\varphi : R^n \to M$ exists as M is finitely generated. Similarly, for some m there exists surjective homomorphism $\phi : R^m \to \ker(\phi)$. As $\ker(\phi) \leq R^n$ we have that $\psi : R^m \to R^n$. In particular, by the first isomorphism theorem, we have that*
$$\operatorname{coker}(\phi) = R^n / \ker(\varphi) \cong M.$$
2. *Note that a finitely presented module is finitely generated, however, a finitely generated module is not necessarily finitely presented.*

Proposition 3.5.16. *Let M be an R-module with $N \leq M$. If M/N is free then $M \cong N \oplus M/N$.*

Proof. As M/N is free it has a basis $S = (s_i + N)_{i \in \mathcal{I}} \subseteq M/N$. Let $\pi : M \to M/N$ be the map given by $m \mapsto m + N$. Let $T = (t_i)_{i \in \mathcal{I}} \subseteq M$ be such that $\pi(t_i) = s_i + N$.
Claim 1: $M = N \oplus R \cdot T$.
Proof of Claim 1: Suppose $m \in N \cap R \cdot T$, so that $m = \sum_{i \in \mathcal{I}} r_i t_i$ for some $r_i \in R$ with only finitely many r_i non-zero. Then $\pi(m) = \sum_{i \in \mathcal{I}} r_i(s_i + N)$. As $m \in N$ we also know that $\pi(m) = 0$, and so $0 = \sum_{i \in \mathcal{I}} r_i(s_i + N)$. Hence, by the linear independence of the basis, we deduce that $m = 0$. Now for any $m \in M$ we can write $\pi(m) = m + N = \sum_{i \in \mathcal{I}} r_i(s_i + N)$. Let $m' = \sum_{i \in \mathcal{I}} r_i t_i$. Then as
$$\pi(m - m') = \pi(m) - \pi(m') = N,$$
it follows that $m - m' \in N$, which implies that $m - m' = n$ for some $n \in N$. Therefore, $m = n + \sum_{i \in \mathcal{I}} r_i t_i$. In conclusion, we have that $M = N \oplus R \cdot T$.
As $R \cdot T \cong M/N$ we deduce that $M \cong N \oplus M/N$. □

Example 3.5.17. *Let M and M' be R-modules with $N \leq M$ and $N' \leq M'$ where $N \cong N'$. One can use Proposition 3.5.16 to deduce that $M \cong M'$ when M/N is free. However, knowing only that N is free is*

insufficient to deduce that $M \cong M'$. Indeed, let $M = \mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ be the $\mathbb{Z}$-module with $\mathbb{Z}$-action given by

$$r \cdot (a, b) = (r \cdot a, r \cdot b).$$

Then, $N = 3\mathbb{Z} \times \{0\}$ is a submodule of M with $M/N = \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. Similarly, let $M' = \mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$ be the $\mathbb{Z}$-module with the same $\mathbb{Z}$-action as M. Then, $N' = 2\mathbb{Z} \times \{0\}$ is a submodule of M' with $M'/N' = \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$. Observe that N is a free module as it is generated by the set $\{(3, 0)\}$ which is also a linearly independent set in $\mathbb{Z}$. Moreover, N and N' are isomorphic as $\mathbb{Z}$-modules through the map $(3k, 0) \mapsto (2k, 0)$. Similarly, M/N is isomorphic to M'/N' as $\mathbb{Z}$-modules through the map $(a, b) \mapsto (b, a)$. However, $M = \mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ and $M' = \mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$ are not isomorphic. Suppose that such an isomorphism $\varphi : M \to M'$ existed and suppose that $\varphi(0, 1) = (a, b)$. Then

$$(a, b) = \varphi(0, 1) = 3 \cdot \varphi(0, 1) = 3 \cdot (a, b) = (3a, 0)$$

which implies that $b = 0$ and $a = 3a$. Thus, $\varphi(0, 1) = (0, 0)$ which contradicts injectivity as we must have $\varphi(0, 0) = (0, 0)$ for φ to be a homomorphism.

Let $u_1, \ldots, u_m \in R^m$ and $v_1, \ldots, v_n \in R^n$ be the standard basis elements. Let $\phi : R^m \to R^n$ be an R-module homomorphism. We can write

$$\phi(u_j) = \sum_{i=1}^{n} a_{ij} v_i$$

for each $j = 1, \ldots, m$. Let $A = (a_{ij}) \in M_{n \times m}(R)$, then we can write

$$\phi(r) = \phi\left(\sum_{j=1}^{m} r_j u_j\right) = \sum_{j=1}^{m} r_j \phi(u_j) = \sum_{j=1}^{m} \sum_{i=1}^{n} r_j a_{ij} v_i = A \cdot r.$$

Thus, ϕ is given by right-multiplication by an appropriate matrix. If R is commutative, then we would equally have $\phi(r) = r \cdot B$. Hence, we arrive at a bijection between $M_{n \times m}(R)$ and the set of R-module homomorphism $R^m \to R^n$. Recall that from linear algebra, for a field F is $F^m \cong F^n$ it follows that $m = n$. We will see that the same is true from commutative rings.

Theorem 3.5.18. *Any commutative ring contains a maximal ideal.*

Theorem 3.5.18 is equivalent to the axiom of choice.

Theorem 3.5.19. *Let R be a commutative ring, and suppose that $R^n \cong R^m$ as R-modules. Then $n = m$.*

Proof. Using Theorem 3.5.18 we can consider $I \subseteq R$ a maximal ideal so that $F = R/I$ is a field. For M an R-module, consider

$$IM := \{i_1 \cdot m_1 + \cdots + i_k \cdot m_k : i_j \in I,\, m_j \in M\} \leq M.$$

Then M/IM is an R-module. In particular, M/IM is an F-module, where for $r + I \in F$ and $m + IM \in M/IM$, we let

$$(r + I) \cdot (m + IM) := (r \cdot m) + IM.$$

If $R^n \cong R^m$ as R-modules, it follows that

$$R^n/IR^n \cong R^m/IR^m \tag{3.5.1}$$

as F-modules. For $k \in \mathbb{N}$ let $\psi : R^k/IR^k \to F^k = (R/I)^k$ be given by

$$(r_1, \ldots, r_k) + IR^k \mapsto (r_1 + I, \ldots, r_k + I)\,.$$

It turns out that ψ is an isomorphism, meaning $R^k/IR^k \cong F^k$. Consequently, from (3.5.1) we get that $F^n \cong F^m$ as F-modules. Since F is a field, this implies that $n = m$. □

Exercise 3.5.20. *Let R be a ring and M the free R-module with basis $(x_i)_{i\in\mathbb{N}}$. Let $S = \mathrm{End}_R(M)$. Show that $S \cong S^2$ as S-modules.*

Remark 3.5.21. *The assumption that the rings are commutative is necessary for Theorem 3.5.19. Indeed Exercise 3.5.20, shows that without this assumption the result no longer holds.*

Definition 3.5.22. *An R-module is stably free if $M \oplus R^n$ is a free module for some n.*

Definition 3.5.23. *An R-module M is projective if $M \oplus N$ is a free R-module for some R-module N.*

Example 3.5.24. *Let R_1 and R_2 be non-trivial rings. Let $R = R_1 \times R_2$. Then R_1 is an R-module with R-action $r \cdot (r_1, r_2) = (r \cdot r_1, r_2)$. Similarly, R_1 is an R-module with R-action $r \cdot (r_1, r_2) = (r_1, r \cdot r_2)$. As R modules we have that $R_1 \oplus R_2 \cong R$ meaning R_1 and R_2 are projective R-modules. However, neither R_1 nor R_2 is free as R-modules. As an explicit example consider $\mathbb{Z}/6\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3$, where $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/3\mathbb{Z}$ are projective modules by our above arguments. However, they are not stably free. To see this note that any stably free $\mathbb{Z}/6\mathbb{Z}$ must have order 6^n for some n. However, $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/3\mathbb{Z}$ are both finitely generated with orders 2 and 3 respectively.*

3.6 Noetherian Modules

Definition 3.6.1. *For a ring R and an R-module M, we call M Noetherian if every increasing chain*

$$N_0 \subseteq N_1 \subseteq N_2 \subseteq \dots$$

of R-submodules of M is eventually constant.

Remark 3.6.2. *Note that R-submodules correspond to ideals of the ring R. Hence, a ring R is Noetherian if and only if it is a Noetherian as an R-module.*

Theorem 3.6.3. *An R-module M is Noetherian if and only if every R-submodule of M is finitely generated.*

Proof. $(\Rightarrow)$. Let N be an R-submodule of M. Let $n_0 \in N$ and let N_0 be the R-submodule of N generated by n_0. If $N_0 = N$ then N is finitely generated. Otherwise, let $n_1 \in N \setminus N_0$ and let N_1 be the R-submodule of N generated by n_1. If N is not finitely generated we can continue this process to generate R-submodules N_k of N that are generated by $n_k \in N$ with $n_k \notin N_{k-1}$. Consequently, we obtain a strictly increasing infinite chain

$$N_0 \subseteq N_1 \subseteq \dots$$

of R-submodules of M which contradicts M being Noetherian.
$(\Leftarrow)$. Let

$$N_0 \subseteq N_1 \subseteq \dots$$

be an increasing chain of R-submodules of M. Let $N = \bigcup_{i=1}^{\infty} N_i$, which we note is also an R-submodule and thus must be finitely generated. Let

$$N = Rn_0 + \dots + Rn_r.$$

Each $n_i \in N_{j_i}$ for some j_i. Let $j = \max_{i\in\{1,\dots,n\}}(j_i)$. Then $\{n_0, \dots, n_r\} \subseteq N_i$ for $i \geq j$ and so $N \subseteq N_j$ which implies that $N_i = N$ for all $i \geq j$. Therefore, the chain is eventually constant meaning N is Noetherian. □

Corollary 3.6.4. *Every principal ideal domain R as an R-module is Noetherian.*

Proof. As R-submodules are ideals of R as a ring. Every R-submodule is finitely generated as the ideal is a principal ideal. Therefore, by Theorem 3.6.3 we have that R is Noetherian. □

Proposition 3.6.5. *If M is a Noetherian R-module, then a submodule $N \leq M$ is Noetherian and the quotient M/N is Noetherian.*

Proof. Since M is Noetherian, any R-submodule is finitely generated by Theorem 3.6.3. Therefore, any R-submodule of N is also finitely generated meaning N is Noetherian by Theorem 3.6.3. Let J be an R-submodule with $\tilde{J}$ its pre-image in N under the map $\varphi : M \to M/N$ given by $m \mapsto m + N$. Then $\tilde{J}$ is finitely generated, with the image of the generating set of $\tilde{J}$ under φ being a generating set for J. Meaning J is finitely generated and so we conclude that M/N is Noetherian using Theorem 3.6.3. □

Proposition 3.6.6. *Let M be an R-module. If there exists a Noetherian submodule N such that M/N is Noetherian then M is Noetherian.*

Proof. Let $J \leq M$, then $J \cap N \leq N$ and so it is finitely generated by Theorem 3.6.3. Let $\{j_1, \ldots, j_n\}$ generated $J \cap N$. Let $\varphi : M \to M/N$ be given by $m \mapsto m + N$ and consider $\bar{J} := \varphi(J)$. As $\bar{J} \leq M/N$ it follows by Theorem 3.6.3 that $\bar{J}$ is finitely generated. Let $\{\bar{j}_{n+1}, \ldots, \bar{j}_m\}$ generate $\bar{J}$ and let $j_k \in J$, for $k \in \{n+1, \ldots, m\}$, be such that $\varphi(j_k) = \bar{j}_k$. Now for any $j \in J$, let $\bar{j} := \varphi(j)$. Then as $\bar{j} \in M/N$ we can write

$$\bar{j} = \sum_{k=n+1}^{m} r_k \bar{j}_k$$

for some $r_k \in R$. Observe that for $j' := j - \bar{j}$ we have $\varphi(j') = 0$ and so $j' \in J \cap N$. Consequently, we can write

$$j' = \sum_{k=1}^{n} r_k j_k$$

for some $r_k \in R$. Therefore,

$$j = \sum_{k=1}^{m} r_k j_k,$$

which shows that J is finitely generated. Using Theorem 3.6.3 we conclude that M is Noetherian. □

Corollary 3.6.7. *If M and N are Noetherian R-modules, then $M \oplus N$ is a Noetherian R-module.*

Proof. The map $\varphi : M \oplus N \to N$ given by $(m, n) \mapsto n$ is a surjection. It is clear that

$$\ker(\varphi) = \{(m, 0) : m \in M\} \cong M,$$

and so $\ker(\varphi)$ is Noetherian. As

$$(M \oplus N)/\ker(\varphi) \cong N,$$

by the first isomorphism theorem, and N is Noetherian, it follows that $(M \oplus N)/K$ is Noetherian. Therefore, using Proposition 3.6.6 we have that $M \oplus N$. □

Remark 3.6.8. *Using Corollary 3.6.7 it is clear that if R is a Noetherian then so is R^n for any $n \in \mathbb{N}$.*

Theorem 3.6.9. *Any finitely generated module over a Noetherian ring is Noetherian.*

Proof. Let M be a finitely generated R-module with a generating set $\{m_1, \ldots, m_s\}$. Let F be a free R-module of rank s with a generating set $\{e_1, \ldots, e_s\}$. Then $\varphi : F \to M$ with $e_i \to m_i$ for each i is a surjection. In particular,

$$F/\ker(\varphi) \cong M.$$

With Corollary 3.6.8 we have that F is Noetherian. Thus, with Proposition 3.6.5 we have that $\ker(\varphi)$ is Noetherian as it is a submodule of F. Therefore, with Proposition 3.6.5 we have that M is Noetherian. □

Corollary 3.6.10. *For a Noetherian ring R, every finitely generated R-module is finitely presented.*

Proof. Let M be a finitely generated R-module, such that there exists a surjective map $f : R^n \to M$ for some n. Since R is Noetherian it follows that R^n is Noetherian by Remark 3.6.8. Therefore, with Proposition 3.6.5 we have that $\ker(f)$ is Noetherian as it is a submodule of R^n. Thus $\ker(f)$ is finitely generated, using Theorem 3.6.3, which means it is finitely presented. □

3.7 Modules over Principal Ideal Domains

Theorem 3.7.15 generalises the classification theorem for finitely generated abelian groups, which are just $\mathbb{Z}$-modules. An outline for the proof Theorem 3.7.15 is as follows.

1. For R a commutative principal ideal domain let M be a finitely generated R-module.
2. Show that M is finitely presented, with $M \cong \operatorname{coker}(\phi_A : R^n \to R^m)$ for some matrix $A \in M_{n\times m}(R)$, where $\phi_A(x) := A \cdot x$.
3. Show that if $B = PAQ$, for invertible matrices P and Q, then ϕ_A and ϕ_B have isomorphic cokernels.
4. Show that for any matrix A over R, there are invertible matrices P and Q such that PAQ is a rectangular diagonal matrix.
5. Combine these steps to prove Theorem 3.7.15.

Lemma 3.7.1. *Let R be a principal ideal domain and $N \leq R^n$ as R-modules. Then $N \cong R^k$ for some $k \leq n$. In particular, N is finitely generated and free.*

Proof. We proceed by induction on n.

- If $n = 0$ then $R^0 = \{0\}$ and so $N = \{0\} \cong R^0$. If $n = 1$, the $N \subseteq R$ is an ideal and so $N = (\alpha)$ for some $\alpha \in R$. Recall that a principal ideal domain is an integral domain, and so using Example 3.3.3 we have that

$$\begin{cases} N \cong R^1 & \alpha \neq 0 \\ N \cong R^0 & \alpha = 0. \end{cases}$$

- Suppose the result holds when $N \leq R^{n-1}$. Now let $N \leq R^n$ and consider the homomorphism $\pi_n : R^n \to R$ where

$$(r_1, \ldots, r_n) \mapsto r_n.$$

  Then $\pi_n(N) \leq R$ as R-modules, and hence $\pi_n(N) = (\alpha)$ for some $\alpha \in R$. Moreover, note that $\ker(\pi_n) \cong R^{n-1}$.

  1. If $\alpha = 0$, then $N \leq \ker(\pi_n) \cong R^{n-1}$, so by the induction hypothesis we are done.

2. If $\alpha \neq 0$, we can pick some $\beta \in N$ such that $\pi_n(\beta) = \alpha$. By the induction hypothesis, there exist $\{x_1, \ldots, x_k\} \subseteq N \cap \ker(\pi_n)$ which freely generate $N \cap \ker(\pi_n)$, with $k \leq n-1$. Let $S = \{x_1, \ldots, x_k, \beta\} \subseteq N$. Let $y \in N$. Then $\pi_n(y) = r\alpha$, for some $r \in R$ which implies that $y - r\beta \in N \cap \ker(\pi_n)$. Consequently, we can write

$$y - r\beta = \sum_{i=1}^{k} r_i x_i$$

and thus

$$y = r\beta + \sum_{i=1}^{k} r_i x_i,$$

which means that S generates $N \cap \ker(\pi_n)$. Now suppose

$$0 = r_1 x_1 + \cdots + r_k x_k + r\beta,$$

for $r_i, r \in R$. Then

$$0 = \pi_n(r_1 x_1 + \cdots + r_k x_k + r\beta) = r\alpha$$

which implies that $r = 0$. Since $\{x_1, \ldots, x_k\}$ is linearly independent, we must have $r_1 = \cdots = r_k = 0$ and hence S is linearly independent. Therefore, N is finitely generated and free.

□

Corollary 3.7.2. *Let M be a finitely generated module over a principal ideal domain R. Then M is finitely presented.*

Proof. As M is finitely generated there exists a surjective homomorphism $\phi : R^n \to M$. As $\ker(\phi) \leq R^n$ as R-modules we can apply Lemma 3.7.1 to deduce that $\ker(\phi) \cong R^k$ for some $k \in \mathbb{N}$. Let $\psi : R^k \to R^n$ have image $\ker(\phi)$, then

$$M \cong R^n / \ker(\phi) = \operatorname{coker}(\psi),$$

which shows that M is finitely presented. □

Definition 3.7.3. *Let R be a ring. Matrices $A, B \in M_{m\times n}(R)$ are equivalent if there are invertible matrices $P \in \mathrm{GL}_n(R)$ and $Q \in \mathrm{GL}_m(R)$ such that $B = PAQ$.*

For a matrix $A \in M_{m\times n}(R)$ we can consider the corresponding homomorphism $\phi_A(x) = A \cdot x$. With this view Definition 3.7.3 is equivalent to say that there are isomorphisms $f = \phi_P : R^n \to R^n$ and $g = \phi_Q : R^m \to R^m$ such that $f \circ \phi_B = \phi_A \circ g$.

