# Algebra III

Thomas Walker

Autumn 2023

## Contents

# 1 Rings

## 1.1 Definitions and examples

There are a few definitions of ring in the literature.

**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 $x, y \in M$.

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

**Example 1.1.3.**

- *The natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$ 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'll always write $x + y$ for $+(x, y)$. Similarly, we will write $x \cdot y$ for $\cdot(x, y)$.*
- *We'll 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 $r, s \in S$ it follows that $-r, r + s, rs \in S$.*

*That is, $S$ is itself 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, note that $\mathbb{N} \not\leq \mathbb{Z}$ as $-1 \notin \mathbb{N}$.*

- *The set $\mathbb{Z}/2\mathbb{Z}$ with standard addition and multiplication is a ring.*

- *The Gaussian integers $\mathbb{Z}[i] := \{a + ib \in \mathbb{C} : a, b \in \mathbb{Z}\}$ 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 $r \cdot 0_R = 0_R \cdot r = 0$.*

*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 $r \cdot 0_R = 0_R$, as desired. □

**Example 1.1.9.** *The trivial ring is a 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.* If $R = \{0\}$, since there's only one element we must have $1_R = 0_R$. Now suppose $1_R = 0_R$, and let $r \in R$. Then

$$\begin{aligned} r &= r \cdot 1_R \\ &= r \cdot 0_R \\ &= 0. \end{aligned}$$

□

Throughout the rest of the course, we will assume any ring we encounter is 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 notation as in Definition 1.1.11 the element $v$ also lies in $R^\times$.*

**Definition 1.1.13.** *A division ring is a non-trivial ring where 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.**

- $\mathbb{Z}$ *isn't a field, as* 2 *isn't invertible. In fact* $\mathbb{Z}^\times = \{\pm 1\}$.
- $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ *are all fields.*
- $\mathbb{Z}[i]$ *isn't a field.*
- $\mathbb{Q}\left[\sqrt{2}\right]$ *is a field.*
- *The quaternions* $\mathbb{H}$ *is the abelian group* $\mathbb{R}^4$ *with standard basis vectors labelled* $\{1, i, j, k\}$. *Moreover,* $\mathbb{H}$ *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* $s \in \mathbb{H}$. *This determines all other products, for instance*

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

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

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

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

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

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

*Proof.* We have that $1 = ru = rv$ which implies that $r(u - v) = 0$. Since $ur = 1$ we note 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 $a, b \in R^\times$, there are some $c, d \in R^\times$ such that $ac = ca = bd = db = 1$. Therefore $(ab)(dc) = a(bd)c = ac = 1$, and similarly $(dc)(ab) = 1$. This shows that $ab \in R^\times$, meaning $R^\times$ is closed under multiplication. Moreover, the element $1 \in R^\times$ as it is a unit. Finally, inverses exist as the inverse of $a$, as above, is $c$ which also lies in $R^\times$. □

## 1.2 Constructions of rings

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

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

and

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

The zero is $(0, 0)$ and the multiplicative unit is $(1, 1)$. This is a ring and 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 $f = (a_0, a_1, a_2, \ldots)$ in $R$ which is eventually $0$. That is, $a_i = 0$ for $i$ sufficiently large.*

**Remark 1.2.2.** *If $a_i = 0$ for $i > N$, we will write this polynomial as*

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

*Note that in this notation, $f$ is also equal to $a_0 + a_1 X + a_2 X^2 + \ldots + a_N X^N + 0X^{N+1}$.*

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

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

*Note that this exists by our definition of a polynomial. If $a_i = 0$ for all $i$, such that $f = 0$, we define $\deg(f)$ to be $-\infty$.*

**Definition 1.2.4.** *Let $R$ be a ring. Then its polynomial ring, $R[X]$, is defined to be 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 $n = m$ 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*

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

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

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

**Remark 1.2.6.** *We can take the polynomial ring of a polynomial ring, to get $(R[X])[Y]$. We will write this as $R[X, Y]$. We can iterate this any number of times, 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.** *$f = 1 + X^2$ is monic whereas $f = 2$ isn't.*

A polynomial $f = a_0 + a_1 X + \ldots \in R[X]$ determines a function $R \to R$, sending $r$ to $f(r) := a_0 + a_1 r + \ldots$. However the polynomial $f$ itself is just a sequence of elements of $R$, and $X$ is some formal symbol. We don't identify the polynomial with the function, in fact, different polynomials can have the same function. For example, let $R = \mathbb{Z}/2\mathbb{Z}$, and let $f = X \in R[X]$ and $g = X^2 \in R[X]$. 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 $R$ consists of a sequence $f = (\ldots, a_{-1}, a_0, a_1, \ldots)$, with only finitely many $a_i$ non-zero. We write this 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 $R$. This has a ring structure defined the same as in $R[X]$.

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

We define the ring structure of $R[\![X]\!]$ in the same way as we did for $R[X]$. Hence, we can observe that $R[\![X]\!]$ contains $R[X]$ as a subring.

**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*

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

*We define addition and multiplication similarly to before. Let $f = \sum_{m \in M} a_m m$ and $g = \sum_{m \in M} b_m m$ be in $R[M]$, then*

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

*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, that is it contains inverses, we call this 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.*

**Example 1.2.14.** *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 $n = 2$, and $R$ is non-trivial, $M_n(R)$ is never 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*

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

*A similar example shows $M_n(R)$ is never 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.*

## 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. *$\phi(0) = 0$, and*
4. *$\phi(1) = 1$.*

**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 $\phi$ is*
$$\ker(\phi) := \{r \in R : \phi(r) = 0\} \subseteq R.$$
- *The image of $\phi$ 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.* Since $0 = \phi(0)$ and $1 = \phi(1)$, we have $0, 1 \in \mathrm{im}(\phi)$. Suppose $s, s' \in \mathrm{im}(\phi)$. Then $s = \phi(r)$ and $s' = \phi(r')$ for some $r, r' \in R$. Then $s + s' = \phi(r + r') \in \mathrm{im}(\phi)$. Similarly $0 = s - s$, but also $0 = \phi(0) = \phi(r - r) = s + \phi(-r)$. So $-s = \phi(-r) \in \mathrm{im}(\phi)$. □

**Remark 1.3.5.** *If $R$ is non-trivial, since $\phi(1) = 1, 1 \notin \ker(\phi)$, so $\ker(\phi)$ isn't a subring of $R$.*

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

*Proof.* If $\phi$ is injective, since $\phi(0) = 0$, there is no other $r \in R$ with $\phi(r) = 0$ so $\ker(\phi) = \{0\}$. If $\ker(\phi) = \{0\}$, then suppose $r, r' \in R$ are such that $\phi(r) = \phi(r')$. Then $\phi(r - r') = 0$ so $r - r' \in \ker(\phi) = \{0\}$, so $r = r'$. □

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

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

We will normally just say an ideal to mean a left ideal. Note that if $R$ is commutative, all three definitions coincide.

**Example 1.3.8.** *For any ring $R, \{0\}, R \subseteq R$ are both ideals.*

Suppose $I \subseteq R$ is an ideal and $1 \in R$. Then for any $r \in R$ we have that $r \cdot 1 = r$ also lies in $I$. So $I = R$. In particular, the only subring which is also an ideal is the whole of $R$. We say an ideal $I \subseteq R$ such that $I \neq R$ is a proper ideal. Suppose $I \subseteq R$ is an ideal containing some unit $u$ with inverse $v$, then $I$ contains $vu = 1$ and hence $I = R$.

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

*Proof.* Since $\phi$ 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 $ir \in \ker(\phi)$. □

**Example 1.3.10.** *Let $R = \mathbb{Z}$, and $n \in \mathbb{Z}$. Then $n\mathbb{Z} \subseteq \mathbb{Z}$ is an ideal. We claim all ideals are of this form. Let $I \subseteq \mathbb{Z}$ be an ideal. If $I = \{0\}$, then $I = 0\mathbb{Z}$, so suppose it's not zero. Let $n \in I$ be its smallest positive element. Then $rn \in I$ for all $r \in \mathbb{Z}$ so $n\mathbb{Z} \subseteq I$. Now let $i \in I$. By the Euclidean algorithm, we can write*

$$i = an + b$$

*with $0 \leq b < n$. Note that $n \in I$ implies that $an \in I$, and we know $i \in I$ and so $b = i - an \in I$. Since $n$ was the smallest positive element of $I$, we must have $b = 0$, and so $i \in n\mathbb{Z}$. We conclude that $I = n\mathbb{Z}$.*

**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.*

**Definition 1.3.12.** *For an element $a \in R$, the ideal generated by $a$ is*

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

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

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

In $\mathbb{Z}$ we have shown that any ideal is of the form $n\mathbb{Z} = (n)$. Therefore, we have shown that all ideals in $\mathbb{Z}$ are principal.

**Exercise 1.3.13.** *Show that the subset*

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

*is an ideal. In particular, show that it is a principal ideal generated by the polynomial $X$.*

**Definition 1.3.14.** *Let $I \subseteq R$ be a two-sided ideal. The quotient ring $R/I$ consists of additive cosets of the form $r + I$, with zero given by $0 + I$, one given by $1 + I$, sum given by*

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

*and product given by*

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

**Proposition 1.3.15.** *The quotient ring $R/I$ is a ring, and the function $R \to R/I$ sending $r$ to $r + I$ is a surjective ring homomorphism.*

The motivation for the definition of a two-sided ideal is that it's exactly the definition that makes this proposition true, much like the definition of a normal subgroup.

*Proof.* We know addition is well-defined, as $(R/I,+)$ is the quotient of $(R,+)$ by a normal subgroup, and so it remains to check that multiplication is well-defined. Let $r,s\in R$, and let $i\in I$. So $r+I=r+i+I$. Then

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

since $is\in I$. Similarly $(r+I)(s+i+I)=(rs+I)$. So multiplication is well-defined. Associativity and distributivity follow from $R$, and one can check that $0+I$ and $1+I$ are zero and one respectively. □

**Example 1.3.16.** *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 $r=0,\ldots,(n-1)$, and addition and multiplication are the usual addition and multiplication and addition modular $n$.*

**Exercise 1.3.17.** *Let $R$ be a ring, and $\phi:G\to H$ a group homomorphism. Then $\phi$ induces a ring homomorphism $\phi_*:R[G]\to R[H]$, defined by*

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

*Note that $N=\ker(\phi)\trianglelefteq G$, 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 $g\in N$, so $(N-1)\subseteq\ker(\phi_*)$. Show that $(N-1)$ is in fact a two-sided ideal, and that $\ker(\phi_*)=(N-1)$.*

**Example 1.3.18.** *Let $R$ be a commutative ring. Consider the ideal $(X)\subseteq R[X]$. Let's compute $R[X]/(X)$. Elements of this quotient ring are of the form*

$$a_0+a_1X+\ldots+a_nX^n+(X).$$

*Everything but the first term is in $(X)$. So any such element is uniquely represented as $a_0+(X)$. So there is an isomorphism $R\to R[X]/(X)$ sending a to $a+(X)$.*

Trying to understand $\mathbb{R}[X]/\left(X^2+1\right)$ is harder. However, it can be made easier with the following result.

> **Proposition 1.3.19.** *Let $F$ be a field and $f,g\in F[X]$, with $g\neq 0$. Then there are some $r,q\in F[X]$ such that*
>
> $$f=gq+r$$
>
> *with $\deg(r)<\deg(g)$.*

*Proof.* □

This says that we can use the Euclidean algorithm on polynomials over a field. We'll revisit this property later.

**Example 1.3.20.** *Let $R=\mathbb{R}[X]/\left(X^2+1\right)$. Elements of $R$ are of the form*

$$a_0+a_1X+\ldots+a_nX^n+\left(X^2+1\right).$$

*Letting $f=a_0+\ldots+a_nX^n$, we can apply the Euclidean algorithm for polynomials, and find $q,r\in R$ such that*

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

*with $\deg(r)<2$. Which implies $r=b_0+b_1X$ 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 $a,b\in\mathbb{R}$. 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 $g=0$ and so $a=a'$ and $b=b'$. So every element of $R$ is of the form $a+bX+\left(X^2+1\right)$, and the element $X+\left(X^2+1\right)$ squares to $-1$ . This looks a lot like the complex numbers. Define a function $\phi:R\to\mathbb{C}$ by*

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

*We can check manually that this is well-defined, bijective and a ring homomorphism. This was a lot of work. The following theorem will make results like this much easier.*

**Theorem 1.3.21** (First isomorphism theorem for rings)**.** *Let $\phi:R\to S$ be a ring homomorphism. Recall that $\ker(\phi)\subseteq R$ is a two-sided ideal. Then there is an isomorphism of rings*

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

*Proof.* Define $\psi:R/\ker(\phi)\to\operatorname{im}(\phi)$ by

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

This is well-defined, since if $r+\ker(\phi)=r'+\ker(\phi)$ then $r-r'\in\ker(\phi)$ so $\phi(r)=\phi(r')$, it is also a ring homomorphisms. This is an isomorphism of abelian groups and so we conclude by the first isomorphism theorem for groups. □

**Example 1.3.22.** *We can define a homomorphism $\phi:\mathbb{R}[X]\to\mathbb{C}$ by setting $\phi(f):=f(i)$. Then we can check that $\phi$ is surjective and that $\ker(\phi)=\left(X^2+1\right)$. Therefore $\mathbb{C}\cong\mathbb{R}(X)/\left(X^2+1\right)$.*

**Theorem 1.3.23** (Second isomorphism theorem)**.** . *Let $R$ and $S$ be rings and suppose $R\leq S$ a subring. Let $I\subseteq S$ be a two-sided ideal. Then the following hold.*

1. *$R+I:=\{r+i\mid r\in R, i\in I\}\leq S$ is a subring.*

2. *$I\subseteq R+I$ and $R\cap I\subseteq R$ are both two-sided ideals.*

3. *$(R+I)/I=\{r+I\mid r\in R\}\leq S/I$ is a subring, and*

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

*Proof.* We first show $R+I\leq S$ is indeed a subring. Since $R\subseteq R+I$, we note that $R+I$ contains 0 and 1. 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 $R+I$, where the second term is in $I$ since $I$ is a two-sided ideal. Note that, $I\subseteq R+I$ is a two-sided ideal since $I\subseteq S$ is a two-sided ideal. We define a homomorphism $\phi:R\to S/I$ 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 $\ker(\phi)=R\cap J$ which is, therefore, a two-sided ideal in $R$. We have that $\operatorname{im}(\phi)=\{r+I:r\in R\}=(R+I)/I$ is a subring of $S/I$. But by the first isomorphism theorem, $\operatorname{im}(\phi)\cong R/(R\cap I)$. So

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

□

**Theorem 1.3.24** (Third isomorphism theorem)**.** *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.* Define a homomorphism $\phi : R/I \to R/J$ 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, $r \in J$. So $\ker(\phi) = J/I$. The result then follows from the first isomorphism theorem. □

**Proposition 1.3.25.** *Let $R$ be a ring, and $I \subseteq R$ a two-sided ideal. Then we have the following natural bijection.*

$$\begin{aligned} \{ \text{ two-sided ideals of } R/I\} &\longleftrightarrow \{ \text{ two-sided ideals of } R \text{ containing } I\} \\ \alpha : J \subseteq R/I &\longleftrightarrow \{r \in R \mid r+I \in J\} \\ K/I \subseteq R/I &\longleftrightarrow I \subseteq K \subseteq R : \beta \end{aligned}$$

*Proof.* Let $\alpha$ be the map going right and $\beta$ the map going left. One can check that $\alpha(J) \subseteq R$ and $\beta(K) \subseteq R/I$ are two-sided ideals, so $\alpha$ and $\beta$ are well-defined functions. We need to check these are inverse to each other. First, let $K \subseteq R$ be a two-sided ideal containing $I$. Then

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

where the last equality is because $I \subseteq K$. Let $J \subseteq R/I$ be a two-sided ideal. Then

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

So $\alpha$ and $\beta$ must be bijections. □

Let $R$ be any ring. Then there is a unqiue homomorphism $\iota : \mathbb{Z} \to R$, where $\iota(n) = \underbrace{1_R + \ldots + 1_R}_{n}$ if $n \geq 0$, and $\iota(n) = -\underbrace{(1_R + \ldots + 1_R)}_{|n|}$ if $n \leq 0$. This is a homomorphism by distributivity. Any homomohprism $\mathbb{Z} \to R$ must be equal to $\iota$, since it must send 1 to $1_R$. As $\ker(\iota) \subseteq \mathbb{Z}$ is an ideal so we must have $\ker(\iota) = n\mathbb{Z}$ for some $n \in \mathbb{Z}$.

