Introduction to Algebra

Prison Mathematics Project

Introduction

Hello and welcome to the module on Introduction to Algebra! What follows is a module intended to support the reader in learning this fascinating topic. The Prison Mathematics Project (PMP) realises that you may be practising mathematics in an environment that is highly restrictive, so this text can both be used independently and does not require a calculator.

What is Algebra?

Abstract algebra is the study of algebraic structures (this includes groups, vector spaces, rings, fields, modules, algebras, etc.). The major motivation in this area was solving systems of equations, formulae for roots of polynomials and integer solutions to polynomial equations.

Learning in this Module

The best way to learn mathematics is to do mathematics. Indeed, education isn't something that happens more than it is something we should all participate in. You will find various exercise questions and worked examples in these notes so that you may try to solve problems and deepen your understanding of this topic. Although the aim is for everything to only require the content of this module, you are encouraged to use any other sources you have at your disposal.

Acknowledgements

These notes are based on lecture courses by P. Walker and J. Truss at the University of Leeds.

Contents

1 Preliminaries

2 Basic Group Theory

Arguably the largest area in abstract algebra, group theory was pioneered by mathematicians like Euler who studied algebraic operations, Lagrange and Cauchy (who studied permutations), and Galois (who made a vital link from symmetry groups to the solving of polynomials, see Chapter **??**). We start of relatively basic but over time, we will advance our group theory a lot.

Definition 2.1 A group is a set G with a binary operation $\cdot : G \times G \to G$ satisfying these:

- The operation is associative, that is for any $g, h, k \in G$, we have $g \cdot (h \cdot k) = (g \cdot h) \cdot k$.
- The set is closed under the operation, that is for any $g, h \in G$, we have $g \cdot h \in G$.
- There is an identity, that is there exists $e \in G$ with $e \cdot g = g = g \cdot e$ for all $g \in G$.
- The set is closed under inverses, that is for any $g \in G$, there exists $h \in G$ such that $g \cdot h = e = h \cdot g$. To emphasise that h is the inverse of g, we relabel it g^{-1}.

Definition 2.2 The number of elements in a group G is the order of G, denoted $|G|$.

Note: To distinguish between the set G and the group, we may write the group $(G, \cdot)$.

Example 2.3 We now look at some examples and non-examples of groups (some will be familiar).

(i) The integers under addition $(\mathbb{Z}, +)$ forms a group.

(ii) The non-zero real numbers under multiplication $(\mathbb{R} \setminus \{0\}, \times)$ forms a group.

(iii) The integers under subtraction $(\mathbb{Z}, -)$ does not form a group.

(iv) The rationals under multiplication $(\mathbb{Q}, \times)$ does not form a group.

Lemma 2.4 *$(\mathbb{R}, +)$ is a group whereas $(\mathbb{R}, \times)$ is not a group.*

Proof: (i) We need only demonstrate each of the group axioms in Definition 2.1:

- Addition of real numbers clearly satisfies $x + (y + z) = (x + y) + z$ for any $x, y, z \in \mathbb{R}$.
- The sum of two real numbers is a real number; this is closure under addition.
- The element $0 \in \mathbb{R}$ is the identity since $x + 0 = x = 0 + x$ for any $x \in \mathbb{R}$.
- For any $x \in \mathbb{R}$, we have an inverse $-x \in \mathbb{R}$ because $x + (-x) = 0 = (-x) + x$.

(ii) We can see that $(\mathbb{R}, \times)$ is not a group because it isn't closed under inverses. Indeed, the identity here is $1 \in \mathbb{R}$ (because multiplication by one doesn't change anything) but there is no

element $x \in \mathbb{R}$ such that $0 \times x = 1$, that is 0 has no inverse. □

Exercise 1 Prove that $(\mathbb{Z}, +)$ is a group.

For small groups, it can help to completely write out the operations between group elements.

Definition 2.5 A group table (or Cayley table) for a group $(G, \cdot)$ with n elements is an $n \times n$ grid containing all possible operations of the group.

Example 2.6 Consider the integers modulo four, that is $\mathbb{Z}_4 = \{0, 1, 2, 3\}$. We can turn this into a group under addition modulo four $+_{\text{mod } 4}$. Although **not** a group, we consider it also under multiplication modulo four $\times_{\text{mod } 4}$. We draw the Cayley tables for them in Table 1 below.

$+_{\text{mod } 4}$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

a The group table for $(\mathbb{Z}_4, +_{\text{mod } 4})$.

$\times_{\text{mod } 4}$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

b The 'group' table for $(\mathbb{Z}_4, \times_{\text{mod } 4})$.

Table 1: The Cayley tables for $(\mathbb{Z}_4, +_{\text{mod } 4})$ and $(\mathbb{Z}_4, \times_{\text{mod } 4})$.

Note: In a group table, the identity is the element whose row is the same as the top row.

Exercise 2 Write out the group table for $(\mathbb{Z}_5, +_{\text{mod } 5})$.

Definition 2.7 The dihedral group D_n is the group of symmetries of a regular n-gon.

Example 2.8 Consider the dihedral group D_3 of symmetries of an equilateral triangle. These are the symmetries of an equilateral triangle ABC centred at O:

(i) Do nothing (the identity).

(ii) Rotate about O by $2\pi/3$.

(iii) Rotate about O by $4\pi/3$.

(iv) Reflect in the line through OA.

(v) Reflect in the line through OB.

(vi) Reflect in the line through OC.

As in Chapter **??**, rotations are taken anti-clockwise. Thus, a rotation about O by $-2\pi/3$ is precisely the same as transformation (iii) above. Also, a rotation about O by 2π is precisely the

same as transformation (i). Hence, (it at least seems plausible that) we have a complete list of symmetries of our equilateral triangle. We use Figure 1 to assist with our thinking.

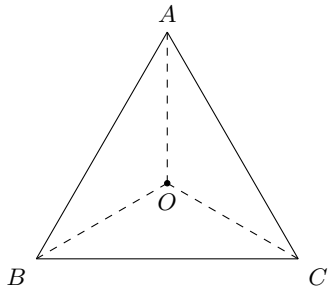
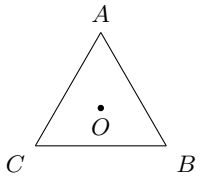
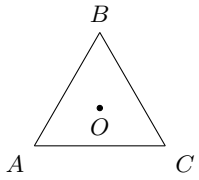


Figure 1: The symmetries of an equilateral triangle ABC.

If we label transformations (i)–(vi) as I, R, S, A, B, C respectively, we can build the group table from this. By virtue of laziness, this is left to you in Exercise 3. To get you started, we show how you can compute members of the group table. Indeed, the operation is denoted $\circ$ and Figure 2 shows why $R \circ A = C$ (which means 'first A, then R', like function composition).



(a) First, we perform the reflection A.



(b) After, we perform the rotation R.

Figure 2: Computing $R \circ A$ in the dihedral group D_3.

Exercise 3 Complete the below group table for the dihedral group D_3.

$\circ$	I	R	S	A	B	C
I	I	R	S	A	B	C
R		S		C		
S			R			
A				I		
B					I	
C						I

Table 2: The group table for $(D_3, \circ)$.

Note: The dihedral group D_n contains $2n$ elements, so some mathematicians prefer to denote it by D_{2n}. I am not a fan of this at all; our (my) preference is to label the dihedral group after the n-gon it applies to, not the number of elements it has.

Definition 2.9 A group $(G, \cdot)$ is Abelian if $g \cdot h = h \cdot g$ for all $g, h \in G$.

Example 2.10 Here are some Abelian and non-Abelian groups.

(i) The groups $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$, $(\mathbb{C}, +)$ are all Abelian.

(ii) The group D_3 is non-Abelian (it is the smallest finite non-Abelian group).

(iii) All groups D_n for $n \geq 3$ are non-Abelian.

Remark 2.11 If given a group table, you will be able to immediately spot that it is Abelian if it is symmetric about the diagonal. Clearly then, Table 2 represents a non-Abelian group whereas Table 1a represents an Abelian group.

Note: From here on out, we abuse notation and say that G is a group (we won't write the operation unless we need it). Also, we write gh in place of the operation notation $g \cdot h$.

Proposition 2.12 *The identity of a group G is unique.*

Proof: Suppose G has two identities, namely $e, f \in G$. Then, $ef = e$ because f is an identity, but $ef = f$ because e is an identity. Consequently, $e = f$, so they are the same. □

Proposition 2.13 *The inverse of an element in a group G is unique.*

Proof: Suppose $g \in G$ has two inverses, namely $h, k \in G$. By definition, this means that $gh = e = hg$ and $gk = e = kg$. But now, $h = he = h(gk) = (hg)k = ek = k$ by associativity, so the 'two' suggested inverses are actually the same. □

Exercise 4 Fully understand the proofs of Propositions 2.12 and 2.13.
[**Note:** Knowing how to use each part of the definition of a group is vitally important.]

Proposition 2.14 *Let G be a group and $g, h \in G$.*
(i) *If $gh = e$, then $g = h^{-1}$ and $h = g^{-1}$.*
(ii) *The inverse $(gh)^{-1} = h^{-1}g^{-1}$.*
(iii) *The inverse $(g^{-1})^{-1} = g$.*

Proof: (i) This is immediate from the definition of an inverse.

(ii) It suffices to show that $(gh)(h^{-1}g^{-1}) = e$ and $(h^{-1}g^{-1})(gh) = e$; this is the definition of an inverse. Indeed, we see that

$$\begin{aligned}(gh)(h^{-1}g^{-1}) &= g(hh^{-1})g^{-1} \\ &= geg^{-1} \\ &= gg^{-1} \\ &= e,\end{aligned} \qquad \begin{aligned}(h^{-1}g^{-1})(gh) &= h^{-1}(g^{-1}g)h \\ &= h^{-1}eh \\ &= h^{-1}h \\ &= e.\end{aligned}$$

(iii) By definition of inverses, we know that $g^{-1}g = e$ and $gg^{-1} = e$. This tells us that the inverse of g^{-1} is g. Written mathematically, this is precisely $(g^{-1})^{-1} = g$. □

Lemma 2.15 (Cancellation Laws) *Let G be a group and $g, h, k \in G$.*
(i) *We have $gh = gk \Rightarrow h = k$.* (Left Cancellation)
(ii) *We have $hg = kg \Rightarrow h = k$.* (Right Cancellation)

Proof: We prove (i) and leave (ii) to you in Exercise 5. Indeed, assume that $gh = gk$. Then, we can multiply by g^{-1} on the left to get $g^{-1}(gh) = g^{-1}(gk) \Rightarrow (g^{-1}g)h = (g^{-1}g)h \Rightarrow h = k$. □

Exercise 5 Prove Lemma 2.15(ii), that is Right Cancellation.

Corollary 2.16 *Group tables form a Latin square, that is an array where each element occurs exactly once in each row and once in each column.*

Proof: Consider the row of the group table of a finite group G which corresponds to the element $g \in G$. By definition of a group table, all elements in this row are of the form gx where $x \in G$. If two elements in this row coincide, that is $gx = gy$ for $x, y \in G$, then the left cancellation law implies that $x = y$. As such, an element cannot appear **more** than once in a row. Suppose the element $h \in G$ appears in the row corresponding to $x \in G$ and the column corresponding to

$y \in G$; this means that $h = xy$. Therefore, it is possible to write $y = x^{-1}h$. As such, an element appears **at least** once in each row. Combining these means it appears **exactly** once.

The proof is near-identical for the columns situation. □

Example 2.17 The contrapositive of Corollary 2.16 implies that $(\mathbb{Z}_4, \times_{\text{mod } 4})$ is **not** a group. Indeed, looking at the row corresponding to 2 in Table 1b, it is clear the elements 1 and 3 do not appear in this row, so the Latin square property fails.

Note: From now, we use the shorthand notation g^n to mean the repeated operation

$$\underbrace{gg\cdots g}_{n \text{ of them}}.$$

We use g^{-n} to represent the product of n copies of g^{-1}. Also, we write the identity $e = 1$.

Lemma 2.18 *In any group G, for an element $g \in G$ and integers $m, n \in \mathbb{Z}$, we have*

$$g^m g^n = g^{m+n} \quad \textit{and} \quad (g^m)^n = g^{mn}.$$

Proof: We proceed inductively. For the base case, suppose that $n = 0$. Then, we can see that the expressions in the statement are true, since $g^0 = 1 = e$. As for the inductive hypothesis, suppose it is true for $n = k$, where $k \in \mathbb{Z}$. Then, we see that

$$g^m g^{k+1} = g^m g^k g = g^{m+k} g = g^{m+k+1} \quad \text{and} \quad (g^m)^{k+1} = (g^m)^k g^m = g^{mk} g^m = g^{m(k+1)}.$$

Thus, by the principal of mathematical induction, the result holds true. □

Definition 2.19 For an integer $n > 1$, we define the group of integers coprime with n as

$$\mathbb{Z}_n^* = \{x \in \mathbb{Z}^+ : x \text{ is coprime with } n\},$$

where the group operation is multiplication modulo n, that is $\times_{\text{mod } n}$.

Remark 2.20 Be careful with the notation; if somebody writes $\mathbb{Z}_n$, then it is natural to think of the *additive* group of integers modulo n, so named because the group operation is addition. Here, we write $\mathbb{Z}_n^*$ to mean the *multiplicative* group of integers modulo n, so named for the obvious reason. Although the notation * means different things depending on the area of mathematics we are in, at least in group theory, it often means a multiplicative group. For example, $\mathbb{Q}^*, \mathbb{R}^*, \mathbb{C}^*$ refer to the groups whose underlying sets miss out 0 and whose operations are multiplication.

Example 2.21 We list a few examples of the groups described in Definition 2.19.

(i) $\mathbb{Z}_2^* = \{1\}$.

(ii) $\mathbb{Z}_3^* = \{1, 2\}$.

(iii) $\mathbb{Z}_4^* = \{1, 3\}$.

(iv) $\mathbb{Z}_5^* = \{1, 2, 3, 4\}$.

In general, if p is prime, then the multiplicative group is given by $\mathbb{Z}_p^* = \{1, 2, ..., p-1\}$.

Exercise 6 Construct the group table of $\mathbb{Z}_{10}^*$.

[**Hint:** To determine $\mathbb{Z}_{10}^*$, recall that *coprime* is introduced in Definition **??**.]

Definition 2.22 A frieze pattern is a strip in the plane $\mathbb{R}^2$ with translational symmetry.

Note: It is common for a frieze to have more symmetries; see Example 2.23, for instance.

Example 2.23 Consider the following pattern consisting of an infinite string of the letter A:

$$\cdots \text{AAAAA} \cdots$$

We will assume that each A is centred at the point $(2n, 0) \in \mathbb{R}^2$. Let T_n be the translation by a shift of $2n$ in the x-direction, that is $T_n(x, y) = (x + 2n, y)$, and let R_n be the reflection in the line $x = n$, that is $R_n(x, y) = (-x + 2n, y)$. Then, the set G of isometries of $\mathbb{R}^2$ that preserve the frieze pattern is given by

$$G = \{T_n : n \in \mathbb{Z}\} \cup \{R_n : n \in \mathbb{Z}\}.$$

In other words, applying any of T_n and R_n will always send the frieze to the frieze. Hence, G is a group under composition. In fact, we can draw the group table below (although there are infinite elements, it suffices to note how the T_n and R_n interact with themselves and each other). We will demonstrate how to get one of the compositions (the rest are left to do in Exercise 7):

$$(T_m \circ T_n)(x, y) = T_m(x + 2n, y) = (x + 2n + 2m, y) = T_{m+n}(x, y).$$

This allows us to conclude that $T_m \circ T_n = T_{m+n}$. The rest are presented in Table 3 below.

$\circ$	T_n	R_n
T_m	T_{m+n}	R_{m+n}
R_m	R_{m-n}	T_{m-n}

Table 3: The group table of the isometries preserving the frieze $\cdots$AAAAA$\cdots$.

Exercise 7 Verify that the other compositions presented in Table 3 are correct.

Definition 2.24 Let G be a group and $g \in G$ some element. The order of g is

$$\text{ord}(g) = \min\{n \in \mathbb{Z}^+ : g^n = e\},$$

the least positive integer where applying the operation that many times to g yields e.

Example 2.25 Consider Table 1a. We can see that $\text{ord}(0) = 1$ because we do not have to apply the operation to itself to get to the identity because 0 already is the identity! However, we see that $\text{ord}(1) = 4$, because $1+1+1+1 \equiv 0 \pmod 4$ is the smallest number of times we successively apply the operation to 1 to get the identity.

Note: We say that $\text{ord}(g) \coloneqq \infty$ when no such minimum n exists in Definition 2.24.

Proposition 2.26 *Let $g \in G$ be an element of a group.*

(i) *If $\text{ord}(g) = \infty$, the g^k are distinct for all $k \in \mathbb{Z}$. Hence, $g^k = 1$ if and only if $k = 0$.*

(ii) *If $\text{ord}(g) = n$, the g^k repeat in cycles of length n. Hence, $g^k = 1$ if and only if $n \mid k$.*

Proof: (i) Assume to the contrary that $g^j = g^k$ where $j < k$. Then, we can invert to get that $g^{k-j} = 1$, which means that $\text{ord}(g) \leq k - j$, a contradiction to it having infinite order.

(ii) The first n powers $g^0, g^1, ..., g^{n-1}$ are distinct, for if they are not, an argument similar to the contradiction in (i) will take place: if $g^j = g^k$ where $0 \leq j < k \leq n - 1$, then inverting gives $g^{k-j} = 1$ but $k - j < n$, contradicting it having order n. By the Division Lemma (Lemma **??**), we can divide any $k \in \mathbb{Z}$ by the integer n to get a quotient q and remainder r, that is we can write $k = qn + r$, for $0 \leq r < n$. Then, we see that $x^k = x^{qn+r} = (x^n)^q x^r = x^r$, where $x^n = 1$ by definition of order. Consequently, this is the identity if and only if $r = 0$, that is $n \mid k$. □

Exercise 8 Consider the frieze pattern drawn in Figure 3 below.

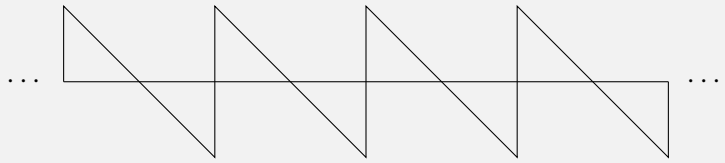


Figure 3: A frieze pattern of triangles.

Suppose one of the triangles has vertices $(0,0)$, $(0,1)$, $(1,0)$. The isometries of the plane that preserve the frieze are translations T_n which shift the diagram n periods to the right and rotations S_n about the point $(n,0)$ by the angle π.

(i) Write down formulae for $T_n(x,y)$ and $S_n(x,y)$.

(ii) Construct the group table for the frieze group.

(iii) State the orders of the elements in the frieze group.

Definition 2.27 Let $(G,\cdot)$ be a group. Then, a subset $H \subseteq G$ is called a subgroup if it becomes a group under the same operation $\cdot$ as G. This is then denoted $H \leq G$.

Example 2.28 Here are some examples and non-examples of subgroups.

(i) We have a chain of additive subgroups $\mathbb{Z} \leq \mathbb{Q} \leq \mathbb{R} \leq \mathbb{C}$.

(ii) We have a chain of multiplicative subgroups $\mathbb{Q}^* \leq \mathbb{R}^* \leq \mathbb{C}^*$.

(iii) The set $\{0,1\}$ is **not** a subgroup of $\mathbb{Z}_4$ because it isn't closed.

(iv) We always have the subgroups $\{e\} \leq G$ and $G \leq G$, where $e \in G$ is the identity.