Exercise 3.7.4. *Let $A, B \in M_{m\times n}(R)$. Show that the relation $A \sim B$ if and only if A and B are equivalent, as formulated in Definition 3.7.3, is an equivalence relation.*

Proposition 3.7.5. *Let $A, B \in M_{m\times n}(R)$. If A and B are equivalent, then $\operatorname{coker}(\phi_A) \cong \operatorname{coker}(\phi_B)$ are isomorphic R-modules.*

Proof. Let $f : R^n \to R^n$ and $g : R^m \to R^m$ be isomorphisms such that $f \circ \phi_B = \phi_A \circ g$. Then

$$\operatorname{im}(\phi_A) = \operatorname{im}(\phi_A \circ g) = \operatorname{im}(f \circ \phi_B) = f(\operatorname{im}(\phi_B)). \tag{3.7.1}$$

Let

$$\psi : \operatorname{coker}(\phi_B) = R^n / \operatorname{im}(\phi_B) \to \operatorname{coker}(\phi_A) = R^n / \operatorname{im}(\phi_A)$$

be given by

$$x + \operatorname{im}(\phi_B) \mapsto f(x) + \operatorname{im}(\phi_A).$$

This is a well-defined homomorphism by (3.7.1) and in particular an isomorphism as f is an isomorphism. □

Remark 3.7.6. *When $R = F$ is a field, the result of Proposition 3.7.5 is an established result in linear algebra. More specifically, in this case, equivalent matrices have the same rank. Hence,* $\operatorname{coker}(\phi_A)$ *and* $\operatorname{coker}(\phi_B)$ *have the same dimension meaning that they are isomorphic.*

Definition 3.7.7. *Let $A \in M_{n\times n}(R)$. Then A is an elementary matrix if it is of one of the following forms.*

1. *A is the identity matrix, with $A_{ij} = c \in R$ for some $i \neq j$.*
2. *A is the identity matrix, with $A_{ii} = A_{jj} = 0$ and $A_{ij} = A_{ji} = 1$ for some $i \neq j$.*
3. *A is the identity matrix, with $A_{ii} = c \in R^\times$ for some $1 \leq i \leq n$.*

Let $A \in M_{m\times n}$. Then an elementary row (column) operation on A is given by left (right) multiplying A by an elementary matrix P.

- Elementary matrices of type 1 correspond to adding $c \in R$ times the i^{th} row (column) to the j^{th} row (column).
- Elementary matrices of type 2 correspond to swapping the i^{th} and j^{th} rows (columns).
- Elementary matrices of type 3 correspond to multiplying the i^{th} row (column) by $c \in R^\times$.

Remark 3.7.8. *Note that elementary matrices are invertible.*

Proposition 3.7.9. *Let $A, B \in M_{m\times n}(R)$. If B is obtained from A by row and column operations, then A and B are equivalent matrices. In particular,* $\operatorname{coker}(\phi_A) \cong \operatorname{coker}(\phi_B)$.

Definition 3.7.10. *A matrix $A \in M_{m\times n}(R)$ is in Smith normal form if A is a rectangular diagonal matrix,*

$$A = \begin{pmatrix} d_1 & & & & & 0 \\ & \ddots & & & & \\ & & d_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ 0 & & & & & 0 \end{pmatrix}$$

with all $d_i \neq 0$ and $d_1 | \ldots | d_r$.

Example 3.7.11. *The matrix*

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 6 & 0 \end{pmatrix}$$

is in Smith normal form.

As R is a principal ideal domain it is also a unique factorisation domain and so the greatest common divisors exist. However, it is only defined up to units.

Theorem 3.7.12. *Let $A \in M_{m\times n}(R)$ be a matrix over a principal ideal domain R. Then A is equivalent to a matrix in Smith normal form.*

Proof. Let $A = (A_{ij}) \in M_{m\times n}(R)$. If $A = 0$ then we are done, and so assume that $A \neq 0$.
Claim 1: Given entries A_{ij} and A_{kl} in the same row, $i = k$, or column, $j = l$, we can modify A so that

$\gcd(A_{ij}, A_{kl})$ appears in A.

Proof of Claim 1. Assume the entries are in the same column. Since invertible 2×2 matrices can be extended to invertible $m \times m$ matrices, it suffices to prove the claim when $m = 2$. Consider a vector $\begin{pmatrix} a \\ b \end{pmatrix} \in R^2$. As R is a principal ideal domain we know that $(a, b) = (d) \subseteq R$ as ideals for some $d \in R$. It follows that $d = \gcd(a, b)$ and there are some $x, y \in R$ such that $xa + yb = d$. Since $\gcd(a, b) = d$, we must have that $\gcd(x, y) = 1$. Hence, there exist elements $u, v \in R$ such that $xu + yv = 1$. Let

$$P := \begin{pmatrix} x & y \\ -v & u \end{pmatrix}.$$

Note that $\det(P) = xu + yv = 1$ and so P is invertible. Moreover,

$$\begin{aligned} P \cdot \begin{pmatrix} a \\ b \end{pmatrix} &= \begin{pmatrix} x & y \\ -v & u \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} \\ &= \begin{pmatrix} xa + yb \\ -va + ub \end{pmatrix} \\ &= \begin{pmatrix} d \\ -va + ub \end{pmatrix}. \end{aligned}$$

If instead A_{ij} and A_{kl} are in the same row then

$$\begin{pmatrix} a & b \end{pmatrix} \cdot P^\top = \begin{pmatrix} d & -va + ub \end{pmatrix}.$$

<u>Claim 2:</u> *We can modify A so that A_{11} divides the rest of the first row and column, that is, $A_{11} \mid A_{i1}, A_{1j}$ for all i, j.*

Proof of Claim 2. For $r \in R \backslash \{0\}$, let $\delta(r) \in \mathbb{Z}_{\geq 0}$ be the number of, possibly repeated, irreducible factors of r. Note that r is a unit if and only if $\delta(r) = 0$. As R is a unique factorisation domain $\delta(\cdot)$ is well-defined. Suppose A is not of the required form. As A is non-zero it must contain some non-zero entry, A_{ij} say. Using row and column operations, we can move this entry to the top left. So $\alpha_1 = A_{11} \neq 0$. If A still is not of the required form, there are some $1 \leq i, j \leq n$, not both equal to 1, such that $A_{11} \nmid A_{ij}$. By Claim 1, we can modify A so that $\alpha_2 := \gcd(A_{11}, A_{ij})$ appears in our matrix. Using row and column operations, we can assume this entry is in the top left. Note that $\alpha_2 \mid \alpha_1$, but α_2 and α_1 are not associates since $\alpha_2 \mid A_{ij}$ but $\alpha_1 \nmid A_{ij}$. So $\delta(\alpha_2) < \delta(\alpha_1)$. Hence, when A is not of the required form we can modify it so that the top left entry has strictly lower δ. Repeating this process it must eventually terminate with $A_{11} \mid A_{ij}$ whenever $i = 1$ or $j = 1$.

Returning to our original A we can modify it to be of the form stated in Claim 2. Since $A_{11} \mid A_{1j}$ for all $j > 1$, subtracting multiples of the first column from the other columns, we can modify A to be of the form

$$A = \begin{pmatrix} A_{11} & 0 & \dots & 0 \\ A_{21} & A_{22} & \dots & \\ \vdots & & \ddots & \\ 0 & & & A_{mn} \end{pmatrix}.$$

Similarly since $A_{11} \mid A_{i1}$ for all $i > 1$, we can modify A to be of the form

$$A = \begin{pmatrix} A_{11} & 0 & \dots & 0 \\ 0 & A_{22} & & \\ \vdots & & \ddots & \\ 0 & & & A_{mn} \end{pmatrix}.$$

In other words, we have that

$$A = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{pmatrix}$$

for some smaller matrix $A' \in M_{(m-1)\times(n-1)}$. We can apply the same process to A' without changing the equivalent class. By repeatedly applying the process, we eventually arrive at a matrix of the form

$$A = \begin{pmatrix} d_1 & & & & & 0 \\ & \ddots & & & & \\ & & d_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ 0 & & & & & 0 \end{pmatrix}.$$

It remains to show that $d_1 | \dots | d_r$. For the case $r = 2$, note that $\gcd(d_1, d_2) = xd_1 + yd_2$ for some $x, y \in R$. Moreover, $d_2 = \lambda \cdot \gcd(d_1, d_2)$ for some $\lambda \in R$. Consequently,

$$\begin{aligned} \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} &\overset{R_1 \mapsto R_1 + \lambda y R_2}{\sim} \begin{pmatrix} d_1 & yd_2 \\ 0 & d_2 \end{pmatrix} \\ &\overset{C_2 \mapsto xC_1 + C_2}{\sim} \begin{pmatrix} d_1 & \gcd(d_1, d_2) \\ 0 & d_2 \end{pmatrix} \\ &\overset{R_2 \mapsto R_2 - \lambda R_1}{\sim} \begin{pmatrix} d_1 & \gcd(d_1, d_2) \\ -\lambda d_1 & 0 \end{pmatrix} \\ &\overset{C_1 \leftrightarrow C_2}{\sim} \begin{pmatrix} \gcd(d_1, d_2) & d_1 \\ 0 & -\lambda d_1 \end{pmatrix} \\ &\overset{R_1 \mapsto R_1 + \lambda R_2}{\sim} \begin{pmatrix} \gcd(d_1, d_2) & 0 \\ 0 & -\lambda d_1 \end{pmatrix} \\ &\overset{R_2 \mapsto -R_2}{\sim} \begin{pmatrix} \gcd(d_1, d_2) & 0 \\ 0 & \lambda d_1 \end{pmatrix}. \end{aligned}$$

Now let $\alpha_1 = d_1$. If $d_1 \nmid d_i$ for some $i \geq 2$, let $\alpha_2 = \gcd(d_1, d_i)$ and use row and column operations to move it to the place of d_1. Now $\alpha_2 \mid \alpha_1$ but $\alpha_1 \nmid \alpha_2$, so we have that $\delta(\alpha_2) < \delta(\alpha_1)$. Once again this process terminates, and we find that $d_1 \mid d_i$ for all $i \geq 2$. Repeating, we can modify A so that $d_2 \mid d_i$ for all $i \geq 3$ whilst preserving $d_1 \mid d_i$ for $i \geq 2$. Iterating this gives the required result. □

Remark 3.7.13. *For the proof of Claim 1 made in the proof of Theorem 3.7.12, we multiplied A by a matrix that is not the product of elementary matrices when defined over an arbitrary principal ideal domain. Therefore, over a general principal ideal domain, we cannot transform a matrix into Smith normal form just by using elementary row or column operations. However, over a Euclidean domain, we can transform any matrix into Smith normal form just by using elementary row and column operations. Consequently, we have that for a Euclidean domain every matrix in $\mathrm{GL}_n(R)$ is the product of elementary matrices. More generally, we say that a ring R is generalised Euclidean if every matrix $\mathrm{GL}_n(R)$ is the product of elementary matrices.*

Proposition 3.7.14. *Let $A \in M_{m\times n}(R)$ be in Smith normal form, that is,*

$$A = \begin{pmatrix} d_1 & & & & & \\ & \ddots & & & & \\ & & d_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

with $d_1, \ldots, d_r \in R$ non-zero and $d_1 | \ldots | d_r$. Then there is an isomorphism of R-modules

$$\operatorname{coker}(\phi_A) \cong \frac{R}{(d_1)} \oplus \cdots \oplus \frac{R}{(d_r)} \oplus R^{m-r}$$

where $\phi_A : R^n \to R^m$ is given by left multiplication by A.

Proof. Let $\psi : R^m \to \frac{R}{(d_1)} \oplus \cdots \oplus \frac{R}{(d_r)} \oplus R^{m-r}$ be the homomorphism given by

$$(x_1, \ldots, x_m) \mapsto (x_1 + (d_1), \ldots, x_r + (d_r), x_{r+1}, \ldots, x_m).$$

Then

$$\operatorname{im}(\phi_A) = \ker(\psi) = (d_1) \oplus \cdots \oplus (d_r) \oplus \underbrace{0 \oplus \cdots \oplus 0}_{m-r}.$$

Therefore,

$$\operatorname{coker}(\phi_A) = R^m / \operatorname{im}(\phi_A) = R^m / \ker(\phi) \overset{\text{Thm 3.3.4}}{\cong} \frac{R}{(d_1)} \oplus \cdots \oplus \frac{R}{(d_r)} \oplus R^{m-r}.$$

□

Theorem 3.7.15. *Let R be a commutative principal ideal domain and M a finitely generated R-module. Then*

$$M \cong R^n \oplus R/(d_1) \oplus \cdots \oplus R/(d_r)$$

for some $d_1, \ldots, d_r \in R$ non-zero, with $d_1 | \ldots | d_r$.

Proof.

1. Since R is a principal ideal domain it follows by Corollary 3.7.2 that M is finitely presented. So $M \cong \operatorname{coker}(\phi_B)$ for some $B \in M_{m \times n}(R)$.
2. By Theorem 3.7.12, B is equivalent to a matrix A that is in Smith normal form. Let

$$A = \begin{pmatrix} d_1 & & & & & \\ & \ddots & & & & \\ & & d_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

   where $d_1 | \ldots | d_r$ are all non-zero.
3. Using Proposition 3.7.14 it follows that

$$\operatorname{coker}(\phi_A) \cong \frac{R}{(d_1)} \oplus \cdots \oplus \frac{R}{(d_r)} \oplus R^{m-r}.$$

4. By Proposition 3.7.9 we know $\operatorname{coker}(\phi_A) \cong \operatorname{coker}(\phi_B)$ and so

$$\operatorname{coker}(\phi_B) \cong \frac{R}{(d_1)} \oplus \cdots \oplus \frac{R}{(d_r)} \oplus R^{m-r}.$$

5. Consequently,

$$M \cong R^n \oplus R/(d_1) \oplus \cdots \oplus R/(d_r).$$

□

Remark 3.7.16. *Another type of decomposition is known as the prime decomposition and represents M in the form,*

$$M \cong R^n \oplus \frac{R}{(p_1^{n_1})} \oplus \cdots \oplus \frac{R}{(p_k^{n_k})}$$

for $n_i \in \mathbb{N}$ and $p_i \in R$ irreducible. This follows from Theorem 3.7.15, along with the fact that if $d = p_1^{n_1} \cdot \cdots \cdot p_k^{n_k}$, then

$$\frac{R}{(d)} \cong \frac{R}{(p_1^{n_1})} \oplus \cdots \oplus \frac{R}{(p_k^{n_k})}.$$

Note that if we know how to compute the greatest common divisors, the proof of Theorem 3.7.15 is constructive.

Example 3.7.17. *Let A be the abelian group generated by the elements a, b and c with relations*

$$\begin{cases} 2a + 3b + c = 0 \\ a + 2b = 0 \\ 5a + 6b + 7c = 0. \end{cases}$$

Then

$$A = \frac{\mathbb{Z}^3}{\left\langle \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix} \right\rangle} = \operatorname{coker}(\phi_X)$$

where

$$X = \begin{pmatrix} 2 & 1 & 5 \\ 3 & 2 & 6 \\ 1 & 0 & 7 \end{pmatrix}$$

We can put X into Smith normal form in the following way

$$\begin{aligned} \begin{pmatrix} 2 & 1 & 5 \\ 3 & 2 & 6 \\ 1 & 0 & 7 \end{pmatrix} &\sim \begin{pmatrix} 1 & 2 & 5 \\ 2 & 3 & 6 \\ 0 & 1 & 7 \end{pmatrix} \\ &\sim \begin{pmatrix} 1 & 2 & 5 \\ 0 & -1 & -4 \\ 0 & 1 & 7 \end{pmatrix} \\ &\sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & -4 \\ 0 & 1 & 7 \end{pmatrix} \\ &\sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 4 \\ 0 & 1 & 7 \end{pmatrix} \\ &\sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 4 \\ 0 & 0 & 3 \end{pmatrix} \\ &\sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}. \end{aligned}$$

Therefore,

$$A \cong \frac{\mathbb{Z}}{1\mathbb{Z}} \oplus \frac{\mathbb{Z}}{1\mathbb{Z}} \oplus \frac{\mathbb{Z}}{3\mathbb{Z}} \cong C_3.$$

Corollary 3.7.18. *Let R be a principal ideal domain. Then finitely generated projective R-modules are free.*

Proof. Let M be a finitely generated R-module. Using Theorem 3.7.15 we have

$$M \cong R^n \oplus R/(d_1) \oplus \dots R/(d_r)$$

for some $d_1, \dots, d_r \in R \setminus \{0\}$ with $d_1 | \dots | d_r$. As M is projective, there exists an R-module N such that $M \oplus N \cong R^{(S)}$ for some set S. If $d_1, \dots, d_r \in R^\times$, then $M \cong R^n$, and thus it is free. Otherwise, for some j there exists an injective R-module homomorphism $i : R/(d_j) \to R^{(S)}$. Let $i(1) = (x_i)_{i \in S}$. Then as $i(1) \neq 0$, as i is injective, it follows that there exists an $s \in S$ such that $x_s \neq 0$. Since, $d_j i(1) = i(d_j) = 0$ we have that $d_j x_s = 0 \in R$. As R is an integral domain it must be the case that $d_j = 0$ which is a contradiction. □

3.8 Jordan Normal Form

In this section, we will study modules over polynomial rings using Theorem 3.7.15. Amongst other things, we will deduce the Jordan normal form. Let F be a field, V an F-vector space and $\alpha : V \to V$ a linear map. Then we can make V into a $F[X]$-module by defining the action

$$f \cdot v := (f(\alpha))(v)$$

for $f \in F[X]$ and $v \in V$. Note that when $f = X$ we recover α. We write V_α when we view V as a $F[X]$-module in this way. The structure of V_α as a $F[X]$-module will help us study the linear map α.

Lemma 3.8.1. *If V is a finite-dimensional F-vector space, then V_α is a finitely generated $F[X]$-module.*

Proof. Let $v_1, \dots, v_n \in V$ be a basis for V as a vector space. Then they generate V_α as an $F[X]$-module. □

Example 3.8.2.

1. *Suppose $V_\alpha \cong F[X]/(X^r)$ as $F[X]$-modules. Then they are isomorphic as F-vector spaces. Note that $1, X, \dots, X^{r-1} \in F[X]/(X^r)$ is a basis for which multiplication by X is represented by the matrix*

$$\begin{pmatrix} 0 & & & 0 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & 1 & 0 \end{pmatrix}.$$

   Hence, as $\alpha(v) = f \cdot v$ for $f = X$, in V_α the linear map α is represented by the same matrix with respect to the corresponding basis.

2. *Suppose that $V_\alpha \cong F[X]/((X-\lambda)^r)$, as $F[X]$-modules, for some $\lambda \in F$. Consider the linear map $\beta : V \to V$ where $\beta = \alpha - \lambda I$. Then $V_\beta \cong F[Y]/(Y^r)$, for $Y = X - \lambda$. So by statement 1 there is a basis for V such that β is given by the matrix*

$$\begin{pmatrix} 0 & & & 0 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & 1 & 0 \end{pmatrix}.$$

Hence, $\alpha = \beta + \lambda I$ has matrix

$$\begin{pmatrix} \lambda & & & 0 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & 1 & \lambda \end{pmatrix},$$

which is called a λ Jordan block.

3. *Suppose $V_\alpha \cong F[X]/(f)$, for $f = a_0 + \cdots + a_{r-1}X^{r-1} + X^r \in F[X]$ some monic polynomial. Then V_α has an F-basis given by $1, X, \ldots, X^{r-1}$, in which α is given by*

$$C(f) := \begin{pmatrix} 0 & & & 0 & -a_0 \\ 1 & \ddots & & & -a_1 \\ & \ddots & \ddots & & \vdots \\ & & \ddots & \ddots & \\ 0 & & & 1 & -a_{r-1} \end{pmatrix}.$$

The matrix $C(f)$ is called the companion matrix of f.