**Definition 1.3.26.** *The characteristic of $R$ is the unique $n \geq 0$ such that $\ker(\iota) = n\mathbb{Z}$, where $\iota$ is the unique ring homomorphism $\mathbb{Z} \to R$.*

**Example 1.3.27.**

- *$\mathbb{Q}$ has characteristic $0$ since $\iota : \mathbb{Z} \to \mathbb{Q}$ is injective so its kernel is $0$ . Similar $\mathbb{R}, \mathbb{C}$ and $\mathbb{Z}[i]$ all have characteristic $0$.*
- *The homomorphism $\iota : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ sends $r$ to $r+n\mathbb{Z}$ and has kernel $n\mathbb{Z}$, so $\mathbb{Z}/n\mathbb{Z}$ has characteristic $n$, for any $n \geq 1$. In particular, any $n \geq 0$ is the characteristic of some commutative ring.*

# 2 Integral Domains

An archetypal ring is the integer, $\mathbb{Z}$, which has lots of special properties, such as unique factorisation, that we'd like to study in a more general context. We'll introduce several types of rings and study their properties. 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, ID, 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, since if $rs = 0$ with $r \neq 0$. Then letting $r^{-1}$ be the inverse of $r$, we see that*
$$0 = r^{-1}rs = s.$$
- *$\mathbb{Z}/6\mathbb{Z}$ isn't an integral domain, since $2 + 6\mathbb{Z}$ and $3 + 6\mathbb{Z}$ are non-zero, but their product is $0$.*

**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.* Suppose $\mathbb{Z}/n\mathbb{Z}$ is an integral domain, and suppose $n \mid rs$, for some $r, s \in \mathbb{Z}$. This means $(r+n\mathbb{Z})(s+n\mathbb{Z}) = 0$, so since $\mathbb{Z}/n\mathbb{Z}$ is an integral domain we have that $r + n\mathbb{Z} = 0$ or $s + n\mathbb{Z} = 0$. This implies that $n \mid r$ or $n \mid s$, so $n$ is prime. Suppose $n$ is prime, and suppose $(r + n\mathbb{Z})(s + n\mathbb{Z}) = 0$. Then $rs + n\mathbb{Z} = 0$ so $n \mid rs$. Which implies that $n \mid r$ or $n \mid s$ so $r + n\mathbb{Z} = 0$ or $s + n\mathbb{Z} = 0$. □

**Example 2.1.5.** *If $R$ is an integral domain and $S \leq R$ is a subring, then $S$ is also an integral domain since a zero divisor in $S$ would also be one in $R$.Therefore, as $\mathbb{C}$ is an integral domain since it's a field, 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
$$g = b_0 + \ldots + b_m X^m$$
where $a_n$ and $b_m$ are non-zero so that $n = \deg(f)$ and $m = deg(g)$. Then the coefficient of $X^{n+m}$ in $fg$ is $a_n b_m$, which is non-zero since $R$ is an integral domain. Therefore, $fg \neq 0$. □

**Remark 2.1.7.** *Iterating this, we see that if $R$ is an integral domain, then $R\left[X_1, \ldots, X_n\right]$ is an integral domain for all $n$.*

In particular, $\mathbb{Z}[X]$ is an integral domain. It is sometimes useful to have alternative descriptions of these notions in terms of ideals.

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

**Remark 2.1.9.** *Since $\{0\}$ and $R$ are always ideals in $R$, this is equivalent to $R$ only having two ideals.*

*Proof.* $(\Rightarrow)$. Suppose $R$ is a field and $I \subseteq R$ is a non-zero ideal, so 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. □

**Definition 2.1.10.** *An ideal $I \subseteq R$ is maximal if $I \neq R$ and any proper ideal $J \subseteq R$ containing $I$ is equal to $I$.*

**Lemma 2.1.11.** *An 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 $R/I$, so $R/I$ is a field.
$(\Leftarrow)$. 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 must be $\{0\}$. So $J = I$. □

We can give a similar description for integral domains.

**Definition 2.1.12.** *An ideal $I \subseteq R$ is prime if $I \neq R$ and whenever there are $r, s \in R$ such that $rs \in I$, we either have $r \in I$ or $s \in I$.*

**Example 2.1.13.** *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. To justify this 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 $n = uv$ where $u, v \notin \{0, \pm 1\}$. Then $n = uv \in n\mathbb{Z}$ but $u, v \notin n\mathbb{Z}$, since $0 < |u|, |v| < |n|$.*

**Lemma 2.1.14.** *An ideal $I \subseteq R$ is prime if and only if $R/I$ is an integral domain.*

*Proof.* $(\Rightarrow)$. First, 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 $I$, so one of $r + I$ and $s + I$ is $0$.
$(\Leftarrow)$. Suppose $R/I$ is an integral domain, and suppose we have $r, s \in R$ such that $rs \in I$. Then $(r+I)(s+I) = 0 + I$, so one of $r + I, s + I$ is $0$ which implies that one of $r, s$ is in $I$. □

**Example 2.1.15.** *Let $R$ be a ring. Then $R[X]/(X) \cong R$, so the ideal $(X) \subseteq R[X]$ is prime if and only if $R$ is an integral domain, and it's maximal if and only if $R$ is a field.*

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

*Proof.* Since $I$ is maximal, $R/I$ is a field, so it is an integral domain, so $I$ is a prime ideal. □

**Example 2.1.17.** *The converse isn't true. $(0) \subseteq \mathbb{Z}$ is a prime ideal but isn't maximal.*

Recall any $n \geq 0$ is the characteristic of some commutative ring.

**Lemma 2.1.18.** *Let $R$ be an integral domain. Then its characteristic is either $0$ or prime.*

*Proof.* Let $\iota : \mathbb{Z} \to R$ be the unique ring homomorphism, so $\ker(\iota) = n\mathbb{Z}$, where $n \geq 0$ is the characteristic of $R$. Then by the first isomorphism theorem, $\operatorname{im}(\iota) \leq R$ is isomorphic to $\mathbb{Z}/n\mathbb{Z}$. Since $R$ is an integral domain, $\mathbb{Z}/n\mathbb{Z}$ is one too. We proved earlier that if $n > 0$ and $\mathbb{Z}/n\mathbb{Z}$ is an integral domain, $n$ must be prime. □

## 2.2 Factorisation

Throughout this subsection, we'll only work with rings that are integral domains. We'll explore some number-theoretic properties of certain classes of rings.

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

- *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, we can say that $(r) = (s)$ or we can say that $r \mid s$ and $s \mid r$.*

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

**Definition 2.2.3.** *Let $R$ be an integral domain. We say that $r \in R$ is irreducible if the following holds.*

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. We $r \in R$ is prime if the following 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.** *These properties depend on the ring, not just the element.*

- *$2$ is prime in $\mathbb{Z}$, but not in $\mathbb{Q}$. In $\mathbb{Q}$ it is a unit.*
- *$2X$ 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 it's an integral domain. We claim that $3$ is irreducible but not prime in $R$. Note that $2$ and $3$ both divide $6$, however, $6 = (1 - \sqrt{-5})(1 + \sqrt{-5})$ and neither $2$ nor $3$ divide either of the factors on the right-hand side. Therefore, $2$ and $3$ cannot be prime. Now suppose that $3 = (a + b\sqrt{-5})(u + v\sqrt{-5})$ for*

*$a, b, u, v \in \mathbb{Z}$. Applying $|\cdot|^2$ to both sides, we see that*

$$9 = \left(a^2 + 5b^2\right)\left(u^2 + 5v^2\right).$$

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

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

*Proof.* $(\Rightarrow)$. Suppose $(r)$ is a prime ideal. If $r = 0$, we're done, so assume $r \neq 0$. Since prime ideals aren't the whole ring, $r$ cannot be a unit. Now suppose $r \mid uv$ which implies that $uv \in (r)$. Since $(r)$ is prime, $u \in (r)$ or $v \in (r)$. So $r \mid u$ 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 $r \mid uv$, so $r \mid u$ or $r \mid v$. So $u \in (r)$ or $v \in (r)$. Hence, $(r)$ is prime. □

**Lemma 2.2.8.** *If $r \in R$ is prime, then $r$ is irreducible.*

*Proof.* Let $r \in R$ be prime, and suppose $r = uv$. Trivially $r \mid uv$, so since $r$ is prime, $r \mid u$ or $r \mid v$. Without loss of generality assume $r \mid u$. So there is some $s \in R$ such that $rs = u$. Hence, $r = uv = r(sv)$. Since $R$ is an integral domain, this implies $sv = 1$, so $v$ is a unit which means that $r$ is irreducible. □

In Example 2.2.6 we saw that $3 \in \mathbb{Z}[\sqrt{-5}]$ 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, ED, 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 $r, s \neq 0 \in R$.*
- *For all $a, b \in R$ with $b \neq 0$, there are $q, r \in R$ such that*

  $$a = qb + r$$

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

**Example 2.2.10.**

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

**Proposition 2.2.11.** *The Gaussian integers $\mathbb{Z}[i]$ is a Euclidean domain with $\theta(r) = |r|^2$.*

*Proof.* First, note that $\theta(rs) = \theta(r)\theta(s)$, and if $r \neq 0 \in \mathbb{Z}[i]$, then $\theta(r) \geq 1$. So if $r, s \neq 0$ then $\theta(rs) \geq \theta(r)$. Suppose we choose $a, b \in \mathbb{Z}[i]$ such that $b \neq 0$. Choose some $q \in \mathbb{Z}[i]$ such that $\left|\frac{a}{b} - q\right| < 1$. We can do this as every complex number has a distance of at most 1 from some Gaussian integer. We can rearrange this to

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

where $|c| < 1$. 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 $\theta(r) = |r|^2 < \theta(b)$, so we're done. □

This strategy doesn't just work for $\mathbb{Z}[i]$. This would have worked for any subring $R \leq \mathbb{C}$ such that for any $r \in \mathbb{C}$, there is a point in $R$ of distance less than 1 from $r$. This doesn't hold for $\mathbb{Z}\left[\sqrt{-5}\right]$, and in fact, this ring isn't a Euclidean domain.

**Definition 2.2.12.** *Let $R$ be an integral domain. Then $R$ is a principal ideal domain, PID, if every ideal is principal. That is, for any ideal $I \subseteq R$ it follows that $I = (r)$ for some $r \in R$.*

**Example 2.2.13.** *$\mathbb{Z}$ is a principal ideal domain.*

**Theorem 2.2.14.** *Let $R$ be a Euclidean domain. Then $R$ is a principal ideal domain.*

*Proof.* Let $\theta$ be the Euclidean function for $R$. Let $I \subseteq R$ be an ideal, which we can assume to be non-zero. Choose $b \neq 0 \in I$ such that $\theta(b)$ is minimised, and let $a \in I$ be another element. Then there are some $q, r \in R$ with $a = qb + r$, with $r = 0$ or $\theta(r) < \theta(b)$. Since $I$ is an ideal and $a, b \in I$ we observe that $r = a - qb$ also lies in $I$. We can't have that $\theta(r) < \theta(b)$ and $r \neq 0$, by how $b$ is defined, so $r = 0$. This implies that $a = qb$, and so $a \in (b)$. Since $a \in I$ was arbitrary, the ideal $I = (b)$ is a principal ideal. □

**Example 2.2.15.** *For any field $F$, as $F[X]$ is a Euclidean domain it is also a principal ideal domain. Similarly, $\mathbb{Z}[i]$ is a principal ideal domain.*

**Proposition 2.2.16.** *The ideal $(2, X) \subset \mathbb{Z}[X]$ 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 $0$, which implies $f$ must be $\pm 1$ or $\pm 2$. We know $f \neq \pm 1$ as $1 \notin (2, X)$. Similarly $f \neq \pm 2$ since $2$ does not divide $X$. So $(2, X)$ can't be a principal ideal. □

**Remark 2.2.17.** *Proposition 2.2.16 tells us that $\mathbb{Z}[X]$ is not a principal ideal domain, and hence not a Euclidean domain.*

**Example 2.2.18.** *Let $F$ be a field and $A \in M_n(F)$. Consider the set*

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

*This is an ideal as if $f, g \in I$ then $(f + g)(A) = f(A) + g(A) = 0$ and if $f \in I$ and $h \in F[X]$ then $(fh)(A) = f(A)h(A) = 0$. Since $F[X]$ is a principal ideal domain, $I = (m)$ for some $m \in F[X]$. Then $m(A) = 0$ and for any $f \in F[X]$ such that $f(A) = 0$ it follows that $m \mid f$ . The polynomial $m$ is called the minimal polynomial of $A$.*

The existence of minimal polynomials can be proven much more directly but would be much longer.

**Definition 2.2.19.** *An integral domain $R$ is a unique factorisation domain, UFD, if the following holds.*

- *Every non-zero and non-unit $r \in R$ 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 $n = m$, and they can be reordered such that each $p_i$ is an associate of $q_i$. That is, they're related by multiplication by a unit.*

In other words, in a UFD the factorisation of elements into irreducibles exists and is unique. Our next goal will be to prove that principal ideal domains are unique factorisation domains.

**Lemma 2.2.20.** *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, and that $r$ is not irreducible. If $r = 0$, then we know $R = R/(r)$ must be a field. Otherwise, we can assume that $r = xy$ for $x, y \in R$ non-units. So $(r) \subseteq (x) \subseteq R$. Since $x$ isn't a unit, $(x) \neq R$. Since $(r)$ is maximal, $(r) = (x)$. So $r = xz$ 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, $y$ isn't a unit whereas $z$ is, which is 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 non-maximal. So there is some proper ideal $(r) \subseteq I \subseteq R$ such that $I \neq (r)$. Since $R$ is a principal ideal domain, $I = (s)$ 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 2.2.21.** *Let $R$ be a principal ideal domain. If $r \in R$ is irreducible, it is prime.*

*Proof.* If $r \in R$ is irreducible, then $(r) \subseteq R$ is a maximal ideal, by Lemma 2.2.20. Therefore it is also a prime ideal. Since the ideal $(r)$ is prime and $r \neq 0$, and so Lemma 2.2.7 implies that $r$ is prime. □

**Remark 2.2.22.** *Recall primes are always irreducible in any integral domain. In Example 2.2.6 we show that $2$ was irreducible but not prime. So Lemma 2.2.21 shows $\mathbb{Z}[\sqrt{-5}]$ cannot be a principal ideal domain.*

**Corollary 2.2.23.** *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 PID, 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. So $r$ is irreducible by Lemma 2.2.21. Therefore by another Lemma 2.2.20 $(r)$ is maximal. □

We already know that maximal ideals are prime. Therefore, Corollary 2.2.23 says that in a principal ideal domain, prime ideals and maximal ideals are the same. Except the zero ideal which is always prime if $R$ is a PID but maximal if and only if $R$ is a field.

**Proposition 2.2.24.** *Let $R$ be a principal ideal domain, and $I_1 \subseteq I_2 \subseteq \ldots \subseteq R$ an increasing sequence of ideals in $R$. Then this sequence is eventually constant. That is, there is some integer $N$ such that $I_n = I_{n+1}$ for all $n \geq N$.*

*Proof.* Let $I = \cup_{i \geq 1} I_i \subseteq R$. One can show that this is an ideal of $R$. Since $R$ is a PID, we have that $I = (r)$ for some $r \in I = \cup_i I_i$. Hence, there is some $N$ such that $r \in I_N$. Then

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

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

**Remark 2.2.25.** *Rings with the property stated in Proposition 2.2.24 are called Noetherian rings.*

**Theorem 2.2.26.** *A principal ideal domain is a unique factorisation domain.*

*Proof.* We first show the uniqueness of a factorisation into irreducibles. Suppose $p_1 \ldots p_n = q_1 \ldots q_m$, where each $p_i, q_j \in R$ is irreducible. In particular, $p_1 \mid q_1 \ldots q_m$. Since $p_1$ is irreducible, it's prime, so $p_1 \mid q_j$ for some $j$. Reordering if necessary, we can assume that $p_1 \mid q_1$, so $p_1 a = q_1$ for some $a$. Since $q_1$ is irreducible, $a$ must be a unit and so $p_1, q_1$ are associates. Since $R$ is an integral domain, we can cancel $p_1$ to obtain

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

and, replacing $q_2$ with $aq_2$, we get

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

Repeating the process, we find that, possibly after reordering, $p_i$ and $q_i$ are associates for all $i$. We also find that $n = m$, as if it were not then without loss of generality we could suppose $n > m$, and thus we find that $p_{m+1} \dots p_n = 1$. This implies that $p_n$ is a unit and so we arrive at a contradiction. We next show the existence of the factorisation into irreducibles. Suppose $r \in R$ is a non-unit which can't be factored as a product of irreducibles. In particular, $r$ isn't irreducible. So we can write $r = r_1 s_1$, with $r_1, s_1 \in R$ non-units. Since $r$ isn't a product of irreducibles, at least one of $r_1, s_1$ isn't 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 $r_2 = r_3 s_3$ with $r_3, s_3 \in R$ non-units. By assumption, we can continue this process indefinitely. Consequently, we have an increasing sequence of ideals,

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

Since $R$ is a PID and it is Noetherian and so there is some $N$ 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. □

**Example 2.2.27.** *$\mathbb{Z}[i]$ is a Euclidean domain, and therefore a PID and so a UFD.*

**Definition 2.2.28.** *Let $R$ be a ring. An element $d \in R$ is a greatest common divisor, GCD, of a finite sequence $a_1, \dots, a_n \in R$ if $d \mid a_i$ for all $i$, and if for any other $d' \in R$ such that $d' \mid a_i$ for all $i$ it follows that $d' \mid d$.*

In general, a GCD may not always exist. If a GCD exists, it is unique up to multiplication by units.

> **Proposition 2.2.29.** *Let $R$ be a unique factorisation domain. Then for any $a_1, \dots, a_n \in R$ not all zero, they 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. 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. We let $l_j = \min_i \{k_{ij}\}$ for all $j$, and define

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

Since $l_j \leq k_{ij}$ for all $i$, it follows that $d \mid a_i$ for all $i$. Suppose $d' \in R$ also satisfies $d' \mid a_i$ for all $i$. 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 all $i$ we must have $c_j \leq l_j$ for all $j$. So $d' \mid d$, and hence $d$ is a GCD. Suppose $d$ and $d'$ are both GCDs of $a_1, \dots, a_n$. Then they must divide each other and hence are associates. □

We can summarise the main relationships we've established so far with the following chain of implications,

$$\mathbb{Z} \Rightarrow \mathrm{ED} \Rightarrow \mathrm{PID} \Rightarrow \mathrm{UFD} \Rightarrow \mathrm{ID} \Rightarrow \text{ Commutative ring } \Rightarrow \text{ Ring.}$$

However, none of these implications go both ways.