Lemma 2.29 *Let $H \leq G$ be a subgroup. Then, the following are true:*

(i) *The identity of H is the identity of G.*

(ii) *The inverses of elements in H are the same as they are as elements in G.*

Proof: (i) Suppose that H and G have identities 1_H and 1_G, respectively. Then, we see that $1_G 1_H = 1_H$ and $1_H 1_H = 1_H$. Applying Right Cancellation to $1_G 1_H = 1_H 1_H$ gives $1_G = 1_H$.

(ii) Let $h \in H$ have inverse $y \in H$, meaning that $hy = 1_H = yh$. But because $1_H = 1_G$ by part (i) and because $H \subseteq G$ implies that $h, y \in G$, we have also that $hy = 1_G = yh$, so y is still the inverse of h when thought of as elements of G. □

Theorem 2.30 (Subgroup Criterion) *Let G be a group. A subset $H \subseteq G$ is a subgroup if and only if it satisfies the following properties:*

(i) *H contains the identity.*

(ii) *H is closed under the operation.*

(iii) *H is closed under inverses.*

Proof: ($\Leftarrow$) Suppose first that (i), (ii), (iii) are satisfied. It is clear that H is a group, because it inherits associativity from G. Thus, it satisfies Definition 2.27 and so we conclude $H \leq G$.

($\Rightarrow$) Suppose now that $H \leq G$ is a subgroup. Because it is closed under the operation, (ii) holds. Finally, the additional conditions (i) and (iii) follow directly from Lemma 2.29. □

Example 2.31 We will exploit the Subgroup Criterion to prove that $\mathbb{R}^{>0} \coloneqq \{x \in \mathbb{R} : x > 0\} \leq \mathbb{R}^*$ is a subgroup. Indeed, it is clear that the identity $1 > 0$, so we have $1 \in \mathbb{R}^{>0}$, satisfying (i) of the Subgroup Criterion. Next, we know the product of two positive numbers is positive, so $xy \in \mathbb{R}^{>0}$ for any $x, y \in \mathbb{R}^{>0}$, satisfying (ii) of the Subgroup Criterion. Finally, the reciprocal of a positive number is positive, so $x^{-1} \in \mathbb{R}^{>0}$ for all $x \in \mathbb{R}^{>0}$, satisfying (iii) of the Subgroup Criterion.

Note: The following are **not** subgroups of $\mathbb{R}^*$: $\mathbb{R}^{\geq 0}$, $\mathbb{R}^{<0}$, $\mathbb{R}^{>1}$, $\mathbb{R}^{\geq 1}$ (think about why).

Exercise 9 Use the Subgroup Criterion to prove that $\{2^n : n \in \mathbb{Z}\} \leq \mathbb{Q}^*$ is a subgroup.

Definition 2.32 Let G be a group and $g \in G$. The subgroup generated by g is the group

$$\langle g \rangle = \{g^n : n \in \mathbb{Z}\}.$$

Lemma 2.33 *The order of $\langle g \rangle$ is precisely the order of g, that is $\left|\langle g \rangle\right| = \operatorname{ord}(g)$.*

Proof: This is immediate from Definition 2.24 and Proposition 2.26. □

Exercise 10 We call G cyclic if it is generated by an element, i.e. $G = \langle g \rangle$ for some $g \in G$. Prove that a group of order n is cyclic if and only if it contains an element of order n.

Example 2.34 Here are some examples and non-examples of cyclic groups.

(i) The group $\mathbb{Z}_n$ is cyclic for all $n \geq 1$, as it is generated by 1.

(ii) The group $\mathbb{Z}$ is cyclic, also generated by 1.

(iii) The group $\mathbb{Z}_5^* = \{1, 2, 3, 4\}$ is cyclic, generated by either 2 or 3.

(iv) The group $\mathbb{R}$ is **not** cyclic, since there is no $x \in \mathbb{R}$ such that $\{nx : n \in \mathbb{Z}\} = \mathbb{R}$.

(v) The group $\mathbb{R}^{>0}$ is **not** cyclic. Indeed, assume to the contrary that there exists $x > 0$ such that $\{x^n : n \in \mathbb{Z}\} = \mathbb{R}^{>0}$. Then, $x \neq 1$, so either $x > 1$ or $x < 1$. Without loss of generality (we can swap to $x \in$ if needed), we can assume $x > 1$. Then, no element of the interval $(1, x)$ can be written as a so-called integer power of x, so $\{x^n : n \in \mathbb{Z}\} \neq \mathbb{R}^{>0}$.

Lemma 2.35 *Every cyclic group is Abelian.*

Proof: In the group $\langle g \rangle$, we have $(g^n)(g^m) = g^{n+m} = g^{m+n} = (g^m)(g^n)$, so it is Abelian. □

Proposition 2.36 *Every subgroup of $\mathbb{Z}$ is of the form $\langle k \rangle = k\mathbb{Z}$.*

Proof: Let $H \leq \mathbb{Z}$ be a subgroup. If $H = \{0\}$, then $H = 0\mathbb{Z}$, so we can now assume that H contains a non-zero element. In particular, let $k \in H$ with $k > 0$ be minimal. Then, $k\mathbb{Z} \subseteq H$. Assume to the contrary that there exists $h \in H$ but $h \notin k\mathbb{Z}$ (meaning that $H \not\subseteq k\mathbb{Z}$). Then, by the Division Algorithm, we can write $h = qk + r$ for $q, r \in \mathbb{Z}$ and $0 < r < k$. By assumption, we have that $r \notin H$, but we can write $r = h - qk \in H$ by closure, a contradiction. □

Theorem 2.37 *Every subgroup of a cyclic group is cyclic.*

Proof: Let $G = \langle g \rangle$ be cyclic and $H \leq G$ be a subgroup. Then, we define $K = \{k \in \mathbb{Z} : g^k \in H\}$. This is a subgroup of $\mathbb{Z}$ (see Exercise 11). Thus, by Proposition 2.36, we know that $K = n\mathbb{Z}$ for some $n \in \mathbb{Z}$. Consequently, $H = \{g^k : k \in K\} = \{g^{na} : a \in \mathbb{Z}\} = \{(g^n)^a : a \in \mathbb{Z}\} = \langle g^n \rangle$, meaning that H is cyclic. □

Exercise 11 Prove that $K = \{k \in \mathbb{Z} : g^k \in H\}$ is a subgroup of $\mathbb{Z}$.

[**Hint:** Remember that $\mathbb{Z}$ is an *additive* group, which means so too is K.]

Recall in Chapter **??** we introduced the notion of the Cartesian product of two sets. We can, instead of considering sets, consider the Cartesian product of two groups and endow upon it a group operation. These are the next types of group to be discussed.

Definition 2.38 Let G and H be groups with their respective operations. Their direct product is the group whose underlying set is $G \times H$ and whose operation is given pointwise, that is for all $g_1, g_2 \in G$ and $h_1, h_2 \in H$,

$$(g_1, h_1)(g_2, h_2) = (g_1 g_2, h_1 h_2).$$

Note: In Definition 2.38, $g_1 g_2$ occurs within G (the G-group operation) and $h_1 h_2$ occurs within H (the H-group operation). To be more careful, let $(G, \cdot_G)$ and $(H, \cdot_H)$ be groups. Then, the operation in their direct product is $(g_1, h_1)(g_2, h_2) = (g_1 \cdot_G g_2, h_1 \cdot_H h_2)$.

Example 2.39 The direct product of $G = \{I, R, S\}$ (with elements the same as in Example 2.8) and $H = \mathbb{Z}_4^*$ is the group with the following underlying set:

$$G \times H = \{(I, 1), (I, 3), (R, 1), (R, 3), (S, 1), (S, 3)\}.$$

To get to grips with the group operation, note $(R, 3)(S, 3) = (R \circ S, 3 \times_{\text{mod } 4} 3) = (I, 1)$. Indeed, as explained in the above note, we look at the operation in G, which is composition, applied to R and S to get the first entry; we look at the operation in H, which is multiplication modulo 4, applied to 3 and 3 to get the second entry.

Note: The cardinality of the direct product is simply $|G \times H| = |G||H|$.

Lemma 2.40 *Let G and H be groups. Then, the order of $(g, h) \in G \times H$ is precisely the lowest common multiple of $\text{ord}(g)$ and $\text{ord}(h)$, or ∞ if either $\text{ord}(g) = \infty$ or $\text{ord}(h) = \infty$.*

Proof: Suppose that $\text{ord}(g) = n$ and $\text{ord}(h) = m$. By Definition 2.24, the order of (g, h) is the smallest $k \in \mathbb{Z}^+$ such that $(g, h)^k = (g^k, h^k) = (1_G, 1_H) = 1_{(G,H)}$. From Proposition 2.26, we know that $g^k = 1_G$ if and only if $n \mid k$ and $h^k = 1_G$ if and only if $m \mid k$. Since the order is the **smallest** such positive integer, it follows that $k = \text{lcm}(n, m)$, as required. On the other hand, if either $\text{ord}(g) = \infty$ or $\text{ord}(h) = \infty$, it is clear that $\text{ord}((g, h)) = \infty$ also. □

Theorem 2.41 *Let G and H be finite cyclic groups. Then, $G \times H$ is cyclic if and only if the group orders $|G|$ and $|H|$ are coprime.*

Proof: ($\Leftarrow$) Since G and H are cyclic, we have $G = \langle g \rangle$ and $H = \langle h \rangle$. Suppose they have orders n and m, respectively. Then, Lemma 2.40 tells us that $\text{ord}((g, h)) = \text{lcm}(n, m)$. However, we

know from Lemma 2.33 that $|G| = n$ and $|H| = m$, so the assumption that they are coprime means $\text{ord}((g,h)) = nm = |G \times H|$. So, Exercise 10 implies $G \times H = \langle (g,h) \rangle$ is cyclic.

($\Rightarrow$) Suppose that $|G| = n$ and $|H| = m$ are **not** coprime, meaning that $\text{lcm}(n,m) =: l < nm$. Then, for any integers $a, b \in \mathbb{Z}^+$, it is true that $(g^a, h^b)^l = (g^{al}, h^{bl}) = (1_G, 1_H) = 1_{(G,H)}$ because $n \mid al$ and $n \mid bl$ (given that $n, m \mid \text{lcm}(n,m)$ by definition). Thus, $\text{ord}((g,h)) \leq l$, meaning that **no** element has order equal to $mn = |G \times H|$. Consequently, $G \times H$ is **not** cyclic. □

> **Exercise 12** Demonstrate Theorem 2.41 by using G and H from Example 2.39.
>
> [**Hint:** Show that $G = \{I, R, S\}$ and $H = \mathbb{Z}_4^*$ are cyclic and give a generator for $G \times H$.]

Often in mathematics, when we have defined ourselves an object (e.g. a group, a set, whatever), we want to be able to discuss when two of these objects are the 'same'. For sets, we introduced the idea of functions (specifically, a bijection tells us when two sets are the 'same'). We have something analogous for groups but we need some extra structure; we have group operations in the mix that we need to worry about.

Definition 2.42 A map between groups $\varphi : G \to H$ is a (group) homomorphism if for all $g_1, g_2 \in G$, we have $\varphi(g_1 g_2) = \varphi(g_1)\varphi(g_2)$ (that is φ respects the group operations). If φ is bijective, then we call it a (group) isomorphism, wherein we say that the groups are isomorphic, denoted $G \cong H$.

Example 2.43 Here are some examples of group homomorphisms.

(i) The map $\varphi : \mathbb{R} \to \mathbb{R}^{>0}$ given by $\phi(x) = \exp(x)$; it is injective **and** surjective (i.e. bijective).

(ii) The map $\psi : \mathbb{Z} \to \mathbb{Z}_3$ given by $\psi(n) = n \pmod 3$; it is **not** injective but it is surjective.

(iii) The map $\phi : \mathbb{Z} \to \mathbb{Z}$ given by $\phi(m) = 2m$; it is injective but **not** surjective.

(iv) The map $\theta : \mathbb{R} \to \mathbb{R}^*$ given by $\theta(y) = \exp(y)$; it is **not** injective and **not** surjective.

> **Note:** If you haven't come across the exponential function $\exp(x) = e^x$ yet, don't worry about it. Just be aware that it allows us to 'translate from addition to multiplication', and its inverse $\log(x)$ allows us to do the reverse of that.

Remark 2.44 Necessarily, for a bijection to exist between groups, then it must be a bijection on the underlying sets. In particular, isomorphic groups have the same cardinality. This allows us to, in principle, look at the group tables of two groups and 'see' an isomorphism if they are the

same size, have elements in corresponding positions, have identities in the same place, etc. (up to reordering the rows).

Lemma 2.45 *A homomorphism $\varphi : G \to H$ preserves the identity and respects inverses, that is $\varphi(1_G) = 1_H$ and $\varphi(g^{-1}) = \varphi(g)^{-1}$ for all $g \in G$.*

Proof: For the first part, $\varphi(1_G)\varphi(1_G) = \varphi(1_G 1_G) = \varphi(1_G)$, by definition of a homomorphism. However, $\varphi(1_G) \in H$ is just some element of the group H, so we can write $\varphi(1_G) = \varphi(1_G)1_H$, by definition of the identity of H. Thus, we have $\varphi(1_G)\varphi(1_G) = \varphi(1_G)1_H$ and Left Cancellation gives us the result.

For the second part, $gg^{-1} = 1_G$ by definition. Therefore, $\varphi(g)\varphi(g^{-1}) = \varphi(gg^{-1}) = \varphi(1_G) = 1_H$ by the first part above. Hence, this tells us that the inverse of $\varphi(g)$ is precisely $\varphi(g^{-1})$. Written out properly, that is $\varphi(g)^{-1} = \varphi(g^{-1})$. □

Proposition 2.46 *Let $\varphi : G \to H$ be an isomorphism. Then, the following properties hold.*

(i) *The elements $g \in G$ and $\varphi(g) \in H$ have the same order.*
(ii) *The groups have the same number of elements of order n, for each $n \in \mathbb{Z}^+$.*
(iii) *G is Abelian if and only if H is Abelian.*
(iv) *H is cyclic if and only if H is cyclic.*

Proof: (i) Suppose $\text{ord}(g) = n$. Then, $g^n = 1_G$ which means that $\varphi(g)^n = \varphi(g^n) = \varphi(1_G) = 1_H$, but because n is minimal such that this occurs, we have $\text{ord}(\varphi(g)) = n$ also.

(ii) This is a trivial consequence of the argument in (i).

(iii) Suppose G is Abelian, meaning that $g_1 g_2 = g_2 g_1$ for all $g_2, g_2 \in G$. Then, $\varphi(g_1 g_2) = \varphi(g_2 g_1)$ which is equivalent to $\varphi(g_1)\varphi(g_2) = \varphi(g_2)\varphi(g_1)$. Thus, H is Abelian. Conversely, because $\varphi^{-1} : H \to G$ is also an isomorphism, H being Abelian implies G being Abelian.

(iv) Suppose $G = \langle g \rangle$. Then, because φ is surjective, for any $h \in H$, there exists $a \in G$ such that $h = \varphi(a)$. By the fact that G is cyclic, we have $a = g^n$ for some $n \in \mathbb{Z}^+$. Substituting this into the previous expression implies that $h = \varphi(g^n) = \varphi(g)^n$. Thus, $H = \langle \varphi(g) \rangle$ is cyclic. Conversely, because $\varphi^{-1} : H \to G$ is surjective, H being cyclic implies G being cyclic. □

Exercise 13 Let $G = \mathbb{R}$ (under $+$) and $H = \mathbb{R}^*$ (under $\times$). Show that H has an element of order two and that G does **not**. Conclude from this that $G \ncong H$ are non-isomorphic.

Theorem 2.47 *Two cyclic groups are isomorphic if and only if they have the same order.*

Proof: ($\Leftarrow$) This is trivial, since they are isomorphic as sets (see Remark 2.44).

($\Rightarrow$) Suppose that $G = \langle g \rangle$ and $H = \langle h \rangle$ have the same order. If said order is infinite, then g^k are distinct for all $k \in \mathbb{Z}^+$ by Proposition 2.26. Thus, we can define $\varphi : G \to H$ via $\varphi(g^k) = h^k$. This is a homomorphism (something which you will prove in Exercise 14) and it is clearly bijective, meaning that $G \cong H$. If said order is finite, n say, then the powers g^k repeat with period n, again a consequence of Proposition 2.26. The same is true of the powers h^k, so the same map $\varphi : G \to H$ given by $\varphi(g^k) = h^k$ implies $G \cong H$. Either way, the groups are isomorphic. □

Exercise 14 Show that $\varphi : G \to H$ in the proof of Theorem 2.47 is a homomorphism.

Corollary 2.48 (Chinese Remainder Theorem) $\mathbb{Z}_n \times \mathbb{Z}_m \cong \mathbb{Z}_{nm}$ *if and only if* n, m *coprime.*

Proof: If n and m are coprime, then $\mathbb{Z}_n \times \mathbb{Z}_m$ is cyclic by Theorem 2.41. Since $\mathbb{Z}_{nm}$ is cyclic and of the same order, they are isomorphic by Theorem 2.47. Conversely, if n and m are **not** coprime, then $\mathbb{Z}_n \times \mathbb{Z}_m$ is **not** cyclic; it is therefore **not** isomorphic to $\mathbb{Z}_{nm}$. □

Example 2.49 $\mathbb{Z}_{60} \cong \mathbb{Z}_{20} \times \mathbb{Z}_3 \cong \mathbb{Z}_5 \times \mathbb{Z}_{12} \cong \mathbb{Z}_5 \times \mathbb{Z}_4 \times \mathbb{Z}_3$ by the Chinese Remainder Theorem.

Definition 2.50 Let $H \leq G$ be a subgroup. A right coset of H in G is a subset of G of the form $Hg \coloneqq \{hg : h \in H\}$, with $g \in G$ fixed. Similarly, a left coset has the form gH.

Note: If a group is Abelian, then the left and right cosets are the same. Even if it is not, we will often abuse nomenclature and just refer to the 'cosets' (at least if it is clear which type of coset we are working with). To be safe, assume that 'cosets' means **right** cosets.

Example 2.51 Here are two examples of cosets.

(i) Consider the dihedral group $G = D_3$ as in Example 2.8 with group table in Exercise 3. We can use the group table to compute the cosets of $H = \{I, R, S\}$ in D_3. Indeed,

$$\begin{aligned} HI &= \{I, R, S\}, & HA &= \{A, C, B\}, \\ HR &= \{R, S, I\}, & HB &= \{B, A, C\}, \\ HS &= \{S, I, R\}, & HC &= \{C, B, A\}. \end{aligned}$$

Because order doesn't matter in sets, we essentially have only two cosets of H in D_3, namely H itself and the coset $\{A, B, C\}$. We will see in Exercise 15 that the cosets really do depend on the choice of subgroup H.

(ii) Consider the group $G = \mathbb{Z}$. We can compute the cosets of $H = 3\mathbb{Z} = \{..., -6, -3, 0, 3, 6, ...\}$ in $\mathbb{Z}$. Indeed,

$$3\mathbb{Z} + 0 = \{..., -6, -3, 0, 3, 6, ...\} = \mathbb{Z} \pm 3 = \mathbb{Z} \pm 6 = \cdots,$$
$$3\mathbb{Z} \pm 1 = \{..., -5, -2, 1, 4, 7, ...\} = \mathbb{Z} \pm 4 = \mathbb{Z} \pm 7 = \cdots,$$
$$3\mathbb{Z} \pm 2 = \{..., -4, -1, 2, 5, 8, ...\} = \mathbb{Z} \pm 5 = \mathbb{Z} \pm 8 = \cdots.$$

Here, we essentially have three cosets: $H = 3\mathbb{Z}$ itself, $3\mathbb{Z} + 1$ and $3\mathbb{Z} + 2$.