Theorem 3.8.3 (Rational Canonical Form)**.** *Let F be a field, V a finite-dimensional F-vector space, and $\alpha : V \to V$ a linear map. Then*

$$V_\alpha \cong \frac{F[X]}{(f_1)} \oplus \cdots \oplus \frac{F[X]}{(f_r)}$$

with $f_1 | \ldots | f_r$ all non-zero polynomials over F. In particular, there is a basis for V in which α is given by the block diagonal matrix

$$\begin{pmatrix} C(f_1) & & 0 \\ & \ddots & \\ 0 & & C(f_r) \end{pmatrix}.$$

Proof. As $F[X]$ is a principal ideal domain, by Theorem 3.7.15 we have

$$V_\alpha \cong \frac{F[X]}{(f_1)} \oplus \cdots \oplus \frac{F[X]}{(f_r)} \oplus F[X]^n$$

with $f_1 | \ldots | f_r$. Since $F[X]$ is not a finite-dimensional F-vector space, but V_α is, we must have $n = 0$. Let $b_i := \deg(f_i)$, and consider the basis

$$\left\{ X_1^0, X_1, \ldots, X_1^{b_1 - 1}, \ldots, X_r^0, \ldots, X_r^{b_r - 1} \right\}$$

where X_j is the X in the j^{th} term on the direct sum. Then α is represented by

$$\begin{pmatrix} C(f_1) & & 0 \\ & \ddots & \\ 0 & & C(f_r) \end{pmatrix}$$

with respect to this basis. Requiring the f_i to be monic ensures this representation is unique. □

We now focus in on the case $F = \mathbb{C}$.

Lemma 3.8.4. *The primes in $\mathbb{C}[X]$ are, up to associates, $X - \lambda$, for $\lambda \in \mathbb{C}$.*

Proof. If $X-\lambda | fg$ for $f,g \in \mathbb{C}[X]$ then $X-\lambda | f$ or $X-\lambda | g$ and so $X-\lambda \in \mathbb{C}[X]$ is prime. Now let $f \in \mathbb{C}[X]$ be prime. As f is non-zero and not a unit it is non-constant, so by the fundamental theorem of algebra f has a root $\lambda \in \mathbb{C}$, which implies that $X-\lambda \mid f$. Since f is also irreducible, we must have that $X-\lambda$ and f are associates. □

Theorem 3.8.5 (Jordan Normal Form)**.** *Let $\alpha : V \to V$ be a linear map, where V is a finite-dimensional complex vector space. Then there is an isomorphism of $\mathbb{C}[X]$-modules*

$$V_\alpha \cong \frac{\mathbb{C}[X]}{((X-\lambda_1)^{n_1})} \oplus \cdots \oplus \frac{\mathbb{C}[X]}{((X-\lambda_r)^{n_r})}$$

for some $n_i \in \mathbb{N}$ and $\lambda_i \in \mathbb{C}$. Furthermore, there is a $\mathbb{C}$-basis for V in which α has the form of the block diagonal matrix

$$\begin{pmatrix} J_{n_1}(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & J_{n_r}(\lambda_r) \end{pmatrix}$$

where $J_n(\lambda)$ is the $n \times n$ Jordan block

$$J_n(\lambda) := \begin{pmatrix} \lambda & & & 0 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & 1 & \lambda \end{pmatrix}.$$

Proof. Using Remark 3.7.16 on V_α, and Lemma 3.8.4 it follows that

$$V_\alpha \cong \frac{\mathbb{C}[X]}{((X-\lambda_1)^{n_1})} \oplus \cdots \oplus \frac{\mathbb{C}[X]}{((X-\lambda_r)^{n_r})}$$

for some $n_i \in \mathbb{N}$ and $\lambda_i \in \mathbb{C}$. Then taking a basis as in the proof of Theorem 3.8.3, and noting that

$$C\left((X-\lambda)^n\right) = J_n(\lambda)$$

we note α has a representation

$$\begin{pmatrix} J_{n_1}(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & J_{n_r}(\lambda_r) \end{pmatrix}.$$

□

Lemma 3.8.6. *Let V be a finite-dimensional F-vector space, and let $\alpha, \beta : V \to V$ be linear maps. Then $V_\alpha \cong V_\beta$ are isomorphic as $F[X]$-modules if and only if α and β are conjugate. That is, there is some isomorphism $\gamma : V \to V$ such that $\gamma \circ \alpha = \beta \circ \gamma$.*

Proof. ($\Rightarrow$). Let $\gamma : V_\alpha \to V_\beta$ be an $F[X]$-module isomorphism, so in particular it is an isomorphism of F-vector spaces. Then for $v \in V$ we have $\beta(v) = X \cdot_{V_\beta} v$, where $\cdot_{V_\beta}$ denotes the $F[X]$-module action on V_β, and similarly $\alpha(v) = X \cdot_{V_\alpha} v$. Then,

$$\begin{aligned} (\beta \circ \gamma)(v) &= X \cdot_{V_\beta} \gamma(v) \\ &= \gamma\left(X \cdot_{V_\alpha} v\right) \\ &= (\gamma \circ \alpha)(v) \end{aligned}$$

and so $\beta \circ \gamma = \gamma \circ \alpha$. That is, α and β are conjugate.
($\Leftarrow$). Suppose α and β are conjugate, so there is an isomorphism $\gamma : V \to V$ such that $\beta \circ \gamma = \gamma \circ \alpha$. Note that $\gamma \circ \alpha^i = \beta^i \circ \gamma$ for all $i \geq 0$, so for $f \in F[X]$ and $v \in V$ we have

$$\begin{aligned} \gamma\left(f \cdot_{V_\alpha} v\right) &= \gamma(f(\alpha)(v)) \\ &= f(\beta(\gamma(v)) \\ &= f \cdot_{V_\beta} \gamma(v). \end{aligned}$$

So γ is an $F[X]$-module homomorphism and a bijection since it is an F-linear isomorphism. □

Reinterpreting this with Theorem 3.8.3 we obtain a classification result for square matrices over a field. It is a weaker classification than Theorem 3.8.5 form but it works over any field.

Corollary 3.8.7. *Let F be a field. There is a bijection between conjugacy classes of $n \times n$ matrices over F and sequences of monic polynomials $f_1, \ldots, f_r \in F[X]$, such that $d_1 | \ldots | d_r$, and $\sum_{i=1}^{r} \deg(f_i) = n$.*

Proof. Let A be an $n \times n$ matrix over F, let $V = F^n$ and let $\alpha : V \to V$ send v to $A \cdot v$. Then by Theorem 3.8.3 we know that

$$V_\alpha \cong \frac{F[X]}{(f_1)} \oplus \cdots \oplus \frac{F[X]}{(f_r)}$$

for some unique sequence $f_i \in F[X]$, with $f_1 | \ldots | f_r$ all non-zero monic polynomials in $F[X]$. By counting dimensions of both sides we note that $\sum_i \deg(f_i) = n$. Furthermore, if A and B are conjugate matrices, then V_α and V_β are isomorphic $F[X]$ modules, by Lemma 3.8.6, and so they correspond to the same sequence of polynomials. If $f_1 | \ldots | f_r$ is any sequence of monic polynomials in $F[X]$ with $\sum_i \deg(f_i) = n$, then

$$V := \frac{F[X]}{(f_1)} \oplus \cdots \oplus \frac{F[X]}{(f_r)}$$

is an n-dimensional F-vector space, with a linear map $\alpha : V \to V$ sending v to $X \cdot v$, which determines a matrix, up to conjugation, upon by fixing a basis. □

Example 3.8.8. *Let us consider the conjugacy classes in $M_{2\times 2}(F)$. This corresponds to classifying $F[X]$-modules of the form*

$$\frac{F[X]}{(f_1)} \oplus \cdots \oplus \frac{F[X]}{(f_r)}$$

with $f_1 | \ldots | f_r$ a sequence of monic polynomials with $\sum_{i=1}^{r} \deg(f_i) = 2$. So either $r = 1$ and $\deg(f_1) = 2$, or $r = 2$ and $\deg(f_1) = \deg(f_2) = 1$. In the latter case, since $f_1 \mid f_2$ and both are monic we must have that $f_1 = f_2 = X - \lambda$ for some $\lambda \in F$. Since the companion matrix of $X - \lambda$ is the 1×1 matrix (λ), the corresponding matrix in $M_{2\times 2}(F)$ is

$$\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}.$$

In the former case, we write $f_1 = X^2 + aX + b$ for some $a, b \in F$ and so

$$C(f_1) = \begin{pmatrix} 0 & -b \\ 1 & -a \end{pmatrix}.$$

If f_1 is irreducible, we cannot simplify this any further for an arbitrary field F. However, if for example $F = \mathbb{Z}/3\mathbb{Z}$ then we could find all irreducible degree 2 polynomials, and arrive at a more detailed classification. If f_1 is reducible then $f_1 = (X - \lambda)(X - \mu)$, for some $\lambda, \mu \in F$. If $\lambda = \mu$, then $f_1 = (X - \lambda)^2$ and so the matrix is conjugate to the Jordan block

$$\begin{pmatrix} \lambda & 0 \\ 1 & \lambda \end{pmatrix}.$$

If $\lambda \neq \mu$, the matrix

$$\begin{pmatrix} 0 & -\lambda\mu \\ 1 & \lambda+\mu \end{pmatrix}$$

has distinct eigenvalues and so it is conjugate to

$$\begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}.$$

Exercise 3.8.9. *Find the conjugacy classes, along with a representative, in $M_{10\times 10}(\mathbb{Q})$ with minimal polynomial $X^7 - 4X^3$.*

3.9 Solution to Exercises

Exercise 3.5.20

Solution. Note that for $s \in S$, the action of s on $\{x_i : i \in \mathbb{N}\}$ fully determines s by the fact that s is a homomorphism and $\{x_i : i \in \mathbb{N}\}$ generates M. Therefore, we can define $s_1, s_2 \in S$ by

$$s_1(x_i) = \begin{cases} x_{\frac{i}{2}} & i \text{ even} \\ 0 & \text{otherwise} \end{cases}$$

and

$$s_2(x_i) = \begin{cases} x_{\frac{i+1}{2}} & i \text{ odd} \\ 0 & \text{otherwise.} \end{cases}$$

For $s \in S$ we can write $s = \alpha_1 \circ s_1 + \alpha_2 \circ s_2$ where $\alpha_1(x_i) = s(x_{2i})$ and $\alpha_2(x_i) = s(x_{2i+1})$. Moreover, suppose that for $\alpha_1, \alpha_2 \in S$ we have

$$\alpha_1 \circ s_1 + \alpha_2 \circ s_2 = 0.$$

Then for all $i \in \mathbb{N}$ it follows that

$$\begin{aligned} 0 &= (\alpha_1 \circ s_1)(x_{2i}) + (\alpha_2 \circ s_2)(x_{2i}) \\ &= \alpha_1(x_i). \end{aligned}$$

Hence, $\alpha_1 = 0$. Similarly, by considering x_{2i+1} we deduce that $\alpha_2 = 0$. Therefore, $\{s_1, s_2\}$ are linearly independent and hence are a basis for S as an S-module. On the other hand, $\{\mathbf{1}\}$, where $\mathbf{1} \in S$ is such that $\mathbf{1}(x_i) = x_i$ for all $i \in \mathbb{N}$, generates S as an S-module. Therefore, we can construct an isomorphism $\varphi : S \to S^2$ where $s \mapsto (\alpha_1, \alpha_2)$. We conclude that $S \cong S^2$. □

Exercise 3.7.4

Solution.

- With $P = Q = I$ we have that $A \sim A$.
- Suppose $A \sim B$ with $B = PAQ$ for $P \in \mathrm{GL}_n(R)$ and $Q \in \mathrm{GL}_m(R)$. Then as P and Q are invertible we have that $A = P^{-1}BQ^{-1}$ which implies that $B \sim A$.
- Suppose $A \sim B$ and $B \sim C$ with $B = P_1AQ_1$ and $C = P_2BQ_2$ for $P_1, P_2 \in \mathrm{GL}_n(R)$ and $Q_1, Q_2 \in \mathrm{GL}_m(R)$. It follows that
$$C = (P_1P_2)A(Q_1Q_2)$$
for $P_1P_2 \in \mathrm{GL}_n(R)$ and $Q_1Q_2 \in \mathrm{GL}_m(R)$, which implies that $A \sim C$.

Therefore, $\sim$ is reflexive, symmetric and transitive making it an equivalence relation. □

Exercise 3.8.9

Solution. Let $A \in M_{10\times 10}(\mathbb{Q})$, and let $\alpha : V \to V$ be its corresponding linear map. Then from Theorem 3.8.3 we have that

$$V_\alpha \cong \frac{\mathbb{Q}[X]}{(f_1)} \oplus \cdots \oplus \frac{\mathbb{Q}[X]}{(f_r)}$$

for $f_1, \ldots, f_r \in \mathbb{Q}[X]$ monic, such that $f_1 | \ldots | f_r$, and $\sum_{k=1}^{r} \deg(f_k) = 10$. The possibilities of such $\{f_1, \ldots, f_r\}$ are in bijection with the conjugacy classes. Note that $g(A) = 0$ if and only if $f_k | g$ for $k = 1, \ldots, r$. Therefore, f_r is the minimal polynomial of A, and so in this case $f_r = X^7 - 4X^3$. Noting that $f_r = X^3\left(X^2 - 2\right)\left(X^2 + 2\right)$, the following represent the only possible cases for $\{f_1, \ldots, f_r\}$.

1. $\left\{X^3, X^7 - 4X^3\right\}$.
2. $\left\{X\left(X^2 - 2\right), X^7 - 4X^3\right\}$.
3. $\left\{X\left(X^2 + 2\right), X^7 - 4X^3\right\}$.
4. $\left\{X, X^2, X^7 - 4X^3\right\}$.
5. $\left\{X, X, X, X^7 - 4X^3\right\}$.

The conjugacy classes above have the following corresponding representatives.

1. $\begin{pmatrix} C\left(X^3\right) & 0 \\ 0 & C\left(X^7 - 4X^3\right) \end{pmatrix}$.
2. $\begin{pmatrix} C\left(X\left(X^2 - 2\right)\right) & 0 \\ 0 & C\left(X^7 - 4X^3\right) \end{pmatrix}$.
3. $\begin{pmatrix} C\left(X\left(X^2 + 2\right)\right) & 0 \\ 0 & C\left(X^7 - 4X^3\right) \end{pmatrix}$.
4. $\begin{pmatrix} C(X) & & 0 \\ & C\left(X^2\right) & \\ 0 & & C\left(X^7 - 4X^3\right) \end{pmatrix}$.
5. $\begin{pmatrix} C(X) & & & \\ & C(X) & & \\ & & C(X) & \\ 0 & & & C\left(X^7 - 4X^3\right) \end{pmatrix}$.

□

4 Matrix Lie Groups

We intend to investigate matrices over $\mathbb{R}$ or $\mathbb{C}$. For ease of notation, we will write $\mathbb{F}$ to mean either $\mathbb{R}$ or $\mathbb{C}$.

4.1 Matrix Groups

Let R be a commutative ring. Recall that $M_n(R)$ is the ring of $n \times n$ matrices over R, which is not a commutative ring for $n \geq 2$. Note that we may identify $M_n(R)$ with R^{n^2}.

Definition 4.1.1. *The general linear group over R denoted $\mathrm{GL}_n(R)$, is the unit group of $M_n(R)$. That is, $\mathrm{GL}_n(R) := M_n(R)^\times$.*

The determinant function $\det : M_n(R) \to R$ is thought of as in the usual sense of linear algebra.

Exercise 4.1.2. *For $M \in M_n(R)$, show that if M is invertible then $\det(M) \in R^\times$.*

Proposition 4.1.3. *The centre of $\mathrm{GL}_n(R)$ is the set of matrices of the form λI, for some $\lambda \in R^\times$.*

Proof. Note that if $A = \lambda I$ for $\lambda \in R^\times$ then $A \in \mathrm{GL}_n(R)$ with $AM = MA$ for all $M \in \mathrm{GL}_n(R)$. Hence, $A \in Z(\mathrm{GL}_n(R))$. Instead, let $A \in Z(\mathrm{GL}_n(R))$. For $1 \leq u, v \leq n$ distinct, let $E^{uv} \in \mathrm{GL}_n(R)$ be the matrix

$$(E^{uv})_{ij} = \delta_{ij} + \delta_{ui}\delta_{vj} = \begin{cases} 1 & i = j \\ 1 & i = u, j = v \\ 0 & \text{otherwise.} \end{cases}$$

By assumption we have $AE^{uv} = E^{uv}A$, and so for all $1 \leq i, j \leq n$, we have

$$\begin{aligned} 0 &= (E^{uv}A - AE^{uv})_{ij} \\ &= E^{uv}_{ik}A_{kj} - A_{ik}E^{uv}_{kj} \\ &= \delta_{ui}A_{vj} - \delta_{vj}A_{ui}. \end{aligned}$$

If $u = i \neq v = j$, this tells us $A_{ii} = A_{jj}$. If $i = j = u \neq v$, this tells us $A_{vj} = 0$. Therefore, $A = \lambda I$, so by Exercise 4.1.2 we know that for $A \in \mathrm{GL}_n(R)$ we need $\lambda \in R^\times$. □

When $R = \mathbb{F}$ we can discuss the topological properties of $\mathrm{GL}_n(R)$. In particular, we endow $M_n(\mathbb{F})$ with the Euclidean topology and $\mathrm{GL}_n(\mathbb{F}) \subseteq M_n(\mathbb{F})$ with the subspace topology. In doing so, topological properties can be considered in the natural sense with the subspace topology on $\mathbb{F}^{n^2}$.

Definition 4.1.4. *Let $(A_m)_{m\in\mathbb{N}} \in M_n(\mathbb{F})$ be a sequence of matrices. Then A_m converges to a matrix A if each entry of A_m converges to the corresponding entry of A.*

Proposition 4.1.5. *The set $\mathrm{GL}_n(\mathbb{R})$ is not path-connected.*

Proof.

- For $n = 1$, consider -1 and 1 in $\mathrm{GL}_1(\mathbb{R}) = \mathbb{R}^\times$. Suppose they are connected by $\gamma : [0, 1] \to \mathbb{R}^\times$. Then by the intermediate value theorem, there exists a $t \in [0, 1]$ such that $\gamma(t) = 0$, but $0 \notin \mathbb{R}^\times$ and so we get a contradiction.
- For $n \geq 2$ consider

$$A := \begin{pmatrix} -1 & \cdots & & 0 \\ \vdots & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix}$$

and I. Suppose $\gamma : [0,1] \to \mathrm{GL}_n(\mathbb{R})$ is a path from I to A. Then $\det \circ \gamma : [0,1] \to \mathbb{R}$ is continuous, and such that $(\det \circ \gamma)(0) = 1$ and $(\det \circ \gamma)(1) = -1$. By the intermediate value theorem, there exists a $t \in [0,1]$ such that $(\det \circ \gamma)(t) = 0$. However, $\gamma(t)$ is an invertible matrix, so we get a contradiction. □

Proposition 4.1.6. *The set $\mathrm{GL}_n(\mathbb{C})$ is path-connected.*

Sketch. Let $A \in \mathrm{GL}_n(\mathbb{C})$. We can write $A = PBP^{-1}$, where B is in Jordan normal form. We assume for simplicity that B is a single Jordan block, that is

$$B = \begin{pmatrix} \lambda & 1 & 0 \\ & \ddots & 1 \\ 0 & & \lambda \end{pmatrix}.$$

As A is invertible we know that $\lambda \neq 0$, and thus we can write $\lambda = e^z$ for some $z \in \mathbb{C}$. Let $\gamma : [0,1] \to \mathrm{GL}_n(\mathbb{C})$ be given by

$$\gamma(t) = \begin{pmatrix} e^{tz} & t & 0 \\ & \ddots & t \\ 0 & & e^{tz} \end{pmatrix}.$$

Note that $\gamma(t)$ is invertible for all $t \in [0,1]$ and so γ is well-defined. Moreover, γ is continuous with $\gamma(0) = I$ and $\gamma(1) = B$. Therefore, $P\gamma(t)P^{-1}$ is a path between I and A, meaning that $\mathrm{GL}_n(\mathbb{C})$ is path-connected. □

4.2 Topological Groups

Definition 4.2.1. *A topological group is a group G with a topology, such that the multiplication map $G \times G \to G$ and the inverse map $G \to G$ are continuous. Moreover, G is Hausdorff.*

Definition 4.2.2. *A homomorphism of topological groups is a continuous group homomorphism.*

An isomorphism of topological groups is an isomorphism of groups which is a homeomorphism.