1. $\mathbb{Q}$ and $\mathbb{Z}[i]$ are Euclidean domains but are clearly not isomorphic to $\mathbb{Z}$ as rings.
2. $\mathbb{Z}\left[\frac{1+\sqrt{-19}}{2}\right]$. One can check this is a PID but not an ED.
3. $\mathbb{Z}[X]$ is not a PID, but we will see it is a UFD.
4. $\mathbb{Z}[\sqrt{-5}]$, 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 Factorisation in 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 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's 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.

**Definition 2.3.1.** *Let $R$ be an integral domain. Its field of fractions $F$ is defined as follows. As a set, let*

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

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

**Lemma 2.3.2.** *The relation $\sim$ as defined 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 $bcf = bde$ respectively. So $adf = bde$. Since $d \neq 0$ and $R$ is an integral domain, we can cancel the $d$, to find that $af = be$ which means $(a,b) \sim (e,f)$. □

We write the equivalence class of $(a,b)$ as $a/b$ or $\frac{a}{b}$. We equip $F$ with a ring structure as follows.

- $0_F := \frac{0}{1}$.
- $1_F := \frac{1}{1}$.
- For $a/b, c/d \in F$ let
$$\frac{a}{b} + \frac{c}{d} := \frac{ad+bc}{bd}.$$
- For $a/b, c/d \in F$ let
$$\frac{a}{b} \cdot \frac{c}{d} := \frac{ac}{bd}.$$

**Lemma 2.3.3.** *The operations $+$ and $\cdot$ on $F$ as defined above are well-defined. That is, they respect the equivalence relation $\sim$, and $F$ is a ring.*

**Example 2.3.4.** *Let $(a,b),(c,d) \in F$. Then*

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

*and so $0_F$ is the additive identity for $F$. 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 $F$ is associative.*

**Proposition 2.3.5.** *The field of fractions $F$ is a field, and the set*

$$\left\{\frac{r}{1} : r \in R\right\} \leq F$$

*is a subring isomorphic to $R$.*

*Proof.* First, let $a/b \in F$ be non-zero. This means that $a \neq 0$, so $b/a \in F$ too. Then

$$\frac{a}{b} \cdot \frac{b}{a} = \frac{ab}{ba} = 1$$

so $(a,b)$ is a unit. Now let $\phi : R \to F$ be the homomorphism such that $r \mapsto (r,1)$. This has kernel

$$\ker(\phi) = \{r \in R : (r,1) = (0,1)\} = \{0\}$$

and so is injective. The set $\left\{\frac{r}{1} : r \in R\right\} = \operatorname{im}(\phi)$ and we know that $\operatorname{im}(\phi) \leq F$ is a subring isomorphic to $R$ by the first isomorphism theorem. □

**Example 2.3.6.** *The field of fractions of $\mathbb{Z}$ is $\mathbb{Q}$. 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 $F[X]$.*

**Definition 2.3.7.** *If $R$ is a ring, and $f \in R[X]$, then $f$ is monic if $f = a_0 + a_1 X + \ldots a_{n-1} X^{n-1} + 1 \cdot X^n$. That is, the coefficient of the highest order term is $1$.*

**Definition 2.3.8.** *Let $R$ be a UFD, and $f = a_0 + \ldots + a_n X^n \in R[X]$. 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 note that the ideal it generates, $(c(f))$, is well-defined.

**Definition 2.3.9.** *Let $R$ be a UFD and $f \in R[X]$. We say $f$ is primitive if $c(f)$ is a unit. In other words, the coefficients of $f$ are all coprime.*

**Example 2.3.10.**

- *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 so $f$ is primitive.*
- *Let $g = 2 + 4X + 6X^2 \in \mathbb{Z}[X]$. Then $c(f) = 2$, up to associates, so $g$ isn't primitive. However, $2$ is a unit in $\mathbb{Q}$ so $g$ is primitive in $\mathbb{Q}[X]$.*

**Lemma 2.3.11.** *Let $R$ be a UFD, and $f \in R[X]$. Then*

$$f = c(f) \cdot f'$$

*for some primitive $f' \in R[X]$.*

*Proof.* By the definition of GCD, $c(f) \mid a_i$ for all $i$. Therefore, $a_i = c(f) b_i$ for some $b_i \in R$ with the GCD of $b_1, \ldots, b_n$ is $1$. Let $f' = b_0 + \ldots b_n X^n$. Then $f'$ is primitive and $f = c(f) \cdot f'$. □

**Lemma 2.3.12.** *Let $R$ be a UFD. If $f, g \in R[X]$ are primitive, so if $f \cdot g$.*

*Proof.* Write $f = a_0 + \ldots a_n X^n$ and $g = b_0 + \ldots b_m X^m$, with $a_n, b_m \neq 0$, and $f, g \in R[X]$ primitive. Suppose $fg$ isn't primitive. So $c(fg)$ isn't a unit. Since $R$ is a UFD, there is some irreducible $p \in R$ with $p \mid c(fg)$. By assumption, $c(f)$ and $c(g)$ are units, so $p$ doesn't divide $c(f)$ or $c(g)$. This implies $p$ doesn't divide all $a_i$ or all $b_j$. So there are some $k, l \geq 0$ such that $p$ divides all of $a_1, \ldots, a_{k-1}$ but $p$ doesn't divide $a_k$, and similarly $p$ divides all $b_1, \ldots, b_{l-1}$ but $p$ doesn't divide $b_l$. Consider the $X^{k+l}$ coefficient in $f \cdot g$ which turns out to be

$$\sum_{i+j=k+l} a_i b_j = (a_{k+l} b_0 + \ldots + a_{k+1} b_{l-1}) + a_k b_l + (a_{k-1} b_{l+1} + \ldots a_0 b_{l+k}) .$$

In the first set of brackets on the right-hand side, $p$ divides each $b_j$, and in the second set of brackets on the right-hand side, $p$ divides each $a_i$. Furthermore, $p$ divides this whole coefficient by assumption. So $p$ must divide the the term $a_k b_l$. 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 so $f \cdot g$ is primitive. □

**Corollary 2.3.13.** *Let $R$ be a UFD. Then for $f, g \in R[X]$ the content $c(f \cdot g)$ is an associate of $c(f) \cdot c(g)$.*

*Proof.* Write $f = c(f) \cdot f'$ and $g = c(g) \cdot g'$ where $f', g' \in R[X]$ are primitive, which we can do by Lemma 2.3.11. Then $f \cdot g = c(f) c(g) (f' \cdot g')$. As $f'$ and $g'$ are primitive we know $f' \cdot g'$ is primitive by Lemma 2.3.12. Therefore, $c(f) c(g)$ is a GCD of the coefficients of $f \cdot g$. □

**Remark 2.3.14.**

- *Note that Corollary 2.3.13 implies Lemma 2.3.12. However, Lemma 2.3.13 is derived from Lemma 2.3.12.*

- *We cannot really say $c(f \cdot g) = c(f) \cdot c(g)$, since both are only well-defined up to associates.*

**Lemma 2.3.15** (Gauss' Lemma)**.** *Let $R$ be a UFD, 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's irreducible in $F[X]$.*

This is useful because checking irreducibility in $F[X]$ is generally harder than checking it in $R[X]$.

**Example 2.3.16.** *Let $f = 1 + X + X^3 \in \mathbb{Z}[X]$ Note that $c(f) = 1$ which implies that $f$ is primitive. Suppose $f$ is reducible in $\mathbb{Q}[X]$. By Gauss' lemma, this means that $f$ is reducible in $\mathbb{Z}[X]$. Let $f = g \cdot h$ for $g, h \in \mathbb{Z}[X]$ non-units. Since $g$ and $h$ are non-units, they cannot be constant since all coefficients of $f$ are $0$ and $1$. 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*

$$1 = b_1 c_2.$$

*Since all $b_i, c_j$ are all integers, all of $b_0, c_0, b_1, c_2$ must be $\pm 1$. In particular, $g = \pm 1 \pm X$. Since $g$ is a factor of $f$, this implies $\pm 1$ is a root of $f$. But we can compute $f(1)$ and $f(-1)$, and they're both non-zero, which is a contradiction. So $f$ is irreducible in $\mathbb{Q}[X]$. In particular $\mathbb{Q}[X]/(f)$ is a field. Note that when we concluded $b_0$ and $b_1$ are $\pm 1$, we needed to be working in $\mathbb{Z}$. If we were in $\mathbb{Q}$ this would tell us a lot less. Gauss' lemma allows us to work over $\mathbb{Z}$ rather than $\mathbb{Q}$.*

*Proof. (Lemma 2.3.15).* Let $R$ be a UFD, $F$ its field of fractions, and $f \in R[X]$ a primitive polynomial.
($\Rightarrow$) 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 $0$ it would be a unit, which it's not. So $\deg(g) > 0$. Similarly $\deg(h) > 0$. Therefore, $g$ and $h$ cannot be units in $F[X]$. Writing $f = gh$ and viewing $f, g$ and $h$ elements of $F[X]$, we see that $f$ is reducible in $F[X]$.
($\Leftarrow$) Suppose $f$ is reducible in $F[X]$. So $f = g \cdot h$, where $g, h \in F[X]$ are non-units. Then we must have $\deg(g), \deg(h) > 0$. We now clear denominators by choosing $a, b \in R \backslash \{0\}$ such that $ag$ and $bh$ lie in $R[X]$. This can be done by taking $a$ to be the product of all denominators of coefficients of $g$ and similarly for $b$. Let $g' = ag$ and $h' = bh$ and note these lie in $R[X]$. As $h$ may not be in $R[X]$ we note that $h' = bh$ isn't a factorisation in $R[X]$, and similarly for $g' = ag$ is also not necessarily a factorisation on $R[X]$. Using $f = gh$ we see that

$$abf = g'h' \in R[X]$$

where now all the factors lie in $R[X]$. So 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 $c(abf) = ab$. But by the above equation, this is equal to $c(g'h') = c(g')\, c(h')$, up to associates. Hence, $ab$ is an associate of $c(g')\, c(h')$. 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 can cancel the factors of $ab$ from both sides to obtain

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

Here $u \in R^\times$is a unit and $g'', h'' \in R[X]$ both lie in $R[X]$. Moreover, as $g''$ and $h''$ have positive degrees they are non-units. Therefore, $f$ is reducible in $R[X]$. □

**Theorem 2.3.17.** *Let $R$ be a UFD. Then $R[X]$ is a UFD.*

*Proof.* We first prove the existence of factorisations. 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 UFD we can factor the content as

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

where each $p_i \in R$ is irreducible in $R$. As $R \subseteq R[X]$ these are irreducible in $R[X]$ as well. Next, suppose $f_1$ is not the product of irreducibles. In particular, $f_1$ itself isn't 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 both $f_2, g_2$ 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 $f_2 = f_3 \cdot g_3$ as a product of non-units. Continuing like this gives a sequence $f_1, f_2, f3 \ldots \in R[X]$ of non-zero elements, with $\deg(f_1) > \deg(f_2) > \ldots > 0$. But we cannot have an infinite sequence of positive integers which is strictly decreasing, 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. Moreover, $f = p_1 \dots p_n \cdot q_1 \dots q_m$, 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. We next prove the uniqueness of factorisations of a non-zero non-unit $f \in R[X]$. Consider, $c(f) = p_1 \dots p_n$ as a factorisation into irreducibles of $R$. This is unique up to reordering and associates since $R$ is a UFD. Consequently, by cancelling out the content, we only need to show that primitive polynomials can be factored uniquely. 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 $c(f')$, which is a unit is a unit as $f'$ is primitive, therefore 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 Gauss' lemma the $q_i$ and $r_j$ are irreducible in $F[X]$. Since $F[X]$ is an ED, it's a PID and hence a UFD. 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 all $i$. That is, $r_i = u_i q_i$ for some units $u_i \in F[X]$. Note each $u_i$ must be a constant polynomial since it's a unit. So $u_i = \frac{a_i}{b_i}$, for some elements $a_i, b_i \in R \setminus \{0\}$. Now $a_i r_i = b_i q_i$, and all factors in this equation are in $R[X]$. Recall that $r_i$ and $q_i$ are primitive, so 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 $r_i = v_i \cdot q_i \cdot v_i$ is a unit in $R^\times$. Hence, it is also a unit in $R[X]^\times$. Consequently, $r_i$ and $q_i$ are associates in $R[X]$ for each $i$, implying factorisations into irreducibles are unique up to associates and reordering. □

**Remark 2.3.18.** *Here are some comments about the proof of Theorem 2.3.17.*

1. *Showing the existence of factorisations was similar to showing a PID is a UFD. First, we found factorisations $f_1 = f_2 g_2, f_2 = f_3 g_3, \ldots$ into non-units, and then we showed this process must terminate; here we did this second step by showing that at each stage we decrease degrees, whereas earlier we did this using the Noetherian property.*

2. *To show uniqueness, we first took out out factors of the contents to turn everything into primitives, used Gauss' lemma to argue reducibility in $R[X]$ and $F[X]$ are the same (noting we're working with primitive polynomials at this stage), and used the fact that $F[X]$ is a UFD.*

**Remark 2.3.19.**

- *Iterating Theorem 2.3.17 we deduce that if $R$ is a UFD then $R[X_1, \ldots, X_n]$ is a UFD, for $n \geq 1$.*

- *Theorem 2.3.17 also shows that $\mathbb{Z}[X]$ is a UFD. Therefore, we have an example of a UFD that is not a PID.*

**Proposition 2.3.20** (Eisenstein's criterion)**.** *Let $R$ be a UFD, and let $f = a_0 + a_1X + \ldots + a_nX^n \in R[X]$ be a primitive polynomial with $a_n \neq 0$. Let $p \in R$ be irreducible such that*

1. *$p$ doesn't divide $a_n$,*
2. *$p$ divides $a_i$ for all $0 \leq i < n$, and*
3. *$p^2$ doesn't divide $a_0$.*

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

*Proof.* □

**Example 2.3.21.** *Consider $f = X^n - p \in \mathbb{Z}[X]$, where $p \in \mathbb{N}$ is prime. All the conditions for Eisenstein's criterion hold and so $\mathbb{Z}[X]$ is irreducible in $\mathbb{Z}[X]$. By Gauss' lemma, it is also irreducible in $\mathbb{Q}[X]$. Consequently, we deduce that $X^n - p$ has no rational zeroes. In particular, $\sqrt[n]{p}$ is irrational for $n > 1$.*

**Example 2.3.22.** *Let $f = X^{p-1} + X^{p-2} + \ldots + 1 \in \mathbb{Z}[X]$, for $p \in \mathbb{N}$ prime. Note that*

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

*Since all the coefficients are $1$, Eisenstein's criterion cannot immediately be applied. Let $Y = X - 1$. Then $f(X) = \hat{f}(Y)$ 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 the prime $p$ as $\hat{f}$ is primitive and $p \mid \binom{p}{i}$ for $1 \leq i \leq p-1$ but $p^2 \nmid \binom{p}{p-1}$. So $\hat{f} \in \mathbb{Z}[Y]$ is irreducible. If we had a factorisation $f = gh$ in $\mathbb{Z}[X]$, then $\hat{f}(Y) = g(Y+1)h(Y+1)$. This contradicts $\hat{f}(Y)$ being irreducible, hence $f \in \mathbb{Z}[X]$ must be irreducible. In particular, this shows that none of the roots of $f$ are rational. Which we already know as the roots are $e^{\frac{2\pi ik}{p}}$ for $1 \leq k \leq p-1$, which aren't even real for $p > 2$.*

## 2.4 Algebraic integers

We defined rings such as $\mathbb{Z}[i]$ and $\mathbb{Z}\left[\sqrt{-5}\right]$ as $\{a + bi \mid a, b \in \mathbb{Z}\}$. How do we instead define rings such as $\mathbb{Z}\left[e^{\frac{2\pi i}{n}}\right]$?

**Definition 2.4.1.** *A complex number $\alpha \in \mathbb{C}$ is called an algebraic integer if it's the root of some monic polynomial $f \in \mathbb{Z}[X]$. That is, $f(\alpha) = 0$.*

**Remark 2.4.2.** *Note that 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 2.4.3.** *For $\alpha$ an algebraic integer, we write $\mathbb{Z}[\alpha] \leq \mathbb{C}$ for the smallest subring containing $\alpha$. More explicitly,*

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

*where we take the intersection over all subrings $S \leq \mathbb{C}$ containing $\alpha$. Equivalently, we can let $\phi_\alpha : \mathbb{Z}[X] \to \mathbb{C}$ be the homomorphism sending $f$ to $f(\alpha)$, and let $\mathbb{Z}[\alpha] = \operatorname{im}(\phi_\alpha) \leq \mathbb{C}$.*

Note that by the first isomorphism theorem,

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

where we know $\ker(\phi_\alpha)$ is non-empty by definition of $\alpha$ being an algebraic integer.

**Proposition 2.4.4.** *Let $\alpha \in \mathbb{C}$ be an algebraic integer, and $\phi_\alpha : \mathbb{Z}[X] \to \mathbb{C}$ 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.*

As we require that $f_\alpha$ is monic, these conditions determine $f_\alpha$ uniquely.

*Proof.* We know 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 $f_\alpha$ isn't 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 $f_\alpha$ is primitive. We claim that $f_\alpha$ generates $I, f_\alpha$ is irreducible, and $f_\alpha$ is monic. 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 $\alpha$ shows that $ar(\alpha) = 0$ and so $ar \in I$. Since $f_\alpha$ has 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

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

where equality makes sense only up to associates and the fact that $f_\alpha$ is primitive. 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 $f_\alpha$ is prime which implies it is irreducible. Since $I$ contained some monic polynomial $f$ and $f_\alpha \mid f$, the leading coefficient of $f_\alpha$ must be a unit in $\mathbb{Z}$, that is $\pm 1$. If it's $-1$, we take $-f_\alpha$. Either way, we may assume $f_\alpha$ is monic. □

**Definition 2.4.5.** *We call the polynomial $f_\alpha$ as above the minimal polynomial of $\alpha$.*

**Example 2.4.6.**

1. *$\alpha = i$ is an algebraic integer with minimal polynomial $X^2 + 1$.*

2. *$\alpha = \sqrt{2}$ is an algebraic integer with minimal polynomial $X^2 - 2$.*

3. $\alpha = \frac{1}{2}(1+\sqrt{-3})$ *is an algebraic integer with minimal polynomial* $X^2 - X - 1$.