Exercise 15 Determine the cosets of the subgroup $K = \{I, A\}$ in D_3.

Proposition 2.52 *Let $H \leq G$ be a subgroup.*

(i) *H is itself a coset in G.*
(ii) *Every element of G belongs to a coset.*
(iii) *If $y \in Hx$ is an element of the coset, then $Hy = Hx$ are the same coset.*
(iv) *Two cosets are either equal or disjoint.*
(v) *Two cosets $Hx = Hy$ if and only if $xy^{-1} \in H$.*

Proof: (i) The subgroup H is itself a coset since $H = H1$, where 1 is the identity of G.

(ii) The element $g \in G$ belongs to the coset Hg, since H is a subgroup and therefore contains the identity. In other words, $g = 1g \in Hg$.

(iii) Since $y \in Hx$, it means that $y = hx$ for some $h \in H$. But then, because H is a subgroup and is therefore closed under forming inverses, this means that $x = h^{-1}y$ by the Cancellation Laws. To show that $Hx = Hy$, we prove the following inclusions: $Hx \subseteq Hy$ and $Hy \subseteq Hx$.

(a) To show that $Hx \subseteq Hy$, we need to pick an arbitrary element of Hx and show that it is also in the coset Hy. Indeed, for $h' \in H$, we know that $h'x \in Hx$ by definition. However, $h'x = (h'h^{-1})y \in Hy$ by associativity and closure of H under the operation. Hence, we have this inclusion.

(b) To show that $Hy \subseteq Hx$, we need to pick an arbitrary element of Hy and show that it is also in the coset Hx. Indeed, for $h' \in H$, we know that $h'y \in Hy$ by definition. However, $h'y = (h'h)x \in Hx$ by associativity and closure of H under the operation. Thus, we also have the other inclusion.

(iv) Suppose that $Hx \cap Hy \neq \emptyset$, that is the cosets are not disjoint. Then, there exists an element $z \in Hx \cap Hy$. By part (iii), this means that $Hx = Hz = Hy$ and so the cosets are equal.

(v) Now, $Hx = Hy$ if and only if $x \in Hy$ by (iii), which is equivalent to saying there exists $h \in H$ such that $x = hy$, that is $xy^{-1} = h$ and therefore $xy^{-1} \in H$. □

Corollary 2.53 *For $H \leq G$ a subgroup, G is partitioned by the cosets of H.*

Proof: (Indirect) This is a direct consequence of Proposition 2.52.

(Direct) An alternate proof is to define an equivalence relation on G as follows: for $x, y \in G$, we say $x \sim y$ if and only if $xy^{-1} \in H$ (the proof that this is an equivalence relation is precisely Exercise 16). From Theorem **??**, we know that the equivalence classes partition (the underlying set of) G. Finally, the equivalence class $[x]$ is precisely the coset Hx. Indeed,

$$[x] = \{y \in G : y \sim x\} = \{y \in G : yx^{-1} \in H\} = \{y \in G : y \in Hx\} = Hx.$$ □

Exercise 16 Show that $\sim$ in the proof of Corollary 2.53 is an equivalence relation.

Definition 2.54 Let $H \leq G$ be a subgroup. The right index of H in G is the number of distinct right cosets of H in G, denoted $[G : H]$. Similarly, the left index is $[H : G]$.

Theorem 2.55 (Lagrange's Theorem) *Let $H \leq G$ be a subgroup. Then, $|G| = |H|[G : H]$.*

Proof: By Corollary 2.53, each coset of H in G has $|H|$ elements. Indeed, if $H = \{h_1, ..., h_k\}$, then $Hg = \{h_1 g, ..., h_k g\}$ and each element is distinct by the Cancellation Laws. Also, every element of G belongs to precisely **one** coset (by definition of a partition). From Definition 2.54, there are precisely $[G : H]$ many cosets in total. □

Note: Lagrange's Theorem really says that $|H|$ divides $|G|$ for any subgroup $H \leq G$.

Example 2.56 Consider Example 2.51(i), wherein $G = D_3$ and $H = \{I, R, S\}$. In this situation, we know that $|G| = 6$ and $|H| = 3$. Thus, Lagrange's Theorem implies the index of H in G is $[G : H] = 2$; this was the case in the aforementioned example.

Corollary 2.57 *The order of an element of a finite group divides the order of the group.*

Proof: For $g \in G$, $\text{ord}(g) = \left|\langle g \rangle\right|$ by Lemma 2.33. Apply Lagrange's Theorem to $H = \langle g \rangle$. □

Corollary 2.58 *Any group of prime order is cyclic.*

Proof: If $|G| = p$ for p prime, then any element $g \in G$ which is **not** the identity will have order p by Corollary 2.57. Consequently, such an element is a generator, i.e. $G = \langle g \rangle$. □

Corollary 2.59 *If $|G| = n$, then $g^n = 1$ for all $g \in G$.*

Exercise 17 Prove Corollary 2.59.

[**Hint:** Use Proposition 2.26(ii) and note that n is not necessarily the order of g.]

We can also finally provide a proof of Fermat's Little Theorem using these group-theoretic means. For reference, we stated the result in Theorem **??** (and alternate versions in Corollary **??**). In particular, we will prove the following result **without** using Theorem **??** (which is what we did in Exercise **??**).

Corollary 2.60 (Fermat's Little Theorem) *Let p be prime. If $a \in \mathbb{Z}$ such that $p \nmid a$, then $a^{p-1} \equiv 1 \pmod{p}$. Otherwise, $a^p \equiv a \pmod{p}$.*

Proof: Consider the multiplicative group $\mathbb{Z}_p^* = \{1, ..., p-1\}$ and let r be the remainder upon dividing $a \in \mathbb{Z}$ by p. Because we assume that $p \nmid a$, it means that the remainder $r \in \mathbb{Z}_p^*$. As a result of Corollary 2.59, we know that $r^{p-1} = 1$. As such, $a^{p-1} \equiv 1 \pmod{p}$. Otherwise, if $p \mid a$, then either both sides of the congruence are zero (so the theorem is trivially satisfied) or multiplying the formula just derived by a gives the desired result. □

Definition 2.61 Let $\varphi : G \to H$ be a group homomorphism.

(i) The kernel is the subset $\ker(\varphi) = \{g \in G : \varphi(g) = 1_H\}$.

(ii) The image is the subset $\text{im}(\varphi) = \{\varphi(g) \in H : g \in G\}$.

Here, the image is the same as in Definition **??** when applied to a map between two sets. The kernel is new; this has a broader definition (which will come later). In the context of group homomorphisms, it's the set of all elements of G that get sent to the identity in H. They are subsets, but we actually have more.

Proposition 2.62 *For $\varphi : G \to H$ a homomorphism, $\ker(\varphi) \leq G$ and $\text{im}(\varphi) \leq H$.*

Proof: For a homomorphism $\varphi : G \to H$, we will show that $\ker(\varphi) \leq G$ is a subgroup and leave the rest of the proof to do in Exercise 18. To do this, we appeal to the Subgroup Criterion. Indeed, let $g_1, g_2 \in \ker(\varphi)$. Then, $\varphi(g_1) = 1_H = \varphi(g_2)$. But now, by properties of homomorphisms, we see that $\varphi(g_1 g_2) = \varphi(g_1)\varphi(g_2) = 1_H 1_H = 1_H$, so $g_1 g_2 \in \ker(\varphi)$. Next, we use Lemma 2.45 to conclude immediately that $1_G \in \ker(\varphi)$ and (almost) immediately that if $g \in \ker(\varphi)$, then $\varphi(g^{-1}) = \varphi(g)^{-1} = 1_H^{-1} = 1_H$, i.e. $g^{-1} \in \ker(\varphi)$. By the Subgroup Criterion, $\ker(\varphi) \leq G$. □

Exercise 18 Prove that for $\varphi : G \to H$ a homomorphism, $\mathrm{im}(\varphi) \leq H$ is a subgroup.

Example 2.63 Here are some examples of kernels and images.

(i) The homomorphism $\exp : \mathbb{C} \to \mathbb{C}^*$ has kernel $2\pi i\mathbb{Z}$ and image $\mathbb{C}^*$.

(ii) The homomorphism $\varphi : \mathbb{Z} \to G$ given by $\varphi(k) = g^k$ has kernel $\mathrm{ord}(g)\mathbb{Z}$ and image $\langle g \rangle$.

(iii) The homomorphism $\psi : \mathbb{Z} \to \mathbb{Z}_n$ given by $\psi(k) = k \pmod n$ has kernel $n\mathbb{Z}$ and image $\mathbb{Z}_n$.

Lemma 2.64 *A group homomorphism $\varphi : G \to H$ is injective if and only if* $\ker(\varphi) = \{1_G\}$.

Proof: ($\Rightarrow$) Let φ be injective and $g \in \ker(\varphi)$. Then, since $\varphi(g) = 1_H = \varphi(1_G)$, the definition of injectivity implies that $g = 1_G$. Hence, $\ker(\varphi) = \{1_G\}$.

($\Leftarrow$) Let $\ker(\varphi) = \{1_G\}$ and $\varphi(g) = \varphi(h)$. Now, $\varphi(gh^{-1}) = \varphi(g)\varphi(h)^{-1} = \varphi(g)\varphi(g)^{-1} = 1_H$, so we conclude $gh^{-1} \in \ker(\varphi)$, meaning $gh^{-1} = 1_G$ and so $g = h$. Hence, φ is injective. □

Exercise 19 (Harder) Let $\varphi : G \to H$ and $\psi : G \to H$ be homomorphisms. Prove that

$$\mathrm{Eq}(\varphi, \psi) := \{g \in G : \varphi(g) = \psi(g)\} \leq G$$

is a subgroup, called the equaliser of φ and ψ, using the Subgroup Criterion.

Classification of Groups

One of the milestones in mathematics research was the classification of finite simple groups. In fact, the proof consists of tens of thousands of pages in several hundred journal articles written by about 100 authors between the years 1955 and 2004. We won't dwell on this but 'simple' is defined in Section 3. There are a few results we can prove pretty much immediately.

Definition 2.65 The cyclic group of order n is the group $C_n = \langle x \rangle$ with $\mathrm{ord}(x) = n$.

Theorem 2.66 *Up to isomorphism, the only group of prime order p is C_p.*

Proof: This is a direct consequence of Corollary 2.58. □

Theorem 2.67 *Up to isomorphism, the groups of order four are C_4 and $C_2 \times C_2$.*

Note: The group $C_2 \times C_2$ is often called the Klein Vierergruppe, denoted V.

Proof: Let G be a group of order four. If it is cyclic, then $G \cong C_4$ by Theorem 2.47. Otherwise, if it isn't cyclic, then every element has either order one or two, meaning that all elements satisfy $g^2 = 1$. This completely determines the group table to be that of the Klein Vierergruppe. □

Exercise 20 Construct the group table for the Klein Vierergruppe, generated by a, b, c.

[**Hint:** You may use the fact it is really only generated by a and b, since $c = ab$.]

Lemma 2.68 *Let G be a group such that $\text{ord}(g) = 2$ for all $g \neq 1$. Then, G is Abelian.*

Proof: Because $g^2 = 1$ for all $g \neq 1$ (well, this is also true for the identity), it means that $g^{-1} = g$. In other words, every element is its own inverse. Consequently,

$$g_1 g_2 g_1^{-1} g_2^{-1} = g_1 g_2 g_1 g_2 = (g_1 g_2)^2 = 1,$$

which is to say $g_1 g_2 = g_2 g_1$, i.e. the operation is commutative and thus G is Abelian. □

Theorem 2.69 *Up to isomorphism, the groups of order six are C_6 and D_3.*

Proof: Let G be a group of order six. If it is cyclic, then $G \cong C_4$ by Theorem 2.47. Otherwise, if it isn't cyclic, then it contains an element of order three. If not, then all non-trivial elements have order two and, by Lemma 2.68, it means G is commutative, so $H = \{1, x, y, xy\} \leq G$ is a subgroup for $x, y \in G$. However, this contradicts Lagrange's Theorem because $4 \nmid 6$. Now, let $h \in G$ be such that $\text{ord}(h) = 3$ and consider the subgroup $H = \{1, h, h^2\}$. Consider now an element $g \in G \setminus H$ which is **not** in H. By Corollary 2.53, we know that $G = H \cup Hg = \{1, h, h^2, g, hg, h^2 g\}$. In the case that $\text{ord}(g) = 3$, it follows that $g^2 \notin \{1, g, hg, h^2 g\}$. Thus, the only option is to have $g^2 \in \{h, h^2\}$. However, we get a contradiction in that $g \in H$ because

$$g = (g^2)^2 \in \{h^2, h^4\} = \{h^2, h\},$$

using the fact that $\text{ord}(h) = 3$. Consequently, the only option is for $\text{ord}(g) = 2$ and, similarly, we know that $\text{ord}(hg) = 2 = \text{ord}(h^2g)$ also. In particular, this means that $hghg = 1$ and, multiplying by h^2 on the left and g on the right, $gh = h^2g$. This is enough information to complete the group table for G, from which it will be the 'same' as Table 2, meaning $G \cong D_3$. □

Definition 2.70 The quaternions are expressions of the form

$$\{a + bi + cj + dk : a, b, c, d \in \mathbb{R} \text{ and } i^2 = j^2 = k^2 = ijk = -1\}.$$

Lemma 2.71 *The quaternion units i, j, k satisfy the following properties:*

(i) $ij = k$ *and* $ji = -k$.
(ii) $jk = i$ *and* $kj = -i$.
(iii) $ki = j$ *and* $ik = -j$.

Proof: Using the facts $ijk = -1$ and $k^2 = -1$, we see that $ijk^2 = -k \Leftrightarrow -ij = -k \Leftrightarrow ij = k$. Similarly, $i^2jk = -i \Leftrightarrow jk = i$, from which it follows that $i^2jki = -i^2 = 1 \Leftrightarrow jki = -1$. Applying a similar procedure, we see that $j^2ki = -j \Leftrightarrow ki = j$. The anti-commutativity, the property that swapping the order in each of the first equations in (i), (ii) and (iii) gives a minus sign, comes from similar manipulations. □

Exercise 21 Prove the anti-commutativity statements in Lemma 2.71, that is

$$ji = -k, \quad kj = -i, \quad ik = -j.$$

Definition 2.72 The quaternion group is the group $Q = \{\pm 1, \pm i, \pm j, \pm k\}$ under the multiplication operation defined by the relations in Definition 2.70 and Lemma 2.71.

Theorem 2.73 *Up to isomorphism, the groups of order eight are*

$$C_8, \quad C_4 \times C_2, \quad C_2 \times C_2 \times C_2, \quad D_4, \quad Q.$$

Proof: Omitted. □

3 Normal Subgroups

The discussion now takes a turn to look at a particular class of subgroup. These are very important in group theory, in part because they allow us to construct 'new groups from old' by taking the so-called *quotient* of a group. First of all, we will get to grips with conjugation in a group (as we will see, this is really the key idea behind a normal subgroup).

Definition 3.1 Let G be a group and $x, y \in G$ be distinct. They are right conjugate if there exists $g \in G$ such that $y = g^{-1}xg$. Similarly, they are left conjugate if $y = gxg^{-1}$. The set of all elements conjugate to a given element is called a conjugacy class. The conjugacy class containing x is the set $\text{conj}_G(x) \coloneqq \{g^{-1}xg : g \in G\}$.

Example 3.2 Here are some examples of conjugacy classes in groups.

(i) For any group G, we have $\text{conj}_G(1) = \{1\}$ since $g^{-1}1g = g^{-1}g1 = 1$.

(ii) For an Abelian group G, we have $\text{conj}_G(x) = \{x\}$ since $g^{-1}xg = g^{-1}gx = x$.

(iii) For the dihedral group D_3, the conjugacy classes are $\{I\}, \{R, S\}, \{A, B, C\}$.

Theorem 3.3 *A group is the disjoint union of its conjugacy classes.*

Proof: It is sufficient to prove that conjugacy is an equivalence relation. In this vein, we define a relation on G as follows: $x \sim y$ if and only if there exists $g \in G$ such that $y = g^{-1}xg$. This is an equivalence relation. (Symmetry) Clearly, $x \sim x$ because we can take $g = 1$. (Reflexivity) If $x \sim y$, then $y = g^{-1}xg$; this is to say $x = gyg^{-1}$ and since $g^{-1} \in G$ by closure under inverses, we have that $y \sim x$. (Transitivity) If $x \sim y$ and $y \sim z$, then $y = g^{-1}xg$ and $z = h^{-1}yh$. But now, we see that $z = h^{-1}(g^{-1}xg)h = (gh)^{-1}x(gh)$ and since $gh \in G$ by closure under the operation, we have that $x \sim z$. Consequently, Theorem **??** implies the result. □

Proposition 3.4 *Conjugate elements have the same order.*

Proof: Suppose that x and y are conjugate in G, meaning $y = g^{-1}xg$ for some $g \in G$. Assume further that $\text{ord}(y) = n$. Then, we see that $y^n = 1$ is equivalent to

$$(g^{-1}xg)^n = 1 \quad \Leftrightarrow \quad (g^{-1}xg)(g^{-1}xg)\cdots(g^{-1}xg) = 1 \quad \Leftrightarrow \quad g^{-1}x^ng = 1 \quad \Leftrightarrow \quad x^n = 1.$$

As n is minimal with $y^n = 1$, it is minimal such that $x^n = 1$. We conclude that $\text{ord}(x) = n$. □

Definition 3.5 Let G be a group and $x \in G$. Then centraliser of x in G is the set

$$\mathrm{C}_G(x) = \{g \in G : gx = xg\}.$$

Note: We call $\mathrm{C}_G(x)$ the centraliser subgroup of x, because it is a subgroup (prove this).

Theorem 3.6 *Let $x \in G$. Then, $\left|\mathrm{conj}_G(x)\right| = [G : \mathrm{C}_G(x)]$.*

Proof: We begin by showing that the cosets of $\mathrm{C}_G(x)$ in G are equal for elements in the same conjugacy class. So, let $g, h \in G$ and suppose that $g^{-1}xg = h^{-1}xh$. If we multiply on the right by h^{-1} and on the left by g, we get $xgh^{-1} = gh^{-1}x$. By definition, this means that $gh^{-1} \in \mathrm{C}_G(x)$ because it commutes with x. Looking back to Proposition 2.52(v), this means that $\mathrm{C}_G(x)g = \mathrm{C}_G(x)h$ are equal cosets. Therefore, the number of conjugacy classes is just the number of these cosets, which is precisely what the statement above says. □

Remark 3.7 There is a significant generalisation to Theorem 3.6 discussed in Chapter **??**; it is clear that the above theorem is analogous to Theorem **??** in Section **??** (where the group action is conjugation). More will be said in the aforementioned chapter.

We can now make the main definition of this section and develop some related theory.

Definition 3.8 A subgroup $N \leq G$ is called normal if $g^{-1}ng \in N$ for all $g \in G$ and $n \in N$. In other words, N is closed under conjugation. This is denoted $N \trianglelefteq G$.

Example 3.9 Here are some examples and non-examples of normal subgroups.

(i) For any group G, we have $\{1\} \trianglelefteq G$ and $G \trianglelefteq G$.

(ii) For an Abelian group G, any subgroup $H \trianglelefteq G$ is normal.

(iii) For D_3, the subgroup $\{I, R, S\} \trianglelefteq D_3$ but the subgroups $\{I, A\}, \{I, B\}, \{I, C\} \not\trianglelefteq D_3$.

Note: If a group's only normal subgroups are trivial, $\{1\}$ and G itself, it is called simple.