Lemma 4.2.3. *The composition of topological group homomorphisms is a homomorphism of topological groups.*

Proof. Since the composition of continuous functions is continuous, and the composition of group homomorphism is a group homomorphism, the composition of topological group homomorphisms is a homomorphism of topological groups. □

Example 4.2.4.

1. *Any group G equipped with the discrete topology is a topological group.*
2. *$(\mathbb{R}, +)$ is a topological group.*
3. *$(\mathbb{Q}, +)$ is a topological group, and the inclusion $\mathbb{Q} \hookrightarrow \mathbb{R}$ is a homomorphism of topological groups.*
4. *$(\mathbb{R}^\times, \cdot)$ and $(\mathbb{C}^\times, \cdot)$ are topological groups. The map $t \mapsto e^{2\pi i t}$ is a continuous homomorphism $(\mathbb{R}, +) \to (\mathbb{C}^\times, \cdot)$.*

Lemma 4.2.5. *The group $\mathrm{GL}_n(\mathbb{F})$ is a topological group.*

Proof. Matrix multiplication $M_n(\mathbb{F}) \times M_n(\mathbb{F}) \to M_n(\mathbb{F})$ is polynomial in each entry, and therefore continuous. For an invertible matrix M, the matrix $\det(M)M^{-1}$ has entries which are polynomials in the entries of M and so is a continuous function of M. As $\det(M)$ is a continuous function of M, the quotient M^{-1} is a continuous function of M. □

Remark 4.2.6. *For a subgroup $G \leq \mathrm{GL}_n(\mathbb{F})$, the multiplication and inverse maps are the restrictions of those on $\mathrm{GL}_n(\mathbb{F})$, and so are also continuous. Hence, a subgroup $G \leq \mathrm{GL}_n(\mathbb{F})$ is also a topological group.*

Lemma 4.2.7. *Let G be a topological group. Any subgroup $H \leq G$, when equipped with the subspace topology, is a topological group.*

Proof. The product $H \times H \to H$ is the restriction of the product $G \times G \to G$, so is continuous in the subspace topology. Similarly, the inverse is continuous, meaning H is a topological group. □

4.3 Matrix Lie Groups

Definition 4.3.1. *A matrix Lie group is a topological group G, which is isomorphic, in the sense of topological groups, to a closed subgroup $H \leq \mathrm{GL}_n(\mathbb{F})$ for some $n \in \mathbb{N}$.*

Since $\mathrm{GL}_n(\mathbb{F})$ is Hausdorff it is metrizable. Thus any subspace is also metrizable meaning a matrix Lie group is also Hausdorff.

Remark 4.3.2. *An equivalent definition of a matrix Lie group is to say that any convergent sequence $(A_m)_{m\in\mathbb{N}} \subseteq G$ either converges to a matrix $A \in G$ or converges to a matrix that is not invertible. That is, if the limit is in $\mathrm{GL}_n(\mathbb{F})$ then it is in G. Convergence of matrices is in the sense of Definition 4.1.4.*

Example 4.3.3.

1. *The topological group $(\mathbb{R}^\times, \cdot) = \mathrm{GL}_1(\mathbb{R})$ is a matrix Lie group.*
2. *Note $\mathbb{Q}^\times \subseteq \mathbb{R}^\times$ is a subgroup but it is not a closed subgroup. However, this does not immediately imply $\mathbb{Q}^\times$ is not a matrix Lie group, since it could be isomorphic to a closed subgroup of $\mathrm{GL}_n(\mathbb{F})$ for some other $n \in \mathbb{N}$. We will see later that $\mathbb{Q}^\times$ is not a matrix Lie group.*

Exercise 4.3.4. *Show that the following are matrix Lie groups.*

1. *The special linear group $\mathrm{SL}_n(\mathbb{F}) := \{M \in \mathrm{GL}_n(\mathbb{F}) : \det(M) = 1\} \leq \mathrm{GL}_n(\mathbb{F})$.*
2. *The set of diagonal matrices in $\mathrm{GL}_n(\mathbb{F})$.*
3. *The orthogonal group $\mathrm{O}(n) := \left\{M \in \mathrm{GL}_n(\mathbb{R}) : M^\top M = I\right\} \leq \mathrm{GL}_n(\mathbb{R})$.*
4. *The special orthogonal group $\mathrm{SO}(n) := \{M \in \mathrm{O}(n) : \det(M) = 1\} \leq \mathrm{GL}_n(\mathbb{R})$.*
5. *The unitary group $\mathrm{U}(n) := \left\{M \in \mathrm{GL}_n(\mathbb{C}) : M^\dagger M = I\right\} \leq \mathrm{GL}_n(\mathbb{C})$.*
6. *The special unitary group $\mathrm{SU}(n) := \{M \in \mathrm{U}(n) : \det(M) = 1\} \leq \mathrm{GL}_n(\mathbb{C})$.*
7. *$\mathbb{Z}$ or any finite group equipped with the discrete topology.*
8. *The projective general linear group $\mathrm{PGL}_n(\mathbb{F}) := \mathrm{GL}_n(\mathbb{F})/\mathrm{Z}(\mathrm{GL}_n(\mathbb{F}))$ with the quotient topology.*

9. The Heisenberg group

$$\mathrm{H}_n(\mathbb{F}) := \left\{ M \in \mathrm{GL}_n(\mathbb{F}) : M = \begin{pmatrix} 1 & & \times \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \right\}.$$

Example 4.3.5. *Note that*

$$\mathrm{SO}(2) = \left\{ \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} : \theta \in \mathbb{R} \right\}.$$

Let $\varphi : \mathrm{SO}(2) \to \mathrm{U}(1)$ be given by $A \mapsto e^{i\theta}$. Clearly, φ continuous. Moreover, if $z \in \mathrm{U}(1)$, it follows that $|z| = 1$ which implies that $z = e^{i\theta} = \cos\theta + i\sin\theta$ for some $\theta \in \mathbb{R}$. Hence, φ is also surjective. Now let

$$A = \begin{pmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{pmatrix}$$

and

$$B = \begin{pmatrix} \cos\theta_2 & \sin\theta_2 \\ -\sin\theta_2 & \cos\theta_2 \end{pmatrix}.$$

Then as

$$AB = \begin{pmatrix} \cos(\theta_1+\theta_2) & \sin(\theta_1+\theta_2) \\ -\sin(\theta_1+\theta_2) & \cos(\theta_1+\theta_2) \end{pmatrix}$$

we see that $\varphi(AB) = e^{i(\theta_1+\theta_2)} = e^{i\theta_1}e^{i\theta_2} = \varphi(A)\varphi(B)$. Hence, φ is a homomorphism. If $\varphi(A) = \varphi(B)$ then $e^{i\theta_1} = e^{i\theta_2}$, which implies that $\theta_1 - \theta_2 = 2k\pi$ for some $k \in \mathbb{Z}$. Which implies that $A = B$. Hence, φ is injective, and thus $\mathrm{SO}(2)$ and $\mathrm{U}(1)$ are isomorphic as topological groups.

Lemma 4.3.6. *Let G be a topological group, and let $H \leq G$. Then the closure $\bar{H}$ of H in G is a closed subgroup.*

Proof. Clearly, $\bar{H} \subseteq G$ is closed. Let $g, h \in \bar{H}$. Then there are sequences $(g_n)_{n\in\mathbb{N}} \subseteq H$ and $(h_n)_{n\in\mathbb{N}} \subseteq H$, such that $g_n \to g$ and $h_n \to h$. As multiplication is continuous it follows that $g_n \cdot h_n \to g \cdot h$. Since $g_n \cdot h_n \in H$ for every $n \in \mathbb{N}$, we have that $g \cdot h \in \bar{H}$. Similarly, as $g_n^{-1} \in H$ for every $n \in \mathbb{N}$ we get that $g^{-1} \in \bar{H}$. Therefore, $\bar{H}$ is also a subgroup. □

Remark 4.3.7. *As a consequence of Lemma 4.3.6 we have that the closure of $G \leq \mathrm{GL}_n(\mathbb{F})$ is a matrix Lie group.*

Example 4.3.8. *The closure of $\mathbb{Q}^\times$ in $\mathbb{R}^\times$ is $\mathbb{R}^\times$, which is a matrix Lie group as expected.*

Proposition 4.3.9. *The group $\mathrm{GL}_n(\mathbb{R})$ is a closed subgroup of $\mathrm{GL}_n(\mathbb{C})$.*

Proof. Consider the continuous map $f : \mathrm{GL}_n(\mathbb{C}) \to M_n(\mathbb{R})$ given by $A \mapsto \mathrm{Im}(A)$. As $\mathrm{GL}_n(\mathbb{R}) = f^{-1}(\{0\})$ it follows that $\mathrm{GL}_n(\mathbb{R}) \subseteq \mathrm{GL}_n(\mathbb{C})$ is a closed subset. As it is a subgroup we conclude that it is a closed subgroup. □

Proposition 4.3.10. *The group $\mathrm{GL}_n(\mathbb{C})$ is isomorphic to a closed subgroup in $\mathrm{GL}_{2n}(\mathbb{R})$.*

Proof. Let $\phi : \mathrm{GL}_n(\mathbb{C}) \to \mathrm{GL}_{2n}(\mathbb{R})$ be given by

$$M \mapsto \begin{pmatrix} \mathrm{Re}(M) & \mathrm{Im}(M) \\ -\mathrm{Im}(M) & \mathrm{Re}(M) \end{pmatrix} \in \mathrm{GL}_{2n}(\mathbb{R}).$$

As ϕ is polynomial in its entry it is continuous. Moreover, ϕ is injective with $\phi(I) = I$. Let $M, N \in \mathrm{GL}_n(\mathbb{C})$ with $M = A + iB$ and $N = C + iD$ for $A, B, C, D \in M_n(\mathbb{R})$. Then,

$$\phi(M)\phi(N) = \begin{pmatrix} A & B \\ -B & A \end{pmatrix} \begin{pmatrix} C & D \\ -D & C \end{pmatrix}$$

and

$$\phi(MN) = \phi((AC - BD) + i(BC + AD)).$$

Expanding these out we see that $\phi(M)\phi(N) = \phi(MN)$. Thus ϕ is a homomorphism. Let

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \in \mathrm{GL}_{2n}(\mathbb{R}).$$

Claim: $\mathrm{im}(\phi) = \{M \in \mathrm{GL}_{2n}(\mathbb{R}) : MJ = JM\}$.
Proof. Writing $M = A + iB \in \mathrm{GL}_n(\mathbb{C})$ with $A, B \in M_n(\mathbb{R})$ we see that

$$\phi(M)J - J\phi(M) = \begin{pmatrix} A & B \\ -B & A \end{pmatrix} \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} - \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} A & B \\ -B & A \end{pmatrix} = 0.$$

Now suppose that $P \in \mathrm{GL}_{2n}(\mathbb{R})$ satisfies $PJ = JP$. Then P is of the form

$$P = \begin{pmatrix} A & B \\ -B & A \end{pmatrix}$$

for some $A, B \in M_n(\mathbb{R})$. Let $M = A + iB \in M_n(\mathbb{C})$. Multiplying $PJ = JP$ on the left and right by P^{-1} gives $P^{-1}J = JP^{-1}$. Therefore, P^{-1} is of the form

$$P^{-1} = \begin{pmatrix} C & D \\ -D & C \end{pmatrix}$$

for some $C, D \in M_n(\mathbb{R})$. Let $N = C + iD \in M_n(\mathbb{C})$. We can check that N is an inverse to M. Hence, $\phi(M)$ is well-defined meaning $P \in \mathrm{im}(\phi)$. Therefore, $\mathrm{im}(\phi) = \{M \in \mathrm{GL}_{2n}(\mathbb{R}) : MJ = JM\}$.
Consequently we have that $\mathrm{im}(\phi)$ is the pre-image of $\{0\}$ under the continuous map $M \mapsto MJ - JM$, so $\mathrm{im}(\phi) \subseteq \mathrm{GL}_{2n}(\mathbb{R})$ is closed. Therefore, we have that ϕ is isomorphic onto its closed image, implying that $\mathrm{GL}_n(\mathbb{C})$ is isomorphic to a closed subgroup of $\mathrm{GL}_{2n}(\mathbb{R})$. □

Exercise 4.3.11. *Let $M \in \mathrm{GL}_n(\mathbb{C})$ with $M = A + iB$ for $A, B \in M_n(\mathbb{R})$. Show that it is not necessarily true that $A, B \in \mathrm{GL}_n(\mathbb{R})$.*

Lemma 4.3.12. *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ be closed and let $H \leq G$ be closed. Then $H \leq \mathrm{GL}_n(\mathbb{F})$ is closed.*

Proof. Being a subgroup is a transitive property and so $H \leq \mathrm{GL}_n(\mathbb{F})$. Similarly, closed subsets of closed subsets are closed subsets, meaning $H \leq \mathrm{GL}_n(\mathbb{R})$ is closed and thus a closed subgroup. □

Corollary 4.3.13.

1. *Any $G \leq \mathrm{GL}_n(\mathbb{R})$ that is closed is also a closed subgroup of $\mathrm{GL}_n(\mathbb{C})$.*
2. *Any $H \leq \mathrm{GL}_n(\mathbb{C})$ that is closed is isomorphic to a closed subgroup of $\mathrm{GL}_{2n}(\mathbb{R})$.*

Proof.

1. Let $G \leq \mathrm{GL}_n(\mathbb{R})$ be closed. Since $\mathrm{GL}_n(\mathbb{R}) \leq \mathrm{GL}_n(\mathbb{C})$ is closed by Proposition 4.3.9, by Lemma 4.3.12 we have that $G \leq \mathrm{GL}_n(\mathbb{C})$ is closed.
2. Let $H \leq \mathrm{GL}_n(\mathbb{C})$ be closed. By Proposition 4.3.10 there exists an isomorphism $\phi : \mathrm{GL}_n(\mathbb{C}) \to K$ where $K \leq \mathrm{GL}_{2n}(\mathbb{R})$ is closed. In particular, $H \cong \phi(H)$ with $\phi(H) \leq K$. So by Lemma 4.3.12 we have that $H \cong \phi(H) \leq \mathrm{GL}_{2n}(\mathbb{R})$ where $\phi(H) \leq \mathrm{GL}_{2n}(\mathbb{R})$ is closed.

□

Corollary 4.3.14. *A closed subgroup of a matrix Lie group is itself a matrix Lie group.*

Proof. Suppose that $G \leq \mathrm{GL}_n(\mathbb{F})$ is a matrix Lie group and let $H \leq G$ be closed. Then by Lemma 4.3.12 we know that $H \leq \mathrm{GL}_n(\mathbb{F})$ is closed and so H is a matrix Lie group. □

Suppose that $G \leq \mathrm{GL}_n(\mathbb{F})$ is a matrix Lie group. Then to be compact, we can apply the Heine-Borel conditions by thinking of $M_n(\mathbb{F})$ as $\mathbb{F}^{n^2}$. That is, G is compact if it is a closed subset of $M_n(\mathbb{F})$ and it is bounded. Where being bounded amounts to there existing a constant C such that for any $A \in G$ the absolute value of its entries is at most C.

Example 4.3.15.

1. *Consider the set*

$$\begin{aligned} U(1) &:= \left\{ \begin{pmatrix} z \end{pmatrix} \in M_1(\mathbb{C}) : \begin{pmatrix} z \end{pmatrix}^\dagger \begin{pmatrix} z \end{pmatrix} = I = \begin{pmatrix} 1 \end{pmatrix} \right\} \\ &= \{ z \in \mathbb{C} : \bar{z} z = 1 \} \\ &= \{ z \in \mathbb{C} : |z| = 1 \}. \end{aligned}$$

   This is bounded as $|z| \leq 1$. Note that $|\cdot|$ is a continuous function and $\{1\}$ is a closed set. So as $U(1)$ is the pre-image of $\{1\}$ under $|\cdot|$ we deduce that $U(1)$ is closed in $M_1(\mathbb{C})$. Hence, $U(1)$ is compact.

2. *The set $\mathrm{GL}_n(\mathbb{F}) \subseteq \mathbb{F}^{n \times n}$ is not bounded, and hence it is not compact. To see this, consider the matrix*

$$\begin{pmatrix} 1 & 0 & \dots & k \\ 0 & 1 & & \\ \vdots & & \ddots & \\ 0 & & & 1 \end{pmatrix} \in \mathrm{GL}_n(\mathbb{F})$$

   for $k \in \mathbb{N}$. As $k \to \infty$, the norm of this matrix goes to infinity and so $\mathrm{GL}_n(\mathbb{F})$ is not a bounded subset of $M_n(\mathbb{F})$.

4.4 Matrix Exponentiation

Definition 4.4.1. *For $A \in M_n(\mathbb{F})$, its exponential, written e^A or $\exp(A)$, is*

$$e^A := \sum_{k=0}^{\infty} \frac{A^k}{k!}. \tag{4.4.1}$$

Remark 4.4.2. *For $n = 1$, Definition 4.4.1 coincides with the exponential on $\mathbb{F}$.*

Theorem 4.4.3. *Equation* (4.4.1) *converges for all* $A \in M_n(\mathbb{F})$. *In particular, the following statements hold.*

1. *Equation* (4.4.1) *converges absolutely and uniformly on compact subsets of* $M_n(\mathbb{F})$.
2. *The partial derivatives in the components of* (4.4.1) *converge absolutely and uniformly on compact subsets of* $M_n(\mathbb{F})$.

Thus, $\exp : M_n(\mathbb{F}) \to M_n(\mathbb{F})$ *is continuously differentiable.*

Example 4.4.4.

1. *If* $A = 0$, *then*
$$e^A = I + 0 + 0 + \cdots = I.$$
2. *If* $A = I$, *then*
$$e^A = I + \frac{1}{2}I + \frac{1}{6}I + \cdots = eI.$$
3. *If* $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, *then*
$$A^n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$$
so that
$$e^A = \begin{pmatrix} \sum_{k=0}^{\infty} \frac{1}{k!} & \sum_{k=0}^{\infty} \frac{k}{k!} \\ 0 & \sum_{k=0}^{\infty} \frac{1}{k!} \end{pmatrix} = \begin{pmatrix} e & e \\ 0 & e \end{pmatrix}.$$
4. *Let* $A = J_n(\lambda)$. *Then,* $A = D + N$ *where* $D = \lambda I$ *and*
$$N = \begin{pmatrix} 0 & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix}.$$
Note that
$$N^k = \begin{pmatrix} 0 & & & & \\ \vdots & \ddots & & & \\ 1 & & \ddots & & \\ & \ddots & & \ddots & \\ & & 1 & \overset{k}{\cdots} & 0 \end{pmatrix}$$
for $1 \leq n - 1$ *meaning*
$$e^N = \begin{pmatrix} 1 & & & \\ \frac{1}{2} & \ddots & & \\ \vdots & \ddots & \ddots & \\ \frac{1}{(n-1)!} & \cdots & \frac{1}{2} & 1 \end{pmatrix}.$$

Therefore, as $DN = ND$ it follows that

$$e^A = e^D e^N = e^\lambda \begin{pmatrix} 1 & & & & \\ 1 & \ddots & & & \\ \frac{1}{2} & \ddots & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & \\ \frac{1}{(n-1)!} & \cdots & \frac{1}{2} & 1 & 1 \end{pmatrix}.$$

Exercise 4.4.5. *Using statement 4 of Example 4.4.4, show that any $A \in \mathrm{GL}_n(\mathbb{C})$ can be written as the exponential of a matrix.*

Definition 4.4.6. *For $A \in M_n(\mathbb{F})$ its logarithm, $\log(A)$, is*

$$\log(A) := \sum_{k=1}^{\infty} \frac{(-1)^{k+1}(A-I)^k}{k}. \tag{4.4.2}$$

Equation (4.4.2) does not converge for all $A \in M_n(\mathbb{F})$. However, for A sufficiently close to I similar results for (4.4.2) hold as those detailed in Theorem 4.4.3 for (4.4.1).