**Lemma 2.4.7.** *Let* $\alpha \in \mathbb{Q}$ *be an algebraic integer. Then* $\alpha \in \mathbb{Z}$.

The terminology "algebraic integer" would be rather odd if this lemma weren't true.

*Proof.* Let $f_\alpha \in \mathbb{Z}[X]$ be the minimal polynomial of $\alpha$. Then $f_\alpha$ is irreducible and primitive and so by Gauss' lemma $f_\alpha$ 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 $f_\alpha$ is irreducible and monic this implies that $f_\alpha = X - \alpha$. But since $f_\alpha \in \mathbb{Z}[X]$, we must have $\alpha \in \mathbb{Z}$. □

**Remark 2.4.8.** *It turns out that 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 UFD for which we will characterise its prime, and hence irreducible, elements.

**Proposition 2.4.9.** *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* $a, b \neq 0 \in \mathbb{Z}$.

*Proof.* $(\Rightarrow)$, If $p = a^2 + b^2$, then $p = (a+ib)(a-ib)$ isn't irreducible, so it isn't 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 $u, v$ aren't units, we must have that $|u|^2, |v|^2 \neq 1$. So $|u|^2 = p$ and $|v|^2 = p$. If $u = a + ib$, this says that $p = a^2 + b^2$. □

Before we say more about primes in $\mathbb{Z}[i]$, we need the following technical lemma.

**Lemma 2.4.10.** *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. Let* $\mathbb{F}_p^\times = \mathbb{F}_p \backslash \{0\}$ *be its group of units. Then* $\mathbb{F}_p^\times \cong C_{p-1}$ *is cyclic of order* $p - 1$.

*Proof.* $\mathbb{F}_p^\times$has order $p-1$, and is abelian. From the classification of finite abelian groups, if $\mathbb{F}_p^\times$isn't cyclic, 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 this subgroup 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 UFD so it cannot have more than $m$ distinct linear factors, which is a contradiction. □

**Theorem 2.4.11.** *The primes in the Gaussian integers* $\mathbb{Z}[i]$ *are one of the following, up to associates.*

1. *Primes 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.* We first show that these are indeed primes. If $p \in \mathbb{N}$ is prime and is equal to $3 \bmod 4$, then $p \neq a^2 + b^2$ for $a, b \in \mathbb{Z}$ as any square number is always $0$ or $1 \bmod 4$. Therefore, $p$ is a prime in $\mathbb{Z}[i]$ by Proposition 2.4.9. 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. Now let $z \in \mathbb{Z}[i]$ be irreducible, and hence prime. 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 $|z|^2 > 1$. So $p \mid z\bar{z}$ in $\mathbb{Z}[i]$. 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 $z$ falls into the first class of primes we found. Otherwise, $p = 2$ or $p$ is $1 \bmod 4$. 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 2.4.10. 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's 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 note that $p \mid (a+i)(a-i)$ for $a = 1$. However, $p \nmid a + i$ for any prime $p$ and so $p$ 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 $z_1, z_2$ aren't 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 UFD we must have that $z = z_1$ or $z = \bar{z}_1$. Either way $|z|^2 = p$. This implies $p$ is a sum of two squares, so we must have $p = 2$ or $p$ is equal to $1 \bmod 4$. □

The upshot of this is that finding primes in $\mathbb{Z}[i]$ is essentially an equivalent problem to finding sums of two squares with are prime.

**Corollary 2.4.12.** *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 2.4.11 each $|\alpha_i|^2$ is either $p^2$ with $p$ being 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 $n = p_1^{n_1} \dots p_k^{n-k}$ as a product of powers of distinct primes. For each $p_i$, if $p$ 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 2.4.13.** *Consider $65 = 5 \times 13$. As $5$ and $13$ are both $1 \bmod 4$, Corollary 2.4.12 tells us that $65$ is a sum of two squares. Moreover, the proof of Corollary 2.4.12 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*

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

*and we can also write*

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

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

# 3 Modules

## 3.1 Definition and Examples

"A module is to a ring what a vector space is to a field". Various parts of the theory of vector spaces will carry over, but many won't, making the theory of modules far richer than that of vector spaces. In this section, we won't assume our rings are commutative. Let $R$ be a ring.

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

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. $1_R \cdot m = m.$

**Remark 3.1.2.** *Here we always multiply elements of $M$ by elements of $R$ on the left. Hence, what we call a module is sometimes called a left module. There is an analogous notion of a right module, but the theory of such things 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 $A \to A$. This 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 $\phi$, 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 3.1.1 are equivalent.*

We can think of an $R$-module like "$R$ acting on an abelian group", just like how groups can act on sets.

**Example 3.1.5.**

- *Let $F$ be a field. Then an $F$-vector space is the same thing as an $F$-module as the axioms for both are the same.*
- *Let $I \subseteq R$ be an ideal. Then $I$ is an $R$-module, via $r \cdot a := r \cdot_R a$, for $r \in R$ and $a \in I$.*
- *$R$ is an $R$-module. For $n \geq 0, R^n$ is an $R$-module, via $r \cdot (r_1, \ldots r_n) := (r \cdot r_1, \ldots, r \cdot r_n)$.*
- *If $I \subseteq R$ is a two-sided ideal, then $R/I$ is an $R$-module via $r \cdot (a + I) := (r \cdot a) + I$.*
- *A $\mathbb{Z}$-module is the same thing as an abelian group. Let $A$ be an abelian group and consider $n \in \mathbb{Z}$ and $a \in A$. Then $n \cdot a := \underbrace{a + \ldots + a}_{n}$, if $n \geq 0$, and $n \cdot a := \underbrace{(-a) + \ldots + (-a)}_{|n|}$ where we sum $|n|$ copies of $-a$ if $n \leq 0$. Alternatively, there's a unique ring homomorphism $\mathbb{Z} \to \mathrm{End}(A)$, so we can take 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 $M_1 \times \ldots \times M_k$ 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 define $R^n = \underbrace{R \oplus \ldots \oplus R}_{n}$ as above.*

**Definition 3.2.3.** *A subset $N \subseteq M$ is an $R$-submodule if it's a subgroup of $M$ and for $r \in R$ and $n \in N$ it follows that $r \cdot n \in N$. If this holds, we write $N \leq M$.*

**Example 3.2.4.**

1. *$\{0\}$ and $M$ are always submodules of a module $M$.*
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 are the same thing.*

**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 given by $r \cdot (m + N) := (r \cdot m) + N$.*

**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. For modules just have submodules.*

**Definition 3.2.7.** *A function $\phi : M \to N$ 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.

**Example 3.2.8.** *The ideals $(2)$, $(3) \subseteq \mathbb{Z}$ are different, but isomorphic as $R$-modules, through the isomorphism $\phi : (2) \to (3)$ sending $2k$ to $3k$. In fact, if $R$ is an integral domain and $r \neq 0 \in R$, then $R$ and $(r)$ are always isomorphic as $R$-modules.*

**Definition 3.2.9.** *Let $R, R'$ be rings, $M$ an $R$-module, and $M'$ an $R'$-module. Then $M \times M'$ is an $R \times R'$-module, with action $(r, r') \cdot (m, m') := (r \cdot m, r' \cdot m')$.*

## 3.3 Some theory

**Theorem 3.3.1** (First isomorphism theorem)**.** *Let $\phi : M \to N$ be an $R$-module homomorphism. Then the following hold.*

1. *$\ker(\phi) \leq M$ is an $R$-submodule.*
2. *$\operatorname{im}(\phi) \leq N$ is an $R$-submodule.*
3. *There is an isomorphism of $R$-modules*

$$\frac{M}{\ker(\phi)} \cong \operatorname{im}(\phi).$$

**Definition 3.3.2.** *The cokernel of $\phi : M \to N$, written $\operatorname{coker}(\phi)$, is the $R$-module $N/\operatorname{im}(\phi)$.*

For rings(groups) as $\operatorname{im}(\phi)$ is a subring(subgroup) rather than ideal(normal subgroup) we cannot make a similar definition.

**Theorem 3.3.3** (Second isomorphism theorem)**.** *Let $A, B \leq M$ be submodules. Then the following hold.*

1. *$A + B := \{a + b : a \in A, b \in B\} \leq M$ is a submodule.*
2. *$A \cap B \leq M$ is a submodule.*
3. *There is an isomorphism of $R$-modules,*

$$\frac{A+B}{A} \cong \frac{B}{A \cap B}.$$

**Remark 3.3.4.** *In general, if $\{A_i\}_{i\in S}$ is a collection of submodules, then*

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

*is a submodule, where the left-hand side is defined to be the set of finite sums of elements of the $A_i$. Similarly*

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

*is a submodule.*

**Theorem 3.3.5** (Third isomorphism theorem)**.** *Let $N \leq L \leq M$ be submodules. Then the following hold.*

1. *$L/N \leq M/N$ is a submodule.*
2. *There is an isomorphism of $R$-modules*

$$M/L \cong \frac{M/N}{L/N}.$$

**Proposition 3.3.6.** *Let $N \leq M$ be a submodule. Then there is a natural bijection*

$$\{\textit{submodules of } M/N\} \longleftrightarrow \{\textit{submodules of } M \textit{ containing } N\}.$$

**Proposition 3.3.7.** *The composition of two $R$-module homomorphisms is an $R$-module homomorphism.*

The proofs of the above results are omitted as they're essentially the same as those for the corresponding results for rings.

**Definition 3.3.8.** *Let $m \in M$. The submodule generated by $m$, denoted $Rm \leq M$, is $\{r \cdot m : r \in R\}$.*

Alternatively, there is a homomorphism $\phi_m : R \to M$ sending $r$ to $r \cdot m$, and $R \cdot m := \operatorname{im}(\phi)$. Also note that $\ker(\phi_m) \leq R$ is a submodule and so we define the annihilator of $m$, written $\operatorname{Ann}(m)$, to be

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

This is a submodule of $R$ and so it is an ideal. By the first isomorphism theorem, we have that $R{\cdot}m \cong R/\operatorname{Ann}(m)$.

**Definition 3.3.9.** *An $R$-module $M$ is finitely generate if there are elements $m_1, \ldots, m_n \in M$, such that*

$$\begin{aligned} M &= Rm_1 + \ldots + Rm_n \\ &= \{r_1 m_1 + \ldots r_n m_n : r_i \in R\}. \end{aligned}$$

**Example 3.3.10.** *If $F$ is a field, and $M$ an $F$-module, then as an $F$-module $M$ is finitely generated if and only it's finite-dimensional as an $F$-vector space.*

**Lemma 3.3.11.** *An $R$-module $M$ is finitely generated if and only if there's a surjective homomorphism $R^n \to M$.*

*Proof.* ($\Rightarrow$). If $M = Rm_1 + \ldots + Rm_n$, let $\phi : R^n \to M$ be the homomorphism sending $(r_1, \ldots, r_n)$ to $r_1 m_1 + \ldots r_n m_n$. This is surjective by assumption.
($\Leftarrow$). Suppose $\phi : R^n \to M$ is a surjective homomorphism. Take $m_1, \ldots, m_n$ to be the images under $\phi$ of the standard basis vectors. Then $M = Rm_1 + \ldots + Rm_n$. □

**Corollary 3.3.12.** *If $N \leq M$ is a submodule and $M$ is finitely generated, then $M/N$ is finitely generated too.*

*Proof.* Let $\phi : R^n \to M$ be a surjective homomorphism. As the quotient map $M \to M/N$ is also surjective, the composition $R^n \to M \to M/N$ is also surjective. □

However, if $N \leq M$ is a submodule and $M$ is finitely generated, then $N$ is not necessarily finitely generated.

**Proposition 3.3.13.** *Let $R = \mathbb{Z}[X_1, X_2, \ldots]$, and let $I$ be the ideal $I = (X_1, X_2, \ldots) \subseteq R$. Then $R$ is a finitely generated $R$-module but $I \leq R$ is a submodule which isn't finitely generated.*

*Proof.* 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 is a contradiction. □

## 3.4 Free modules

**Definition 3.4.1.** *Let $S$ be a set. The free module over $S$, written $R^{(S)}$, is*

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

*with coordinate-wise addition and $R$-action.*

**Proposition 3.4.2.** *Let $R$ be a ring, and $S$ a set. Then $R^{(S)}$ 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 $R^n \to R^{(S)}$.
($\Rightarrow$). Suppose $S$ is infinite and $R^{(S)} = Rm_1 + \ldots + Rm_n$. Write each $m_k = \left(x_i^{(k)}\right)_{i \in S} \in R^{(S)}$. Define the subset $T \subseteq S$ by

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

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

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

Then $a \notin Rm_1 + \ldots + Rm_n$, which is a contradiction. □

**Theorem 3.4.3.** *Let $F$ be a field. Any $F$-module is free.*

This is equivalent to the axiom of choice. We can give an alternative characterisation of freeness.

**Definition 3.4.4.** *A subset $S \subseteq M$ generates $M$ freely if the following hold.*

1. *$S$ generates $M$ as an $R$-module. The set of finite sums of elements of the form $rs$ with $r \in R$ and $s \in S$, denoted $R \cdot S$, equals $M$.*
2. *For any other $R$-module $N$, and function $\psi : S \to N$ can be extended to an $R$-module homomorphism $\phi : M \to N$.*

**Remark 3.4.5.** *Suppose point 2 of Definition 3.4.4 holds and $\phi, \phi' : M \to N$ are extensions of a function $\psi : S \to N$. Then $\phi - \phi' : M \to N$ is a homomorphism sending all of $S$ to $0$. So $S \subseteq \ker(\phi - \phi')$. But since $S$ generates $M$, this implies $\ker(\phi - \phi') = M$. In other words, $\phi = \phi'$ and so the extension $\phi$ is unique.*

**Definition 3.4.6.** *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 3.4.7.** *Definition 3.4.1 and Definition 3.4.6 are equivalent.*

*Proof.* First suppose $M \cong R^{(S)}$. We view $S$ as a subset in $M$ by identifying $s \in S$ with the element $(x_i^s)_{i \in S}$ defined by

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

This subset generates $M$ as an $R$-module. Let $\psi : S \to N$ be any function, for $N$ an $R$-module. Define a homomorphism $\phi : M \to N$ to send $(x_i)_{i \in S}$ to

$$\phi\left((x_i)_{i \in S}\right) := \sum_{i \in S} x_i \cdot \psi(i).$$

This is a homomorphism extending $\psi$, so $S$ generates $M$ freely. Conversely, now suppose some subset $S \subseteq M$ freely generates $M$. We claim $M \cong R^{(S)}$. Let $\psi : S \to R^{(S)}$ send $s \in S$ to the element $(x_i^s)_{i \in S}$ as defined above. This extends to a homomorphism $\phi : M \to R^{(S)}$. This is an isomorphism, with inverse sending $(x_i)_{i \in S}$ to $\sum_{i \in S} x_i \cdot i$. □

**Definition 3.4.8.** *A set of elements $m_1, \ldots, m_n \in M$ are linearly independent, LI, if whenever $r_1 \cdot m_1 + \ldots + r_n \cdot m_n = 0$ for $r_i \in R$, we have $r_1 = \ldots = r_n = 0$.*

**Proposition 3.4.9.** *For a subset $S = \{m_1, \ldots, m_n\} \subseteq M$, the following are equivalent.*

1. *$S$ generates $M$ freely.*
2. *$S$ generates $M$ and $S$ is linearly independent.*
3. *Every element of $M$ is uniquely expressible as $r_1 \cdot m_1 + \ldots + r_n m_n$, for $r_i \in M$.*

**Example 3.4.10.** *Consider $\mathbb{Z}/2\mathbb{Z}$ as a $\mathbb{Z}$-module. This isn't free as if it were freely generated by $S \subseteq \mathbb{Z}/2\mathbb{Z}$, we would have $S = \{1\}$, since any set containing 0 can't be linearly independent. But this set isn't linearly* *independent.*

**Example 3.4.11.** *The subset $\{2,3\} \subseteq \mathbb{Z}$ generates $\mathbb{Z}$. However, it is not linearly independent as $3 \cdot 2 + (-2) \cdot 3 = 0$ and so doesn't generate $\mathbb{Z}$ freely. Recall that if $S$ spans an $F$-vector space $V$ and $S$ isn't linearly independent, we can throw away some of $S$ and find some basis $T \subseteq S$. However no subset of $\{2,3\}$ freely generates $\mathbb{Z}$.*

**Definition 3.4.12.** *Let $M$ be a finitely generated $R$-module. Then there is some surjective homomorphism $\phi : R^n \to M$. We call the submodule $\ker(\phi) \leq R^n$ the relation module for these generators.*

**Definition 3.4.13.** *A finitely generated $R$-module $M$ is finitely presented if there's a surjective homomorphism $R^n \to M$, with a finitely generated kernel. In other words, $M$ is isomorphic to the cokernel of some homomorphism $\phi : R^m \to R^n$ for some $n$ and $m$.*

**Remark 3.4.14.** *Finitely presented modules are finitely generated, by definition. However, the converse isn't always true.*

We can make the following construction, similar to vector spaces. Let $\phi : R^m \to R^n$ be an $R$-module homomorphism. Let $u_1, \ldots, u_m \in R^m$ and $v_1, \ldots, v_n \in R^n$ be the standard bases. Then we can write $\phi(u_j) = \sum_{i=1}^n A_{ij} v_i$, for some matrix $A \in M_{m\times n}(R)$. That is, $\phi(r) = r \cdot A$ and so $\phi$ is given by multiplying by some matrix on the right. If $R$ is commutative, we could similarly write $\phi(r) = B \cdot r$, that is we could have represented $\phi$ by multiplying by some matrix on the left. This provides a bijection

$$\begin{aligned} M_{m\times n}(R) &\leftrightarrow \{R\text{-module homomorphisms } R^m \to R^n\} \\ A &\mapsto (r \mapsto r \cdot A). \end{aligned}$$

Recall from linear algebra, that if $F$ is a field and $F^n \cong F^m$, then $n = m$. The same is true for any commutative ring.