Proposition 3.10 *For a homomorphism $\varphi : G \to H$, the kernel $\ker(\varphi) \trianglelefteq G$ is normal.*

Proof: We already showed that $\ker(\varphi) \leq G$ in Proposition 2.62. It remains to show that it is closed under conjugation by elements of G. Indeed, if $x \in \ker(\varphi)$ and $g \in G$, we see that $\varphi(g^{-1}xg) = \varphi(g)^{-1}\varphi(x)\varphi(g) = \varphi(g)^{-1}1_H\varphi(g) = \varphi(g)^{-1}\varphi(g) = 1_H$, meaning $g^{-1}xg \in \ker(\varphi)$. □

Exercise 22 For a homomorphism $\varphi : G \to H$, is $\mathrm{im}(\varphi) \trianglelefteq H$ normal? Justify your claim.

Lemma 3.11 *Let $N \leq G$. Then, N is normal if and only if $Ng = gN$ for all $g \in G$.*

Proof: ($\Rightarrow$) If $N \trianglelefteq G$, then $g^{-1}ng \in N$ for all $n \in N$ and $g \in G$. But this means that $ng = g(g^{-1}ng) \in gN$ (since $g^{-1}ng \in N$), which means that $Ng \subseteq gN$. For the other inclusion, we can see that $gng^{-1} = (g^{-1})^{-1}n(g^{-1}) \in N$, from which we get $gn = (gng^{-1})g \in Ng$, implying that $gN \subseteq Ng$. Combining both inclusions gives the desired equality.

($\Leftarrow$) If the cosets $Ng = gN$ coincide, this means that any element of the right coset can be expressed as an element of the left coset. In particular, where $n \in N$, this means $ng = gm$ for some $m \in N$. Hence, $g^{-1}ng = m \in N$ and this is what it means to be normal. □

Corollary 3.12 *Any subgroup of index two is normal.*

Proof: Let $H \leq G$ be a subgroup such that $[G : H] = 2 = [H : G]$. By definition, there are only two right cosets; because H is always a coset, it follows from Corollary 2.53 that the other coset is $G \setminus H$, i.e. everything **not** in H. This is identical for the left cosets. By Lemma 3.11, the left and right cosets coincide, so $H \trianglelefteq G$ is normal. □

Corollary 3.13 *Let $N \trianglelefteq G$. For $x, y \in G$, it follows that $(Nx)(Ny) = N(xy)$.*

Proof: Well, associativity tells us that $(Nx)(Ny) = N(xN)y = N(Nx)y = N(xy)$, where we applied Lemma 3.11 to get the second equality. Finally, note that $NN \coloneqq \{nm : n, m \in N\}$ is just the subgroup N itself because it is closed under the operation and contains the identity. □

We are nearly ready to define the next important object in this section, but first an example.

Example 3.14 We again consider the dihedral group D_3 and its group table, presented in Table 2. For completion (spoilers for Exercise 3), we write out the table in full below, which we separate into blocks. In fact, we can see that this defines a natural multiplication of cosets as in Table 4b, where $H = \{I, R, S\} \trianglelefteq D_3$ as in Example 3.9 and Example 2.51(i).

$\circ$	I	R	S	A	B	C
I	I	R	S	A	B	C
R	R	S	I	C	A	B
S	S	I	R	B	C	A
A	A	B	C	I	R	S
B	B	C	A	S	I	R
C	C	A	B	R	S	I

a The completed group table for D_3.

	HI	HA
HI	HI	HA
HA	HA	HI

b The coset group table for D_3.

Table 4: The multiplication of cosets in D_3.

Definition 3.15 If $N \trianglelefteq G$ is a normal group, then the set of cosets of N in G equipped with the operation from Corollary 3.13 is the quotient group of G by N, denoted G/N. We call the homomorphism $\pi : G \to G/N$ given by $\pi(g) = Ng$ the canonical projection.

Exercise 23 Show that the operation from Corollary 3.13 gives a group structure on G/N.

Lemma 3.16 *The canonical projection is a surjective homomorphism with* $\ker(\pi) = N$.

Proof: Showing that π is a homomorphism is standard: $\pi(gh) = N(gh) = (Ng)(Nh) = \pi(g)\pi(h)$ for $g, h \in G$, where we use Corollary 3.13 in the second equality. Next, the pre-image of the element $Ng \in G/N$ is simply g (that is to say we have found an element of G to get to any element of G/N), so π is surjective. Finally, $g \in \ker(\pi)$ if and only if $Ng = N$, which is to say that $g \in N$ by Proposition 2.52. In other words, $\ker(\pi) = N$, as required. □

Example 3.17 Consider the additive group $G = \mathbb{Z}$. Because it is Abelian by Example 2.10(i), we know that all subgroups are normal by Example 3.9. In particular, the subgroup $n\mathbb{Z} \trianglelefteq \mathbb{Z}$ of multiples of n is normal. The cosets of $n\mathbb{Z}$ in $\mathbb{Z}$ are precisely $n\mathbb{Z} + 0, ..., n\mathbb{Z} + (n-1)$, a generalisation of Example 2.51(ii). The canonical projection is the map $\pi : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ whereby $k \mapsto k \pmod n$. This looks awfully similar to $\mathbb{Z}_n$, the integers under addition modulo n....

Theorem 3.18 (First Isomorphism Theorem) *For a group homomorphism* $\varphi : G \to H$,

$$G/\ker(\varphi) \cong \operatorname{im}(\varphi).$$

Proof: The isomorphism is $f : G/\ker(\varphi) \to \operatorname{im}(\varphi)$, where $f\big(\ker(\varphi)g\big) = \varphi(g)$. It is well-defined and injective since $\ker(\varphi)g_1 = \ker(\varphi)g_2 \Leftrightarrow g_1 g_2^{-1} \in \ker(\varphi) \Leftrightarrow \varphi(g_1 g_2^{-1}) = 1_H \Leftrightarrow \varphi(g_1) = \varphi(g_2)$. It remains to prove surjectivity: $\operatorname{im}(f) = \{f\big(\ker(\varphi)g\big) : g \in G\} = \{\varphi(g) : g \in G\} = \operatorname{im}(\varphi)$. □

Note: In the above proof, we may write f as $\overline{\varphi}$ to emphasise the dependence on φ.

Exercise 24 Explain the equivalences for the injectivity in the proof of Theorem 3.18.

Example 3.63 (Continued) Here are some uses of the First Isomorphism Theorem.

(i) We have an isomorphism $\mathbb{C}/2\pi i\mathbb{Z} \cong \mathbb{C}^*$.

(ii) We have an isomorphism $\mathbb{Z}/\operatorname{ord}(g)\mathbb{Z} \cong \langle g \rangle$.

(iii) We have an isomorphism $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}_n$.

In fact, (ii) and (iii) suggest an additional isomorphism between cyclic groups and the integers under addition modulo some number. In particular, the map $f : \mathbb{Z} \to \langle g \rangle$ given by $f(k) = g^k$ will **almost** work. However, if we say that $\operatorname{ord}(g) = n$, then we can see that $\ker(f) = n\mathbb{Z}$ but it is a surjective map. Hence, the First Isomorphism Theorem implies that $\mathbb{Z}_n \cong \langle g \rangle$.

Remark 3.92 Since we know that the canonical projection π is surjective, the First Isomorphism Theorem implies that a subgroup $N \leq G$ is normal if and only if it is the kernel of some group homomorphism from G to another group (said other group is isomorphic to G/N).

Corollary 3.93 *Let G and H be finite groups and $\varphi : G \to H$ be a homomorphism. Then, $\big|\operatorname{im}(\varphi)\big|$ divides both the orders $|G|$ and $|H|$.*

Proof: By the First Isomorphism Theorem, we know $G/\ker(\varphi) \cong \operatorname{im}(\varphi)$. Thus, it follows that $|G|/\big|\ker(\varphi)\big| = \big|\operatorname{im}(\varphi)\big|$ since bijections, in general, preserve cardinalities (Definition **??**). An immediate consequence is that $\big|\operatorname{im}(\varphi)\big|$ divides $|G|$. Next, since $\operatorname{im}(\varphi) \leq H$, by Proposition 2.62, we directly apply Lagrange's Theorem to conclude that $\big|\operatorname{im}(\varphi)\big|$ divides $|H|$. □

Corollary 3.94 *Let G and H be finite groups whose orders $|G|$ and $|H|$ are coprime. Then, any homomorphism $\varphi : G \to H$ is trivial.*

Proof: By Corollary 3.93, $\big|\operatorname{im}(\varphi)\big|$ divides both $|G|$ and $|H|$. Because the aforementioned numbers are coprime, it must be that $\big|\operatorname{im}(\varphi)\big| = 1$. The only option for a homomorphism $\varphi : G \to H$ is that which is defined by $\varphi(g) = 1_H$. In other words, the homomorphism is trivial. □

Exercise 25 For an arbitrary group G, prove that (i) $G/\{1\} \cong G$ and (ii) $G/G \cong \{1\}$.

[**Hint:** For each, construct a homomorphism and apply the First Isomorphism Theorem.]

There are an additional two isomorphism theorems that are widely known in group theory. They concern isomorphisms involving normal subgroups and even getting to their statements is more challenging. Bear with and we will soon be computing isomorphisms like nobody's business.

Lemma 3.95 *Let G be a group and $\mathcal{H}$ be a (possibly infinite) family of subgroups of G. Then, the intersection of all these subgroups $\bigcap_{H\in\mathcal{H}} H \leq G$ is itself a subgroup.*

Proof: We will show that the intersection of **two** subgroups $H, K \leq G$ is itself a subgroup and then we can extend this arbitrarily. Indeed, we apply the Subgroup Criterion to $H \cap K$. First, if $x, y \in H \cap K$, then $x, y \in H$ and $x, y \in K$ by definition of the intersection, but because these are both subgroups, they are closed under the operation and forming inverses. Thus, $xy \in H$ and $xy \in K$ (which means that $xy \in H \cap K$) as well as $x^{-1} \in H$ and $x^{-1} \in K$ (which means that $x^{-1} \in H \cap K$). Lastly, $1 \in H$ and $1 \in K$, so $1 \in H \cap K$. The Subgroup Criterion implies that $H \cap K \leq G$. A near-identical argument gives the result for a generic-sized family $\mathcal{H}$. □

Corollary 3.96 *Let G be a group and $\mathcal{N}$ be a (possibly infinite) family of normal subgroups of G. Then, the intersection of all these subgroups $\bigcap_{N\in\mathcal{N}} N \trianglelefteq G$ is itself normal.*

Proof: By Lemma 3.95, we know that $\bigcap_{N\in\mathcal{N}} N \leq G$ is at least a subgroup. For normality, suppose we have $n \in \bigcap_{N\in\mathcal{N}} N$ and $g \in G$. By definition of the intersection, $n \in N$ for every normal subgroup in our family, $N \in \mathcal{N}$. By normality in each of these subgroups, we know that $g^{-1}ng \in N$ for every $N \in \mathcal{N}$. Hence, it is true that $g^{-1}ng \in \bigcap_{N\in\mathcal{N}} N$. □

Lemma 3.97 *Let G be a group, $H \leq G$ a subgroup and $N \trianglelefteq G$ a normal subgroup. Then, $N \cap H \trianglelefteq H$ is normal in H.*

Proof: First, Lemma 3.95 implies that $N \cap H \leq H$ is a subgroup of H. This is because $N \cap H \subseteq H$ is at least a subset, by definition of intersection (Definition **??**), and the group structure comes from the fact that $N, H \leq G$ are both subgroups. It remains to show normality. Well, let $x \in N \cap H$ (meaning $x \in N$ and $x \in H$) and we will conjugate by $h \in H$. But we know that $h^{-1}xh \in N$ because $N \trianglelefteq G$ is normal (and $h \in G$ since $H \subseteq G$). Since H is closed under the operation and forming inverses, we also have $h^{-1}xh \in H$. Therefore, $h^{-1}xh \in N \cap H$. □

Exercise 26 (Longer) Let G be a group, $H \leq G$ and $N \trianglelefteq G$. Prove the following:

(i) $NH = HN$, where $NH \coloneqq \{nh : n \in N \text{ and } h \in H\}$.

[**Hint:** You may find it useful to work with $NH = \bigcup_{h \in H} Nh$.]

(ii) $NH \leq G$ is a subgroup.

[**Hint:** Be aware of Remark 3.98.]

(iii) $N \trianglelefteq NH$ is a normal subgroup.

Remark 3.98 If $AB = BA$ are equal sets, it does **not** imply that $ab = ba$ (this might be true, but this isn't true in general). The only thing that can be concluded is $ab = \beta\alpha$ for some $\alpha \in A$ and $\beta \in B$. In words, this says the elements aren't necessarily commutative, but we can always re-write the elements so that they are of the form

$$[\text{something in } A][\text{something in } B] = [\text{something (else) in } B][\text{something (else) in } A].$$

Theorem 3.99 (Second Isomorphism Theorem) *For G a group, $H \leq G$ and $N \trianglelefteq G$,*

$$H/(H \cap N) \cong NH/N.$$

Proof: Define the map $f : H \to NH/N$ by $f(h) = Nh$. Now, Lemma 3.97 and Exercise 26 imply that this map is well-defined. It remains to show that it is a homomorphism and to compute its kernel and image. Well, for $h_1, h_2 \in H$, we see $f(h_1h_2) = N(h_1h_2) = (Nh_1)(Nh_2) = f(h_1)f(h_2)$, by using Corollary 3.13. Next, suppose that $N(nh) \in NH/N$, where $n \in N$ and $h \in H$. Then, $N(nh) = (Nn)h = Nh = f(h)$, so f is surjective. Finally,

$$\begin{aligned} \ker(f) &= \{h \in H : Nh = N\} \\ &= \{h \in H : h \in N\} \\ &= H \cap N, \end{aligned}$$

where the second equality comes from Proposition 2.52 and the third equality is the definition of an intersection. Hence, applying the First Isomorphism Theorem means that

$$H/\ker(f) \cong \operatorname{im}(f) \qquad \Leftrightarrow \qquad H/(H \cap N) \cong NH/N. \qquad \square$$

Example 3.100 Let $G = \mathbb{Z}$, $N = 10\mathbb{Z}$ and $H = 4\mathbb{Z}$. By the Second Isomorphism Theorem,

$$4\mathbb{Z}/20\mathbb{Z} \quad = \quad 4\mathbb{Z}/(4\mathbb{Z} \cap 10\mathbb{Z}) \quad \cong \quad (10\mathbb{Z} + 4\mathbb{Z})/10\mathbb{Z}.$$

Be aware that because G here is an *additive* group, it follows that the operations on the cosets is also additive, so in the statement of the Second Isomorphism Theorem, instead of NH (which is *multiplicative* notation), we will have $N + H$, which is exactly what we have above.

Lemma 3.101 *Let G be a group with normal subgroups $M, N \trianglelefteq G$ such that $M \leq N$.*
(i) *$M \trianglelefteq N$ is a normal subgroup.*
(ii) *$N/M \trianglelefteq G/M$ is a normal subgroup.*

Proof: (i) We already assume that $M \leq N$, so now we take $m \in M$ and $n \in N$. Because $M \trianglelefteq G$, it is true $g^{-1}mg \in M$ for **any** $g \in G$. In particular, this holds for $n \in N \subseteq G$, as required.

(ii) Because $N \leq G$, it follows that $N/M \leq G/M$, so now we take $Mn \in N/M$ and $Mg \in G/M$, where $n \in N$ and $g \in G$. Then, $(Mg)^{-1}(Mn)(Mg) = M(g^{-1}ng) \in N/M$ by the fact $N \trianglelefteq G$. □

Theorem 3.102 (Third Isomorphism Theorem) *For G a group, $M, N \trianglelefteq G$ and $M \leq N$,*

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

Exercise 27 Prove the Third Isomorphism Theorem.

[**Hint:** Apply the First Isomorphism Theorem to $f : G/M \to N/M$ where $f(Mg) = Ng$.]

Example 3.103 Let $G = \mathbb{Z}$, $M = 12\mathbb{Z}$ and $N = 3\mathbb{Z}$. By the Third Isomorphism Theorem,

$$(\mathbb{Z}/12\mathbb{Z})/(3\mathbb{Z}/12\mathbb{Z}) \quad \cong \quad \mathbb{Z}/3\mathbb{Z} \quad = \quad \mathbb{Z}_3.$$

4 Vector Spaces

A very important related notion to a group is that of a *vector space*. It is likely you will have encountered *vectors* before in high school. Here, the meaning is more general than it (likely) was back then. We will omit a little bit of generality; really we can only define a vector space 'over a field', whatever that means. We will define a field but a more valid introduction is presented in Chapter **??**. We then look at a number of vector spaces consisting of so-called *matrices*.

Definition 4.1 A field is a set K with two binary operations $+$ and $\times$ satisfying these:

(i) $(K, +)$ is an Abelian group with additive identity 0.

(ii) $(K^*, \times)$ is an Abelian group with multiplicative identity 1.

(iii) For all $a, b, c \in K$, we have $a \times (b + c) = (a \times b) + (a \times c)$. (Distributivity)

Example 4.2 Here are some examples and non-examples of fields.

(i) $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ **are** all fields with their usual addition and multiplication operations.

(ii) $\mathbb{Z}$ is **not** a field because $(\mathbb{Z}^*, \times)$ isn't a group.

(iii) $\mathbb{Z}_p$ for p prime **is** a field, with addition and multiplication modulo p.

Note: When discussing the *field* $\mathbb{Z}_p$, we denote it by $\mathbb{F}_p$ to emphasise that we are talking about the field and not the group. This is then called the field with p elements.

Exercise 28 Give an example of some $\mathbb{Z}_n$ which is **not** a field and explain why.

[**Hint:** By Example 4.2(iii), it must be that n isn't a prime number.]

Definition 4.3 Let K be a field, whose elements are called scalars. A K-vector space (or vector space over K) is a set V, whose elements are called vectors, with the operations vector addition $+ : V \times V \to V$ and scalar multiplication $\cdot : K \times V \to V$ satisfying these:

(i) $(V, +)$ is an Abelian group, with additive identity $\mathbf{0}$.

(ii) For all $\lambda, \mu \in K$ and $\mathbf{v} \in V$, we have $(\lambda\mu)\mathbf{v} = \lambda(\mu\mathbf{v})$. (Compatibility)

(iii) For all $\mathbf{v} \in V$, we have $1\mathbf{v} = \mathbf{v}$. (Scalar Identity)

(iv) For all $\lambda \in K$ and $\mathbf{u}, \mathbf{v} \in V$, we have $\lambda(\mathbf{u} + \mathbf{v}) = \lambda\mathbf{u} + \lambda\mathbf{v}$. (Distributivity)

(v) For all $\lambda, \mu \in K$ and $\mathbf{v} \in V$, we have $(\lambda + \mu)\mathbf{v} = \lambda\mathbf{v} + \mu\mathbf{v}$. (Distributivity)

Example 4.4 Here are some examples of vector spaces.

(i) The space $\mathbb{R}^n$ of columns of n real numbers forms an $\mathbb{R}$-vector space, with pointwise addition

and scalar multiplication. In general, the space K^n of columns of n entries of the field K forms a K-vector space. Even more generally, the space K^∞ of columns of infinitely-many entries of the field K forms a K-vector space.

(ii) The set $\mathbb{Q}[x]$ of polynomials with coefficients in $\mathbb{Q}$ in the indeterminate x (expressions of the form $a_0+a_1x+\cdots+a_nx^n$ with each $a_i \in \mathbb{Q}$) is a $\mathbb{Q}$-vector space, with addition given by the addition of polynomials and scalar multiplication simply multiplication by a constant.