Theorem 4.4.7. *There is an open neighbourhood U of I in $M_n(\mathbb{F})$ such that (4.4.2) converges for all $A \in U$. In particular, the following statements hold.*

1. *Equation (4.4.2) converges absolutely on U, and uniformly on compact subsets of U.*
2. *The partial derivatives in the components of (4.4.2) converge absolutely on U and uniformly on compact subsets of U.*

Thus, $\log : U \to M_n(\mathbb{F})$ is continuously differentiable.

Example 4.4.8.

1. *When $n = 1$, Definition 4.4.6 coincides with the logarithm on $\mathbb{F}$. In particular, for this case, $\log(z)$ converges for $|z-1| < 1$. Similarly, for $n \geq 1$ we have that (4.4.2) is guaranteed to converge on*

$$U = \left\{ M : |M_{ij} - \delta_{ij}| < \frac{1}{n}, \text{ for each } i, j = 1, \ldots, n \right\}.$$

2. *Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Note that $(A-I)^k = 0$ for $k \geq 2$. Hence, (4.4.2) converges with*

$$\log(A) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} =: B.$$

Moreover, we can check that $e^B = A$ and so in this case $e^{\log(A)} = A$.

When $\exp$ and $\log$ converge they are inverses of each other.

Proposition 4.4.9. *Suppose (4.4.2) is absolutely convergent for $A \in M_n(\mathbb{F})$. Then $\exp(\log(A)) = A$. Similarly, if the (4.4.2) converges for $\exp(B)$, where $B \in M_n(\mathbb{F})$, then $\log(\exp(B)) = B$.*

The proof of Proposition 4.4.9 amounts to writing the composition of $\exp(\log(A))$ as a double power series in A

and showing that it simplifies to A.

Proposition 4.4.10. *Let $A, B \in M_n(\mathbb{F})$. If $AB = BA$ then $e^{A+B} = e^A e^B$.*

Proof. Observe that

$$\begin{aligned} e^{A+B} &= \sum_{k=0}^{\infty} \frac{(A+B)^k}{k!} \\ &\overset{(1)}{=} \sum_{k=0}^{\infty} \sum_{l=0}^{k} \frac{1}{k!} \binom{k}{l} A^{k-l} B^l \\ &= \sum_{i,j=0}^{\infty} \frac{1}{i!} \frac{1}{j!} A^i B^j \\ &= \left(\sum_{i=0}^{\infty} \frac{A^i}{i!} \right) \left(\sum_{j=0}^{\infty} \frac{B^j}{j!} \right) \\ &= e^A e^B, \end{aligned}$$

where in (1) we use the assumption that $AB = BA$. Moreover, when we rearrange the infinite series we use the fact the series absolutely converges. □

Remark 4.4.11. *The requirement that $AB = BA$ in Proposition 4.4.10 is necessary.*

Exercise 4.4.12. *Give an example of $A, B \in M_n(\mathbb{F})$ where $e^{A+B} \neq e^A e^B$.*

Corollary 4.4.13. *For $A \in M_n(\mathbb{F})$ we have that $e^A e^{-A} = e^0 = I$. In particular, e^A is invertible.*

Proof. As A and $-A$ commute we can use Proposition 4.4.10 to deduce that

$$e^A e^{-A} = e^{A-A} = e^0 = I.$$

□

Remark 4.4.14. *Using Corollary 4.4.13, it makes sense to write* $\exp : M_n(\mathbb{F}) \to \mathrm{GL}_n(\mathbb{F})$.

Example 4.4.15. *As A and A^k always commute, we deduce that A and e^A commute. Hence,*

$$e^{A+e^A} = e^A e^{e^A}.$$

Lemma 4.4.16. *Let $A, B \in M_n(\mathbb{F})$ with $A \in \mathrm{GL}_n(\mathbb{F})$. Then,*

$$e^{ABA^{-1}} = Ae^B A^{-1}.$$

Proof. Note that $\left(ABA^{-1}\right)^k = AB^kA^{-1}$ for all $k \in \mathbb{N}$. So,

$$\begin{aligned} e^{ABA^{-1}} &= \sum_{k=0}^{\infty} \frac{\left(ABA^{-1}\right)^k}{k!} \\ &= \sum_{k=0}^{\infty} \frac{AB^kA^{-1}}{k!} \\ &= Ae^BA^{-1}. \end{aligned}$$

□

Lemma 4.4.17. *For $A \in M_n(\mathbb{F})$ we have that $\left(e^A\right)^\top = e^{A^\top}$.*

Proof. Note that $\left(A^\top\right)^k = \left(A^k\right)^\top$ for all $k \in \mathbb{N}$. So,

$$\begin{aligned} e^{A^\top} &= \sum_{k=0}^{\infty} \frac{\left(A^\top\right)^k}{k!} \\ &= \left(\sum_{k=0}^{\infty} \frac{A^k}{k!}\right)^\top \\ &= \left(e^A\right)^\top. \end{aligned}$$

□

Example 4.4.18. *Consider the upper-triangular matrix*

$$A = \begin{pmatrix} \lambda_1 & & \times \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Then

$$A^k = \begin{pmatrix} \lambda_1^k & & \times \\ & \ddots & \\ 0 & & \lambda_n^k \end{pmatrix}.$$

Therefore,

$$e^A = \begin{pmatrix} \sum_{k=0}^{\infty} \frac{\lambda_1^k}{k!} & & \times \\ & \ddots & \\ 0 & & \sum_{k=0}^{\infty} \frac{\lambda_n^k}{k!} \end{pmatrix}$$

which is also upper-triangular. From this, we observe that

$$\det\left(e^A\right) = e^{\lambda_1} \dots e^{\lambda_n} = e^{\operatorname{tr}(A)}.$$

Lemma 4.4.19. *For $A \in M_n(\mathbb{F})$ we have*

$$\det\left(e^A\right) = e^{\operatorname{tr}(A)}.$$

Proof. Write $A = PBP^{-1}$ where $P \in \mathrm{GL}_n(\mathbb{F})$, even for $\mathbb{F} = \mathbb{R}$ we can do this, and B is in Jordan normal form. Then

$$\begin{aligned}\det\left(e^A\right) &= \det\left(e^{PBP^{-1}}\right)\\ &= \det\left(Pe^BP^{-1}\right)\\ &= \det\left(e^B\right)\end{aligned}$$

and

$$\begin{aligned}e^{\mathrm{tr}(A)} &= e^{\mathrm{tr}\left(PBP^{-1}\right)}\\ &= e^{\mathrm{tr}(B)}.\end{aligned}$$

Using Example 4.4.18 we know that $\det\left(e^B\right) = e^{\mathrm{tr}(B)}$ and so $\det\left(e^A\right) = e^{\mathrm{tr}(A)}$. □

Recall from Theorem 4.4.3 that $\exp : M_n(\mathbb{F}) \to \mathrm{GL}_n(\mathbb{F})$ is continuously differentiable. Thus, for any $A \in M_n(\mathbb{F})$ the function $\gamma_A : \mathbb{R} \to \mathrm{GL}_n(\mathbb{F})$ given by $\gamma_A(t) = e^{tA}$ is continuously differentiable.

Lemma 4.4.20. *For any $A \in M_n(\mathbb{F})$, let $\gamma_A : \mathbb{R} \to \mathrm{GL}_n(\mathbb{F})$ be given by $\gamma_A(t) = e^{tA}$. Then $\dot{\gamma}_A(0) = A$.*

Proof. Proceeding from first principles we get that

$$\begin{aligned}\dot{\gamma}_A(0) &= \lim_{t\to 0}\frac{e^{tA} - e^0}{t}\\ &= \lim_{t\to 0}\frac{\sum_{k=0}^{\infty}\frac{t^kA^k}{k!} - I}{t}\\ &= \lim_{t\to 0}\sum_{k=1}^{\infty}\frac{t^{k-1}A^k}{k!}\\ &= A.\end{aligned}$$

□

Lemma 4.4.21. *For $A \in M_n(\mathbb{F})$ the map $\gamma_A : \mathbb{R} \to \mathrm{GL}_n(\mathbb{F})$ given by $\gamma_A(t) = e^{tA}$ is a continuous homomorphism.*

Proof. Since $\exp$ is continuous, γ_A is continuous. Let $s, t \in \mathbb{R}$ then

$$\begin{aligned}\gamma_A(s+t) &= e^{(s+t)A}\\ &\overset{(1)}{=} e^{sA}e^{tA}\\ &= \gamma_A(s)\gamma_A(t)\end{aligned}$$

where in (1) we are using Proposition 4.4.10 as sA and tA commute. Therefore, γ_A is a continuous homomorphism. □

Corollary 4.4.22. *For $A \in M_n(\mathbb{F})$ let $\gamma_A : \mathbb{R} \to \mathrm{GL}_n(\mathbb{F})$ be given by $\gamma_A(t) = e^{tA}$. Then for any $t_0 \in \mathbb{R}$ we have that $\dot{\gamma}_A(t_0) = A\gamma_A(t_0)$.*

Proof. Since γ_A is a homomorphism,

$$\gamma_A(t) = \gamma_A(t - t_0)\gamma_A(t_0).$$

Differentiating both sides at $t = t_0$ and using Lemma 4.4.20, it follows that

$$\dot{\gamma}_A(t_0) = \dot{\gamma}_A(0)\gamma_A(t_0) = Ae^{t_0A}.$$

□

The derivative of $\exp$ at zero can be thought of as a linear map $\mathrm{d}\exp_0 : M_n(\mathbb{F}) \to M_n(\mathbb{F})$.

Corollary 4.4.23. *The map $\mathrm{d}\exp_0 : M_n(\mathbb{F}) \to M_n(\mathbb{F})$ is the identity map.*

Proof. For $A \in M_n(\mathbb{F})$, note that the chain rule shows that the derivative of γ_A at $t = 0$ is $\mathrm{d}\exp_0(A)$. Therefore, using Lemma 4.4.20 we have that $A = \mathrm{d}\exp_0(A)$ for all $A \in M_n(\mathbb{F})$ which implies that $\mathrm{d}\exp_0$ is the identity map. □

Proposition 4.4.24. *The map $\exp$ restricts to a continuously differentiable homomorphism from an open neighbourhood of $0 \in M_n(\mathbb{F})$ to an open neighbourhood of $I \in \mathrm{GL}_n(\mathbb{F})$, with a continuously differentiable inverse.*

Proof. Note that $\exp(0) = I$. By Corollary 4.4.23 we know that $\exp$ is differentiable at zero with a derivative that is an invertible linear map. Therefore, we conclude by applying Theorem 5.2.2. □

Example 4.4.25. *For $n = 1$ we can take $U = (-1, 1)$ and observe that*

$$\exp|_U : (-1, 1) \to \left(e^{-1}, e\right)$$

is a homeomorphism with inverse being $\log$, which is continuously differentiable.

Proposition 4.4.26. *Any continuous homomorphism $\psi : (\mathbb{R}, +) \to (\mathbb{R}^\times, \cdot)$ is of the form $\psi(t) = e^{\lambda t}$ for some $\lambda \in \mathbb{R}$.*

Proof. Let $\psi : (\mathbb{R}, +) \to (\mathbb{R}^\times, \cdot)$ be a continuous homomorphism. Then $\psi(1) = 1 > 0$, and so by the intermediate value theorem we know that $\psi(t) > 0$ for all $t \in \mathbb{R}$ as otherwise there would be some $s \in \mathbb{R}$ for which $\psi(s) = 0$. Let $\phi : \mathbb{R} \to \mathbb{R}$ be given by $t \mapsto \log(\psi(t))$. Note that ϕ is continuous and for $s, t \in \mathbb{R}$ we have that

$$\begin{aligned}\phi(s + t) &= \log(\psi(s + t)) \\ &= \log(\psi(s)\psi(t)) \\ &= \phi(s) + \phi(t)\end{aligned}$$

and so ϕ is a homomorphism. As any continuous homomorphism $(\mathbb{R}, +) \to (\mathbb{R}, +)$ is of the form $t \mapsto \lambda t$ for some $\lambda \in \mathbb{R}$, we deduce that $\phi(t) = \lambda t$ which implies that $\psi(t) = e^{\lambda t}$. □

Note that in the proof of Proposition 4.4.26 we use the fact that $\log$ is defined for all positive real numbers. This is not the case for $n > 0$ and so studying homomorphism $\mathbb{R} \to \mathrm{GL}_1(\mathbb{C}) = \mathbb{C}^\times$ requires extra care.

Theorem 4.4.27. *Let $\gamma : \mathbb{R} \to \mathrm{GL}_n(\mathbb{R})$ be a homomorphism of topological groups. Then $\gamma(t) = e^{tA}$ for some $A \in M_n(\mathbb{F})$.*

Proof. Using Proposition 4.4.24 we can choose open neighbourhoods U of zero in $M_n(\mathbb{F})$ and V of I in $\mathrm{GL}_n(\mathbb{F})$ such that $V = \exp(U)$ and $\exp|_U : U \to V$ is a homeomorphism. By the continuity of γ, there is some $\delta > 0$ such that $\gamma([-\delta, \delta]) \subseteq V$. Let $\beta : [-\delta, \delta] \to U$ be given by $t \mapsto \exp^{-1}|_U(\gamma(t))$.
Claim 1: $\beta(r\delta) = r\beta(\delta)$ for all $r \in [-1, 1]$.
Proof. Let $r = \frac{p}{q} \in [0, 1] \cap \mathbb{Q}$. Then

$$\begin{aligned}\exp(\beta(\delta)) &= \gamma(\delta) \\ &\overset{(1)}{=} \gamma\left(\frac{\delta}{q}\right)^q \\ &= \exp\left(\beta\left(\frac{\delta}{q}\right)\right)^q \\ &= \exp\left(q\beta\left(\frac{\delta}{q}\right)\right)\end{aligned}$$

where in (1) we used the fact that γ is a homomorphism. As this equality holds in V we deduce that $q\beta\left(\frac{\delta}{q}\right) = \beta(\delta)$. Similarly,

$$\begin{aligned}\exp\left(\beta\left(\frac{p\delta}{q}\right)\right) &= \gamma\left(\frac{p\delta}{q}\right)\\ &= \gamma\left(\frac{\delta}{q}\right)^p\\ &= \exp\left(\beta\left(\frac{\delta}{q}\right)\right)^p\\ &= \exp\left(p\frac{\beta(\delta)}{q}\right)\end{aligned}$$

and so

$$\beta\left(\frac{p\delta}{q}\right) = \frac{p}{q}\beta(\delta).$$

This can be extended to any $r \in [-1,1]\cap\mathbb{Q}$. For $r \in [-1,1]$ we choose a sequence $(r_n)_{n\in\mathbb{N}} \subseteq [-1,1]\cap\mathbb{Q}$, where $r_n = \frac{p_n}{q_n}$, converging to r and use the continuity of β to deduce that $\beta(r_n\delta) \to \beta(r\delta)$. As $\beta(r_n\delta) = r_n\beta(\delta)$ and $r_n\beta(\delta) \to r\beta(\delta)$, it follows by the uniqueness of limits that $\beta(r\delta) = r\beta(\delta)$.

Claim 2: For any $t \in \mathbb{R}$ we have $\gamma(t) = \exp\left(\frac{t\beta(\delta)}{\delta}\right)$.

Proof. Let $t \in \mathbb{R}$. Then there exists some integer $N > 0$ such that $\left|\frac{t}{N}\right| \leq \delta$. That is, $\frac{t}{N} = r\delta$ for some $r \in [-1,1]$. Therefore, by Claim 1 it follows that

$$\gamma\left(\frac{t}{N}\right) = \gamma(r\delta) = e^{r\beta(\delta)}.$$

Therefore,

$$\begin{aligned}\gamma(t) &= \gamma\left(\frac{t}{N}\right)^N\\ &= e^{rN\beta(\delta)}\\ &= \exp\left(\frac{t\beta(\delta)}{\delta}\right).\end{aligned}$$

□

Theorem 4.4.27 shows that γ is continuously differentiable.

Corollary 4.4.28. *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ be a closed. Then any continuous homomorphism $\gamma : \mathbb{R} \to G$ is of the form γ_A for some $A \in \mathrm{GL}_n(\mathbb{F})$.*

Proof. Since $G \leq \mathrm{GL}_n(\mathbb{F})$, any continuous homomorphism $\gamma : \mathbb{R} \to G$ is a continuous homomorphism $\mathbb{R} \to \mathrm{GL}_n(\mathbb{F})$. Hence, we can conclude by applying Theorem 4.4.27. □

Example 4.4.29. *For $A \in M_n(\mathbb{R})$, we have that $e^A \in \mathrm{SL}_n(\mathbb{R})$ if and only if $\mathrm{tr}(A) = 0$. Indeed, recall by Lemma 4.4.19 we have that $\det\left(e^A\right) = e^{\mathrm{tr}(A)}$. So since $\mathrm{tr}(A) \in \mathbb{R}$ it follows that $\det\left(e^A\right) = 1$ if and only if $\mathrm{tr}(A) = 0$. In particular, consider $\gamma : \mathbb{R} \to \mathrm{SL}_n(\mathbb{R})$ a continuous homomorphism. Then $\gamma = \gamma_A$ for some $A \in M_n(\mathbb{R})$. Hence, $e^{tA} \in \mathrm{SL}_n(\mathbb{R})$ for all $t \in \mathbb{R}$ if and only if $\mathrm{tr}(tA) = 0$ which happens if and only if $\mathrm{tr}(A) = 0$.*

4.5 The Lie Algebra of a Matrix Lie Group

Definition 4.5.1. *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ be closed and let $A \in G$. Then the tangent space T_AG of G at A is*

$$T_AG := \{\dot{\gamma}(0) \in M_n(\mathbb{F}) : \gamma : \mathbb{R} \to G \text{ continuously differentiable with } \gamma(0) = A\}.$$

Remark 4.5.2. *The γ in Definition 4.5.1 need not be a homomorphism. A group homomorphism has to satisfy $\gamma(0) = I$, whereas γ in Definition 4.5.1 is such that $\gamma(0) = A$.*

Proposition 4.5.3. *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ be a closed subgroup. Then for any $A \in G$ we have that*

$$T_AG = A \cdot T_IG = T_IG \cdot A.$$

Proof. Let $M \in T_IG$ with $M = \dot{\gamma}(0)$ for $\gamma : \mathbb{R} \to G$ a continuously differentiable function with $\gamma(0) = I$. Consider the function $A \cdot \gamma : \mathbb{R} \to G$ given by $t \mapsto A \cdot \gamma(t)$. This is continuously differentiable with $\gamma(0) = A$. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t}(A\gamma(t))|_{t=0} = A\dot{\gamma}(0) \in T_AG.$$

Hence, $A \cdot T_IG \subseteq T_AG$. A similar argument with A^{-1} shows that

$$A^{-1} \cdot T_AG \subseteq T_IG$$

which implies that $T_AG \subseteq A \cdot T_IG$ and so $T_AG = A \cdot T_IG$. Similarly, one shows that $T_AG = T_IG \cdot A$. □

Proposition 4.5.4. *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ be closed. Then for any A, the set T_AG is a real vector space.*