**Theorem 3.4.15.** *Any commutative ring contains a maximal ideal.*

We will assume Theorem 3.4.15 without proof to prove an upcoming result. It's equivalent to the axiom of choice.

**Theorem 3.4.16.** *Let $R$ be a commutative ring, and suppose that $R^n \cong R^m$ as $R$-modules. Then $n = m$.*

*Proof.* Let $I \subsetneq R$ be a maximal ideal so that $F = R/I$ is a field. For $M$ an $R$-module, we define $IM$ to be the submodule of $M$ generated by elements $i \cdot m$ for $I \in I$ and $m \in M$. That is,

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

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

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

If $R^n \cong R^m$ as $R$-modules, this then implies that

$$R^n/IR^n \cong R^m/IR^m$$

as $F$-modules. We claim that $R^k/IR^k \cong F^k$ are isomorphic $F$-modules. Indeed there is an isomorphism $\psi : R^k/IR^k \to F^k = (R/I)^k$ sending $(r_1, \ldots, r_k) + IR^k$ to $(r_1 + IR, \ldots, r_k + IR)$. From this, it follows that $F^n \cong F^m$ as $F$-modules. Since $F$ is a field, this implies that $n = m$. □

**Remark 3.4.17.** *This can fail for non-commutative rings. There exist rings $R$ such that, for example, $R^2 \cong R^1$ are isomorphic $R$-modules.*

## 3.5 Modules over Principal Ideal Domains

Our goal now is to prove Theorem 3.5.13 which generalises the classification theorem for finitely generated abelian groups. Since abelian groups are just $\mathbb{Z}$-modules. An outline for the proof is as follows.

1. Show $M$ 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 $\phi_A(x) := A \cdot x$.
2. Show that if $B = PAQ$, for invertible matrices $P$ and $Q$, then $\phi_A$ and $\phi_B$ have isomorphic cokernels.
3. Show that for any matrix $A$ over a principal ideal domain $R$, there are invertible matrices $P$ and $Q$ such that $PAQ$ is a rectangular diagonal matrix.
4. Combine these steps to prove Theorem 3.5.13.

**Lemma 3.5.1.** *Let $R$ be a principal ideal domain, and $N \leq R^n$ a submodule. Then $N \cong R^k$ for some $k \leq n$. Hence, $N$ is finitely generated and free.*

*Proof.* We induct 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 $\alpha \in R$. Since principal ideal domains are integral domains, $N \cong R^1$ if $\alpha \neq 0$ and $N \cong R^0$ if $\alpha = 0$. $R^{n-1}$. Now let $\pi_n : R^n \to R$ be the homomorphism sending $(r_1, \ldots, r_n)$ to $r_n$. Then $\ker(\pi_n) \cong \pi_n(N) \leq R$ is a submodule, so it is equal to $(\alpha)$ for some $\alpha \in R$. If $\alpha = 0$, this means $N \leq \ker(\pi_n) \cong R^{n-1}$, so by the induction hypothesis we're done. If $\alpha \neq 0$, we can pick some $\beta \in N$ such that $\pi_n(\beta) = \alpha$. By the induction hypothesis, we can choose some $x_1, \ldots, x_k \in N \cap \ker(\pi_n)$ which freely generate $N \cap \ker(\pi_n)$, with $k \leq n-1$. Let $S = \{x_1, \ldots, x_k, \beta\} \subseteq N$. We claim that $S$ freely generates $N$ to complete the proof. We first show $S$ generates $N$. Let $y \in N$. Then $\pi_n(y) = r\alpha$, for some $r \in R$. So $y - r\beta \in N \cap \ker(\pi_n)$ is spanned by the $x_1, \ldots, x_k$. Therefore, $y$ is generated by $S$. Now suppose $0 = r_1 \cdot x_1 + \ldots + r_k \cdot x_k + r' \cdot \beta$, for $r_i, r' \in R$. Then $0 = \pi_n(0) = r'\alpha$ which implies that $r' = 0$. Since $x_1, \ldots, x_k$ are linearly independent, we must have $r_1 = \ldots = r_k = 0$ and hence $S$ is linearly independent. □

**Corollary 3.5.2.** *Let $M$ be a finitely generated module over a principal ideal domain $R$. Then $M$ is finitely presented.*

*Proof.* There's a surjective homomorphism $\phi : R^n \to M$. As $\ker(\phi) \leq R^n$ is a submodule we can apply Lemma 3.5.1 to deduce that $\ker(\phi) \cong R^k$ for some $k$. Let $\psi : R^k \to R^n$ have image $\ker(\phi)$. Then $M \cong \operatorname{coker}(\psi)$. Hence, $M$ is finitely presented. □

This completes the first step of the proof of the theorem.

**Definition 3.5.3.** *Two matrices $A, B \in M_{m\times n}(R)$ are equivalent if there are invertible matrices $P \in \mathrm{GL}_n(R), Q \in \mathrm{GL}_m(R)$, such that $B = PAQ$.*

By considering the homomorphisms $\phi_A, \phi_B : R^m \to R^n$, this is equivalent to there being isomorphisms $f = \phi_P : R^n \to R^n$ and $g = \phi_Q : R^m \to R^m$ such that $f \circ \phi_B = \phi_A \circ g$. In other words, $A$ and $B$ are "the same after changing bases". One can check that this defines an equivalence relation.

**Proposition 3.5.4.** *Let $A, B \in M_{m\times n}(R)$ be matrices. 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 $f \circ \phi_B = \phi_A \circ g$. Note that $\operatorname{im}(\phi_A) = \operatorname{im}(\phi_A \circ g) = \operatorname{im}(f \circ \phi_B) = f(\operatorname{im}(\phi_B))$. Let $\psi : \operatorname{coker}(\phi_B) = R^n/\operatorname{im}(\phi_B) \to \operatorname{coker}(\phi_B) = R^n/\operatorname{im}(\phi_A)$ send $x + \operatorname{im}(\phi_B)$ to $f(x)+ \operatorname{im}(\phi_A)$. This is a well-defined homomorphism and is an isomorphism since $f$ is. □

**Remark 3.5.5.** *We already knew Proposition 3.5.4 for the case when $R = F$. As equivalent matrices have the same rank. Therefore, $\operatorname{coker}(\phi_A)$ and $\operatorname{coker}(\phi_B)$ have the same dimension and so they're isomorphic.*

This completes the second step of the proof of the theorem.

**Definition 3.5.6.** *An elementary matrix is an $n \times n$ matrix $A$ over $R$ of one of the following forms.*

1. *$A$ has $1$s along the diagonal, and $A_{ij} = c$ for some fixed $c \in R$ and $1 \leq i, j \leq n$. All other entries of $A$ are $0$.*
2. *Let $1 \leq i, j \leq n$ be distinct and fixed. Then $A$ is $I$ except in, $A_{ii} = A_{jj} = 0$ and $A_{ij} = A_{ji} = 1$.*
3. *Let $1 \leq i$ and $c \in R^\times$ a unit be fixed. Then $A$ is $I$ except in 1 entry: $A_{ii} = c$.*

Now let $A \in M_{m\times n}$ be any matrix. An elementary row/column operation on $A$ is given by replacing $A$ with $PA$ or $AP$ respectively, for $P$ an elementary matrix.

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

Note that all elementary matrices are invertible.

**Proposition 3.5.7.** *Let $A \in M_{m\times n}(R)$ be a matrix. If $B$ is obtained from $A$ by row and column operations, then $A$ and $B$ are equivalent matrices.*

In particular, if this holds, then $\operatorname{coker}(\phi_A) \cong \operatorname{coker}(\phi_B)$.

**Definition 3.5.8.** *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 $d_1 | \dots | d_r$.*

**Example 3.5.9.** *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 and so a unique factorisation domain, greatest common divisors exist. These are only defined up to units, so each time we take the greatest common divisor, we'll just fix some choices.

**Theorem 3.5.10.** *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 two 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 they're in the same column. Since invertible $2\times 2$ matrices can be extended to invertible $m \times m$ matrices, it suffices to prove the claim when $m = 2$. So instead we consider a vector $\begin{pmatrix} a \\ b \end{pmatrix} \in R^2$, and attempt to left-multiply by some matrix $P \in M_2(R)$ to get a vector containing $\gcd(a, b)$. As $R$ is a principal ideal domain we know that $(a, b) = (d) \subseteq R$ are equal ideals, for some $d \in R$. Therefore, $d = \gcd(a, b)$ 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 $xu + yv = 1$. Now let $P$ be the matrix $P := \begin{pmatrix} x & y \\ -v & u \end{pmatrix}$. Note that $\det(P) = xu + yv = 1$ and so $P$ is invertible. Now

$$\begin{aligned} P \cdot \begin{pmatrix} a \\ b \end{pmatrix} &= \begin{pmatrix} x & y \\ -v & u \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} \\ &= \begin{pmatrix} xa + yb \\ -va + ub \end{pmatrix} \\ &= \begin{pmatrix} d \\ -va + ub \end{pmatrix} \end{aligned}$$

and so we're done. If instead $A_{ij}$ and $A_{kl}$ are in the same row, then we can do the same but with the row matrix, $\begin{pmatrix} a & b \end{pmatrix}$ and multiplying with $P^\top$ on the right.

<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 $r$. For example, if $r = p_1 \dots p_k$, with all $p_i$ irreducible, then $\delta(r) = k$. 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 $\alpha_2$ and $\alpha_1$ aren't 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 $\delta$. Repeating this process it must eventually terminate with $A_{11} \mid A_{ij}$ whenever $i = 1$ or $j = 1$. This finishes the proof of Claim $2$.

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

$$A = \begin{pmatrix} A_11 & 0 & \dots & 0 \\ A_{21} & A_{22} & \dots & \\ \vdots & & \ddots & \end{pmatrix}.$$

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 apply the same process to $A'$, and repeating this process which does not change the equivalence class, 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 $\gcd(d_1, d_2) = x d_1 + y d_2$ for some $x, y \in R$. Moreover, $d_2 = \lambda \cdot \gcd(d_1, d_2)$ for some $\lambda \in R$. Consequently,

$$\begin{aligned} \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} &\overset{R_1 \mapsto R_1 + \lambda y R_2}{\sim} \begin{pmatrix} d_1 & y d_2 \\ 0 & d_2 \end{pmatrix} \\ &\overset{C_2 \mapsto x C_1 + C_2}{\sim} \begin{pmatrix} d_1 & \gcd(d_1, d_2) \\ 0 & d_2 \end{pmatrix} \\ &\overset{R_2 \mapsto R_2 - \lambda R_1}{\sim} \begin{pmatrix} d_1 & \gcd(d_1, d_2) \\ -\lambda d_1 & 0 \end{pmatrix} \\ &\overset{C_1 \leftrightarrow C_2}{\sim} \begin{pmatrix} \gcd(d_1, d_2) & d_1 \\ 0 & -\lambda d_1 \end{pmatrix} \\ &\overset{R_1 \mapsto R_1 + \lambda R_2}{\sim} \begin{pmatrix} \gcd(d_1, d_2) & 0 \\ 0 & -\lambda d_1 \end{pmatrix} \\ &\overset{R_2 \mapsto -R_2}{\sim} \begin{pmatrix} \gcd(d_1, d_2) & 0 \\ 0 & \lambda d_1 \end{pmatrix}. \end{aligned}$$

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 3.5.11.** *When we modified $A$ to an equivalent matrix, we mostly did this by applying row and column operations, but not always. In proving Claim $1$, we used an equivalence of matrices that wasn't necessarily of this form. If $R$ is a Euclidean domain, any $A \in \mathrm{GL}_n(R)$ is a product of elementary matrices. So if $R$ is a Euclidean domain, and $A \in M_{m \times n}(R)$ can be put into Smith normal form, just using elementary matrices.*

To complete the classification theorem, we just need the following,

**Proposition 3.5.12.** *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 $d_1, \dots, d_r \in R$ non-zero, and $d_1 | \dots | d_r$. Then there is an isomorphism of $R$-modules*

$$\operatorname{coker}(\phi_A) \cong \frac{R}{(d_1)} \oplus \dots \oplus \frac{R}{(d_r)} \oplus R^{m-r}$$

*where $\phi_A$ is given by left multiplication by $A$.*

*Proof.* Let $\psi : R^m \to \frac{R}{(d_1)} \oplus \ldots \oplus \frac{R}{(d_r)} \oplus R^{m-r}$ be the homomorphism sending $(x_1, \ldots, x_m)$ to

$$(x_1 + (d_1), \ldots, x_r + (d_r), x_{r+1}, \ldots, x_m).$$

Then $\operatorname{im}(\phi_A) = \ker(\psi) = (d_1) \oplus \ldots \oplus (d_r) \oplus \underbrace{0 \ldots \oplus 0}_{m-r}$. So by the first isomorphism theorem, the result follows. □

Let's put all of this together to prove the theorem.

**Theorem 3.5.13** (Classification of finitely generated modules over a PID)**.** *Let $R$ be a principal ideal domain (in particular, $R$ is commutative), and $M$ a finitely generated $R$-module. Then*

$$M \cong R^n \oplus R/(d_1) \oplus \ldots \oplus R/(d_r)$$

*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, $M$ is in fact finitely presented. So $M \cong \operatorname{coker}(\phi_A)$ for some $A \in M_{m\times n}(R)$.
2. If $A, B \in M_{m\times n}(R)$ are equivalent, then $\operatorname{coker}(\phi_A) \cong \operatorname{coker}(\phi_B)$.
3. $A$ is equivalent to a matrix in Smith normal form, that is, we can assume that

$$A = \begin{pmatrix} d_1 & & & & & \\ & \ddots & & & & \\ & & d_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}.$$

4. Therefore

$$M \cong R^n \oplus R/(d_1) \oplus \ldots \oplus R/(d_r)$$

and $d_1 | \ldots | d_r$ are all non-zero.

□

**Remark 3.5.14.** *There's another type of decomposition that we could consider, called prime decomposition,*

$$M \cong R^n \oplus \frac{R}{(P_1^{n_1})} \oplus \ldots \oplus \frac{R}{(p_k^{n_k})}$$

*for $n_i \in \mathbb{N}$ and $p_i \in R$ irreducible. This follows Theorem 3.5.13, along with the fact that if $d = p_1^{n_1} \cdot \ldots \cdot p_k^{n_k}$, then*

$$\frac{R}{(d)} \cong \frac{R}{(p_1^{n_1})} \oplus \ldots \oplus \frac{R}{(p_k^{n_k})}.$$

Note that if we know how to compute the greatest common divisors, our proof is entirely constructive.

**Example 3.5.15.** *Let $A$ be the abelian group generated by the elements a,b,c, with relations*

$$\begin{cases} 2a + 3b + c = 0 & (1) \\ a + 2b = 0 & (2) \\ 5a + 6b + 7c = 0. & (3) \end{cases}$$

*This means that*

$$A = \frac{\mathbb{Z}^3}{\left\langle \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix} \right\rangle} = \operatorname{coker}(\phi_X)$$

*where $X$ is the matrix*

$$X = \begin{pmatrix} 2 & 1 & 5 \\ 3 & 2 & 6 \\ 1 & 0 & 7 \end{pmatrix}$$

*We can put $X$ into Smith normal form in the following way*

$$\begin{aligned} \begin{pmatrix} 2 & 1 & 5 \\ 3 & 2 & 6 \\ 1 & 0 & 7 \end{pmatrix} &\sim \begin{pmatrix} 1 & 2 & 5 \\ 3 & 2 & 6 \\ 0 & 1 & 7 \end{pmatrix} \\ &\sim \begin{pmatrix} 1 & 2 & 5 \\ 0 & -1 & -4 \\ 0 & 1 & 7 \end{pmatrix} \\ &\sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & -4 \\ 0 & 1 & 7 \end{pmatrix} \\ &\sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 4 \\ 0 & 1 & 7 \end{pmatrix} \\ &\sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 4 \\ 0 & 0 & 3 \end{pmatrix} \\ &\sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}. \end{aligned}$$

*Therefore,*

$$A \cong \frac{\mathbb{Z}}{1\mathbb{Z}} \oplus \frac{\mathbb{Z}}{1\mathbb{Z}} \oplus \frac{\mathbb{Z}}{3\mathbb{Z}} \cong C_3.$$

## 3.6 Jordan Normal Form

In this section, we will study modules over polynomial rings using Theorem 3.5.13. Amongst other things, we will deduce the Jordan normal form.

Let $F$ be a field, $V$ a $F$-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 $v \in V$. We write $V_\alpha$ for $V$, when we view $V$ as a $F[X]$-module in this way. The structure of $V_\alpha$ as a $F[X]$-module will help us study the linear map $\alpha$.