(iii) The field $\mathbb{C}$ is actually an $\mathbb{R}$-vector space, where addition is the usual addition of complex numbers and scalar multiplication is the usual product λz, for $\lambda \in \mathbb{R}$ and $z \in \mathbb{C}$.

Exercise 29 Give an example of a field K and a set V where V is **not** a K-vector space.

Notation 4.5 Hence, we denote scalars by non-bold letters (often Greek) and vectors by bold letters (often Latin). In later chapters, it may be convenient to just call a vector v but, here and where possible, we will write $\mathbf{v}$. Note that it is common to also see vectors denoted by $\underline{v}$ and $\vec{v}$.

We will spend tome time looking at the vector spaces $\mathbb{R}^n$ (which we will always, unless otherwise stated, assume are studied over the field $\mathbb{R}$). This will allow us to get to grips with a very understandable and concrete idea before we jump into the abstract world once again.

The Real Vector Space $\mathbb{R}^2$

The vector space $\mathbb{R}^2$ consists of elements with two rows and one column:

$$\mathbb{R}^2 = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\}.$$

We can give a geometric picture to these vectors. Indeed, if we are stood at the origin $(0,0)$, then the vector above tells us to 'move' x_1 in the x-direction and 'move' x_2 in the y-direction. In this way, whether the entries x_i are positive or negative tell us to move right/left, respectively, and up/down, respectively. This is captured in Figure 4 below (in which $x_1 > 0$ and $x_2 > 0$).

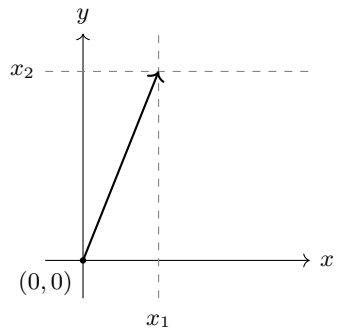


Figure 4: The geometric interpretation of vectors in $\mathbb{R}^2$.

This essentially allows us to identify the vector space $\mathbb{R}^2$ with the plane $\mathbb{R}^2$, as in Section **??**.

> **Note:** There is something **much** deeper happening here; what the above says is that columns $\binom{x}{y}$ are 'the same' as rows $(x\ y)$. In fact, this is identifying the vector space $\mathbb{R}^2$ with its so-called *dual space*; this consists of the rows, so-called *covectors*. In this way, we write $\binom{x}{y} = (x\ y)^T$ (the superscript T stands for **transpose**, defined shortly).

Remark 4.6 A vector space is defined in terms of a group-theoretic object (Chapter **??**), which is an extension of the theory of sets (Chapter **??**); they can also have a geometric interpretation (Chapter **??**). This is a major overlap between the three areas studied thus far.

The Real Vector Space $\mathbb{R}^3$

In a near-identical way, the vector space $\mathbb{R}^3$ consists of elements with three rows and one column:

$$\mathbb{R}^3 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : x_1, x_2, x_3 \in \mathbb{R} \right\}.$$

The corresponding geometric picture to these vectors is similar to that of $\mathbb{R}^2$. Indeed, if we stand at the origin $(0, 0, 0)$, then the vector above tells us to 'move' x_1 in the x-direction, 'move' x_2 in the y-direction and 'move' x_3 in the z-direction. We can think of this as first moving in $\mathbb{R}^2$ by the vector $\binom{x_1}{x_2}$ (where we view $\mathbb{R}^2$ as lying flat; think of it like the top of a horizontal table) and then we move x_3 above/below the table. This idea is captured in Figure 5 below, where we 'lift' the two-dimensional vector to the three-dimensional one.

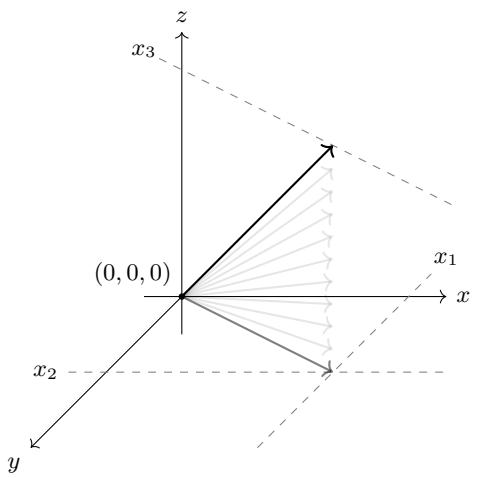


Figure 5: The geometric interpretation of vectors in $\mathbb{R}^3$.

The Real Vector Space $\mathbb{R}^n$

We can extend the ideas discussed above to work in any dimension. The only problem is that it is difficult (essentially impossible) to visualise what is happening above three dimensions. Sometimes, a picture can help whereby the plane (table top) represents $\mathbb{R}^{n-1}$ (all dimensions except one) and the protruding axis (z-axis) represents $\mathbb{R}$ (the missing dimension); this it is very abstract and really just a guide.

Now that we have some intuition, we will return to the abstract meaning of a vector space as in Definition 4.3. But we can apply any of what comes hence to $\mathbb{R}^n$, so the picture to have in mind is $\mathbb{R}^2$ or $\mathbb{R}^3$.

Proposition 4.7 *Let V be a vector space over a field K.*

(i) *For all $\lambda \in K$, we have $\lambda\mathbf{0} = \mathbf{0}$.*
(ii) *For all $\mathbf{v} \in V$, we have $0\mathbf{v} = \mathbf{0}$.*
(iii) *If $\lambda \in K$ and $\mathbf{v} \in V$ such that $\lambda\mathbf{v} = \mathbf{0}$, then either $\lambda = 0$ or $\mathbf{v} = \mathbf{0}$.*
(iv) *For all $\lambda \in V$ and $\mathbf{v} \in V$, we have $(-\lambda)\mathbf{v} = -(\lambda\mathbf{v})$.*

Proof: (i) Since $\mathbf{0}$ is the additive identity, we have $\mathbf{0} + \mathbf{0} = \mathbf{0}$. Therefore, $\lambda(\mathbf{0} + \mathbf{0}) = \lambda\mathbf{0}$, which is to say $\lambda\mathbf{0} + \lambda\mathbf{0} = \lambda\mathbf{0}$ by distributivity. Subtracting $\lambda\mathbf{0}$ from both sides gives the result.

The proofs of (ii), (iii) and (iv) also use the vector space axioms and are thus similar. □

Exercise 30 Complete the proof of Proposition 4.7.

By definition, a vector space is essentially an Abelian group with some extra structure (the scalar multiplication map considered in the definition of a vector space). We know that groups have subgroups, so we should expect an analogy here.

Definition 4.8 Let V be a K-vector space. A subspace of V is a subset $U \subseteq V$ such that U is also a vector space under the same operations of vector addition and scalar multiplication as are defined for V.

Theorem 4.9 (Subspace Criterion) *Let V be a K-vector space. A subset $U \subseteq V$ is a subspace if and only if it satisfies the following properties:*

(i) $\mathbf{0} \in U$.

(ii) *For all* $\mathbf{u}, \mathbf{w} \in U$, *we have* $\mathbf{u} + \mathbf{w} \in U$.

(iii) *For all* $\lambda \in K$ *and* $\mathbf{u} \in U$, *we have* $\lambda\mathbf{u} \in U$.

Proof: This is near-identical to that of the Subgroup Criterion, so we omit it. □

Note: We can state the conditions of the Subgroup Criterion in words as follows:

(i) U contains the zero vector.

(ii) U is closed under vector addition.

(iii) U is closed under scalar multiplication.

Example 4.10 Here are some examples and non-examples of subspaces.

(i) For any vector space V, we have the trivial subspace $\{\mathbf{0}\}$ and V itself as subspaces.

(ii) The subspaces of $\mathbb{R}$ are the trivial subspace $\{\mathbf{0}\}$ and the space itself.

(iii) The subspaces of $\mathbb{R}^2$ are the $\{\mathbf{0}\}$, $\mathbb{R}^2$ itself and any line which passes through the origin.

(iv) The subspaces of $\mathbb{R}^3$ are the $\{\mathbf{0}\}$, $\mathbb{R}^3$ itself, any line which passes through the origin and any plane which contains the origin.

(v) The subset $\{(x, y) : y = 2x + 1\} \subseteq \mathbb{R}^2$ is **not** a subspace because it doesn't contain the zero vector (i.e. it describes a line which does not pass through the origin).

(vi) The subspace $P_n := \{a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 : a_i \in \mathbb{R}\} \subseteq \mathbb{R}[x]$ is a subspace of the vector space of polynomials in one variable with coefficients in $\mathbb{R}$. This subspace P_n contains polynomials of degree at most n (we will discuss this in more depth later).

Definition 4.11 Let V and W each be K-vector spaces. A map $f : V \to W$ is called a linear map (or linear transformation) if it satisfies the following two properties:

(i) For all $\mathbf{v}_1, \mathbf{v}_2 \in V$, we have $f(\mathbf{v}_1 + \mathbf{v}_2) = f(\mathbf{v}_1) + f(\mathbf{v}_2)$.

(ii) For all $\lambda \in K$ and $\mathbf{v} \in V$, we have $f(\lambda \mathbf{v}) = \lambda f(\mathbf{v})$.

If a linear map is bijective, we call it a linear isomorphism and denote this by $V \cong W$.

Note: We can restate the conditions of Definition 4.11 as the following single condition:

$$f(\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2) = \lambda_1 f(\mathbf{v}_1) + \lambda_2 f(\mathbf{v}_2).$$

Example 4.12 We will show that the map $f : \mathbb{R}^3 \to \mathbb{R}^2$ given by

$$f\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2x + y \\ y - z \end{pmatrix}$$

is a linear map. Indeed, if we sum two vectors and apply the map f, then we get

$$\begin{aligned} f\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} u \\ v \\ w \end{pmatrix}\right) &= f\begin{pmatrix} x + u \\ y + v \\ z + w \end{pmatrix} \\ &= \begin{pmatrix} 2(x + u) + (y + v) \\ (y + v) - (z + w) \end{pmatrix} \\ &= \begin{pmatrix} 2x + y \\ y - z \end{pmatrix} + \begin{pmatrix} 2u + v \\ v - w \end{pmatrix} \\ &= f\begin{pmatrix} x \\ y \\ z \end{pmatrix} + f\begin{pmatrix} u \\ v \\ w \end{pmatrix}. \end{aligned}$$

As for the second property, let $\lambda \in \mathbb{R}$ be some scalar. Then,

$$f\left(\lambda \begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = f\begin{pmatrix} \lambda x \\ \lambda y \\ \lambda z \end{pmatrix} = \begin{pmatrix} 2\lambda x + \lambda y \\ \lambda y - \lambda z \end{pmatrix} = \lambda \begin{pmatrix} 2x + y \\ y - z \end{pmatrix},$$

as required. Hence, we can conclude that f is indeed a linear map.

Notation 4.13 As in the note before Remark 4.6, we can instead write a column vector as a row vector, where we use ^{T} to denote the transpose. In this way, we could write the linear map in Example 4.12 as follows: $f((x,y,z)^T) = (2x+y, y-z)^T$. This is **much** better for saving space, but we can abuse notation and forget about (i) the transpose symbol ^{T} and (ii) avoid double brackets. We will henceforth write linear maps in this abused way. For instance, the linear map in Example 4.12 would be denoted $f(x,y,z) = (2x+y, y-z)$.

Exercise 31 Let $f : V \to W$ be an arbitrary linear map between K-vector spaces. Prove that $\ker(f) \subseteq V$ and $\mathrm{im}(f) \subseteq W$ are both subspaces in their respective vector spaces.

[**Hint:** Here, the kernel contains elements that get sent to $\mathbf{0}_W$, the zero vector of W.]

Example 4.14 Let $p \in P_n$, that is a polynomial whose degree is at most n; this was defined in Example 4.10(vi). We can define the **evaluation map** as follows:

$$f_\lambda : P_n \to \mathbb{R}, \qquad f(p) = p(\lambda).$$

What does this do? Well, we first choose a fixed number $\lambda \in \mathbb{R}$. Then, the map f_λ just substitutes $x = \lambda$ into the polynomial $p \in P_n$. For a specific example, consider $p(x) = x^2 + 1$. Then, we have

$$f_1(p) = p(1) = 1^2 + 1 = 2, \qquad f_2(p) = p(2) = 2^2 + 1 = 5,$$
$$f_4(p) = p(4) = 4^2 + 1 = 17, \qquad f_\pi(p) = p(\pi) = \pi^2 + 1 \approx 10.869.$$

By Exercise 31, we know that $\ker(f_\lambda) = \{p \in P_n : p(\lambda) = 0\}$ is a subspace of P_n.

Note: Once more, something rather deep is hinted at by Example 4.14. It is actually a huge area of study to look at so-called 'zero sets' of polynomials and 'vanishing ideals' of a set of points. In this language, $\ker(f_\lambda)$ is the 'vanishing idea' of λ. This lies in an area called algebraic geometry (see Chapter **??**) and it has been a hot topic for a while.

Lemma 4.15 *Consider a finite set of vectors $S = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ in a K-vector space V. Then, the map $\varphi_S : K^n \to V$ given by $\varphi_S(a_1, ..., a_n) = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$ is a linear map.*

Proof: We appeal to the definition to show that φ_S is linear. Indeed,

$$\begin{aligned}\varphi_S\left((a_1, ..., a_n) + (b_1, ..., b_n)\right) &= \varphi_S(a_1 + b_1, ..., a_n + b_n) \\ &= (a_1 + b_1)\mathbf{v}_1 + \cdots + (a_n + b_n)\mathbf{v}_n \\ &= (a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) + (b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n)\end{aligned}$$

$$= \varphi_S(a_1, ..., a_n) + \varphi_S(b_1, ..., b_n)$$

demonstrates the first property. Next, let $\lambda \in K$. Then,

$$\begin{aligned}\varphi_S\left(\lambda(a_1, ..., a_n)\right) &= \varphi_S(\lambda a_1, ..., \lambda a_n) \\ &= \lambda a_1 \mathbf{v}_1 + \cdots + \lambda a_n \mathbf{v}_n \\ &= \lambda(a_1 \mathbf{v}_1 + \cdots + a_n \mathbf{v}_n) \\ &= \lambda \varphi_S(a_1, ..., a_n).\end{aligned}$$

□

Definition 4.16 The span of a finite set of vectors $S = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ in a K-vector space V is the set of all linear combinations of them, i.e. $\text{span}(S) = \{a_1 \mathbf{v}_1 + \cdots + a_n \mathbf{v}_n : a_i \in K\}$.

Note: By convention, we declare the span of an empty set of vectors $\text{span}(\emptyset) = \{\mathbf{0}\}$.

Corollary 4.17 *For any finite subset $S \subseteq V$ in a K-vector space V, $\text{span}(S)$ is a subspace.*

Proof: We can see that $\text{span}(S) = \text{im}(\varphi_S)$, where φ_S is as in Lemma 4.15. As φ_S is linear, we can apply Exercise 31 to conclude that $\text{span}(S) \subseteq V$ is a subspace. □

Notation 4.18 If given a set of vectors $S = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$, we would often write $\text{span}(\mathbf{v}_1, ..., \mathbf{v}_n)$ for their span. What's the problem? Strictly speaking, we should actually write $\text{span}(\{\mathbf{v}_1, ..., \mathbf{v}_n\})$ but we omit the curly brackets from the set. This is far from a major deal but be aware of it.

Exercise 32 For a vector $\mathbf{v} \in \mathbb{R}^n$, give a geometric interpretation of $\text{span}(\mathbf{v}) \subseteq \mathbb{R}^n$.

[**Hint:** It is a subspace, by Corollary 4.17, so look to Example 4.10 for inspiration.]

Definition 4.19 Let V be a K-vector space. We call a finite set of vectors $S = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ is linearly independent if there is **no** linear relation between them, that is the only way we can have $a_1 \mathbf{v}_1 + \cdots + a_n \mathbf{v}_n = \mathbf{0}$, for $a_i \in K$, is to have every scalar $a_i = 0$. Otherwise, we say that S is linearly dependent

The linear relation in Definition 4.19 can be interpreted as follows: it is a collection of scalars $(a_1, ..., a_n) \in \ker(\varphi_S)$, where φ_S is the map in Lemma 4.15. Thus, S is linearly independent if and only if $\ker(\varphi_S) = \{\mathbf{0}\}$.

Note: By convention, we declare the empty set to be linearly independent.

Example 4.20 Here are some examples and non-examples of linearly independent sets.

(i) The set $\{\mathbf{v}\}$ is linearly independent if and only if $\mathbf{v} \neq \mathbf{0}$.

(ii) The set $\{\mathbf{v}, \mathbf{w}\}$ is linearly independent if and only if $\mathbf{v} \neq \lambda\mathbf{w}$ for some scalar λ.

(iii) If any $\mathbf{v}_i = \mathbf{0}$, then the set $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ is linearly dependent.

Proposition 4.21 *The set $\{\mathbf{v}_1, ..., \mathbf{v}_n\} \subseteq V$ is linearly dependent if and only if some $\mathbf{v}_i$ is a linear combination of its predecessors, that is $\mathbf{v}_i \in \text{span}(\mathbf{v}_1, ..., \mathbf{v}_{i-1})$.*

Proof: ($\Rightarrow$) By assumption, there exists a linear relation $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$ where at least one coefficient is non-zero. Let i be the largest such that $a_i \neq 0$. This means that all higher-indexed scalars are zero ($a_{i+1} = a_{i+2} = \cdots = a_n = 0$). We can therefore re-write the linear relation as

$$a_1\mathbf{v}_1 + \cdots + a_i\mathbf{v}_i = \mathbf{0} \quad \Rightarrow \quad \mathbf{v}_i = \left(-\frac{a_1}{a_i}\right)\mathbf{v}_1 + \cdots + \left(-\frac{a_{i-1}}{a_i}\right)\mathbf{v}_{i-1} \in \text{span}(\mathbf{v}_1, ..., \mathbf{v}_{i-1}).$$

($\Leftarrow$) Suppose $\mathbf{v}_i \in \text{span}(\mathbf{v}_1, ..., \mathbf{v}_{i-1})$. By definition, there are some scalars $b_1, ..., b_{i-1}$ such that $\mathbf{v}_i = b_1\mathbf{v}_1 + \cdots + b_{i-1}\mathbf{v}_{i-1}$. This easily rearranges to $b_1\mathbf{v}_1 + \cdots + b_{i-1}\mathbf{v}_{i-1} - \mathbf{v}_i = \mathbf{0}$; the final coefficient is -1 so we have a non-trivial linear relation, as required. □

Exercise 33 Determine if $\{(1, 2, 3), (2, 3, 0), (0, 4, 0)\} \subseteq \mathbb{R}^3$ is a linearly independent set.

Definition 4.22 Let V be a K-vector space. We say that a finite subset of vectors S is a basis for V if it is linearly independent and it spans V, that is $\text{span}(S) = V$.

Example 4.23 The so-called standard basis (or canonical basis) for the vector space $\mathbb{R}^3$ is the set $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$. However, there are other bases for this same vector space. For instance, $\{(1, 2, 4), (0, 1, 2), (0, 0, 3)\}$ is another basis for $\mathbb{R}^3$.

Theorem 4.24 *For a K-vector space V and $S = \{\mathbf{v}_1, ..., \mathbf{v}_n\} \subseteq V$, these are equivalent:*

(i) *The set S is a basis for V.*

(ii) *The map $\varphi_S : K^n \to V$ is an isomorphism.*

(iii) *Every $\mathbf{v} \in V$ can be written uniquely as $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$ for $a_i \in K$.*

Proof: $\big($(i) $\Leftrightarrow$ (ii)$\big)$ Because $\text{span}(S) = \text{im}(\varphi_S)$, as mentioned in the proof of Corollary 4.17, we know that S spans V if and only if φ_S is surjective. But by the discussion immediately after Definition 4.19, we know that S is linearly independent if and only if φ_S is injective (has trivial kernel). Hence, S is a basis for V if and only if φ_S is a bijective linear map, i.e. an isomorphism of vector spaces.