Proof. For $A = I$, let $M, N \in T_IG$ and $\lambda \in \mathbb{R}$. Let $\alpha, \beta : \mathbb{R} \to G$ be continuously differentiable functions such that $\alpha(0) = \beta(0) = I$, $\dot{\alpha}(0) = M$ and $\dot{\beta}(0) = N$. Let $\gamma(t) = \alpha(\lambda t)$. Then since $\gamma : \mathbb{R} \to G$ is continuously differentiable with $\gamma(0) = I$ and $\dot{\gamma}(0) = \lambda\dot{\alpha}(0) = \lambda M$ it follows that $\lambda M \in T_IG$. Now let $\delta(t) = \alpha(t)\beta(t)$. Then $\delta : \mathbb{R} \to G$ is continuously differentiable with $\delta(0) = I$. Using the product rule we deduce that

$$\dot{\delta}(0) = \dot{\alpha}(0)\beta(0) + \alpha(0)\dot{\beta}(0) = MI + IN = M + N \in T_IG.$$

Therefore, T_IG is a real vector space. We generalise the arguments to T_AG by using Proposition 4.5.3. □

Definition 4.5.5. *The dimension of a closed subgroup $G \leq \mathrm{GL}_n(\mathbb{F})$ closed is given by the dimension of T_AG as a real vector space.*

Recall that the commutator of $A, B \in M_n(\mathbb{F})$ is

$$[A, B] = AB - BA. \tag{4.5.1}$$

Proposition 4.5.6. *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ be closed, and let $M, N \in T_IG$. Then $[M, N] \in T_IG$.*

Proof. Let $\alpha, \beta : \mathbb{R} \to G$ be continuously differentiable with $\alpha(0) = \beta(0) = I$, $\dot{\alpha}(0) = M$ and $\dot{\beta}(0) = N$. For $s, t \in \mathbb{R}$ let $\delta_s(t) = \alpha(s)\beta(t)\alpha(s)^{-1}$. Note that $\delta_s(t) : \mathbb{R}^2 \to G$ is a continuously differentiable function. In particular, for fixed s the function $\delta_s : \mathbb{R} \to G$ is continuously differentiable with $\delta_s(0) = I$ and so $\dot{\delta}_s(0) \in T_IG$. By the product rule we have that

$$\dot{\delta}_s(0) = \alpha(s)\dot{\beta}(0)\alpha(s)^{-1} = \alpha(s)N\alpha(s)^{-1}.$$

Letting s vary we observe that $\dot{\delta}_s(0) : \mathbb{R} \to T_IG$ is a continuously differentiable path with $\dot{\delta}_0(0) = I$. So its derivative at zero lies in T_IG, that is

$$\frac{\mathrm{d}}{\mathrm{d}s}\dot{\delta}_s(0)\big|_{s=0} \in T_IG.$$

Therefore, as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\dot{\delta}_s(0)\big|_{s=0} &= \left(\frac{\mathrm{d}}{\mathrm{d}s}\alpha(s)\big|_{s=0}\right)\cdot N\cdot\alpha(0)^{-1} + \alpha(0)\cdot N\cdot\left(\frac{\mathrm{d}}{\mathrm{d}s}\alpha(s)^{-1}\big|_{s=0}\right)\\ &= \dot{\alpha}(0)\cdot N\cdot\alpha(0)^{-1} - \alpha(0)\cdot N\cdot\left(-\dot{\alpha}(0)\alpha(0)^{-2}\right)\\ &= M\cdot N\cdot I - I\cdot N\cdot M\\ &= [M,N]\end{aligned}$$

we deduce that $[M,N]\in T_IG$. □

Proposition 4.5.7. *For $A,B,C\in M_n(\mathbb{F})$ and the commutator bracket $[\cdot,\cdot]$, as given in* (4.5.1)*, the Jacobi identity holds. That is,*

$$[A,[B,C]]+[B,[C,A]]+[C,[A,B]]=0.$$

Proof. Expanding the commutators we get that

$$\begin{aligned}[A,[B,C]]+[B,[C,A]]+[C,[B,A]] &= ABC-ACB-BCA+CBA\\ &\quad+BCA-BAC-CAB+ACB\\ &\quad+CAB-CBA-ABC+BAC\\ &= 0.\end{aligned}$$

□

Definition 4.5.8. *A real Lie algebra is a pair $(V,[\cdot,\cdot])$, where V is a real vector space, and $[\cdot,\cdot]:V\times V\to\mathbb{R}$ is a bilinear map such that the following statements hold.*

1. *$[\cdot,\cdot]:V\times V\to\mathbb{R}$ is anti-symmetric. That is, $[A,B]=-[B,A]$ for all $A,B\in V$.*
2. *The Jacobi identity holds. That is,*

$$[A,[B,C]]+[B,[C,A]]+[C,[A,B]]=0$$

for all $A,B,C\in V$.

Remark 4.5.9. *Often a Lie algebra will just be referred to as V, where the $[\cdot,\cdot]$ is dropped from the notation.*

Definition 4.5.10. *Let $(V,[\cdot,\cdot]_V)$ and $(W,[\cdot,\cdot]_W)$ be Lie algebras. Then $f:V\to W$ is a homomorphism of Lie algebras if it is linear and*

$$f\left([A,B]_V\right)=[f(A),f(B)]_W$$

for all $A,B\in V$.

Definition 4.5.11. *For $G\leq\mathrm{GL}_n(\mathbb{F})$ closed, its Lie algebra is $(T_IG,[\cdot,\cdot])$ where $[\cdot,\cdot]$ is the commutator of matrices as given in* (4.5.1)*.*

Remark 4.5.12.

1. *Definition 4.5.11 makes sense as T_IG is a real-vector space by Proposition 4.5.4, $[\cdot,\cdot]$ on T_IG is well-defined by Proposition 4.5.6 and satisfies the Jacobi identity by Proposition 4.5.7.*
2. *The convention is to denote Lie algebras using lowercase Franktur font. So $\mathrm{GL}_n(\mathbb{R})$ as a Lie algebra is*

denoted $\mathfrak{gl}_n(\mathbb{R})$*, similarly* $\mathrm{SL}_n(\mathbb{R})$ *is denoted* $\mathfrak{sl}(\mathbb{R})$*.*

Proposition 4.5.13. *Let* $\gamma : \mathbb{R} \to \mathrm{GL}_n(\mathbb{F})$ *be a continuously differentiable function with* $\gamma(0) = I$*. For* $t \in \mathbb{R}$ *we have*

$$e^{t\dot{\gamma}(0)} = \lim_{n\to\infty} \gamma\left(\frac{t}{n}\right)^n.$$

Proof. By the same arguments as those made at the beginning of the proof of Theorem 4.4.27, we can write $\gamma(t) = e^{\beta(t)}$ for $|t| \leq \delta$ for some $\delta > 0$ and $\beta : [-\delta, \delta] \to M_n(\mathbb{F})$ a continuously differentiable function. Recall $\mathrm{d}\exp_0 : M_n(\mathbb{F}) \to M_n(\mathbb{F})$ is the identity map, so by the chain rule it follows that $\dot{\beta}(0) = \dot{\gamma}(0)$. Therefore,

$$\begin{aligned}
\lim_{n\to\infty} \gamma\left(\frac{t}{n}\right)^n &\overset{(1)}{=} \lim_{n\to\infty} \left(\exp\left(\beta\left(\frac{t}{n}\right)\right)^n\right) \\
&\overset{(2)}{=} \lim_{n\to\infty} \left(\exp\left(\frac{t}{n}\dot{\gamma}(0) + o\left(\frac{t}{n}\right)\right)^n\right) \\
&= \lim_{n\to\infty} \exp\left(t\dot{\gamma}(0) + no\left(\frac{t}{n}\right)\right) \\
&= e^{t\dot{\gamma}(0)}
\end{aligned}$$

where in (1) for have the fact that for large n we have $\left|\frac{t}{n}\right| < \delta$, and in (2) we have used a Taylor expansion for γ. □

Remark 4.5.14. *Note that if* γ *is additionally a homomorphism in the setting of Proposition 4.5.13, it follows that* $e^{t\dot{\gamma}(0)} = \gamma(t)$*.*

Example 4.5.15. *For* $n = 1$ *and* $\gamma = 1 + t$ *the result of Proposition 4.5.13 gives the expected*

$$e^t = \lim_{n\to\infty} \left(1 + \frac{t}{n}\right)^n.$$

Proposition 4.5.16. *Let* $G \leq \mathrm{GL}_n(\mathbb{F})$ *be closed, and let* $M \in \mathfrak{g} = T_I G$*. Then* $e^M \in G$*.*

Proof. Let $\gamma : \mathbb{R} \to G$ be a continuously differentiable curve with $\gamma(0) = I$, and $\dot{\gamma}(0) = M$. Then by Proposition 4.5.13 the limit

$$e^M = \lim_{n\to\infty} \gamma\left(\frac{1}{n}\right)^n$$

exists. Therefore, as $\gamma\left(\frac{1}{n}\right)^n \in G$ and $G \subseteq \mathrm{GL}_n(\mathbb{F})$ is closed, we deduce that $e^M \in G$. □

Corollary 4.5.17. *For* $G \leq \mathrm{GL}_n(\mathbb{F})$*, the continuous homomorphism* $\gamma : \mathbb{R} \to G$ *are exactly those given by* $\gamma_A(t) = e^{tA}$ *for some* $A \in \mathfrak{g} = T_I G$*.*

Proof. Any continuous homomorphism $\gamma : \mathbb{R} \to G$ is also a continuous homomorphism into $\mathrm{GL}_n(\mathbb{R})$. Hence, $\gamma(t) = \gamma_A(t) = e^{tA}$ for some $A \in M_n(\mathbb{F})$ by Theorem 4.4.27. As γ_A is continuously differentiable it follows that $A = \dot{\gamma}_A(0) \in \mathfrak{g}$. Conversely, if $A \in \mathfrak{g}$ then $tA \in \mathfrak{g}$ for all $t \in \mathbb{R}$. Then, by Proposition 4.5.16 $\gamma_A(t) = e^{tA} \in G$ for all $t \in \mathbb{R}$. Moreover, by Lemma 4.4.21, $\gamma_A : \mathbb{R} \to \mathrm{GL}_n(\mathbb{F})$ is a continuous homomorphism. Since $\mathrm{im}(\gamma_A) \subseteq G$ when $A \in \mathfrak{g}$ it follows that $\gamma_A : \mathbb{R} \to G$ is a continuous homomorphism. □

Remark 4.5.18.

1. *From Corollary 4.5.17 we see that there is a bijection between $\mathfrak{g}$ and the continuous homomorphism $\mathbb{R} \to G$. Where $A \in \mathfrak{g}$ is mapped to γ_A in one direction, and γ to $\dot{\gamma}(0)$ in the other.*

2. *Let $\gamma, \delta : \mathbb{R} \to G$ be continuous homomorphism with $G \leq \mathrm{GL}_n(\mathbb{F})$ being closed. Then they are continuously differentiable with $\dot{\gamma}(0) + \dot{\delta}(0) \in \mathfrak{g} = T_I G$, which corresponds to some continuous homomorphism $\xi : \mathbb{R} \to G$ given by $t \mapsto \exp\left(t\left(\dot{\gamma}(0) + \dot{\delta}(0)\right)\right)$. Note that $\eta(t) = \gamma(t) \cdot \delta(t) : \mathbb{R} \to G$ is a continuously differentiable map with*

$$\dot{\eta}(0) = \dot{\gamma}(0) + \dot{\delta}(0) \in \mathfrak{g},$$

but η is not necessarily a homomorphism as G may not be abelian. Therefore, we cannot deduce that ξ and η are equal, but by using Proposition 4.5.16 we can write

$$\begin{aligned} \xi(t) &= e^{t\dot{\eta}(0)} \\ &= \lim_{n\to\infty} \left(\gamma\left(\frac{t}{n}\right)\delta\left(\frac{t}{n}\right)\right)^n. \end{aligned}$$

Theorem 4.5.19. *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ be closed. Then there are open neighbourhoods U of zero in $\mathfrak{g}$ and $V = \exp(U)$ of I in G such that $\exp|_U : U \to V$ is a homeomorphism.*

Proof. Let $W \leq M_n(\mathbb{F})$ be a subspace such that $\mathfrak{g} \oplus W = M_n(\mathbb{F})$. Let $\widetilde{\exp} : \mathfrak{g} \oplus W \to \mathrm{GL}_n(\mathbb{F})$ be given by $(X, Y) \mapsto e^X e^Y$. As X and Y might not commute this is not the modification of $\exp$ given by $(X, Y) \mapsto e^{X+Y}$. However, $\exp$ and $\widetilde{\exp}$ agree on $\mathfrak{g}$. Hence, it suffices to prove the statement for $\widetilde{\exp}$.
Step 1: The derivative of $\widetilde{\exp}$ at zero $\mathrm{d}\widetilde{\exp}_0 : M_n(\mathbb{F}) \to M_n(\mathbb{F})$, is the identity.
Let $(X, Y) \in \mathfrak{g} \oplus W$. Let $\gamma(t) = e^{tX}e^{tY} = \widetilde{\exp}(tX, tY)$. Then

$$\begin{aligned} \dot{\gamma}(0) &= \left(\frac{\mathrm{d}}{\mathrm{d}t}e^{tX}\big|_{t=0}\right)e^{0\cdot Y} + e^{0\cdot X}\left(\frac{\mathrm{d}}{\mathrm{d}t}e^{tY}\big|_{t=0}\right) \\ &= X \cdot I + I \cdot Y \\ &= X + Y. \end{aligned}$$

By the chain rule $\dot{\gamma}(0) = \mathrm{d}\widetilde{\exp}_0(X, Y)$ and so $\mathrm{d}\widetilde{\exp}_0(X, Y) = X + Y$.
Using step 1 we can apply Theorem 5.2.2 to obtain a neighbourhood $\tilde{U}$ of zero in $M_n(\mathbb{F})$ and a neighbourhood $\tilde{V} = \widetilde{\exp}\left(\tilde{U}\right)$ of I in $\mathrm{GL}_n(\mathbb{F})$ such that $\widetilde{\exp}|_{\tilde{U}} : \tilde{U} \to \tilde{V}$ is a homeomorphism.
Step 2: There is a neighbourhood Z of zero in $\tilde{U} \subseteq M_n(\mathbb{F})$ such that $\widetilde{\exp}^{-1}(G) \cap Z = \mathfrak{g} \cap Z$.
It suffices to show that for $(X, Y) \in Z \subseteq \mathfrak{g} \oplus W$ we have $\widetilde{\exp}(X, Y) \in G$ if and only if $Y = 0$. Suppose that this is not true, so that for every neighbourhood of zero in $M_n(\mathbb{F})$ there is some $(X, Y) \in \mathfrak{g} \oplus W$ with $Y \neq 0$ and $\widetilde{\exp}(X, Y) = e^X e^Y \in G$. Then there is a sequence of vectors $((X_i, Y_i))_{i\in\mathbb{N}} \subseteq \mathfrak{g} \oplus W$ converging to 0 such that $Y_i \neq 0$ and $\widetilde{\exp}(X_i, Y_i) \in G$ for all $i \in \mathbb{N}$. As $-X_i \in \mathfrak{g}$ it follows that $e^{-X_i} \in G$ and so $e^{-X_i}e^{X_i}e^{Y_i} = e^{Y_i} \in G$. Since each $Y_i \neq 0 \in W$ we can consider

$$\frac{Y_i}{|Y_i|} \in S^d := \{y \in W : |y| = 1\}$$

the normalised Y_i. By the sequential compactness of S^d we can pass to a convergent subsequence, such that $\frac{Y_i}{|Y_i|} \to Y \in W$ for some $Y \in S^d$. In particular, $Y \neq 0$. Now fix $t \in \mathbb{R}$, and choose an integer m_i such that

$$m_i|Y_i| \leq t \leq (m_i + 1)|Y_i|$$

for all $i \in \mathbb{N}$. Note that $e^{m_i Y_i} = \left(e^{Y_i}\right)^{m_i} \in G$. Moreover, as

$$m_i Y_i = (m_i|Y_i|)\frac{Y_i}{|Y_i|} \to tY$$

we deduce that $e^{m_i Y_i} \to e^{tY}$. As $G \subseteq \mathrm{GL}_n(\mathbb{F})$ is closed it follows that $e^{tY} \in G$ for all $t \in \mathbb{R}$ which implies that $Y \in \mathfrak{g}$. However, $\mathfrak{g} \cap W = \{0\}$ and $Y \neq 0$ and so we get a contradiction.
Assume $U = \mathfrak{g} \cap Z$, where Z is given by step 1. Then by step 1 we have that $\widetilde{\exp}|_Z : Z \to \mathrm{im}\,(\widetilde{\exp}|_Z)$ is a homeomorphism to a neighbourhood of I in $\mathrm{GL}_n(\mathbb{F})$. Restricting to U, which is an open neighbourhood of zero, we see that

$$\widetilde{\exp}|_U : U \to \mathrm{im}\,(\widetilde{\exp}|_U)$$

is a homeomorphism. By step 2 we know that $\mathrm{im}\,(\widetilde{\exp}|_U) = G \cap \mathrm{im}\,(\widetilde{\exp}|_Z)$ where $\mathrm{im}\,(\widetilde{\exp}|_Z)$ is an open neighbourhood of I. Hence $G \cap \mathrm{im}\,(\widetilde{\exp}|_Z)$ is an open neighbourhood of I in G. □

Corollary 4.5.20. *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ be closed. Then for all $A \in G$ there is an open neighbourhood U of A in G which is homeomorphic to an open subset of $\mathbb{R}^d$, where d is the dimension of G.*

Proof. When $A = I$ this follows from Theorem 4.5.19 since $\mathfrak{g}$ is homeomorphic to $\mathbb{R}^d$. When $A \neq I$ we note that multiplication by A is a homeomorphism, and sends an open neighbourhood of I, which is homeomorphic to an open subset of $\mathbb{R}^d$, to an open neighbourhood of A in G. □

Corollary 4.5.21. *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ be closed. Then G is a discrete topological space if and only if its dimension is zero.*

Proof. ($\Leftarrow$). If the dimension of G is zero, then every $A \in G$ has a neighbourhood homeomorphic to $\mathbb{R}^0$ by Corollary 4.5.20, which is just a point. Hence, all points in G are open, implying that G has the discrete topology.
($\Rightarrow$). As there is an open neighbourhood U of I in G that is homeomorphic to an open neighbourhood of $\mathbb{R}^d$, where d is the dimension of G, it follows that G is discrete as U is discrete. Hence, $d = 0$. □

Example 4.5.22.

1. *For $G \leq \mathrm{GL}_n(\mathbb{F})$ closed, there is an open neighbourhood U of I in G which is homeomorphic to a ball in $\mathbb{R}^d$. We can say this as we can restrict open sets to open balls. In particular, this implies that U is path-connect. For the topological group $(\mathbb{Q}, +)$, no open neighbourhood of zero in $\mathbb{Q}$ is path-connected. So $(\mathbb{Q}, +)$ is not isomorphic to a closed subgroup of any $\mathrm{GL}_n(\mathbb{R})$. This shows that $(\mathbb{Q}, +)$ is not a matrix Lie group.*

2. *Let*

$$G := U(1) \times U(1) = \left\{ \begin{pmatrix} z & 0 \\ 0 & w \end{pmatrix} : z, w \in \mathbb{C},\ |z| = |w| = 1 \right\}.$$

Let $\gamma : \mathbb{R} \to G$ be the continuous homomorphism given by

$$t \mapsto \begin{pmatrix} e^{it} & 0 \\ 0 & e^{i\sqrt{2}t} \end{pmatrix}.$$

Then γ is injective, but no open neighbourhood of $I \in \mathrm{im}(\gamma)$ is path-connected, due to the zeros on the cross-diagonal. Hence, $\mathrm{im}(\gamma) \leq G$ is a subgroup but cannot be closed as it cannot be a matrix Lie group. Hence, images of homomorphism are not necessarily matrix Lie groups, which differs from the theory of other abstract objects we have discussed so far. Note that injectivity follows for any other irrational number in the position of $\sqrt{2}$.