**Lemma 3.6.1.** *If $V$ is a finite-dimensional $F$-vector space, then $V_\alpha$ is a finitely generated $F[X]$-module.*

*Proof.* Let $v_1, \ldots, v_n \in V$ be a basis for $V$ as a vector space. Then they generate $V_\alpha$ as an $F[X]$-module. □

**Example 3.6.2.** *Suppose $V_\alpha \cong F[X]/(X^r)$ as $F[X]$-modules. In particular, they are isomorphic as $F$-vector* *spaces. Note that $1, X, \ldots, X^{r-1} \in F[X]/\left(X^r\right)$ is a $F$-basis, and the action of multiplication by $X$ has matrix*

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

*with respect to this basis. So in $V_\alpha$ the linear map $\alpha$ is represented by the same matrix with respect to the corresponding basis.*

**Example 3.6.3.** *Suppose that $V_\alpha \cong F[X]/\left((X-\lambda)^r\right)$, as $F[X]$-modules, for some $\lambda \in F$. Consider the linear map $\beta : V \to V$ where $\beta = \alpha - \lambda I$. Then $V_\beta \cong F[Y]/\left(Y^r\right)$, for $Y = X - \lambda$. So by Example 3.6.2 there is a basis for $V$ such that $\beta$ 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 $\lambda$ Jordan block.*

**Example 3.6.4.** *Suppose $V_\alpha \cong F[X]/(f)$, for $f = a_0 + \ldots + a_{r-1}X^{r-1} + X^r \in F[X]$ some monic polynomial. Then $V_\alpha$ has an $F$-basis given by $1, X, \ldots, X^{r-1}$, in which $\alpha$ 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 3.6.5** (Rational canonical form)**.** *Let $F$ be any field, $V$ a finite-dimensional $F$-vector space, and $\alpha : V \to V$ some linear map. Then*

$$V_\alpha \cong \frac{F[X]}{(f_1)} \oplus \ldots \oplus \frac{F[X]}{(f_r)}$$

*with $f_1 | \ldots | f_r$ all non-zero polynomials over $F$. Therefore there is a basis for $V$ in which $\alpha$ is given by the* *block diagonal matrix*

$$\begin{pmatrix} C(f_1) & & 0 \\ & \ddots & \\ 0 & & C(f_r) \end{pmatrix}.$$

*Proof.* $F[X]$ is a principal ideal domain and so by Theorem 3.5.13 we must have

$$V_\alpha \cong \frac{F[X]}{(f_1)} \oplus \ldots \oplus \frac{F[X]}{(f_r)} \oplus F[X]^n$$

with $f_1 | \ldots | f_r$. Since $F[X]$ is not a finite-dimensional $F$-vector space, but $V_\alpha$ is, we must have $n = 0$. If we write $b_i$ for each $\deg(f_i)$, then with respect to the basis

$$X_1^0, X_1, \ldots, X_1^{b_1 - 1}, \ldots, X_r^0, \ldots, X_r^{b_r - 1},$$

where $X_j$ is the $X$ in the $j^{\text{th}}$ term in the direct sum $F[X]/(f_j)$, we see that $\alpha$ is in the required form. If we required all the $f_i$ to be monic, this is completely canonical and there is no ambiguity. □

We now focus in on the case $F = \mathbb{C}$.

**Lemma 3.6.6.** *The primes in $\mathbb{C}[X]$ are, up to associates, $X - \lambda$, for $\lambda \in \mathbb{C}$.*

*Proof.* Note any linear polynomial $X - \lambda \in \mathbb{C}[X]$ is prime. Now let $f \in \mathbb{C}[X]$ be prime. As $f$ is non-zero and not a unit it's non-constant. So by the fundamental theorem of algebra, $f$ has some root $\lambda \in \mathbb{C}$. So $(X - \lambda) \mid f$. Since $f$ is irreducible, we must have that $X - \lambda$ and $f$ are associates. □

Combining previous results of this section, we have the following.

**Theorem 3.6.7** (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*

$$V_\alpha \cong \frac{\mathbb{C}[X]}{((X - \lambda_1)^{n_1})} \oplus \ldots \oplus \frac{\mathbb{C}[X]}{((X - \lambda_r)^{n_r})}$$

*for some $n_i \in \mathbb{N}$ and $\lambda_i \in \mathbb{C}$. Furthermore there is a $\mathbb{C}$-basis for $V$ in which $\alpha$ 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.* For the first part, use Remark 3.5.14 on $V_\alpha$, and Lemma 3.6.6. Then we take a basis as in the proof of Theorem 3.6.5, and use the fact that

$$C((X - \lambda)^n) = J_n(\lambda).$$

□

**Lemma 3.6.8.** *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 $\alpha$ and $\beta$ 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 $V_\beta$, and similarly $\alpha(v) = X \cdot_{V_\alpha} v$. Then

$$\begin{aligned}\beta \circ \gamma(v) &= X \cdot_{V_\beta} \gamma(v) \\ &= \gamma\left(X \cdot_{V_\beta} v\right) \\ &= \gamma \circ \alpha(v)\end{aligned}$$

and so $\beta \circ \gamma = \gamma \circ \alpha$. That is, $\alpha$ and $\beta$ are conjugate.
($\Leftarrow$) Suppose $\alpha$ and $\beta$ are conjugate, so there is an isomorphism $\gamma : V \to V$ such that $\beta \circ \gamma = \gamma \circ \alpha$. We claim that $\gamma : V_\alpha \to V_\beta$ is an $F[X]$-module isomorphism. Indeed for $f \in F[X]$ and $v \in V$

$$\begin{aligned}\gamma\left(f \cdot V_\alpha\right) &= \gamma(f(\alpha)(v)) \\ & f(\beta)(\gamma(v))\end{aligned}$$

where we note that $\gamma \circ \alpha^i = \alpha^i \circ \beta$ for all $i \geq 0$. So $\gamma$ is an $F[X]$-module homomorphism and it is a bijection since it is an $F$-linear isomorphism. □

Reinterpreting this with Theorem 3.6.5 we obtain a classification result for square matrices over a field. It is a weaker classification than Theorem 3.6.7 form but it works over any field.

> **Corollary 3.6.9.** *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 $\sum_i \deg(f_i) = n$.*

*Proof.* Let $A$ be an $n \times n$ matrix over $F$, $V = F^n$ and $\alpha : V \to V$ send $v$ to $A \cdot v$. Then by Theorem 3.6.5 we know that

$$V_\alpha \cong \frac{F[X]}{(f_1)} \oplus \ldots \oplus \frac{F[X]}{(f_r)}$$

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, $\sum_i \deg(f_i) = n$. Furthermore, if $A$ and $B$ are conjugate matrices, then $V_\alpha$ and $V_\beta$ are isomorphic $F[X]$ modules, by Lemma 3.6.8, and so they define 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

$$V := \frac{F[X]}{(f_1)} \oplus \ldots \oplus \frac{F[X]}{(f_r)}$$

is a finite-dimensional $F$-vector space of dimension $n$, with a linear map $\alpha : V \to V$ sending $v$ to $X \cdot v$, which determines a matrix up to conjugation, by picking a basis. □

> **Example 3.6.10.** *Let's look at conjugacy classes in $M_{2\times 2}(F)$. This corresponds to classifying $F[X]$-modules of the form*
>
> $$\frac{F[X]}{(f_1)} \oplus \ldots \oplus \frac{F[X]}{(f_r)}$$
>
> *with $f_1 | \ldots | f_r$ a sequence of monic polynomials, with $\sum_i \deg(f_i) = 2$. 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 $\lambda \in F$. So since the companion matrix of $X - \lambda$ is the $1 \times 1$ matrix $(\lambda)$, 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*

$$C\left(f_1\right) = \begin{pmatrix} 0 & -b \\ 1 & -a \end{pmatrix}.$$

*If* $f_1$ *is irreducible, we can't 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 get a better 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*

$$\begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}.$$

# 4 Matrix Lie Groups

Our intention for the next few sections is 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 $R$, and is a non-commutative ring for $n \geq 2$. 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 defined as the unit group of $M_n(R)$. That is, $\mathrm{GL}_n(R) := M_n(R)^\times$.*

One can define the determinant function $\det : M_n(R) \to R$ in the usual sense given by linear algebra.

**Exercise 4.1.2.** *Show that $M$ is invertible if and only if $\det(M) \in R^\times$.*

**Proposition 4.1.3.** *The centre of $\mathrm{GL}_n(R)$ is the set of matrices of the form $\lambda I$, for some $\lambda \in R^\times$.*

*Proof.* 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

$$(E^{uv})_{ij} = \delta_{ij} + \delta_{ui}\delta_{vj} = \begin{cases} 1 & i = j \\ 1 & i = u, j = v \\ 0 & \text{otherwise.} \end{cases}$$

We must have that $AE^{uv} = E^{uv}A$, and so for all $1 \leq i, j \leq n$, we have

$$\begin{aligned} 0 &= (E^{uv}A - AE^{uv})_{ij} \\ &= E^{uv}_{ik}A_{kj} - A_{ik}E^{uv}_{kj} \\ &= \delta_{ui}A_{vj} - \delta_{vj}A_{ui}. \end{aligned}$$

For $u = i \neq v = j$, this tells us $A_{ii} = A_{jj}$ and so all diagonal entries are the same. If $i = j = u \neq v$, this tells us $A_{vj} = 0$ and so all other entries are $0$. So $A$ must be of the form $\lambda I$. By Exercise 4.1.2 we know that for $A \in \mathrm{GL}_n(R)$ we need $\lambda \in R^\times$. □

We now start to discuss some topological properties of $\mathrm{GL}_n(R)$ when $R = \mathbb{F}$. In particular, we endow $M_n(\mathbb{F})$ with the Euclidean topology and hence define $\mathrm{GL}_n(\mathbb{F}) \subseteq M_n(\mathbb{F})$ with the subspace topology. In doing so, topological properties can be thought considered in the natural sense with the subspace topology on $\mathbb{F}^{n^2}$.

**Definition 4.1.4.** *Let $(A_m) \in M_n(\mathbb{F})$ be a sequence of matrices. Then $A_m$ converges to a matrix $A$ if each entry of $A_m$ converges to the corresponding matrix of $A$.*

**Proposition 4.1.5.** *For any $n$ the set $\mathrm{GL}_n(\mathbb{R})$ is not path-connected.*

*Proof.* For $n = 1$, we can show that $-1$ and $1$ are not connected. Suppose they were connected by $\gamma : [0, 1] \to \mathrm{GL}_1(\mathbb{R}) = \mathbb{R}^\times$. Then by the intermediate value theorem there would exist a $t \in [0, 1]$ such that $\gamma(t) = 0$, but $0 \notin \mathbb{R}^\times$ and so we get a contradiction. For $n \geq 2$ we will show there is no path in $\mathrm{GL}_n(\mathbb{R})$ between

$$A := \begin{pmatrix} -1 & \cdots & & 0 \\ \vdots & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix}$$

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 is such that $(\det \circ\gamma)(0) = 1$ and $(\det \circ\gamma)(1) = 1$. So by the intermediate value theorem, there would be some $t$ such that $(\det \circ\gamma)(t) = 0$. But $\gamma(t)$ was an invertible matrix, so this is a contradiction. □

**Proposition 4.1.6.** *The set $\mathrm{GL}_n(\mathbb{C})$ is always 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

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

As $\lambda \neq 0$ 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 $t$, $\gamma(0) = A$ and $\gamma(1) = I$. The case where $B$ is the direct sum of Jordan blocks is similar. □

## 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 both continuous. Moreover, $G$ is Hausdorff.*

**Definition 4.2.2.** *A homomorphism of topological groups is a continuous group homomorphism.*

An isomorphism is an isomorphism of groups which is a homomorphism.

**Lemma 4.2.3.** *The composition of topological group homomorphism 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 result follows. □

**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 set $\mathrm{GL}_n(\mathbb{F})$ 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 hence is 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 also a continuous function of $M$, their quotient, $M^{-1}$, is a continuous function of $M$. □

Note that 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 4.2.6.** *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 also continuous and so $H$ 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 $n$.*

Since $\mathrm{GL}_n(\mathbb{F})$ is Hausdorff it is metrizable, and so is any subspace. Hence, 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 $(A_m) \in G$ either converges to a matrix $A$ that is invertible or converges to a matrix that is not invertible. In particular, this means that $G$ may not be a closed subset of $M_n(\mathbb{F})$.*

**Example 4.3.3.** $(\mathbb{R}^\times, \cdot) = \mathrm{GL}_1(\mathbb{R})$ *is a matrix Lie group. 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 $n$. We will see later on that it is in fact 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* $\mathrm{O}(n) := \left\{M \in \mathrm{GL}(\mathbb{R}) : M^\top M = I\right\} \leq \mathrm{GL}_n(\mathbb{R})$.
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.** *The map $\phi : \mathbb{R}/\mathbb{Z} \to \mathrm{U}(1)$ given by $t \mapsto e^{2i\pi t}$ is an isomorphism of topological groups.*

**Lemma 4.3.6.** *Let $G$ be a topological group, and let $H \leq G$ be any subgroup. 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 $\{g_i\}_i$ and $\{h_i\}_i$ in $H$, such that $g_i \to g$ and $h_i \to h$. As multiplication is continuous we know that $g_i \cdot h_i \to g \cdot h$. Since each $g_i \cdot h_i \in H$ the limit of this sequence, $g \cdot h$, lies in $\bar{H}$. Similarly each $g_i^{-1} \in H$ so $g^{-1} \in \bar{H}$ too. Therefore, $\bar{H}$ is also a subgroup. □

Consequently, taking the closure of any subgroup $G \leq \mathrm{GL}_n(\mathbb{F})$ generates a matrix Lie group.

**Example 4.3.7.** *The closure of $\mathbb{Q}^\times$ in $\mathbb{R}^\times$ is the entirety of $\mathbb{R}^\times$.*

**Proposition 4.3.8.** *the set $\mathrm{GL}_n(\mathbb{R})$ is a closed subgroup in $\mathrm{GL}_n(\mathbb{C})$.*

*Proof.* Consider the continuous map $f : \mathrm{GL}_n(\mathbb{C}) \to M_n(\mathbb{R})$ sending $A$ to $\mathrm{Im}(A)$. As $\mathrm{GL}_n(\mathbb{R}) = f^{-1}(\{0\})$ we know it is a closed subset. Hence, as it is clearly a subgroup we deduce it is a closed subgroup. □

**Proposition 4.3.9.** $\mathrm{GL}_n(\mathbb{C})$ *is isomorphic to a closed subgroup in* $\mathrm{GL}_{2n}(\mathbb{R})$.

*Proof.* We will need to define an injective group homomorphism $\phi : \mathrm{GL}_n(\mathbb{C}) \hookrightarrow \mathrm{GL}_{2n}(\mathbb{R})$ which is a homeomorphism onto its image. Then $\phi$ is an isomorphism of topological groups between $\mathrm{GL}_n(\mathbb{C})$ and $\mathrm{Im}(\phi)$. For $M \in \mathrm{GL}_n(\mathbb{C})$, we define $\phi(M)$ to be the block matrix

$$\begin{pmatrix} \mathrm{Re}(M) & \mathrm{Im}(M) \\ -\mathrm{Im}(M) & \mathrm{Re}(M) \end{pmatrix} \in \mathrm{GL}_{2n}(\mathbb{R}).$$

It is immediate that $\phi$ is continuous, injective, and $\phi(I) = I$. Let's check that it is a homomorphism. Let $M, N \in \mathrm{GL}_n(\mathbb{C})$ with $M = A + iB$ and $N = C + iD$ for $A, B, C, D \in M_n(\mathbb{R})$. Note that we cannot assume that $A, B, C, D \in \mathrm{GL}(\mathbb{R})$. 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 $\phi(M)\phi(N) = \phi(MN)$. It remains to check that $\mathrm{Im}(\phi) \subseteq \mathrm{GL}_{2n}(\mathbb{R})$ is closed. Let

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \in \mathrm{GL}_n(\mathbb{R}).$$

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 $M = A + iB \in M_n(\mathbb{C})$. If $M$ is invertible then $\phi(M) = P$ and so $P \in \mathrm{Im}(\phi)$. Thus we need to show that $M$ is invertible. Multiplying $PJ = JP$ on the left and right by $P^{-1}$ we get that $P^{-1}J = JP^{-1}$. 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 $M$. Therefore, $\mathrm{Im}(\phi) = \{M \in \mathrm{GL}_{2n}(\mathbb{R}) : MJ = JM\}$.
As $M \mapsto MJ - JM$ is continuous, and we have just shown the pre-image of $\{0\}$ under this map is $\mathrm{Im}(\phi)$ we conclude that $\mathrm{Im}\phi$ is closed as $\{0\}$ is closed. □

**Exercise 4.3.10.** *If $M \in \mathrm{GL}_n(\mathbb{C})$ with $M = A + iB$ for $A, B \in M_n(\mathbb{R})$. Show that it is not necessarily true that $A, B \in \mathrm{GL}_n(\mathbb{R})$.*

Proposition 4.3.9 tells us that a topological group $G$ is a matrix Lie group with respect to $\mathbb{R}$ if and only if it is a topological group with respect to $\mathbb{C}$.

**Lemma 4.3.11.** *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ be a closed subgroup, and $H \leq G$ also a closed subgroup. Then $H \leq \mathrm{GL}_n(\mathbb{F})$ is a closed subgroup.*

*Proof.* Subgroups of subgroups are subgroups, and so $H \leq \mathrm{GL}_n(\mathbb{F})$ is also a subgroup. Similarly, closed subsets of closed subsets are closed subsets, and so $H \leq \mathrm{GL}_n(\mathbb{R})$ is also closed and thus a closed subgroup. □

**Corollary 4.3.12.**

1. *Any closed subgroup $G \leq \mathrm{GL}_n(\mathbb{R})$ is a closed subgroup on $\mathrm{GL}_n(\mathbb{C})$.*
2. *Any closed subgroup $H \leq \mathrm{GL}_m(\mathbb{C})$ is isomorphic to a closed subgroup in $\mathrm{GL}_{2m}(\mathbb{R})$.*

*Proof.*

1. Let $G \leq \mathrm{GL}_n(\mathbb{R})$ be a closed subgroup. Since $\mathrm{GL}_n(\mathbb{R}) \leq \mathrm{GL}_n(\mathbb{R})$ is a closed subgroup, by Lemma 4.3.11 we have that $G \leq \mathrm{GL}_n(\mathbb{C})$ is a closed subgroup.
2. Let $H \leq \mathrm{GL}_m(\mathbb{C})$ be a closed subgroup and $\phi : \mathrm{GL}_m(\mathbb{C}) \to K$ an isomorphism, where $K \leq \mathrm{GL}_{2m}(\mathbb{R})$ is some closed subgroup. Then $H \cong \phi(H)$ and $\phi(H) \leq K$ is a closed subgroup. So by Lemma 4.3.11 we have that $H \cong \phi(H) \leq \mathrm{GL}_{2m}(\mathbb{R})$ where $\phi(H) \leq \mathrm{GL}_{2m}(\mathbb{R})$ is a closed subgroup.