$\big($(ii) $\Leftrightarrow$ (iii)$\big)$ This is obvious by definition of φ_S. □

The goal is to determine that every basis of a given vector space will contain the same number of elements (this is then a numerical invariant of that vector space, that is a number associated to the vector space which will not change if we change bases). We will see time and again that invariants are of fundamental importance in many areas of mathematics. To convincingly prove this fact, we can appeal to matrices to make life easier.

5 Matrices

As mentioned, matrices are an incredibly helpful object in algebra. They can encode information about systems of linear equations, differential equations, vector spaces and many more things. At long last, let's finally introduce and discuss them.

Definition 5.1 An $m \times n$ matrix A is a grid with m rows and n columns of the form

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

where a_{ij} is the entry in the i^{th} row and j^{th} column. The entries are usually elements of a field K (e.g. real numbers or complex numbers, but others are possible). The space of $m \times n$ matrices with entries in K is denoted $\mathbb{M}_{m\times n}(K)$. A matrix is square if $m = n$.

Note: The identity matrix is the square matrix whose main diagonal (top-left to bottom-right) consists of ones; all other entries are zero. We denote the identity matrix by $\mathbb{1}_n$.

Exercise 34 Explain how we can view each of a column and row vector as a matrix.

We can define addition of matrices (of the same size), multiplication of matrices (of the relevant size) and scalar multiplication of a matrix (of any size). This, along with the fact that we have an identity matrix, suggests we should get some group-like or vector space-like structure for $\mathbb{M}_{m\times n}(K)$. We **almost** get this, but we will see that the sizes of the matrices (plural of matrix) really inhibits what can be achieved.

Lemma 5.2 *We can view $\mathbb{M}_{m\times n}(K)$ as a K-vector space.*

Proof: We won't go through all the details, but the main idea is to define a natural way to add two matrices from this space and to multiply a matrix by a scalar. Indeed, let $A, B \in \mathbb{M}_{m\times n}(K)$. Then, the sum of matrices is $A + B$, with ij^{th} entry $a_{ij} + b_{ij}$. This means that we simply add corresponding elements. Similarly, for $\lambda \in K$, the scalar multiple of a matrix is λA, with ij^{th} entry λa_{ij}. The zero vector is the $m \times n$ matrix consisting of all-zeros. □

Definition 5.3 For two matrices $A = (a_{ij}) \in \mathbb{M}_{m\times n}(K)$ and $B = (b_{jk}) \in \mathbb{M}_{n\times p}$, we can define matrix multiplication as AB, with ij^{th} entry given by $\sum_{k=1}^{n} a_{ik}b_{kj}$.

Remark 5.4 Again, we have encountered an operation which is non-commutative. In general, we have that $AB \neq BA$ for two matrices. In fact, even if AB is well-defined, it might not even be the case that BA is defined.

Definition 5.5 Let $A \in \mathbb{M}_{m\times n}(K)$ be a matrix. The transpose of this matrix is the matrix $A^T \in \mathbb{M}_{n\times m}(K)$ such that, if $A = (a_{ij})$, then $A^T = (a_{ji})$.

Example 5.6 Let's consider the following matrix $A \in \mathbb{M}_{2\times 3}(\mathbb{R})$:

$$A = \begin{pmatrix} 3 & 2 & 8 \\ 7 & \pi & 1 \end{pmatrix}.$$

By definition, the transpose occurs by swapping the roles of the columns and rows (i.e. so the rows of A are the columns of A^T and the columns of A are the rows of A^T). In this way, we get

$$A^T = \begin{pmatrix} 3 & 7 \\ 2 & \pi \\ 8 & 1 \end{pmatrix}.$$

Exercise 35 For A in Example 5.6, compute the transpose of the transpose $(A^T)^T$.

Lemma 5.7 *Let A and B be matrices such that their product AB is well-defined. Then,*

$$(A^T)^T = A \qquad \text{and} \qquad (AB)^T = B^T A^T.$$

Proof: The fact the transpose of the transpose gives the original matrix is trivial, by Definition 5.5. As for the product property, note that the ij^{th} entry of $(AB)^T$ is the ji^{th} entry of AB, by definition, and this is given in Definition 5.3 as $\sum_k a_{jk}b_{ki}$. If we denote the entries of A^T and B^T by a^T_{ij} and b^T_{ij}, respectively, then we see that

$$\sum_k b^T_{ik}a^T_{kj} = \sum_k b_{ki}a_{jk} = \sum_k a_{jk}b_{ki},$$

but the left-hand side is simply the ij^{th} entry of the matrix $B^T A^T$. Hence, all corresponding

entries of B^TA^T and $(AB)^T$ are identical, meaning they are equal as matrices. □

Exercise 36 For matrices A, B, C of sizes to ensure well-definedness, prove the following:

(i) $(A+B)C = AC + BC$. (Distributivity)

(ii) $A(B+C) = AB + AC$. (Distributivity)

(iii) $A(BC) = (AB)C$. (Associativity)

Lemma 5.8 *For $A \in \mathbb{M}_{m\times n}(\mathbb{R})$, the map $f : \mathbb{R}^n \to \mathbb{R}^m$ given by $f(\mathbf{x}) = A\mathbf{x}$ is linear.*

Proof: Let $\lambda, \mu \in \mathbb{R}$ be scalars and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ be (column) vectors. Then, since we can treat the vectors as $n \times 1$ matrices by Exercise 34, we can use the properties in Exercise 36 to see that

$$f(\lambda\mathbf{x} + \mu\mathbf{y}) = A(\lambda\mathbf{x} + \mu\mathbf{y}) = A\lambda\mathbf{x} + A\mu\mathbf{y} = \lambda A\mathbf{x} + \mu A\mathbf{y} = \lambda f(\mathbf{x}) + \mu f(\mathbf{y}).$$ □

Note: One can adapt the proof of Lemma 5.8 to give the same result for $f : V \to W$, where V and W are two K-vector spaces. In this general case, the matrix $A \in \mathbb{M}_{m\times n}(K)$.

Theorem 5.9 *Let V and W be two K-vector spaces and $f : V \to W$. Then, f is linear if and only if there exists $A \in \mathbb{M}_{m\times n}(K)$ such that $f(\mathbf{v}) = A\mathbf{v}$.*

Proof: ($\Leftarrow$) This is (the note about) Lemma 5.8 above.

($\Rightarrow$) Let $\{\mathbf{v}_1, ..., \mathbf{v}_n\} \subseteq V$ be a basis for V. Therefore, any $\mathbf{v} \in V$ can be written uniquely as $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$ for $a_i \in K$. By linearity, $T(\mathbf{v}) = a_1T(\mathbf{v}_1) + \cdots + a_nT(\mathbf{v}_n)$. Consequently, the image of $\mathbf{v}$ under f is given as a linear combination of the vectors $\{T(\mathbf{v}_1), ..., T(\mathbf{v}_n)\} \subseteq W$. As such, we can take A to have columns $\{T(\mathbf{v}_1), ..., T(\mathbf{v}_n)\}$, that is

$$A = \begin{pmatrix} \uparrow & & \uparrow \\ T(\mathbf{v}_1) & \cdots & T(\mathbf{v}_n) \\ \downarrow & & \downarrow \end{pmatrix}.$$ □

To motivate the next subtopic of discussion regarding matrices, we consider $A \in \mathbb{M}_{2\times 2}(\mathbb{R})$ 'acting' on $\mathbb{R}^2$ (the topic of an action is discussed later in Chapter **??** but here, we just mean how a matrix interacts with a two-dimensional vector). Let $\{\mathbf{e}_1, \mathbf{e}_2\}$ be the standard basis for $\mathbb{R}^2$. Then, we can get the first column of A by $A\mathbf{e}_1$; similarly, we get the second column of A by $A\mathbf{e}_2$. We can consider also the action of such a matrix on the vector $\mathbf{e}_1 + \mathbf{e}_2$.

Geometrically, $\mathbf{0}, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_1 + \mathbf{e}_2$ are the corners of a square of side length one and the matrix A transforms this shape into a parallelogram. Specifically, if we say that

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

the area of the parallelogram can be expressed as follows:

$$P_{\text{area}} = |A\mathbf{e}_1||A\mathbf{e}_2| \sin(\theta),$$

where θ is the angle between the vectors $A\mathbf{e}_1$ and $A\mathbf{e}_2$. We can use the dot product formula in Theorem **??** to get an explicit formula for the area of the parallelogram. Indeed, recall that

$$\mathbf{x} \cdot \mathbf{y} = |\mathbf{x}||\mathbf{y}| \cos(\theta),$$

where θ is the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$. We will apply this formula for $\mathbf{x} = A\mathbf{e}_1$ and $\mathbf{y} = A\mathbf{e}_2$. Now, (the two-dimensional analogue of) Definition **??** gives $(A\mathbf{e}_1) \cdot (A\mathbf{e}_2) = ab + cd$, which we make use in the following argument:

$$\begin{aligned} P_{\text{area}}^2 &= |A\mathbf{e}_1|^2 |A\mathbf{e}_2|^2 \sin^2(\theta) \\ &= |A\mathbf{e}_1|^2 |A\mathbf{e}_2|^2 (1 - \cos^2\theta) \\ &= |A\mathbf{e}_1|^2 |A\mathbf{e}_2|^2 - |A\mathbf{e}_1|^2 |A\mathbf{e}_2|^2 \cos^2(\theta) \\ &= (a^2 + c^2)(b^2 + d^2) - (ab + cd)^2 \\ &= (ad - bc)^2. \end{aligned}$$

Taking the square root reveals that the area of the parallelogram is $P_{\text{area}} = |ad - bc|$.

Note: The number $ad - bc$ is the so-called determinant of the matrix $A \in \mathbb{M}_{2\times 2}(\mathbb{R})$.

Definition 5.10 Let $A \in \mathbb{M}_{m\times n}(K)$ be a matrix. The ij^{th} minor of A is the matrix $A_{ij} \in \mathbb{M}_{(m-1)\times(n-1)}(K)$ obtained from A by removing the i^{th} row and the j^{th} column.

Exercise 37 Write out all minors of the following matrices:

$$\text{(i)}\ \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \quad \text{and} \quad \text{(ii)}\ \begin{pmatrix} 3 & 2 & 2 \\ -1 & 0 & 8 \\ 0 & 7 & 5 \end{pmatrix} \quad \text{and} \quad \text{(iii)}\ \begin{pmatrix} 4 \end{pmatrix}.$$

[**Note:** The final part is a bit of a cruel question as it really is a matter of convention.]

We can now define the *determinant* for any square matrix (not just 2×2).

Definition 5.11 The determinant of a matrix $A \in \mathbb{M}_{n\times n}(K)$ is the alternating sum

$$\det(A) = \sum_{k=1}^{n} (-1)^{k+1} a_{1k} \det(A_{1k}),$$

where we note that the determinant of a 1×1 minor is simply the entry of that minor and the determinant of a 2×2 minor is computed using the formula in the note above Definition 5.10. We also denote the determinant by $|A|$.

Remark 5.12 As we can see in Definition 5.11, the determinant has been defined based on fixing row one (as we can see, all the minors we consider are ones where the first row is removed and we move along the columns). We can equally define the determinant by fixing **any** row, e.g.

$$\det(A) = \sum_{k=1}^{n} (-1)^{k+4} a_{4k} \det(A_{4k}).$$

Moreover, we could instead fix any **column** and move along the rows, e.g.

$$\det(A) = \sum_{k=1}^{n} (-1)^{k+1} a_{k1} \det(A_{k1}).$$

For the sake of consistency, we will use Definition 5.11 in its written form **most** of the time, but we will occasionally take the determinant in a different way if it is convenient to do so.

Note: We saw the geometric interpretation of the determinant of a 2×2 matrix as the area of a parallelogram achieved when applying the matrix transformation to the unit square. Similarly, the determinant of an $n \times n$ matrix is the n-dimensional volume of the parallelotope when applying the matrix transformation to the n-dimensional cube.

Example 5.13 Consider the following matrix $A \in \mathbb{M}_{3\times 3}(\mathbb{Z})$:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

We can compute the determinant by using the formula in Definition 5.11:

$$\begin{aligned} \det(A) &= 1\det(A_{11}) - 2\det(A_{12}) + 3\det(A_{13}) \\ &= 1\begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - 2\begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + 3\begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} \\ &= 1(45-48) - 2(36-42) + 3(32-35) \\ &= 0. \end{aligned}$$

Exercise 38 Compute the determinant of the matrices in Exercise 37.

In order to prove some useful formulae regarding the determinant, it will be helpful to consider so-called row operations applied to matrices and prove some results in more generality. This will also help with the discussion in Section 7.

Definition 5.14 A matrix is in row echelon form (REF) if the following are true:

(i) The first non-zero number in each row is 1, called the leading one.

(ii) Each leading one appears in a column to the right of the leading one before it.

(iii) Any all-zero rows appear at the bottom of the matrix.

Example 5.15 The following matrix is written in row echelon form:

$$\begin{pmatrix} 1 & 2 & 2 & 1 \\ 0 & 1 & -9 & 6 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

However, the following matrix is **not** in row echelon form:

$$\begin{pmatrix} 1 & 3 & 2 \\ 0 & 6 & 1 \\ 2 & 0 & 0 \end{pmatrix}.$$

Definition 5.16 The elementary row operations applied to a matrix are as follows:

(i) Swap two rows, denoted $R_i \leftrightarrow R_j$.

(ii) Multiply a row by a non-zero scalar, denoted $R_i \mapsto \lambda R_i$.

(iii) Add a non-zero multiple of one row to another, denoted $R_i \mapsto R_i + \lambda R_j$.

Note: Applying a single row operation to the identity matrix gives an elementary matrix. We will soon see that these allow us to translate between doing elementary row operations and performing matrix multiplication.

Exercise 39 Determine the elementary 3×3 matrices corresponding to these operations:

(i) $R_1 \mapsto R_1 + 2R_2$.

(ii) $R_3 \leftrightarrow R_2$.

(iii) $R_2 \mapsto -3R_2$.

Remark 5.17 Analogous to Definition 5.16, we can define the elementary column operations applied to a matrix in a near-identical way to the row operations. In this way, we have $C_i \leftrightarrow C_j$, $C_i \mapsto \lambda C_i$ and $C_i \mapsto C_i + \lambda C_j$ as the column operations.

Theorem 5.18 *Let A be an $m \times n$ matrix and $E_{\mathcal{R}}$ be the elementary matrix corresponding to some row operation $\mathcal{R}$. Then, applying $\mathcal{R}$ to matrix A is equivalent to computing $E_{\mathcal{R}}A$.*

Proof: We will consider cases depending on if $\mathcal{R}$ is each elementary row operation. First, let

$$A = \begin{pmatrix} \leftarrow & \mathbf{a}_1 & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{a}_m & \rightarrow \end{pmatrix},$$

so the rows of A are the (row) vectors $\mathbf{a}_1, ..., \mathbf{a}_m$, and suppose also that

$$E_{\mathcal{R}} = \begin{pmatrix} \leftarrow & \mathbf{r}_1 & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{r}_m & \rightarrow \end{pmatrix}.$$

(i) Let $\mathcal{R}$ be the operation $R_i \leftrightarrow R_j$. By definition, the i^{th} row of $E_{\mathcal{R}}$ has a one in the j^{th} column and zeros everywhere else; similarly, the j^{th} row of $E_{\mathcal{R}}$ has a one in the i^{th} column

and zeros everywhere else. Hence, the i^{th} and j^{th} rows of $E_{\mathcal{R}}A$ are given as follows:

$$\mathbf{r}_i A = \mathbf{a}_j \qquad \text{and} \qquad \mathbf{r}_j A = \mathbf{a}_i.$$

All other rows of $E_{\mathcal{R}}A$ are the same as the rows of A. Hence, applying the row operation and considering $E_{\mathcal{R}}A$ are one in the same.

(ii) Let $\mathcal{R}$ be the operation $R_i \mapsto \lambda R_i$. By definition, the i^{th} row of $E_{\mathcal{R}}$ has a λ in the i^{th} column and zeros everywhere else. Thus, the i^{th} row of $E_{\mathcal{R}}A$ is given as follows:

$$\mathbf{r}iA = \lambda \mathbf{a}_i.$$

All other rows of $E_{\mathcal{R}}A$ are the same as the rows of A. Thus, once more, we see that the row operation and the multiplication with the elementary matrix yield the same result.

(iii) Let $\mathcal{R}$ be the operation $R_i \mapsto R_i + \lambda R_j$. By definition, the i^{th} row of $E_{\mathcal{R}}$ has a one in the i^{th} column and a λ in the j^{th} column and zeros everywhere else. So, the i^{th} row of $E_{\mathcal{R}}$ is give as follows:

$$\mathbf{r}_i A = \mathbf{a}_i + \lambda \mathbf{a}_j.$$

All other rows of $E_{\mathcal{R}}$ are the same as the rows of A. Once again, the row operations approach is the same as the matrix multiplication approach. □

Exercise 40 Attempt the proof of Theorem 5.18 in the following specific example: the row operation $\mathcal{R}$ is the operation $R_1 \mapsto R_1 + 2R_2$, which is applied to the matrix

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 8 & 4 \\ 0 & 6 & 5 \end{pmatrix}.$$

We now describe an algorithm to follow to transform a matrix into a row echelon form.

Method – Gaussian Elimination: Let A be an $m \times n$ matrix.

(i) If **every** column of A is zero, we are done. Otherwise, go to Step (ii).

(ii) In the first non-zero column, choose a non-zero element; this is called the pivot.

  (a) If the pivot is at the top of the column, proceed to Step (ii)(b). Otherwise, swap the top row with the pivot row.

  (b) If the pivot is a one, proceed to Step (iii). Otherwise, divide the pivot row by the pivot value.

(iii) If the pivot from Step (ii) is in the n^{th} row, we are done. Otherwise, use the pivot to make the entries below it zero by adding/subtracting multiples of the pivot row to/from the rows below it.

(iv) Now, the first non-zero row has a leading one with zeros below it. Fix the first row and repeat Steps (i)-(iii) with each of the following rows.

Example 5.19 We will transform the following into row echelon form via Gaussian Elimination:

$$\begin{pmatrix} 2 & 2 \\ 3 & 4 \end{pmatrix}.$$

The first step is to choose a pivot in the first column, say 2. By Method **??**, the pivot is already at the top of the column but we do have to divide the pivot row by two. Indeed,

$$\begin{pmatrix} 2 & 2 \\ 3 & 4 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 \\ 3 & 4 \end{pmatrix}, \qquad \text{via } R_1 \mapsto \frac{1}{2}R_1.$$

We can now use the pivot to clear the entries below it, that is we need a zero in the bottom-left position. This can be achieved by subtracting the first row from the second three times, that is

$$\sim \begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix}, \qquad \text{via } R_2 \mapsto R_2 - 3R_1.$$

We are done with the first row and so proceed to the second. The pivot here is 3 by default; this is not a leading one, so dividing the row by three will do the trick. Well,

$$\sim \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad \text{via } R_2 \mapsto \frac{1}{3}R_2.$$

This is all we need to do, and we have found a row echelon form.