Example 4.5.23.

1. *From Example 4.4.29, we see that*

$$\mathfrak{sl}_n(\mathbb{R}) = \{A \in M_n(\mathbb{R}) : \mathrm{tr}(A) = 0\}.$$

2. *Consider the Lie algebra of $O(n)$, namely $\mathfrak{o}(n) = T_I O(n)$. Let $A \in \mathfrak{o}(n)$, then $tA \in \mathfrak{o}(n)$ for all $t \in \mathbb{R}$. Hence, $e^{tA} \in O(n)$ for all $t \in \mathbb{R}$ which implies that*

$$e^{tA}\left(e^{tA}\right)^\top = I.$$

As $e^{tA^\top} = \left(e^{tA}\right)^\top$ and we know that $e^{-tA^\top}$ is the inverse of $e^{tA^\top}$, it follows that $e^{tA} = e^{-tA^\top}$. As $\exp$ is not injective we cannot immediately deduce that $tA = -tA^\top$. However, for sufficiently small $t > 0$, we have that tA and $-tA^\top$ lie in an open neighbourhood of zero in $M_n(\mathbb{F})$ on which $\exp$ is injective. Hence, $tA = -tA^\top$ and so $A = -A^\top$ which says that A is skew-symmetric. Conversely, suppose that A is skew-symmetric. Then as A and $A^\top = -A$ commute we have that

$$\begin{aligned} e^{tA}\left(e^{tA}\right)^\top &= e^{tA}e^{-tA} \\ &= e^0 \\ &= I \end{aligned}$$

for all $t \in \mathbb{R}$. Hence, $e^{tA} \in O(n)$. Letting $\gamma_A : \mathbb{R} \to O(n)$ be the continuous homomorphism given by $t \mapsto e^{tA}$ it follows that $A \in \mathfrak{o}(n)$. In conclusion, we have shown that $\mathfrak{o}(n) \leq M_n(\mathbb{R})$ is the set of skew-symmetric matrices.

   - *In particular, we can say that $O(n)$ and $O(m)$ are isomorphic if and only if $n = m$, as $\mathfrak{o}(n)$ has dimension $\frac{1}{2}n(n-1)$.*

3. *As $\mathrm{SL}_n(\mathbb{R}) = O(n) \cap \mathrm{SL}_n(\mathbb{R})$ it follows that*

$$\mathfrak{so}(n) = \mathfrak{o}(n) \cap \mathfrak{sl}_n(\mathbb{R}) = \left\{A \in M_n(\mathbb{R}) : A = -A^\top, \operatorname{tr}(A) = 0\right\}.$$

Definition 4.5.24. *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ and $H \leq \mathrm{GL}_n(\mathbb{F})$ be closed, and let $\phi : G \to H$ be a continuous homomorphism. Let $\phi_* : \mathfrak{g} \to \mathfrak{h}$ be defined as follows. For $A \in \mathfrak{g}$, let $\gamma_A(t) : \mathbb{R} \to G$ be the continuous homomorphism given by $\gamma_A(t) = e^{tA}$. Then $\phi \circ \gamma : \mathbb{R} \to H$ is also a continuous homomorphism and so $(\phi \circ \gamma)(t) = e^{tB}$ for some unique $B \in \mathfrak{h}$. Consequently, we let*

$$\phi_*(A) = B.$$

Equivalently, $\phi_*(A)$ is characterised by the equation

$$\gamma_{\phi_*(A)} = \phi \circ \gamma_A. \tag{4.5.2}$$

Remark 4.5.25. *Letting $t = 1$ in (4.5.2) we deduce that*

$$\left(\exp \circ \phi_*\right)(A) = (\phi \circ \exp)(A).$$

Lemma 4.5.26. *The map $\phi_* : \mathfrak{g} \to \mathfrak{h}$, as given in Definition 4.5.24, is linear.*

Proof. Let $A, B \in \mathfrak{g}$ and $\lambda \in \mathbb{R}$. Then for any $t \in \mathbb{R}$ we have that

$$\begin{aligned} e^{t\phi_*(\lambda A)} &= \phi\left(e^{t\lambda A}\right) \\ &= e^{(t\lambda)\phi_*(A)} \\ &= e^{t(\lambda\phi_*(A))}. \end{aligned}$$

Since this holds for all $t \in \mathbb{R}$ we deduce that $\phi_*(\lambda A) = \lambda \phi_*(A)$. Note that for any $t \in \mathbb{R}$ we have

$$\begin{aligned}\gamma_{A+B}(t) &= e^{t(A+B)} \\ &= \lim_{n\to\infty} \left(\gamma_A\left(\frac{t}{n}\right)\gamma_B\left(\frac{t}{n}\right)\right)^n\end{aligned}$$

where we have used Proposition 4.5.16 and that fact that if $\delta(t) = \gamma_A(t)\gamma_B(t)$ then $\dot{\delta}(0) = A + B$. Hence,

$$\begin{aligned}\gamma_{\phi_*(A)+\phi_*(B)} &= \lim_{n\to\infty} \left(\gamma_{\phi_*(A)}\left(\frac{t}{n}\right)\gamma_{\phi_*(B)}\left(\frac{t}{n}\right)\right)^n \\ &= \lim_{n\to\infty} \left(\phi\left(\gamma_A\left(\frac{t}{n}\right)\right)\phi\left(\gamma_B\left(\frac{t}{n}\right)\right)\right)^n \\ &\overset{(1)}{=} \lim_{n\to\infty} \phi\left(\left(\gamma_A\left(\frac{t}{n}\right)\gamma_B\left(\frac{t}{n}\right)\right)^n\right) \\ &\overset{(2)}{=} \phi\left(\lim_{n\to\infty}\left(\gamma_A\left(\frac{t}{n}\right)\gamma_B\left(\frac{t}{n}\right)\right)^n\right) \\ &= \phi\left(\gamma_{A+B}(t)\right) \\ &= \gamma_{\phi_*(A+B)}(t),\end{aligned}$$

where in (1) we use the fact that ϕ is a homomorphism, and in (2) we use the continuity of ϕ. □

Lemma 4.5.27. *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ and $H \leq \mathrm{GL}_n(\mathbb{F})$ be closed, and let $\phi : G \to H$ be a continuous homomorphism. Then $\phi_* : \mathfrak{g} \to \mathfrak{h}$ satisfies*

$$\phi_*([A, B]) = [\phi_*(A), \phi_*(B)]$$

for all $A, B \in \mathfrak{g}$. In other words, ϕ_* is a homomorphism of Lie algebras.

Proof. Let

$$\delta_s(t) = \gamma_A(s)\gamma_B(t)\gamma_A(-s)$$

so that $\dot{\delta}_s(0) = e^{sA}Be^{-sA}$. As s varies, $\dot{\delta}_s(0)$ traces a continuously differentiable path in $\mathfrak{g}$. More specifically,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\left(\dot{\delta}_s(0)\right)\Big|_{s=0} &= \frac{\mathrm{d}}{\mathrm{d}s}\left(e^{sA}\right)\Big|_{s=0} Be^{-sA} + e^{sA}B\frac{\mathrm{d}}{\mathrm{d}s}\left(e^{-sA}\right)\Big|_{s=0} \\ &= ABI - IBA \\ &= AB - BA \\ &= [A, B].\end{aligned}$$

Similarly, let

$$\beta_s(t) = \gamma_{\phi_*(A)}(s)\gamma_{\phi_*(B)}(t)\gamma_{\phi_*(A)}(-s)$$

so that as s varies $\dot{\beta}_s(0)$ traces a continuously differentiable path in $\mathfrak{h}$. As before, we note that

$$\frac{\mathrm{d}}{\mathrm{d}s}\left(\dot{\beta}_s(0)\right)\Big|_{s=0} = \left[\phi_*(A), \phi_*(B)\right].$$

Since, $\phi \circ \delta_s = \beta_s$ it follows that $\phi\left(\dot{\delta}_s(0)\right) = \dot{\beta}_s(0)$. Taking the derivative with respect to s and evaluating it at zero gives

$$\phi_*([A, B]) = [\phi_*(A), \phi_*(B)]$$

for all $A, B \in \mathfrak{g}$. In conjunction with Lemma 4.5.26 we conclude that ϕ_* is a homomorphism of Lie algebras. □

Lemma 4.5.28. *Let G, H and K be closed subgroups of $\mathrm{GL}_n(\mathbb{R})$. Let $\phi : G \to H$ and $\psi : H \to K$ be continuous homomorphism. Then $\psi_* \circ \phi_* = (\psi \circ \phi)_*$.*

Proof. Let $A \in \mathfrak{g}$. Recall that $\phi_*(A)$ is determined by the (4.5.2). Consequently, $\psi_*(\phi_*(A))$ is determined by the equation

$$\gamma_{(\psi_* \circ \phi_*)(A)} = \psi \circ \gamma_{\phi_*(A)} = \psi \circ \phi \circ \gamma_A.$$

Similarly, $(\psi \circ \phi)_*(A)$ is determined by the equation

$$\gamma_{(\psi \circ \phi)_*(A)} = \psi \circ \phi \circ \gamma_A.$$

Therefore, $(\psi \circ \phi)_*(A) = \psi_*(\phi_*(A))$ for all $A \in \mathfrak{g}$. □

Corollary 4.5.29. *If $G, H \leq \mathrm{GL}_n(\mathbb{F})$ are closed and isomorphic as topological groups, then their respective Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$ are isomorphic as Lie algebras.*

Proof. As G and H are isomorphic as topological groups, there exists continuous homomorphisms $f : G \to H$ and $g : H \to G$ that are inverses to each other. Using Lemma 4.5.28 it follows that

$$f_* \circ g_* = (f \circ g)_* = (\mathrm{Id}_H)_* = \mathrm{Id}_{\mathfrak{h}}.$$

Similarly, $g_* \circ f_* = \mathrm{Id}_{\mathfrak{g}}$. Hence, f_* and g_* are inverses to each other. Thus, as f_* is a Lie algebra homomorphism by Lemma 4.5.27, it follows that $\mathfrak{g}$ and $\mathfrak{h}$ are isomorphic as Lie algebras. □

4.6 Solution to Exercises

Exercise 4.1.2

Solution. Suppose M is invertible, with inverse M^{-1}. Then $MM^{-1} = I$ which implies that $1 = \det\left(MM^{-1}\right) = \det(M)\det\left(M^{-1}\right)$. Therefore, $\det(M) \in R^\times$ with inverse $\det\left(M^{-1}\right)$. □

Exercise 4.3.4

Solution.

1. Let $M_1, M_2 \in \mathrm{SL}_n(\mathbb{F})$, then
$$\det(M_1 M_2) = \det(M_1)\det(M_2) = 1,$$
which implies that $M_1 M_2 \in \mathrm{SL}_n(\mathbb{F})$. Moreover, for $M \in \mathrm{SL}_n(\mathbb{F})$ we have
$$1 = \det(I) = \det\left(MM^{-1}\right) = \det(M)\det\left(M^{-1}\right)$$
and so $1 = \det\left(M^{-1}\right)$ meaning $M^{-1} \in \mathrm{SL}_n(\mathbb{F})$. Therefore, as $I \in \mathrm{SL}_n(\mathbb{F})$ we conclude that $\mathrm{SL}_n(\mathbb{F}) \leq \mathrm{GL}_n(\mathbb{F})$. Moreover, as $\det(\cdot)$ is continuous and $\mathrm{SL}_n(\mathbb{F})$ is the pre-image of $\{1\}$ under this map we deduce that $\mathrm{SL}_n(\mathbb{F})$ is closed. Therefore, $\mathrm{SL}_n(\mathbb{F})$ is a matrix Lie group.

2. Let
$$\mathrm{D}_n = \{M \in \mathrm{GL}_n(\mathbb{F}) : M \text{ diagonal}\}.$$
Consider $M \in \mathrm{D}_n$ given by $M = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. As $M \in \mathrm{GL}_n(\mathbb{F})$, it follows that $\lambda_k \neq 0$ for $k = 1, \ldots, n$. In particular,
$$M\mathrm{diag}\left(\frac{1}{\lambda_1}, \ldots, \frac{1}{\lambda_n}\right) = I.$$
Therefore, $M^{-1} \in \mathrm{D}_n$. Similarly, for $M_1, M_2 \in \mathrm{D}_n$ with $M_1 = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and $M_2 = \mathrm{diag}(\mu_1, \ldots, \mu_n)$ we have
$$M_1 M_2 = \mathrm{diag}(\lambda_1\mu_1, \ldots, \lambda_n\mu_n).$$
Hence, $M_1 M_2 \in \mathrm{D}_n$. As $I \in \mathrm{D}_n$ we have that $\mathrm{D}_n \leq \mathrm{GL}_n(\mathbb{F})$. Note that matrices converge if and only if each of the entries converges to the entries of the corresponding limit matrix. Hence, a sequence in D_n converges to an element of $\mathrm{GL}_n(\mathbb{F})$ if and only if that element is in D_n. Therefore, $\mathrm{D}_n \leq \mathrm{GL}_n(\mathbb{F})$ is closed and thus a matrix Lie group.

3. Let $M_1, M_2 \in \mathrm{O}(n)$. Note that

$$(M_1M_2)^\top (M_1M_2) = M_2^\top M_1^\top M_1 M_2 = M_2^\top M_2 = I.$$

Moreover, if $M \in \mathrm{O}(n)$ then $M^{-1} = M^\top$ which implies that $MM^\top = I$. Hence, $\mathrm{O}(n)$ is a subgroup of $\mathrm{GL}_n(\mathbb{R})$. It is closed as $M \mapsto M^\top M$ is a continuous map and $\mathrm{O}(n)$ is the pre-image of the closed set $\{I\}$ under this map. Therefore, $\mathrm{O}(n)$ is a matrix Lie group.

4. $\mathrm{SO}(n)$ is a subgroup of $\mathrm{O}(n)$. It is closed as $\det(\cdot)$ is a continuous map. Therefore, $\mathrm{SO}(n)$ is a matrix Lie group by Corollary 4.3.14.

5. $\mathrm{U}(n)$ is a matrix Lie group by similar arguments as made for $\mathrm{O}(n)$.

6. $\mathrm{SU}(n)$ is a subgroup of $\mathrm{U}(n)$. It is closed as $\det(\cdot)$ is a continuous map. Therefore, $\mathrm{SU}(n)$ is a matrix Lie group by Corollary 4.3.14.

7. Let G be a finite group of cardinality $n \in \mathbb{N}$. Then by Cayley's theorem, G is isomorphic to a subgroup of S_n. Let $\phi : G \to S_n$ be an injective homomorphism. This is continuous on the discrete topology. Let $\psi : S_n \to \mathrm{GL}_n(\mathbb{R})$ be given by $\sigma \mapsto P_\sigma$ where P_σ is the corresponding unique permutation matrix. Note that $P_\sigma^{-1} = P_{\sigma^{-1}}$. Suppose $\psi(\sigma_1) = \psi(\sigma_2)$, then $P_{\sigma_1} = P_{\sigma_2}$ and so

$$I = P_{\sigma_2}^{-1} P_{\sigma_1} = P_{\sigma_2^{-1}} P_{\sigma_1}$$

which implies that $\sigma_2^{-1} \circ \sigma_1 = e$ meaning $\sigma_1 = \sigma_2$. Hence, ψ is injective, and it is also continuous on the discrete topology. Therefore, $\varphi = \psi \circ \phi : G \to \mathrm{GL}_n(\mathbb{R})$ is an injective and continuous homomorphism. Therefore, G is isomorphic to a subgroup of $\mathrm{GL}_n(\mathbb{R})$. Moreover, this subgroup is closed as all discrete groups in the discrete topology are closed.

8. Recall from Proposition 4.1.3 that

$$Z\left(\mathrm{GL}_n(\mathbb{F})\right) = \left\{\lambda I : \lambda \in \mathbb{F}^\times\right\}.$$

As the centre of a group is a normal subgroup, we have $\mathrm{PGL}_n(\mathbb{F}) = \mathrm{GL}_n(\mathbb{F})/Z\left(\mathrm{GL}_n(\mathbb{F})\right)$ is well-defined and in particular a group. With $Z\left(\mathrm{GL}_n(\mathbb{F})\right)$ being closed it follows that $\mathrm{PGL}_n(\mathbb{F})$ is closed in the quotient topology. Therefore, $\mathrm{PGL}_n(\mathbb{F})$ is a matrix Lie group.

9. The determinant of any matrix in $\mathrm{H}_n(\mathbb{F})$ is one, and so $\mathrm{H}_n(\mathbb{F}) \subseteq \mathrm{GL}_n(\mathbb{F})$. Clearly, $\mathrm{H}_n(\mathbb{F})$ is closed under matrix multiplication. Let $A \in \mathrm{H}_n(\mathbb{F})$, then we know that A^{-1} exists and is upper triangular as A is upper triangular. Moreover, as $AA^{-1} = I$ it is clear that the diagonal of A^{-1} must contain all ones and so $A^{-1} \in \mathrm{H}_n(\mathbb{F})$. Thus, $\mathrm{H}_n(\mathbb{F})$ is closed under inverses. As $I \in \mathrm{H}_n(\mathbb{F})$ we deduce that $\mathrm{H}_n(\mathbb{F}) \leq \mathrm{GL}_n(\mathbb{F})$ is a subgroup. Now let $(A_m)_{m\in\mathbb{N}} \subseteq \mathrm{H}_n(\mathbb{F})$ be a sequence converging to $A \in M_n(\mathbb{F})$. Then it must be the case that $(A_m)_{ij} \to A_{ij}$ for $1 \leq i, j \leq n$ and so $A \in \mathrm{H}_n(\mathbb{F})$ meaning $\mathrm{H}_n(\mathbb{F}) \leq \mathrm{GL}_n(\mathbb{F})$ is a closed subgroup. Therefore, $\mathrm{H}_n(\mathbb{F})$ is a matrix Lie group.