□

**Corollary 4.3.13.** *A closed subgroup of a matrix Lie group is itself a matrix Lie group.*

*Proof.* Suppose that $G \leq \mathrm{GL}_n(\mathbb{F})$ be a matrix Lie group and let $H \leq G$ be a closed subgroup. Then by Lemma 4.3.11 we know that $H \leq \mathrm{GL}_n(\mathbb{F})$ is a closed subgroup 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 the property that there exists a constant $C$ such that for any $A \in G$ the absolute value of its entries are at most $C$.

**Example 4.3.14.**

1. *Consider the set*

$$\begin{aligned} U(1) &:= \left\{(z) \in M_1(\mathbb{C}) : (z)^\dagger (z) = I = (1)\right\} \\ &= \{z \in \mathbb{C} : \bar{z}z = 1\} \\ &= \{z \in \mathbb{C} : |z| = 1\}. \end{aligned}$$

   *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 also closed. Hence, it is also compact.*

2. *The $\mathrm{GL}_n(\mathbb{F}) \subseteq \mathbb{F}^{n\times n}$ 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 $\infty$ 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 $A \in M_n(F)$, its exponential, written $e^A$ or $\exp(A)$, is*

$$e^A := \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

**Remark 4.4.2.** *For $n = 1$ this coincides with the usual exponential.*

**Theorem 4.4.3.** *The power series $\exp(A)$ converges for all $A$. Moreover, we have the following.*

1. *The power series for $\exp : M_n(\mathbb{F}) \to M_n(\mathbb{F})$ converges absolutely and uniformly on compact subsets of $M_n(\mathbb{F})$.*
2. *The power series for its partial derivatives, in the components of $A$, of $\exp$ all converge absolutely and uniformly on compact subsets of $M_n(\mathbb{F})$.*

*In particular, $\exp : M_n(\mathbb{F}) \to M_n(\mathbb{F})$ 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 $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, note that*

$$A^n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$$

*so that*

$$e^A = \begin{pmatrix} \sum_{k=0}^{\infty} \frac{1}{k!} & \sum_{k=0}^{\infty} \frac{k}{k!} \\ 0 & \sum_{k=0}^{\infty} \frac{1}{k!} \end{pmatrix} = \begin{pmatrix} e & e \\ 0 & e \end{pmatrix}.$$

There is also a logarithm for matrices defined similarly.

**Definition 4.4.5.** *For $A \in M_n(\mathbb{F})$ its logarithm, $\log(A)$, is*

$$\log(A) := \sum_{k=1}^{\infty} \frac{(-1)^{k+1}(A-I)^k}{k}.$$

Unfortunately, this does not converge for all $A \in M_n(\mathbb{F})$. However, when $A$ is close enough to $I$, then all the same properties as we had for the exponential hold.

**Theorem 4.4.6.** *There is an open neighbourhood of $I$ in $M_n(\mathbb{F})$ such that the power series for $\log$ converges for all $A \in U$. Moreover, the following hold.*

1. *The power series for $\log$ converges absolutely on $U$, and uniformly on compact subsets of $U$.*
2. *The power series for its partial derivatives, in the components of $A$, of $\log$ converges absolutely on $U$, and uniformly on compact subsets of $U$.*

*In particular, this implies that $\log : U \to M_n(\mathbb{F})$ is continuously differentiable.*

**Example 4.4.7.**

1. *When $n = 1$ we have the usual power series for $\log$, which converges for $|z-1| < 1$. Similarly, for $n \geq 1$ the set $U = \left\{M : |M_{ij} - \delta_{ij}| < \frac{1}{n}, \text{ for all } i, j\right\}$ ensures the convergence of $\log$.*
2. *Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Note that $(A-I)^k = 0$ for $k \geq 2$. Hence, the series $\log(A)$ 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 both $\exp$ and $\log$ converge they are inverses of each other.

**Proposition 4.4.8.** *Suppose the power series for $\log(A)$ is absolutely convergent, then $\exp(\log(A)) = A$. Similarly, if the power series for $\log(\exp(B))$ is convergent, then $\log(\exp(B)) = B$.*

The proof amounts to writing the composition of $\exp(\log(A))$ as a double power series in $A$ and showing that this power series is just $A$.

**Proposition 4.4.9.** *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 4.4.10.** *The requirement that $AB = BA$ in Proposition 4.4.9 is necessary.*

**Corollary 4.4.11.** *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 4.4.9 to deduce that

$$e^A e^{-A} = e^{A-A} = e^0 = I.$$

□

Hence, we can think of $\exp$ as a map $M_n(\mathbb{F}) \to \mathrm{GL}_n(\mathbb{F})$.

**Example 4.4.12.** *As $A$ and $A^k$ always commute, we deduce that $A$ and $e^A$ commute. Hence,*

$$e^{Ae^A} = e^A e^{e^A}.$$

**Lemma 4.4.13.** *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 $\left(ABA^{-1}\right)^k = AB^k A^{-1}$. So

$$\begin{aligned} e^{ABA^{-1}} &= \sum_{k=0}^{\infty} \frac{\left(ABA^{-1}\right)^k}{k!} \\ &= \sum_{k=0}^{\infty} \frac{AB^k A^{-1}}{k!} \\ &= Ae^B A^{-1}. \end{aligned}$$

□

**Lemma 4.4.14.** *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 $k$. Hence,

$$\begin{aligned} e^{A^\top} &= \sum_{k=0}^{\infty} \frac{\left(A^\top\right)^k}{k!} \\ &= \left(\sum_{k=0}^{\infty} \frac{A^k}{k!}\right)^\top \\ &= \left(e^A\right)^\top. \end{aligned}$$

□

**Example 4.4.15.** *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 see a relationship between the trace, determinant and eigenvalues, namely*

$$\det\left(e^A\right) = e^{\lambda_1} \dots e^{\lambda_n} = e^{\operatorname{tr}(A)}.$$

**Lemma 4.4.16.** *For $A \in M_n(\mathbb{F})$ it follows that*

$$\det\left(e^A\right) = e^{\operatorname{tr}(A)}.$$

*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

$$\begin{aligned} e^{\operatorname{tr}(A)} &= e^{\operatorname{tr}\left(PBP^{-1}\right)} \\ &= e^{\operatorname{tr}(B)}. \end{aligned}$$

Using Example 4.4.15 we know that $\det\left(e^B\right) = e^{\operatorname{tr}(B)}$ and so $\det\left(e^A\right) = e^{\operatorname{tr}(A)}$. □

Recall that $\exp : M_n(\mathbb{F}) \to \mathbb{M}_n(\mathbb{F})$ is differentiable. Thus for any $A \in M_n(\mathbb{F})$ the function $\gamma_A(t) : \mathbb{R} \to \mathrm{GL}_n(\mathbb{F})$ given by $\gamma_A(t) = e^{tA}$ is continuously differentiable.

**Lemma 4.4.17.** *For any $A \in M_n(\mathbb{F})$, the derivative of $\gamma_A$ at $0$ is $\dot{\gamma}_A(0) = A$.*

*Proof.* Proceeding from first principles we get that

$$\begin{aligned}\dot{\gamma}_A(0) &= \lim_{t\to 0} \frac{e^{tA} - e^0}{t} \\ &= \lim_{t\to 0} \frac{\sum_{k=0}^{\infty} \frac{t^k A^k}{k!} - I}{t} \\ &= \lim_{t\to 0} \sum_{k=1}^{\infty} \frac{t^{k-1}A^k}{k!} \\ &= A.\end{aligned}$$

□

**Lemma 4.4.18.** *For any $A \in M_n(\mathbb{F})$ 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, $\gamma_A$ is also continuous. Let $t, s \in \mathbb{R}$ 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 the fact that $sA$ and $tA$ commute. Therefore, $\gamma_A$ is also a homomorphism. □

**Corollary 4.4.19.** *For any $t_0 \in \mathbb{R}$ the derivative of $\gamma_A$ at $t_0$ is $\dot{\gamma}_A(t_0) = A\gamma_A(t_0)$.*

*Proof.* Since $\gamma_A$ is a homomorphism,

$$\gamma_A(t) = \gamma_A(t - t_0)\gamma_A(t_0).$$

Differentiating both sides at $t = t_0$ and using the $t_0 = 0$ case, it follows that

$$\dot{\gamma}_A(t_0) = \dot{\gamma}_A(0)\gamma_A(t_0) = Ae^{t_0 A}.$$

□

Recall that the derivative of $\exp$ at $0$ can be thought of as a linear map $d\exp_0 : M_n(\mathbb{F}) \to M_n(\mathbb{F})$.

**Corollary 4.4.20.** *The map $d\exp_0 : M_n(\mathbb{F}) \to M_n(\mathbb{F})$ is the identity map.*

*Proof.* By the chain rule, the derivative of $\gamma_A$ at $t = 0$ is $d\exp_0(A)$. Therefore, as $A = d\exp_0(A)$ for all $A \in M_n(\mathbb{F})$ it follows that $d\exp_0$ is the identity map. □

**Proposition 4.4.21.** *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.* Observe that $\exp$ sends $0$ to $I$. Moreover, by Corollary 4.4.20 we know that $\exp$ is differentiable at $0$ with a derivative that is an invertible linear map. Therefore, we conclude by applying Theorem 5.2.2. □

**Example 4.4.22.** *For $n = 1$ we can take $U = (-1, 1)$ and see that*

$$\exp|_U : (-1, 1) \to \left(e^{-1}, e\right)$$

*is a homomorphism, with inverse being* $\log$*, which is continuously differentiable.*

**Proposition 4.4.23.** *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 $\phi$ 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 $\phi$ is also a homomorphism. As any continuous homomorphism $(\mathbb{R}, +) \to (\mathbb{R}, +)$ is of the form $\lambda t$ 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 4.4.23 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 4.4.24.** *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.* Choose open neighbourhoods $U$ of $0$ 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 $\gamma$, there is some $\delta > 0$ such that $\gamma([-\delta, \delta]) \in V$. Let $\beta : [-\delta, \delta] \to U$ be given by $t \mapsto \exp^{-1}|_U(\gamma(t))$.

Claim 1: $\underline{\beta(r\delta) = r\beta(\delta) \text{ for all } r \in [-1, 1].}$

*Proof.* Let $r = \frac{p}{q} \in [0, 1]$ be rational. Then

$$\begin{aligned}e^{\beta(\delta)} &= \gamma(\delta)\\ &\overset{(1)}{=} \gamma\left(\frac{\delta}{q}\right)^q\\ &= \left(\exp\left(\beta\left(\frac{\delta}{q}\right)\right)\right)^q\\ &= \exp\left(q\beta\left(\frac{\delta}{q}\right)\right)\end{aligned}$$

where in $(1)$ we used the fact that $\gamma$ is a homomorphism. As this equality holds in $V$ we deduce that $q\beta\left(\frac{\delta}{q}\right) = \beta(\delta)$. Similarly,

$$\begin{aligned}e^{\beta\left(\frac{p\delta}{q}\right)} &= \gamma\left(\frac{p\delta}{q}\right)\\ &= \gamma\left(\frac{\delta}{q}\right)^p\\ &= \left(\exp\left(\beta\left(\frac{\delta}{q}\right)\right)\right)^p\\ &= \exp\left(p\frac{\beta(\delta)}{q}\right)\end{aligned}$$

and so

$$\beta\left(\frac{p\delta}{q}\right) = \frac{p}{q}\beta(\delta).$$

This can be extended to any rational $r \in [-1,1]$. For general $r \in [-1,1]$ we choose a sequence of rational points $r_i = \frac{p_i}{q_i} \in [-1,1] \cap \mathbb{Q}$ converging to $r$ and use the continuity of $\beta$ to deduce that $\beta(r_i\delta) \to \beta(r\delta)$. As $\beta(r_i\delta) = r_i\beta(\delta)$ and $r_i\beta(\delta) \to r\beta(\delta)$, 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 4.4.24 implies that $\gamma$ is continuously differentiable.

> **Corollary 4.4.25.** *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ be a closed subgroup. Then any continuous homomorphism $\gamma : \mathbb{R} \to G$ is of the form $\gamma_A$ 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 4.4.24 to conclude. □

> **Lemma 4.4.26.** *For $B \in M_n(\mathbb{R})$, we have that $e^B \in \mathrm{SL}_n(\mathbb{R})$ if and only if $\mathrm{tr}(B) = 0$.*

*Proof.* Recall that by Lemma 4.4.16 we have that $\det\left(e^B\right) = e^{\mathrm{tr}(B)}$. So since $\mathrm{tr}(B) \in \mathbb{R}$ it follows that $\det\left(e^B\right) = 1$ if and only if $\mathrm{tr}(B) = 0$. □

Let $\gamma : \mathbb{R} \to \mathrm{SL}_n(\mathbb{R})$ be 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 $t \in \mathbb{R}$. Therefore, by Lemma 4.4.26 this happens 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 a closed subgroup 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.** *Note that the $\gamma$ in Definition 4.5.1 need not be a homomorphism. In particular, $\gamma(0) = A$ and a group homomorphism needs to satisfy $\gamma(0) = I$.*

**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*

$$T_A G = A \cdot T_I G = T_I G \cdot A.$$

*Proof.* Let $M \in T_I G$ 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 and is such that $\gamma(0) = A$. Therefore,

$$\frac{d}{dt}(A\gamma(t))|_{t=0} = A\dot{\gamma}(0) \in T_A G.$$

Hence, $A \cdot T_I G \subset T_A G$. A similar argument with with $A^{-1}$ shows that

$$A^{-1} \cdot T_A G \subset T_I G$$

which implies that $T_A G = A \cdot T_I G$. Similarly, one shows that $T_A G = T_I G \cdot A$. □

**Proposition 4.5.4.** *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ be a closed subgroup. Then for any $A$, the set $T_A G$ is a real vector space.*

*Proof.* For $A = I$, let $M, N \in T_I G$ 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 $\lambda M \in T_I G$. 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

$$\dot{\delta}(0) = \dot{\alpha}(0)\beta(0) + \alpha(0)\dot{\beta}(0) = MI + IN = M + N \in T_I G.$$

Therefore, $T_I G$ is a real vector space. We can generalise this $T_A G$ by using Proposition 4.5.3. □

**Definition 4.5.5.** *The dimension of $G \leq \mathrm{GL}_n(\mathbb{F})$ a closed subgroup is given by the dimension of $T_A G$ as a real vector space.*

Recall the commutator of matrices $A$ and $B$ is $[A, B] = AB - BA$.

**Proposition 4.5.6.** *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ be a closed subgroup, and let $M, N \in T_I G$. Then $[M, N] \in T_I G$.*

*Proof.* Let $\alpha, \beta : \mathbb{R} \to I$ 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 $\dot{\delta}_s(0) \in T_I G$. By the product rule we have that

$$\dot{\delta}_s(0) = \alpha(s)\dot{\beta}(0)\alpha(s)^{-1} = \alpha(s) \cdot N \cdot \alpha(s)^{-1}.$$

Letting $s$ vary we observe that $\dot{\delta}_s(0) : \mathbb{R} \to T_I G$ defines a continuously differentiable path. So its derivative at $0$ must also lie in $T_I G$, that is

$$\frac{d}{ds}\dot{\delta}_s(0)\big|_{s=0} \in T_I G.$$

Therefore as

$$\begin{aligned}
\frac{d}{ds}\dot{\delta}_s(0)\big|_{s=0} &= \left(\frac{d}{ds}\alpha(s)\big|_{s=0}\right) \cdot N \cdot \alpha(0)^{-1} + \alpha(0) \cdot N \cdot \left(\frac{d}{ds}\alpha(s)^{-1}\big|_{s=0}\right) \\
&= \dot{\alpha}(0) \cdot N \cdot \alpha(0)^{-1} - \alpha(0) \cdot N \cdot \left(-\dot{\alpha}(0)\alpha(0)^{-2}\right) \\
&= M \cdot N \cdot I - I \cdot N \cdot M \\
&= [M, N]
\end{aligned}$$

we deduce that $[M, N] \in T_I G$. □

**Proposition 4.5.7.** *For $A, B, C \in M_n(\mathbb{F})$ and the commutator bracket $[\cdot,\cdot]$, 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 following 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.*

A homomorphism of Lie algebras from $V$ to $W$ consists of a linear map $f; V \to W$ such that

$$f\left([A,B]_V\right) = [f(A), f(B)]_W$$

for all $A, B \in V$.

**Definition 4.5.10.** *For a closed subgroup $G \leq \mathrm{GL}_n(\mathbb{F})$, its Lie algebra is the vector $(T_I G, [\cdot,\cdot])$ where $[\cdot,\cdot]$ is the commutator of matrices.*

**Remark 4.5.11.**

1. *Definition 4.5.10 makes sense as $T_I G$ is a real-vector space by Proposition 4.5.4, $[\cdot,\cdot]$ 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 $\mathfrak{su}(2)$.*

**Proposition 4.5.12.** *Let $\gamma : \mathbb{R} \to \mathrm{GL}_n(\mathbb{F})$ be a continuously differentiable function with $\gamma(0) = I$, and let $t \in \mathbb{R}$. Then*
$$e^{t\dot{\gamma}(0)} = \lim_{n\to\infty} \gamma\left(\frac{t}{n}\right)^n.$$
*In particular, the limit exists.*

*Proof.* 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 $d\exp_0 : M_n(\mathbb{F}) \to M_n(\mathbb{F})$ is the identity map, so by the chain rule it follows that $\dot{\beta}(0) = \dot{\gamma}(0)$.