Lemma 5.20 *A matrix does* **not** *have a unique row echelon form.*

Proof: We need only exhibit an example of this for the statement to be true. Indeed, we turn to Example 5.19. We found (at least) one row echelon form by following Method **??**. We can, however, perform the row operation $R_1 \mapsto R_1 - R_2$ to the final matrix in the example to get

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

which is also in row echelon form; we can reach (at least) two different row echelon forms. □

Exercise 41 Use Gaussian Elimination to transform the following into row echelon form:

$$\begin{pmatrix} 0 & 7 & 21 \\ 2 & 10 & -6 \\ 2 & 6 & 5 \end{pmatrix}.$$

Definition 5.21 A matrix is in reduced row echelon form (RREF) if these are true:

(i) The matrix is in row echelon form.

(ii) All columns containing a leading one have zeros everywhere else.

Method – Gauss-Jordan Elimination: Let A be an $m \times n$ matrix.

(i) Transform A into a row echelon form by using Gaussian Elimination.

(ii) Starting at the right-most column, add/subtract multiples of the leading one row to/from the rows above it.

(iii) We have now transformed the last column so that the only non-zero entry is the leading one; repeat Step (ii) with the next column along (working right-to-left).

Example 5.22 We will transform the matrix in Example 5.19 into reduced row echelon form via Gauss-Jordan Elimination. Well, much of the hard work was done whilst following Method **??**, but it remains to clear the leading one columns by Method **??**. Indeed, we can see that subtracting the second row from the first will clear the end column:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \text{via } R_1 \mapsto R_1 - R_2.$$

We are now done, since the only other leading one satisfies Definition 5.21(ii) already.

Exercise 42 Two matrices are row equivalent if it is possible to transform one to the other using row operations. Determine if row equivalence is an equivalence relation.

Theorem 5.23 *A matrix* **does** *have a unique reduced row echelon form.*

Proof: Suppose that matrix A has two reduced row echelon forms, namely B and C; we aim to prove that $B = C$. Note that the three matrices A, B, C are row equivalent by definition. Thus, using the elementary matrix approach, every row of A is a linear combination of the rows of B and vice versa (and similar for C). Assume to the contrary that $B \neq C$. We will choose the first (from the left) column where they differ, along with all pivot columns to the left of this column. We can form a matrix from these columns, say R and S (corresponding to B and C respectively). Because B and C are row equivalent, so too are R and S. In fact, are can swap columns so that they have the form

$$R = \left(\begin{array}{c|c} \mathbb{1}_k & \mathbf{r} \\ \hline \mathbf{0} & 0 \end{array}\right) \qquad \text{and} \qquad S = \left(\begin{array}{c|c} \mathbb{1}_k & \mathbf{s} \\ \hline \mathbf{0} & 0 \end{array}\right).$$

As they are row equivalent, we have $\mathbf{r} = \mathbf{s}$. We thus have $R = S$, a contradiction to $B \neq C$. □

We aim to simplify how we compute the determinant beyond the alternating sum given in Definition 5.11 (because doing that for a large matrix is quite tedious). Well, we can use elementary row operations to make our lives easier, but proving this fact requires a bit of work. We begin with some auxiliary results which can be helpful in their own right.

Lemma 5.24 *Let A be an $n \times n$ matrix. Then,* $\det\left(A^T\right) = \det(A)$.

Proof: For notation, let $A = (a_{ij})$ and $A^T = (a^T_{ij})$ where $a^T_{ij} = a_{ji}$. By definition,

$$\begin{aligned} \det\left(A^T\right) &= \sum_{k=1}^{n} (-1)^{k+1} a^T_{1k} \det\left(A^T_{1k}\right) \\ &= \sum_{k=1}^{n} (-1)^{k+1} a_{k1} \det(A_{k1}) \\ &= \det(A), \end{aligned}$$

where we use Remark 5.12 to see that the determinant can also be given by fixing a column and moving along the rows (this is how we pass from the second line to the third line above). □

Definition 5.25 Let A be a square matrix.

(i) We say A is upper triangular if all entries **below** the main diagonal are zero.

(ii) We say A is lower triangular if all entries **above** the main diagonal are zero.

(iii) We say A is diagonal if all entries **not on** the main diagonal are zero.

Remark 5.26 If we apply the transpose to an upper triangular matrix, we get a lower triangular matrix (and vice versa). Moreover, all diagonal matrices are fixed by the transpose, that is it will not change the matrix at all.

Exercise 43 Prove that the product of two upper triangular matrices is upper triangular. What can we say about the product of k upper triangular matrices, for some $k \in \mathbb{Z}^+$? Use Lemma 5.7 to conclude the analogue for the product of lower triangular matrices .

Proposition 5.27 *Let $A = (a_{ij})$ be $n \times n$ upper triangular. Then,* $\det(A) = a_{11}a_{22}\cdots a_{nn}$.

Proof: Proceed by induction on the size of the matrix. The base case is clear. Assume that the result holds for an $(n-1) \times (n-1)$ upper triangular matrix. Note that $a_{i1} = 0$ for every $i > 1$ (i.e. everything in the first column is zero except for the top entry). By definition, we know that $\det(A) = a_{11}\det(A_{11})$. But the minor A_{11} is also upper triangular and is size $(n-1) \times (n-1)$. Thus, the inductive hypothesis tells us that $\det(A_{11}) = a_{22}a_{33}\cdots a_{nn}$. Substituting this into the formula we just wrote for $\det(A)$ gives the result. □

Corollary 5.28 *Let $A = (a_{ij})$ be $n \times n$ lower triangular. Then,* $\det(A) = a_{11}a_{22}\cdots a_{nn}$.

Proof: If A is lower triangular, then A^T is upper triangular by Remark 5.26. Hence, we see that $\det\left(A^T\right) = a_{11}a_{22}\cdots a_{nn}$ by Proposition 5.27, but $\det\left(A^T\right) = \det(A)$ by Lemma 5.24. □

It now seems that, if we can determine the effect performing row operations has on a determinant, then we can do row operations to transform a matrix to upper/lower triangular form and pretty much read-off the determinant fully. This is the true aim of this part of the discussion. We still have some work to do, however.

Lemma 5.29 *Any square matrix A is row equivalent to an upper/lower triangular matrix.*

Proof: We prove the result for upper triangular matrices, but the argument is near-identical for lower triangular matrices. Recall from Exercise 42 that the statement means we can perform elementary row operations to get from A to an upper triangular matrix. We will prove this by

induction. Again, the base case is trivial. Assume the result holds for any $(n-1)\times(n-1)$ matrix and let A be an $n\times n$ matrix. There are two cases to consider.

(i) Suppose the first column of A is all-zero. Then, A is upper triangularisable if and only if the minor A_{11} is upper triangularisable. Indeed, because A_{11} has size $(n-1)\times(n-1)$, the inductive hypothesis says that this is indeed the case. Note that the elementary row operations which achieve this will not change the fact that the first column of A is all-zero. Hence, A is indeed upper triangularisable.

(ii) Suppose the first column of A is not all-zero.

  (a) Let $a_{11}=0$. Then, suppose i is the smallest such that $a_{i1}\neq 0$ (so we are picking the first non-zero element in the first column from top-to-bottom). Said i will exist because we are assuming there is an element of the first column which is non-zero. If we apply the operation $R_1\to R_1+R_i$, this will mean that the new entry in the top-left is non-zero, that is $a'_{11}\neq 0$. We can then move to Case (b) below.

  (b) Let $a_{11}\neq 0$. If we apply the operation $R_i\to R_i-\frac{a_{i1}}{a_{11}}R_1$ for every $i\in\{2,3,...,n\}$, this will ensure the first column is all-zero **except** for the top-left entry. It again follows that A is upper triangularisable if and only if A_{11} is; this is just another use of the inductive hypothesis.

Hence, all cases are exhausted and A is always row equivalent to an upper triangular matrix. □

Exercise 44 Perform row operations to the following so the result is upper triangular:

$$\begin{pmatrix} 0 & 2 & 0 & 4 \\ 1 & 0 & 7 & 9 \\ 8 & 3 & 0 & 5 \\ 0 & 0 & 1 & 6 \end{pmatrix}.$$

Lemma 5.30 *Let $A\in\mathbb{M}_{n\times n}(K)$ and $\mathcal{R}$ be a sequence of elementary row operations which are applied to A; call the resulting matrix A'. Then, there exists $\lambda\in K$ such that*

$$\det(A)=\lambda\det\big(A'\big).$$

Proof: Omitted; it is an induction proof which isn't very illuminating. □

Theorem 5.31 *Let A and B be $n\times n$ matrices. Then,* $\det(AB)=\det(A)\det(B)$.

Proof: From Lemma 5.29, we can convert A into an upper triangular matrix by a sequence of row operations; call the resulting matrix A'. If we let $C = AB$, then we call $C' = A'B$. By Lemma 5.7, we know that $(C')^T = (A'B)^T = B^T(A')^T$. Again by applying Lemma 5.29, we can convert B^T into a lower triangular matrix by a sequence of row operations; call the resulting matrix $(B^T)'$. Then, we call $C'' = (B^T)'(A')^T$. Because A' is upper triangular, we know that $(A')^T$ is lower triangular. Hence, by Exercise 43, we know that C'' is also lower triangular (it is the product of two lower triangular matrices). Hence, applying Corollary 5.28 tells us that

$$\det\left((B^T)'(A')^T\right) = \det\left((B^T)'\right)\det\left((A')^T\right).$$

But by Lemma 5.30, we know that

$$\begin{aligned}\det(A) &= \alpha\det\big(A'\big),\\ \det\Big(B^T\Big) &= \beta\det\left((B^T)'\right),\end{aligned}$$

from which it follows that

$$\begin{aligned}\det(C) &= \alpha\det\big(C'\big),\\ \det\Big((C')^T\Big) &= \beta\det\big(C''\big).\end{aligned}$$

Combining all this together gives us

$$\begin{aligned}\det(C) &= \alpha\det\big(C'\big)\\ &= \alpha\det\left((C')^T\right)\\ &= \alpha\beta\det\big(C''\big)\\ &= \alpha\beta\det\left((B^T)'(A')^T\right)\\ &= \alpha\det\left((A')^T\right)\beta\det\left((B^T)'\right)\\ &= \alpha\det\big(A'\big)\det\Big(B^T\Big)\\ &= \det(A)\det(B).\end{aligned}$$

□

Note: Truthfully, the proof of Theorem 5.31 is also not exactly illuminating; don't worry if you can't follow all the details. The statement is the most important thing about this.

We can now state one of the most important results about determinants and row operations.

Corollary 5.32 *Let A be a square matrix.*

(i) *Applying $R_i \leftrightarrow R_j$ has the effect of multiplying* $\det(A)$ *by* -1.

(ii) *Applying $R_i \mapsto \lambda R_i$ has the effect of multiplying* $\det(A)$ *by* λ.

(iii) *Applying $R_i \mapsto R_i + \lambda R_j$ has* **no effect** *to* $\det(A)$.

Sketch of Proof: Compute the determinant of the corresponding elementary matrix. Then, we can appeal to Theorems 5.18 and 5.31 to conclude that the statements above are true. □

Note: An analogous result to Corollary 5.32 holds for elementary column operations.

Remark 5.33 We can now state a number of properties of the determinant of a square matrix A that are a consequence of Corollary 5.32:

(i) If A has two (or more) identical rows, then $\det(A) = 0$.

(ii) If A has two (or more) identical columns, then $\det(A) = 0$.

(iii) If A has an all-zero row, then $\det(A) = 0$.

(iv) If A has an all-zero column, then $\det(A) = 0$.

Method – Finding the Determinant: Let A be a square matrix.

(i) If A is triangular, we are done. Otherwise, go to Step (ii).

(ii) If A is not triangular, perform a sequence of row operations to transform it to an triangular matrix, which we now call A'.

(iii) Compute the changes made to the determinant of A by performing the sequence of row operations in Step (ii); this is done by using Corollary 5.32.

(iv) Finally, we can write $\det(A)$ by **dividing** $\det\big(A'\big)$ by the number we find in Step (iii).

Example 5.34 We will use Method **??** to compute the determinant of the following:

$$A = \begin{pmatrix} 0 & 3 & 9 & 12 \\ 1 & 5 & 6 & 7 \\ 1 & 7 & 10 & 11 \\ 2 & 9 & 12 & 23 \end{pmatrix}.$$

Indeed, we will first apply row operations to get the matrix into upper triangular form, so

$$\begin{pmatrix} 0 & 3 & 9 & 12 \\ 1 & 5 & 6 & 7 \\ 1 & 7 & 10 & 11 \\ 2 & 9 & 12 & 23 \end{pmatrix} \sim \begin{pmatrix} 1 & 5 & 6 & 7 \\ 0 & 3 & 9 & 12 \\ 1 & 7 & 10 & 11 \\ 2 & 9 & 12 & 23 \end{pmatrix}, \qquad \text{via } R_1 \leftrightarrow R_2,$$

$$\sim \begin{pmatrix} 1 & 5 & 6 & 7 \\ 0 & 3 & 9 & 12 \\ 0 & 2 & 4 & 4 \\ 2 & 9 & 12 & 23 \end{pmatrix}, \qquad \text{via } R_3 \mapsto R_3 - R_1,$$

$$\sim \begin{pmatrix} 1 & 5 & 6 & 7 \\ 0 & 3 & 9 & 12 \\ 0 & 2 & 4 & 4 \\ 0 & -1 & 0 & 9 \end{pmatrix}, \qquad \text{via } R_4 \mapsto R_4 - 2R_1,$$

$$\sim \begin{pmatrix} 1 & 5 & 6 & 7 \\ 0 & 1 & 3 & 4 \\ 0 & 2 & 4 & 4 \\ 0 & -1 & 0 & 9 \end{pmatrix}, \qquad \text{via } R_2 \mapsto \frac{1}{3}R_2,$$

$$\sim \begin{pmatrix} 1 & 5 & 6 & 7 \\ 0 & 1 & 3 & 4 \\ 0 & 0 & -2 & -4 \\ 0 & -1 & 0 & 9 \end{pmatrix}, \qquad \text{via } R_3 \mapsto R_3 - 2R_2,$$

$$\sim \begin{pmatrix} 1 & 5 & 6 & 7 \\ 0 & 1 & 3 & 4 \\ 0 & 0 & -2 & -4 \\ 0 & 0 & 3 & 13 \end{pmatrix}, \qquad \text{via } R_4 \mapsto R_4 + R_2,$$

$$\sim \begin{pmatrix} 1 & 5 & 6 & 7 \\ 0 & 1 & 3 & 4 \\ 0 & 0 & -2 & -4 \\ 0 & 0 & 0 & 7 \end{pmatrix}, \qquad \text{via } R_4 \mapsto R_4 + \frac{3}{2}R_3.$$

The resulting matrix, denoted A', is such that $\det(A') = -14$ (multiply the diagonal entries). Now, if we look at the row operations performed above, we see that they corresponding to the constant $-1/3$ (because we did one swap and we multiplied one row by $1/3$; all the other

operations we did don't change the determinant). Hence, we conclude that

$$\det(A) = \frac{\det\big(A'\big)}{-1/3} = -3\det\big(A'\big) = 42.$$

Note: To transform a matrix into upper triangular form, one can almost follow the method of Gaussian Elimination directly; we can be a little bit more relaxed here and not make the pivots into ones (e.g. we didn't change the -2 to a 1 in the last step above) but this is just dealer's choice.

Exercise 45 Compute the determinant of the matrix given in Exercise 44.

Theorem 5.35 *Any bases of the K-vector space V have the same number of elements.*

Proof: Suppose $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ and $\{\mathbf{w}_1, ..., \mathbf{w}_m\}$ are both bases of V. □

6 Permutation Groups

7 Solving Linear Equations

8 Eigenvalues and Eigenvectors

9 Exercise Solutions

We provide detailed solutions to the exercises interwoven within each section of the module. Hopefully you have given these questions a try whilst on your learning journey with the module. But mathematics is difficult, so don't feel disheartened if you had to look up an answer before you knew where to begin (we have all done it)!

Solutions to Exercises in Section 2

> **Exercise 1** Prove that $(\mathbb{Z}, +)$ is a group.

Solution: This is much the same as the proof of Lemma 2.4.

(i) It is clear that $a + (b + c) = a + (b + c)$ for any $a, b, c \in \mathbb{Z}$, giving us associativity.

(ii) The sum of two integers is obviously an integer, so $\mathbb{Z}$ is closed under $+$.

(iii) The element $0 \in \mathbb{Z}$ is the identity since it is an integer and $a + 0 = a = 0 + a$ for any $a \in \mathbb{Z}$.

(iv) Finally, we have an inverse $-a \in \mathbb{Z}$ for any integer $a \in \mathbb{Z}$ because $a + (-a) = 0 = (-a) + a$.

Consequently, we can see that $(\mathbb{Z}, +)$ is a group. □

> **Note:** We haven't really proved such, but we noted in Lemma 2.4 that addition in $\mathbb{R}$ is associative. Because $\mathbb{Z} \subseteq \mathbb{R}$ is a subset, and the addition in the group $(\mathbb{Z}, +)$ is really the same as the addition in the group $(\mathbb{R}, +)$, we already get the associative condition for free.

> **Exercise 2** Write out the group table for $(\mathbb{Z}_5, +_{\text{mod } 5})$.

Solution: This is similar to Example 2.6 in the notes. The solution is Table 5 below. □

$+_{\text{mod } 5}$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

Table 5: The group table for $(\mathbb{Z}_5, +_{\text{mod } 5})$.

Exercise 3 Complete the below group table for the dihedral group D_3.

$\circ$	I	R	S	A	B	C
I	I	R	S	A	B	C
R		S		C		
S			R			
A				I		
B					I	
C						I

Table 2: The group table for $(D_3, \circ)$.

Solution: The idea here is to see how the different symmetries in D_3 interact with each other. The best way to do so is to draw some pictures of a triangle ABC and apply these symmetries, much as we did in Example 2.8. Here, we provide explanation but we won't draw any pictures.

First, note that $I \circ$ [anything] = anything because it is the identity. Hence, this means the first column in Table 2 will be trivial. Now, rotating by both $2\pi/3$ and $4\pi/3$ gives a rotation by 2π, which is again equivalent to doing nothing.

The composition of two reflections will give a rotation and the composition of a rotation and reflection will give another reflection. This will mean that the group table should contain four 3×3 grids. where the top left and bottom right are rotations and the top right and bottom left are reflections. We can see the finished table in Table 2 below. □

$\circ$	I	R	S	A	B	C
I	I	R	S	A	B	C
R	R	S	I	C	A	B
S	S	I	R	B	C	A
A	A	B	C	I	R	S
B	B	C	A	S	I	R
C	C	A	B	R	S	I

Table 2: The group table for $(D_3, \circ)$.

Exercise 4 Fully understand the proofs of Propositions 2.12 and 2.13.
[**Note:** Knowing how to use each part of the definition of a group is vitally important.]

Solution: In the proof of Proposition 2.12, we use the fact that anything multiplied by f will give the original thing back (this is what is means for f to be an identity) and we use the fact that anything multiplied by e will give the original thing back (again because e is an identity).

In the proof of Proposition 2.13, we use the fact that k is an inverse of g (which means that $e = gk$ by definition) to re-write he and then we use associativity of the group operation to change the brackets: $h(gk) = (hg)k$. From here, we use the fact that h is also an inverse of g (which means that $e = hg$). □

Exercise 5 Prove Lemma 2.15(ii), that is Right Cancellation.

Solution: We practically replicate the proof of Lemma 2.15(i). Indeed, assume $hg = kg$. Then, we can multiply by g^{-1} on the right to get $(hg)g^{-1} = (kg)g^{-1} \Rightarrow h(gg^{-1}) = k(gg^{-1}) \Rightarrow h = k$. □

Exercise 6 Construct the group table of $\mathbb{Z}_{10}^*$.