□

Exercise 4.3.11

Solution. Consider $M = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$. Then M is invertible as

$$\begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -i \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

However,

$$A = \mathrm{Re}(M) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

and

$$B = \mathrm{Im}(M) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

are not invertible. □

Exercise 4.4.5

Solution. For $A \in \mathrm{GL}_n(\mathbb{C})$ we can write it in Jordan normal form as

$$A = P\left(J_{n_1}(\lambda_1) \oplus \cdots \oplus J_{n_r}(\lambda_r)\right) P^{-1}$$

for some $\lambda_1, \ldots, \lambda_r \in \mathbb{C}$, $P \in \mathrm{GL}_n(\mathbb{C})$ and with $J_{n_i}(\lambda_i)$ denoting the $n_i \times n_i$ λ_i-Jordan block. As A is invertible, each $\lambda_i \neq 0$ and so we can write $\lambda_i = e^{z_i}$ for some $z_i \in \mathbb{C}$. By statement . of Example 4.4.4 we have

$$e^{J_{n_i}(z_i)} = \begin{pmatrix} \lambda_i & & & & \\ \lambda_i & \ddots & & & \\ \frac{\lambda_i}{2} & \ddots & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & \\ \frac{\lambda_i}{(n-1)!} & \cdots & \frac{\lambda_i}{2} & \lambda_i & \lambda_i \end{pmatrix}.$$

As the minimal polynomial of $e^{J_{n_i}(z_i)}$ is $(X-\lambda_i)^{n_i}$ we deduce that it has Jordan normal form $J_{n_i}(\lambda_i)$ and so we can write $J_{n_i}(\lambda_i) = P_i \exp\left(J_{n_i}(z_i)\right) P_i^{-1} = \exp\left(P_i J_{n_i}(z_i) P_i^{-1}\right)$ for some $P_i \in \mathrm{GL}_n(\mathbb{C})$. Therefore,

$$\begin{aligned} A &= P\left(\exp\left(P_1 J_{n_1}(z_1) P_1^{-1}\right) \oplus \cdots \oplus \exp\left(P_r J_{n_r}(z_r) P_r^{-1}\right)\right) P^{-1} \\ &= \exp\left(P\left(P_1 J_{n_1}(z_1) P_1^{-1} \oplus \cdots \oplus P_r J_{n_r}(z_r) P_r^{-1}\right) P^{-1}\right). \end{aligned}$$

Hence, we conclude that any $A \in \mathrm{GL}_n(\mathbb{C})$ is the exponential of some matrix. □

Exercise 4.4.12

Solution. Let $A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Then

$$e^A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

and

$$e^B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

so that

$$e^A e^B = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

On the other hand,

$$(A+B)^k = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^k = \begin{cases} I & k \text{ even} \\ A+B & k \text{ odd.} \end{cases}$$

Therefore,

$$e^{A+B} = \begin{pmatrix} \cosh(1) & \sinh(1) \\ \sinh(1) & \cosh(1) \end{pmatrix} \neq e^A e^B.$$

Indeed

$$AB = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \neq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = BA.$$

□

5 Appendix

5.1 Topology

Throughout let $\mathbb{F}$ be $\mathbb{R}$ or $\mathbb{C}$.

Definition 5.1.1. *A topology $\mathcal{T}$ on a set X is a collection of subsets of X, which are called the open subsets, such that the following statements hold.*

- $\emptyset, X \in \mathcal{T}$.
- *If $(U_i)_{i\in I} \subseteq \mathcal{T}$ is a collection of open subsets of X, where I is a possibly infinite set, then*
$$\bigcup_{i\in I} U_i \in \mathcal{T}.$$
- *If $\{U_1, \dots, U_n\} \subseteq \mathcal{T}$ are open, then*
$$\bigcap_{i=1}^{n} U_i \in \mathcal{T}.$$

A subset $C \subseteq X$ is said to be closed if $X \setminus C \in \mathcal{T}$.

Example 5.1.2. *The Euclidean topology on $\mathbb{F}^n$ says that $U \subseteq X$ is open if, for all $x \in X$, there is some $\varepsilon > 0$ such that for any $y \in X$ with $|y - x| < \varepsilon$ it follows that $y \in U$. Henceforth, any topological reference involving $\mathbb{F}^n$ will implicitly endow $\mathbb{F}^n$ with the Euclidean topology.*

Definition 5.1.3. *A function $f : X \to Y$ between topological spaces is continuous if for any open set $U \subseteq Y$ the set $f^{-1}(U)$ is open in X.*

Note that if $f : X \to Y$ is continuous and $C \subseteq Y$ is closed, then $f^{-1}(C) \subseteq X$ is closed.

Definition 5.1.4. *Let $(X, \mathcal{T})$ be a topology. Then for $V \subseteq X$ the collection*
$$\mathcal{T}_V := \{V \cap U : U \in \mathcal{T}\}$$
defines the subspace topology on V.

Lemma 5.1.5. *If $X \subseteq \mathbb{F}^n$ and $Y \subseteq \mathbb{F}^m$ have the subspace topology, then $f : X \to Y$ is continuous if and only if for all $x \in X$ and $\varepsilon > 0$, there is some $\delta > 0$ such that whenever $y \in X$ is such that $|x - y| < \delta$, we have $|f(x) - f(y)| < \varepsilon$.*

Remark 5.1.6. *The notion of continuity established in Lemma 5.1.5 is the ϵ-δ definition of continuity. Intuitively, it says that f is continuous at x if $f(y)$ is close to $f(x)$ for y sufficiently close to x.*

Lemma 5.1.7. *The composition of continuous functions is continuous.*

Lemma 5.1.8. *A function $f = (f_1, \dots, f_n) : \mathbb{F}^m \to \mathbb{F}^n$ is continuous if and only if each component $f_i : \mathbb{F}^m \to \mathbb{F}$ is continuous.*

Example 5.1.9. *Projection maps $\pi_j : \mathbb{F}^n \to \mathbb{F}$, given by $(x_1, \ldots, x_n) \mapsto x_j$ are continuous. Moreover, if $f, g : \mathbb{F}^n \to \mathbb{F}$ are continuous, then so are $f + g$ and $f \cdot g$. Combining these observations, we deduce that any polynomial, in multiple variables, is continuous. Hence, $\mu : M_n(\mathbb{F}) \times M_n(\mathbb{F}) \to M_n(\mathbb{F})$ given by $(A, B) \mapsto A \cdot B$ is continuous, since each entry of $A \cdot B$ is a polynomial in the entries of A and B.*

Definition 5.1.10. *Let $Y \subseteq \mathbb{F}^n$ have the Euclidean topology and consider $X \subseteq Y$. We say X is dense in Y if $\bar{X}$, the closure of X in Y, is equal to Y. Equivalently, we have that for all $y \in Y$ and for all $\varepsilon > 0$, there is some $x \in X$ with $|x - y| < \varepsilon$.*

Example 5.1.11.

1. *$\mathbb{Q} \subseteq \mathbb{R}$ is dense.*
2. *$\mathbb{Q}[i] \subseteq \mathbb{C}$ is dense.*

Definition 5.1.12. *A topological space X is path-connected if for all $x, y \in X$, there is a path from x to y. Where a path is a continuous function $f : [0, 1] \to X$ such that $f(0) = x$ and $f(1) = y$.*

Lemma 5.1.13. *Let $x, y, z \in X$. If there is a path γ from x to y and a path δ from y to z, then there is a path from x to z.*

Proof. Let $\xi : [0, 1] \to X$ be given by

$$\xi(t) = \begin{cases} \gamma(2t) & t \leq \frac{1}{2} \\ \delta(2t - 1) & t \geq \frac{1}{2}. \end{cases}$$

This is a path from x to z. □

Definition 5.1.14. *A set $X \subseteq \mathbb{F}^n$ is bounded if there is some $C > 0$ such that for all $x \in X$ we have $|x| \leq C$.*

Definition 5.1.15. *A set $X \subseteq \mathbb{F}^n$ is compact if it is a closed and bounded subset of $\mathbb{F}^n$.*

5.2 Differentiability

Definition 5.2.1.

- *A function $\gamma : \mathbb{R} \to \mathbb{F}^m$ is continuously differentiable, written $\gamma \in \mathcal{C}^1$, if its derivative $\dot{\gamma}(t) : \mathbb{R} \to \mathbb{F}^m$ exists and is continuous for all $t \in \mathbb{R}$.*
- *A function $f : \mathbb{R}^m \to \mathbb{R}^n$ is continuously differentiable, written $f \in \mathcal{C}^1$, if all partial derivatives $\frac{\partial f_i}{\partial x_j}$ exist and are continuous. The derivative at p is the matrix*

$$Df(p) = \left(\frac{\partial f_i}{\partial x_j}(p) \right)_{ij}.$$

Theorem 5.2.2 (Inverse Function Theorem)**.** *Let $f : \mathbb{R}^m \to \mathbb{R}^m$ be continuously differentiable. Suppose $x_0 \in \mathbb{R}^m$ is such that $Df(x_0)$ is an invertible matrix. Then there exists an open neighbourhood $U \subseteq \mathbb{R}^m$ of x_0 such that $f|_U : U \to f(U)$ is a homeomorphism, and is continuously differentiable with a continuously differentiable inverse.*

Remark 5.2.3. *Theorem 5.2.2 says that if the derivative of f is invertible at x_0 then f is also invertible in a neighbourhood of x_0.*

Example 5.2.4.

1. *Consider $\exp : \mathbb{R} \to \mathbb{R}$ given by $x \mapsto e^x$. This is not a homeomorphism as it is not surjective. However, the derivative at zero is $1 \in M_1(\mathbb{R})$, which is invertible. Hence, using Theorem 5.2.2 we can find an open neighbourhood, say $(-1, 1)$ such that*

$$\exp|_{(-1,1)} : (-1, 1) \to \left(e^{-1}, e^1\right)$$

   is a homeomorphism. Namely, it is continuously differentiable and its inverse, $\log$, is continuously differentiable.

2. *Consider $\exp : \mathbb{C} \to \mathbb{C}$, sending $z \mapsto e^z$. The derivative at zero is the identity matrix, $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, which is invertible. It turns out, in this case, we can take $U = \left\{z \in \mathbb{C} : |z|^2 = 1\right\}$.*

5.3 Abelian Matrix Lie Groups

Just as we can classify finitely generated abelian groups, we can also classify connected abelian matrix lie groups.

Lemma 5.3.1. *Let $\phi : G \to H$ be a homomorphism between matrix Lie groups. Then $\ker(\phi)$ is a closed topological subgroup of G, meaning $\ker(\phi)$ is itself a matrix lie group. Moreover,*

$$T_I \ker(\phi) = \ker\left(\phi_* : T_I G \to T_I H\right).$$

Proof. Let $(x_n)_{n\in\mathbb{N}} \subset \ker(\phi)$ be a sequence converging to x in G. As ϕ is continuous it follows that $\phi(x) = I$ which implies that $x \in \ker(\phi)$, meaning $\ker(\phi)$ is a closed subgroup of G. Recall, that we have a bijection between $T_I G$ and continuous homomorphism $\gamma : \mathbb{R} \to G$. Note that $\ker(\phi_*)$ consists of the γ for which $\phi \circ \gamma$ is constant, which holds if and only if $\mathrm{im}(\gamma) \subseteq \ker(G)$, and so $\gamma : \mathbb{R} \to \ker(\phi)$ is a continuous homomorphism. Therefore, using the associated bijection, this happens if and only if $\gamma \in T_I \ker(\phi)$. □

Lemma 5.3.2. *Let G be a matrix Lie group with $K \triangleleft G$ a discrete normal subgroup. Then G/K is a topological group. In particular, the quotient map $\pi : G \to G/K$ is a continuous homomorphism, for which there exists an open neighbourhood U of $I \in G$ such that $\pi|_U$ is a homeomorphism into its image.*

Proof. Note that multiplication, $G \times G \to G$, when restricted to $K \times K$ maps $K \times K \to K$ as K is a subgroup. Therefore, $(G \times G)/(K \times K) \to G/K$ defines a continuous map. As $(G/K) \times (G/K) = (G \times G)/(K \times K)$ it follows that the product on G/K is continuous. Similarly, one argues that the inverse is continuous meaning that G/K is a topological group. Now as K is discrete, we can choose a compact neighbourhood C of $I \in G$ such that $\pi|_C$ is an injective. Then $\pi|_C$ is a continuous bijection onto its image. As C and $\pi(C)$ are both Hausdorff we have that $\pi|_C$ is homeomorphic onto its image. Taking U to be the interior of C we find an open neighbourhood U of $I \in G$ such that $\pi|_U$ is homeomorphic onto its image. □

For each of the abstract objects we have considered, there have been corresponding isomorphism theorem. Similar isomorphism results hold for matrix Lie groups.

Proposition 5.3.3. *Let $\phi : G \to H$ be a surjective homomorphism of matrix Lie groups, with a discrete kernel. Then $G/\ker\phi$ and H are isomorphic topological groups. In particular, $G/\ker\phi$ is a matrix Lie group.*

Proof. Note that ϕ respects the equivalence classes of $G/\ker(\phi)$. Hence, ϕ can be restricted to a continuous homomorphism $\tilde{\phi} : G/\ker(\phi) \to H$. In particular, $\tilde{\phi}$ is injective and surjective. Since $\ker(\phi)$ is discrete it follows that ϕ_* is injective. Moreover, since ϕ is surjective it follows that ϕ_* is surjective and thus defines an isomorphism.

Therefore, there exists a neighbourhood U of $I \in G$ such that $\phi|_U$ is homeomorphic onto its image. Using Lemma 5.3.2 we have a neighbourhood V of $I \in G/\ker(\phi)$ such that $\tilde{\phi}|_V$ is homeomorphic onto its image. Implying the $\tilde{\phi}^{-1}$ is continuous in an open neighbourhood of I. Let $A \in H$. As multiplication by A is homeomorphic, it follows that $\tilde{\phi}^{-1}$ is continuous in a neighbourhood of A. Thus $\tilde{\phi}$ is a homeomorphism meaning $G/\ker(\phi)$ and H are isomorphic topological groups. □

Example 5.3.4. *Let $G = \mathbb{R}^2/\mathbb{Z}^2$ and $\phi : \mathbb{R} \to G$ be given by $t \mapsto (t, \pi t)$. One can associate G with $[0,1)^2$, where $(x, y) \in \mathbb{R}$ is represented by $(x \mod 1, y \mod 1$. It is clear then that ϕ is injective due to the irrationality of π. Specifically if $(t_1, \pi t_1) \sim (t_2, \pi t_2)$ it follows that $t_1 - t_2 \in \mathbb{Z}$ and $\pi(t_1 - t_2) \in \mathbb{Z}$, which is clearly a contradiction. Moreover, ϕ is a continuous homomorphism, however, ϕ is not a homeomorphism. To see why it is not a homeomorphism, take U as an open neighbourhood of $0 \in \mathrm{im}(\phi)$. Then U contains no path-connected neighbourhood of 0 due to the lattice lines. However, every neighbourhood of $0 \in \mathbb{R}$ contains a path-connected neighbourhood of 0. Therefore, these spaces cannot even be homeomorphic. Therefore, it is not always true that $G/\ker(\phi)$ is isomorphic to $\mathrm{im}(\phi)$ when $\ker(\phi)$ is discrete.*

Lemma 5.3.5. *Let G be an abelian matrix Lie group. Then the commutator vanishes on $\mathfrak{g}$. Furthermore, $\exp : \mathfrak{g} \to G$ is a homomorphism of topological groups.*

Proof. Let $A, B \in \mathfrak{g}$ and let $\delta_s(t) = e^{sA}e^{tB}e^{-sA}$ be a path in G for each s. Recall that $\dot{\delta}_s(0) \in \mathfrak{g}$ for all s, and is differentiable in s with derivative

$$\frac{\mathrm{d}}{\mathrm{d}s}\dot{\delta}_s(0)\Big|_{s=0} = [A, B].$$

As G is abelian it is clear that $\delta_s(t) = e^{tB}$ and so we also have that $\dot{\delta}_s(0) = B$ for all s which implies that

$$[A, B] = \frac{\mathrm{d}}{\mathrm{d}s}\dot{\delta}_s(0)\Big|_{s=0} = 0.$$

In particular, this means that A and B commute and so we can write $e^{A+B} = e^A e^B$. Meaning $\exp$ is a homomorphism as we already know it is continuous. □

Lemma 5.3.6. *Let G be a connected abelian matrix Lie group. Then $\exp : \mathfrak{g} \to G$ has a discrete kernel.*

Proof. Suppose that the kernel is not discrete. Then there exists a sequence $(x_i)_{i\in\mathbb{N}} \subseteq \ker(\exp)$ such that $x_i \to x \in \ker(\exp)$ with $x_i \neq x$ for any $i \in \mathbb{N}$. By replacing x_i with $x_i - x$ we can suppose without loss of generality that $x = 0$. As $x_i \in \ker(\exp)$ we have $\exp(x_i) = I$ for all $i \in \mathbb{N}$, with the added property that $x_i \to 0$. However, as $\exp$ is injective on some open neighbourhood of 0. Therefore, for some $i, j \in \mathbb{N}$ large enough the points x_i and x_j lie in this neighbourhood and it follows that $x_i = x_j$ by the injectivity of $\exp$, which is a contradiction. □

Lemma 5.3.7. *Let G be a connected abelian matrix Lie group. Then $\exp : \mathfrak{g} \to G$ is surjective.*

Proof. Since G is connected, for any $g \in G$ we can write $g = e^{A_1} \dots e^{A_k}$ for some $A_1, \dots, A_k \in \mathfrak{g}$. As $\exp$ is a homomorphism it follows that $g = e^{A+1+\dots+A_k} \in \mathrm{im}(\exp)$. Hence, $\exp : \mathfrak{g} \to G$ is surjective. □

Definition 5.3.8. *For a finite-dimensional real vector space V, a lattice is a discrete subgroup Λ.*

Lemma 5.3.9. *Let $\Lambda \leq V$ be a lattice. Then there is a basis $\{e_i\}_{i=1}^d$ of V such that*

$$\Lambda = \mathbb{Z}e_1 + \dots + \mathbb{Z}e_k$$

for some $k \leq d$. In particular, Λ is finitely generated.

Proof. We proceed by induction on the dimension of V. The case for $\dim(V) = 0$. Moreover, the case for $\Lambda = 0$ is clear, so we assume $\Lambda \neq 0$. Define some inner product on V and let $e \in \Lambda \setminus \{0\}$ be an element which minimises $|e|$. We can do this as Λ is discrete. Now consider $V' := V/\langle e \rangle$, the quotient map $\pi : V \to V'$ and $\Lambda' = \pi(\Lambda)$. It is clear that Λ' is a subgroup of V'. Suppose that Λ' is not discrete. Then there is some sequence $(\lambda'_i)_{i\in\mathbb{N}} \subseteq V'$ such that $\lambda'_i \to 0$ and $\lambda'_i \neq 0$ for any $i \in \mathbb{N}$. For $i \in \mathbb{N}$, let $\lambda_i \in \Lambda$ be such that $\pi(\lambda_i) = \lambda'_i$. By adding multiples of e where necessary we can assume that $\lambda_i \to 0$. However, this contradicts the existence of e. Therefore, Λ' is discrete. So by the inductive hypothesis, there is a basis $\{e'_i\}_{i=1}^d$ of V' such that

$$\Lambda' = \mathbb{Z}e'_1 + \cdots + \mathbb{Z}e'_k.$$

For each i let e_i be such that $\pi(e_i) = e'_i$. Then $\{e\} \cup \{e_i\}_{i=1}^d$ is a basis for V such that

$$\Lambda = \mathbb{Z}e + \mathbb{Z}e_1 + \cdots + \mathbb{Z}e_k.$$

□

Theorem 5.3.10. *Let G be a connected abelian matrix Lie group. Then G is isomorphic to $(\mathbb{R}/\mathbb{Z})^i \times \mathbb{R}^j$ for some $i, j \geq 0$. Moreover, G is compact if and only if $j = 0$.*

Proof. Consider the homomorphism $\exp : \mathfrak{g} \to G$. This is a surjective homomorphism with a discrete kernel, and so it follows that G is isomorphic to $\mathfrak{g}/\ker(\exp)$. As $\ker(\exp)$ is a lattice in $\mathfrak{g}$ we can choose a basis $\{e_i\}_{i=1}^d \subseteq \mathfrak{g}$ such that

$$\ker(\exp) = \mathbb{Z}e_1 + \cdots + \mathbb{Z}e_k$$

for some $k \leq d$. Therefore, G is isomorphic to $\mathbb{R}^d/\mathbb{Z}^k$ which is isomorphic to $(\mathbb{R}/\mathbb{Z})^k \times \mathbb{R}^{d-k}$. Moreover, we note that $(\mathbb{R}/\mathbb{Z})^i \times \mathbb{R}^j$ is compact if and only if $j = 0$ to complete the proof. □