Therefore,

$$\begin{aligned}\lim_{n\to\infty}\gamma\left(\frac{t}{n}\right)^n &\overset{(1)}{=} \lim_{n\to\infty}\left(\exp\left(\beta\left(\frac{t}{n}\right)\right)\right)^n\\ &\overset{(2)}{=} \lim_{n\to\infty}\left(\exp\left(\frac{t}{n}\dot\gamma(0)+o\left(\frac{t}{n}\right)\right)\right)^n\\ &= \lim_{n\to\infty}\exp\left(t\dot\gamma(0)+no\left(\frac{t}{n}\right)\right)\\ &= e^{t\dot\gamma(0)}\end{aligned}$$

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 $\gamma$. □

**Example 4.5.13.** *For $n = 1$ and $\gamma = 1 + t$ the result of Proposition 4.5.12 tells us that*

$$e^t = \lim_{n\to\infty}\left(1+\frac{t}{n}\right)^n$$

*as expected.*

**Proposition 4.5.14.** *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ be a closed subgroup, 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

$$e^M = \lim_{n\to\infty}\gamma\left(\frac{t}{n}\right)^n$$

which exists. Moreover, as $\gamma\left(\frac{t}{n}\right)^n \in G$ and $G \subseteq \mathrm{GL}_n(\mathbb{F})$ is a closed subset we deduce that $e^M \in G$. □

**Corollary 4.5.15.** *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 $A \in M_n(\mathbb{F})$. As $\gamma_A$ is always 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 $t \in \mathbb{R}$. Therefore, by Proposition 4.5.14 $\gamma_A(t) = e^{tA} \in G$ for all $t \in \mathbb{R}$. Moreover, we showed earlier that $\gamma_A : \mathbb{R} \to \mathrm{GL}_n(\mathbb{F})$ defines a continuous homomorphism. Since $\gamma_A$ has an image in $G$ when $A \in \mathfrak{g}$ it follows that $\gamma_A : \mathbb{R} \to G$ is a continuous homomorphism. □

**Remark 4.5.16.**

1. *From Corollary 4.5.15 we see that there is a bijection*

   $$\mathfrak{g} \leftrightarrow \{\textit{continuous homomorphism } \mathbb{R} \to G\}$$

   *that sends $A \in \mathfrak{g}$ to $\gamma_A$ in one direction, and $\gamma$ to $\dot\gamma(0)$ in the other.*

2. *Let $\gamma, \delta : \mathbb{R} \to G$ be continuous homomorphism with $G \leq \mathrm{GL}_n(\mathbb{F})$ a closed subgroup. 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 $t \mapsto \exp\left(t\dot\gamma(0) + \dot\delta(0)\right)$. Recall 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 $\eta$ is not necessarily a homomorphism as $G$ may not be abelian. Therefore, we cannot deduce that $\xi$ and $\eta$ are equal, but by using Proposition 4.5.14 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 4.5.17.** *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ be a closed subgroup. Then there are open neighbourhoods $U$ of $0$ 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 a modification of $\exp$ which is given by $(X, Y) \mapsto e^{X+Y}$. However, $\exp$ and $\widetilde{\exp}$ agree on $\mathfrak{g}$. Hence, it suffices to prove the statement of the theorem for $\widetilde{\exp}$.
Claim 1: The derivative of $\widetilde{\exp}$ at $0$ denoted $d\widetilde{\exp}_0 : M_n(\mathbb{F}) \to M_n(\mathbb{F})$, is the identity.
*Proof.*Let $(X, Y) \in \mathfrak{g} \oplus W$. Let $\gamma(t) = e^{tX}e^{tY} = \widetilde{\exp}(X, Y)$. Then

$$\begin{aligned}\dot{\gamma}(0) &= \left(\frac{d}{dt}e^{tX}|_{t=0}\right)e^{0\cdot Y} + e^{0\cdot X}\left(\frac{d}{dt}e^{tY}|_{t=0}\right)\\ &= X\cdot I + I \cdot Y\\ &= X + Y.\end{aligned}$$

By the chain rule $\dot{\gamma}(0) = d\widetilde{\exp}_0(X, Y)$ and so $d\widetilde{\exp}_0(X, Y) = X + Y$.
Claim 2: There is a neighbourhood $Z$ of $0$ in $\tilde{U} \subseteq M_n(\mathbb{F})$ such that $\widetilde{\exp}^{-1}(G) \cap Z = \mathfrak{g} \cap Z$.
*Proof.* It suffices to show that for $(X, Y) \in Z \subseteq \mathfrak{g} \oplus W$ we have that $\widetilde{\exp}(X, Y) \in G$ if and only if $Y = 0$. Suppose that this is not true, so that for every neighbourhood of $0$ 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 $(X_i, Y_i) \in \mathfrak{g} \oplus W$ converging to $0$ such that $Y_i \neq 0$ and $\widetilde{\exp}(X_i, Y_i) \in G$ for all $i$. 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, that is $\frac{Y_i}{|Y_i|} \to Y \in W$ 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 $i$. Note that $e^{m_iY_i} = \left(e^{Y_i}\right)^m_i \in G$. Moreover, as

$$m_iY_i = (m_i|Y_i|)\frac{Y_i}{|Y_i|} \to tY$$

we deduce that $e^{m_iY_i} \to e^{tY}$. As $G \subset \mathrm{GL}_n(\mathbb{F})$ is closed it follows that $e^{tY} \in G$ for all $t$ 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 Claim 2. Then by Claim 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 $0$, we see that

$$\widetilde{\exp}|_U : U \to \mathrm{im}\,(\widetilde{\exp}|_U)$$

is a homeomorphism. By Claim 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 $I$, and hence $G \cap \mathrm{im}\,(\widetilde{\exp}|_Z)$ is an open neighbourhood of $I$ in $G$. □

**Corollary 4.5.18.** *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ be a closed subgroup. 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 4.5.17 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 itself homeomorphic to an open subset of $\mathbb{R}^d$, to an open neighbourhood of $A$ in $G$. □

**Corollary 4.5.19.** *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ be a closed subgroup. Then $G$ is a discrete topological space if and only if its dimension is $0$.*

*Proof.* ($\Leftarrow$). If the dimension of $G$ is $0$, then every $A \in G$ has a neighbourhood homeomorphic to $\mathbb{R}^0$, which is just a point. Hence, all points in $G$ are open which implies 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 $G$, this implies that $G$ is discrete as $U$ is discrete. Hence, $d = 0$. □

**Example 4.5.20.**

1. *For any closed subgroup $G \leq \mathrm{GL}_n(\mathbb{F})$, there is an open neighbourhood $U$ of $I$ in $G$ which is homeomorphic to a ball in $\mathbb{R}^d$. We say this as we can make open sets smaller if necessary. In particular, this implies that $U$ is path-connect. For the topological group $(\mathbb{Q}, +)$, no open neighbourhood of $0$ 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*
$$t \mapsto \begin{pmatrix} e^{it} & 0 \\ 0 & e^{i\sqrt{2}t} \end{pmatrix}.$$
*Note that $\sqrt{2}$ can be replaced with any other irrational number. Then $\gamma$ is injective, but no open neighbourhood of $I \in \mathrm{im}(\gamma)$ is path-connected. Hence, $\mathrm{im}(\gamma) \leq G$ is a subgroup but it is not closed. Hence, images of homomorphism are not necessarily matrix Lie groups, which differs from the theory of other abstract objects we have discussed so far.*

**Example 4.5.21.** *Consider the Lie algebra of $O(n)$, that is $\mathfrak{o}(n) = T_I O(n)$. 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 $t > 0$ small enough $tA$ and $-tA^\top$ lie in an open neighbourhood of $0$ in $M_n(\mathbb{F})$ on which $\exp$ is injective. Hence, for $t > 0$ small enough we have that $tA = -tA^\top$ and so $A = -A^\top$ which says that $A$ is skew-symmetric. Conversely, suppose that $A$ is skew-symmetric. Then $A$ and $A^\top = -A$ commute which implies 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.*

**Definition 4.5.22.** *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ and $H \leq \mathrm{GL}_n(\mathbb{F})$ be closed subgroups, 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, so $(\phi \circ \gamma)(t)e^{tB}$ for some unique $B \in \mathfrak{h}$. Consequently, we let*

$$\phi_*(A) = B.$$

*Equivalently, $\phi_*(A)$ is defined by the equation*

$$\gamma_{\phi_*(A)} = \phi \circ \gamma_A. \tag{4.5.1}$$

**Remark 4.5.23.** *Substituting $1$ into (4.5.1) we deduce that*

$$\exp \circ \phi_* = \phi \circ \exp$$

*as maps $\mathfrak{g} \to H$.*

**Lemma 4.5.24.** *The map $\phi_* : \mathfrak{g} \to \mathfrak{h}$, as given in Definition 4.5.22, 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 is true for all $t \in \mathbb{R}$ we deduce that $\phi_*(\lambda A) = \lambda \phi_*(A)$. Moreover, for any $t \in \mathbb{R}$ we have that

$$\begin{aligned} \gamma_{A+B}(t) &= e^{t(A+B)} \\ &= \lim_{n \to \infty} \left(\gamma_A\left(\frac{t}{n}\right)\gamma_B\left(\frac{t}{n}\right)\right)^n \end{aligned}$$

where we have used Proposition 4.5.14 and that fact that if $\delta(t) = \gamma_A(t) \cdot \gamma_B(t)$ then $\dot{\delta}(0) = A + B$. Similarly,

$$\begin{aligned} \gamma_{\phi_*(A)+\phi_*(B)} &= \lim_{n \to \infty} \left(\gamma_{\phi_*(A)}\left(\frac{t}{n}\right)\gamma_{\phi_*(B)}\left(\frac{t}{n}\right)\right)^n \\ &= \lim_{n \to \infty} \left(\phi\left(\gamma_A\left(\frac{t}{n}\right)\right)\phi\left(\gamma_B\left(\frac{t}{n}\right)\right)\right)^n \\ &\overset{(1)}{=} \lim_{n \to \infty} \phi\left(\left(\gamma_A\left(\frac{t}{n}\right)\gamma_B\left(\frac{t}{n}\right)\right)^n\right) \\ &\overset{(2)}{=} \phi\left(\lim_{n \to \infty}\left(\gamma_A\left(\frac{t}{n}\right)\gamma_B\left(\frac{t}{n}\right)\right)^n\right) \\ &= \phi\left(\gamma_{A+B}(t)\right) \\ &= \gamma_{\phi_*(A+B)}(t), \end{aligned}$$

where in $(1)$ we use the fact that $\phi$ is a homomorphism, and in $(2)$ we use the continuity of $\phi$. □

**Lemma 4.5.25.** *Let $G \leq \mathrm{GL}_n(\mathbb{F})$ and $H \leq \mathrm{GL}_n(\mathbb{F})$ be closed subgroups, 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, $\phi_*$ is a homomorphism of Lie algebras.*

*Proof.* Let

$$\delta_s(t) = \gamma_A(t)\gamma_B(t)\gamma_A(-s)$$

so that $\dot{\delta}_s(0) = e^{sA}Be^{-sA}$. As $s$ varies it is clear that $\dot{\delta}_s(0)$ traces a continuously differentiable path in $\mathfrak{g}$ whose derivative at $0$ is given by $[A, B]$. 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 $\mathfrak{h}$ with derivative at $0$ given by $[\phi_*(A), \phi_*(B)]$. 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 $0$ gives

$$\phi_*([A, B]) = [\phi_*(A), \phi_*(B)]$$

for all $A, B \in \mathfrak{g}$. In conjunction with Lemma 4.5.24 we conclude that $\phi_*$ is a homomorphism of Lie algebras. □

**Lemma 4.5.26.** *Let $G, H$ and $K$ be closed subgroups of $\mathrm{GL}_n(\mathbb{R})$. Furthermore, let $\phi : G \to H$ and $\psi : H \to K$ be continuous homomorphism. Then it follows that $\psi_* \circ \phi_* = (\psi \circ \phi)_*$.*

*Proof.* Let $A \in \mathfrak{g}$. Recall that $\phi_*(A)$ is determined by the equation

$$\gamma_{\phi_*(A)} = \phi \circ \gamma_A.$$

Consequently, $\psi_*(\phi_*(A))$ is determined by the equation

$$\psi_*(\phi_*(A)) = \psi \circ \gamma_{\phi_*(A)} = \psi \circ \phi \circ \gamma_A.$$

Similarly, $(\psi \circ \phi)_*(A)$ is determined by the equation

$$\gamma_{(\psi\circ\phi)_*(A)} = \psi \circ \phi \circ \gamma.$$

Thus we deduce that $(\psi \circ \phi)_*(A) = \psi_*(\phi_*(A))$. □

**Corollary 4.5.27.** *If $G, H \leq \mathrm{GL}_n(\mathbb{F})$ are closed subgroups which are 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 we can consider $f : G \to H$ and $g : H \to G$ continuous homomorphisms that are inverses to each other. Using Lemma 4.5.26 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 4.5.25 it follows that $\mathfrak{g}$ and $\mathfrak{h}$ are isomorphic as Lie algebras. □

# 5 Appendix

## 5.1 Topology

We adopt the notation of Section 4 by letting $\mathbb{F}$ refer to $\mathbb{R}$ 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 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*

$$\bigcup_i U_i \in \mathcal{T}.$$

- *If $\{U_1, \ldots, U_n\} \subseteq \mathcal{T}$ are open, then*

$$\bigcap_{k=1}^{n} U_k \in \mathcal{T}.$$

*A subset $C \subseteq X$ is said to be closed if $X \setminus C \in \mathcal{T}$.*

**Example 5.1.2.** *For $X \subseteq \mathbb{F}^n$, we give it the Euclidean topology by saying 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 $|y - x| < \varepsilon$ we have that $y \in U$.*

1. *Consider $X = \mathbb{Q} \subseteq \mathbb{R}$. Then open intervals $(a, b) \cap \mathbb{Q}$ are open in $\mathbb{Q}$ and any open subset $U \subseteq \mathbb{Q}$ is a union of these.*
2. *The open square $(0, 1)^2 \subseteq \mathbb{R}^2$ is open.*

**Definition 5.1.3.** *A function $f : X \to Y$ between two topological spaces is continuous if for 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 with respect to the above definition if and only if it's continuous with respect to the $\varepsilon$ - $\delta$ definition of continuity. That is, 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 that $|f(x) - f(y)| < \varepsilon$.*

**Remark 5.1.6.** *The $\epsilon$-$\delta$ definition of continuity 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 $f = (f_1, \ldots, f_n) : \mathbb{F}^m \to \mathbb{F}^n$ is continuous if and only if each component $f_i : \mathbb{F}^m \to \mathbb{F}$ are continuous.*

**Example 5.1.9.** *Projection maps $\pi_j : \mathbb{F}^n \to \mathbb{F}$, sending $(x_1, \ldots, x_n)$ to $x_j$ are continuous. Moreover, if $f, g : \mathbb{F}^n \to \mathbb{F}$ are continuous, then so are $f + g$ and $f \cdot g$. Combining this, 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 $X \subseteq Y$. We say $X$ is dense in $Y$ if $\bar{X}$, the closure of $X$ in $Y$, is the whole of $Y$. 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 $\gamma$ from $x$ to $y$ and a path $\delta$ from $y$ to $z$, then there is a path from $x$ to $z$.*

*Proof.* Let $\xi : [0, 1] \to X$ be given by

$$\xi(t) = \begin{cases} \gamma(2t) & \text{if } t \leq \frac{1}{2} \\ \delta(2t - 1) & \text{if } t \geq \frac{1}{2}. \end{cases}$$

This is a path from $x$ to $z$. □

**Definition 5.1.14.** *A set $X \subseteq \mathbb{F}^n$ is said to be bounded if there is some $C > 0$ such that 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*

$$\left(\frac{\partial f_i}{\partial x_j}(p)\right)_{ij}.$$

**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 the derivative, $df_{x_0}$, of $f$ at $x_0$ 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 invertible near $x_0$ as well.*

**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, 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 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 $\phi$ 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 $\gamma$ 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 $\phi$ respects the equivalence classes of $G/\ker(\phi)$. Hence, $\phi$ 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 $\phi_*$ is injective. Moreover, since $\phi$ is surjective it follows that $\phi_*$ 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 $\phi$ is injective due to the irrationality of $\pi$. 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, $\phi$ is a continuous homomorphism, however, $\phi$ 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 $\Lambda$.*

**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, $\Lambda$ 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 $\Lambda$ is discrete. Now consider $V' := V/\langle e \rangle$, the quotient map $\pi : V \to V'$ and $\Lambda' = \pi(\Lambda)$. It is clear that $\Lambda'$ is a subgroup of $V'$. Suppose that $\Lambda'$ 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, $\Lambda'$ is discrete. So by the inductive hypothesis, there is a basis $\{e_i'\}_{i=1}^d$ of $V'$ such that

$$\Lambda' = \mathbb{Z}e_1' + \dots + \mathbb{Z}e_k'.$$

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

$$\Lambda = \mathbb{Z}e + \mathbb{Z}e_1 + \dots + \mathbb{Z}e_k.$$

□

**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

$$\ker(\exp) = \mathbb{Z}e_1 + \dots + \mathbb{Z}e_k$$

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. □