[**Hint:** To determine $\mathbb{Z}_{10}^*$, recall that *coprime* is introduced in Definition **??**.]

Solution: First, note that $\mathbb{Z}_{10}^* = \{1, 3, 7, 9\}$, with multiplication defined modulo 10. We can easily multiply the numbers and look at their remainders on division by ten; see Table 3 below. □

$\times_{\text{mod } 10}$	1	3	7	9
1	1	3	7	9
3	3	9	1	7
7	7	1	9	3
9	9	7	3	1

Table 3: The group table for $\mathbb{Z}_{10}^*$.

Exercise 7 Verify that the other compositions presented in Table 3 are correct.

Solution: We must show (i) $T_m \circ R_n = R_{m+n}$, (ii) $R_m \circ T_n = R_{m-n}$ and (iii) $R_m \circ R_n = T_{m-n}$. It is done by using the formulae; $(T_m \circ R_n)(x, y) = T_m(-x+2n, y) = (-x+2n+2m, y) = R_{m+n}(x, y)$ means (i) is true. Similarly, $(R_m \circ T_n)(x, y) = R_m(x + 2n, y) = (-x - 2n + 2m, y) = R_{m-n}(x, y)$ gives (iii) and, lastly, $(R_m \circ R_n)(x, y) = R_m(-x + 2n, y) = (x - 2n + 2m, y) = T_{m-n}(x, y)$. □

Exercise 8 Consider the frieze pattern drawn in Figure 3 below.

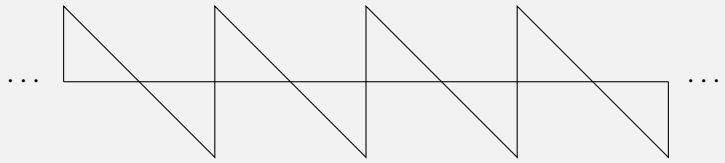


Figure 3: A frieze pattern of triangles.

Suppose one of the triangles has vertices $(0,0)$, $(0,1)$, $(1,0)$. The isometries of the plane that preserve the frieze are translations T_n which shift the diagram n periods to the right and rotations S_n about the point $(n,0)$ by the angle π.

(i) Write down formulae for $T_n(x,y)$ and $S_n(x,y)$.

(ii) Construct the group table for the frieze group.

(iii) State the orders of the elements in the frieze group.

Solution: (i) The translation $T_n(x,y) = (x+2n, y)$ and the rotation $S_n(x,y) = (2n-x, -y)$. note that it is $2n$ (and not just n) because the pattern in Figure 3 repeats 'every two places'.

(ii) We can proceed similarly to Exercise 7 to construct the group table. Indeed, we just need to apply the formulae for T_n and S_n to see how they interact; this gives Table 4.

$\circ$	T_n	S_n
T_m	T_{n+m}	S_{n+m}
S_m	S_{n-m}	T_{n-m}

Table 4: The group table of the isometries preserving the frieze in Figure 3.

(iii) The elements of the frieze group are $T_0, T_1, T_2, \ldots$ and $R_0, R_1, R_2, \ldots$. Well, $\text{ord}(T_0) = 1$ as it is the identity. It is easy to notice $\text{ord}(T_i) = \infty$ for all $i \geq 1$ and $\text{ord}(R_j) = 2$ for all $j \geq 0$. In words, all non-trivial translations have infinite order and all rotations have order two. □

Exercise 9 Use the Subgroup Criterion to prove that $\{2^n : n \in \mathbb{Z}\} \leq \mathbb{Q}^*$ is a subgroup.

Solution: Let $H = \{2^n : n \in \mathbb{Z}\}$. We are to prove that $H \leq \mathbb{Q}^*$.

(i) The identity in $\mathbb{Q}^*$ is 1, but $1 = 2^0 \in H$.

(ii) Let $x, y \in H$ be of the form $x = 2^n$ and $y = 2^m$ for some $m, n \in \mathbb{Z}$. Then, $xy = 2^{n+m} \in H$ because $n + m \in \mathbb{Z}$ by the fact that $\mathbb{Z}$ is closed under addition.

(iii) Let $x \in H$ have the form $x = 2^n$ as in (ii). Because $-n \in \mathbb{Z}$ since $\mathbb{Z}$ is closed under forming (additive) inverses, we know $2^{-n} \in H$ also. However, we see that $x(2^{-n}) = 1 = (2^{-n})x$, so we have found the inverse of x. Namely, $x^{-1} = 2^{-n} \in H$.

By the Subgroup Criterion, one can conclude that $H \leq \mathbb{Q}^*$. □

Exercise 10 We call G cyclic if it is generated by an element, i.e. $G = \langle g \rangle$ for some $g \in G$. Prove that a group of order n is cyclic if and only if it contains an element of order n.

Solution: ($\Rightarrow$) Suppose that G is a cyclic group of order n. By Lemma 2.33, we know that any generator of the group will have order n, so it clearly contains an element of order n.

($\Leftarrow$) Conversely, suppose that G contains an element of order n, g say. Then, $\langle g \rangle \leq G$ is a subgroup of order n, but this is true only when $\langle g \rangle = G$. Thus, G is cyclic. □

Exercise 11 Prove that $K = \{k \in \mathbb{Z} : g^k \in H\}$ is a subgroup of $\mathbb{Z}$.

[**Hint:** Remember that $\mathbb{Z}$ is an *additive* group, which means so too is K.]

Solution: Recall that $G = \langle g \rangle$ is a cyclic group and $H \leq G$ is a subgroup (this exercise relates to proving something we used in the proof of Theorem 2.37).

(i) Because $g^0 \in H$, since $g^0 = 1_H$ is the identity and H is a subgroup, this means that $0 \in K$. Therefore, K contains the identity of $\mathbb{Z}$.

(ii) Let $k, n \in K$, meaning that $g^k, g^n \in H$ by definition. Because H is a subgroup, it is closed under the operation. Hence, $g^k g^n = g^{k+n} \in H$. As such, we conclude that $k + n \in K$ so it too is closed under its operation.

(iii) Let $k \in K$, meaning that $g^k \in H$. Again, by the fact that H is a subgroup, it is closed under forming inverses, so $(g^k)^{-1} = g^{-k} \in H$. But this implies that $-k \in K$, so it too is closed under forming inverses.

Once again, the Subgroup Criterion implies the result. □

Exercise 12 Demonstrate Theorem 2.41 by using G and H from Example 2.39.

[**Hint:** Show that $G = \{I, R, S\}$ and $H = \mathbb{Z}_4^*$ are cyclic and give a generator for $G \times H$.]

Solution: The group G is cyclic since it is generated by R, or by S. This is clear when looking at Table 2 (or at least the upper left block). The group $H = \mathbb{Z}_4^* = \{1, 3\}$ is also cyclic as it is

generated by 3. Both are finite and so satisfy the hypotheses of Theorem 2.41. By this result, the group $G \times H$ is cyclic. In full, this is the group whose underlying set is

$$G \times H = \{(I,1),(R,1),(S,1),(I,3),(R,3),(S,3)\}.$$

Because the group operation for $G \times H$ is inherited from those of the constituent groups, we know that $G \times H = \langle (R,3) \rangle$ since this consists of the generators of each of the other groups. □

Exercise 13 Let $G = \mathbb{R}$ (under $+$) and $H = \mathbb{R}^*$ (under $\times$). Show that H has an element of order two and that G does **not**. Conclude from this that $G \ncong H$ are non-isomorphic.

Solution: The element $-1 \in H$ has order two, because $(-1)^2 = 1$ and 2 is the least positive integer such that this occurs. In G, we have $\text{ord}(0) = 1$ because it is the identity and, for any $x \neq 0$, the elements $x, x+x, x+x+x, \ldots$ are all distinct. Thus, $\text{ord}(x) = \infty$. In particular, there is no element of order two. If an isomorphism $G \cong H$ existed, then Proposition 2.46(ii) is contradicted. Therefore, there is no such isomorphism. □

Exercise 14 Show that $\varphi : G \to H$ in the proof of Theorem 2.47 is a homomorphism.

Solution: So, $G = \langle g \rangle$ and $H = \langle h \rangle$ have the same order and $\varphi : G \to H$ is defined by $\varphi(g^k) = h^k$. To show this is a homomorphism, it suffices to show φ respects the group operations. For $g^k, g^n \in G$, we have $\varphi(g^k g^n) = \varphi(g^{k+n}) = h^{k+n} = h^k h^n = \varphi(g^k)\varphi(g^n)$ and we are done. □

Exercise 15 Determine the cosets of the subgroup $K = \{I, A\}$ in D_3.

Solution: This can be done relatively easily by using Table 2. Indeed, we just need to see how the elements of K interact with the other elements of the group. For example, $KI = \{I \circ I, A \circ I\} = \{I, A\}$ and $KA = \{I \circ A, A \circ A\} = \{A, I\}$, so we know that the cosets $KI = KA$. This can be done similarly and the punchline is written below:

$$\begin{aligned} KI = \{I, A\} &= KA, \\ KR = \{R, B\} &= KB, \\ KS = \{S, C\} &= KC. \end{aligned}$$

Hence, there are only the three cosets written above in this situation. □

Exercise 16 Show that $\sim$ in the proof of Corollary 2.53 is an equivalence relation.

Solution: Recall that the equivalence relation is defined as follows: $x \sim y$ if and only if $xy^{-1} \in H$, where $x, y \in G$ are elements of a group and $H \leq G$ is a subgroup. The justification for the things below is simply the Subgroup Criterion.

(i) Clearly, $x \sim x$ because $xx^{-1} = 1_G \in H$.

(ii) Let $x \sim y$, so $xy^{-1} \in H$. Then, $(xy^{-1})^{-1} = yx^{-1} \in H$, and so $y \sim x$.

(iii) Let $x \sim y$ and $y \sim z$, so $xy^{-1}, yz^{-1} \in H$. Then, $(xy^{-1})(yz^{-1}) = xz^{-1} \in H$, and so $x \sim z$.

As the relation is (i) reflexive, (ii) symmetric and (iii) transitive, it is an equivalence relation. □

Exercise 17 Prove Corollary 2.59.

[**Hint:** Use Proposition 2.26(ii) and note that n is not necessarily the order of g.]

Solution: Let $|G| = n$ and $g \in G$ be a general element. By Proposition 2.26(ii), we know that $\text{ord}(g) \mid n$. Let's relabel $\text{ord}(g) =: m$, so we know that there exists $k \in \mathbb{Z}$ such that $n = km$ (by definition of divisors). Thus, $g^n = g^{km} = (g^k)^m = 1^m = 1$, as required. □

Exercise 18 Prove that for $\varphi : G \to H$ a homomorphism, $\text{im}(\varphi) \leq H$ is a subgroup.

Solution: We once again apply the Subgroup Criterion.

(i) We know from Lemma 2.45 that $\varphi(1_G) = 1_H$. Therefore, $1_H \in \text{im}(\varphi)$.

(ii) Let $x, y \in \text{im}(\varphi)$, meaning there exist elements $g, h \in G$ such that $x = \varphi(g)$ and $y = \varphi(h)$. Because φ is a homomorphism, we know that $xy = \varphi(g)\varphi(h) = \varphi(gh) \in \text{im}(\varphi)$ since $gh \in G$ by closure under the operation.

(iii) Let $x \in \text{im}(\varphi)$, meaning there is an element $g \in G$ such that $x = \varphi(g)$. By Lemma 2.45, we have $x^{-1} = \varphi(g)^{-1} = \varphi(g^{-1}) \in \text{im}(\varphi)$, as $g^{-1} \in G$ by closure under forming inverses.

Again, this suffices to show that $\text{im}(\varphi) \leq H$ is a subgroup. □

Exercise 19 (Harder) Let $\varphi : G \to H$ and $\psi : G \to H$ be homomorphisms. Prove that

$$\mathrm{Eq}(\varphi, \psi) := \{g \in G : \varphi(g) = \psi(g)\} \leq G$$

is a subgroup, called the equaliser of φ and ψ, using the Subgroup Criterion.

Solution: The difficulty here really comes from understanding how $\mathrm{Eq}(\varphi, \psi)$ is defined.

(i) By Lemma 2.45, we know that any homomorphism sends the identity to the identity. In particular then, $\varphi(1_G) = 1_H = \psi(1_G)$. By definition, $1_G \in \mathrm{Eq}(\varphi, \psi)$.

(ii) Let $g, h \in \mathrm{Eq}(\varphi, \psi)$; this means that $\varphi(g) = \phi(g)$ and $\varphi(h) = \psi(h)$. But now, we can see that $\varphi(gh) = \varphi(g)\varphi(h) = \psi(g)\psi(h) = \psi(gh)$ by the defining property of a homomorphism. Thus, $gh \in \mathrm{Eq}(\varphi, \psi)$.

(iii) Let $g \in \mathrm{Eq}(\varphi, \psi)$; this means that $\varphi(g) = \phi(g)$. By Lemma 2.45 again, we can invert this equation to see that $\varphi(g)^{-1} = \psi(g)^{-1} \Leftrightarrow \varphi(g^{-1}) = \psi(g^{-1})$. Thus, $g^{-1} \in \mathrm{Eq}(\varphi, \psi)$.

Consequently, the equaliser of φ and ψ is a subgroup of G by the Subgroup Criterion. □

Exercise 20 Construct the group table for the Klein Vierergruppe, generated by a, b, c.

[**Hint:** You may use the fact it is really only generated by a and b, since $c = ab$.]

Solution: We are told that this group V is generated by three elements, but of course we need an identity also. Hence, we are considering a 4×4 table with entries consisting of applying the operation to elements in $\{1, a, b, c\}$. Because 1 is the identity, we at least know the first row immediately. Now, per the note after Theorem 2.67, we have that $V = C_2 \times C_2$, the direct product of two cyclic groups of order two. Well, any cyclic group of order two has two elements, an identity and an element of order two. Thus, we can write $V = \langle a \rangle \times \langle b \rangle$ where $a^2 = 1$ and $b^2 = 1$. From this, we know that the diagonal in our group table will consist of ones.

Note: The hint says $c = ab$, from which it follows that $c^2 = (ab)^2 = abab$. However, this must be the identity; if it is not, as $|V| = 4$ and the order of an element divides the order of the group, we get $\mathrm{ord}(c) = 4$ and this makes V cyclic. This contradicts Theorem 2.41.

Given we now know that $c^2 = 1$, this implies also that $ab = ba$, making V an Abelian group. We can now fill in the rest of the group table (e.g. $ca = aba = aab = a^2b = b$ because $a^2 = 1$, and so forth). The conclusion to this is Table 5 below. □

$\cdot$	1	a	b	c
1	1	a	b	c
a	a	1	c	b
b	b	c	1	a
c	c	b	a	1

Table 5: The group table for the Klein Vierergruppe V.

Exercise 21 Prove the anti-commutativity statements in Lemma 2.71, that is

$$ji = -k, \quad kj = -i, \quad ik = -j.$$

Solution: We already proved the commutativity part of Lemma 2.71, and these (along with the definition of quaternions) will near-on immediately give us the anti-commutativity rules.

- To prove $ji = -k$, multiply the equation $ki = j$ by i. Indeed, $ki^2 = ji \Leftrightarrow ji = -k$.
- To prove $kj = -i$, multiply the equation $ij = k$ by j. Indeed, $ij^2 = kj \Leftrightarrow kj = -i$.
- To prove $ik = -j$, multiply the equation $jk = i$ by k. Indeed, $jk^2 = ik \Leftrightarrow ik = -j$. □

Exercise 22 For a homomorphism $\varphi : G \to H$, is $\mathrm{im}(\varphi) \trianglelefteq H$ normal? Justify your claim.

Solution: This is **not** true in general. Let $G = \{I, A\}$, a subgroup of D_3, and take $H = D_3$ to be the whole of the dihedral group. We can define a homomorphism $\varphi : G \to H$ by $\varphi(g) = g$. This essentially 'places' the subgroup $\{I, A\}$ inside the dihedral group D_3. From Example 3.9(iii), we can see that $\mathrm{im}(\varphi) = \{I, A\} \not\trianglelefteq D_3$ is not normal. □

Exercise 23 Show that the operation from Corollary 3.13 gives a group structure on G/N.

Solution: This amounts to demonstrating the axioms of a group where the operation is 'coset multiplication'. Throughout, we consider the quotient G/N of a group G by a normal $N \trianglelefteq G$.

(i) Well, $(Ng)\big((Nh)(Nk)\big) = (Ng)\big(N(hk)\big) = N(ghk) = \big(N(gh)\big)(Nk) = \big((Ng)(Nh)\big)(Nk)$ establishes the associativity of the operation on the quotient group; this implicitly uses the associativity of G when splitting $N(ghk)$ into a product of two cosets.

(ii) By Corollary 3.13, we know that the quotient group is closed under this operation.

(iii) The identity is the trivial coset $N(1_G)$, which is clear by Corollary 3.13.

(iv) We can define the inverse coset $(Ng)^{-1} \coloneqq N(g^{-1})$, which works by Corollary 3.13. □

Exercise 24 Explain the equivalences for the injectivity in the proof of Theorem 3.18.

Solution: In the proof of the First Isomorphism Theorem, we have three equivalence statements in quick succession: $\ker(\varphi)g_1 = \ker(\varphi)g_2 \Leftrightarrow g_1g_2^{-1} \in \ker(\varphi) \Leftrightarrow \varphi(g_1g_2^{-1}) = 1_H \Leftrightarrow \varphi(g_1) = \varphi(g_2)$.

(i) $\ker(\varphi)g_1 = \ker(\varphi)g_2 \Leftrightarrow g_1g_2^{-1} \in \ker(\varphi)$ is a direct consequence of Proposition 2.52(v).

(ii) $g_1g_2^{-1} \in \ker(\varphi) \Leftrightarrow \varphi(g_1g_2^{-1}) = 1_H$ is a direct consequence of Definition 2.61.

(iii) $\varphi(g_1g_2^{-1}) = 1_H \Leftrightarrow \varphi(g_1) = \varphi(g_2)$ is a direct consequence of Definition 2.42. □

Exercise 25 For an arbitrary group G, prove that (i) $G/\{1\} \cong G$ and (ii) $G/G \cong \{1\}$.

[**Hint:** For each, construct a homomorphism and apply the First Isomorphism Theorem.]

Solution: (i) Let $\varphi : G \to G$ be given by $\varphi(g) = g$. Then, it is clear that the only element of the kernel is the identity, that is $\ker(\varphi) = \{1\}$. By the First Isomorphism Theorem, we see that $G/\{1\} \cong G$, since $\mathrm{im}(\varphi)$ is quite clearly all of G (this homomorphism is already an isomorphism).

(ii) Let $\psi : G \to \{1\}$ be given by $\psi(g) = 1$. It is obvious that everything gets sent to the identity, so $\ker(\psi) = G$. Moreover, there is only one element in the image and it is clearly achieved, so $\mathrm{im}(\psi) = \{1\}$. By the First Isomorphism Theorem, we conclude that $G/G \cong \{1\}$. □

Exercise 26 (Longer) Let G be a group, $H \leq G$ and $N \trianglelefteq G$. Prove the following:

(i) $NH = HN$, where $NH \coloneqq \{nh : n \in N \text{ and } h \in H\}$.

[**Hint:** You may find it useful to work with $NH = \bigcup_{h \in H} Nh$.]

(ii) $NH \leq G$ is a subgroup.

[**Hint:** Be aware of Remark 3.98.]

(iii) $N \trianglelefteq NH$ is a normal subgroup.

Solution: (i) Per the hint, we write $NH = \bigcup_{h \in H} Nh = \bigcup_{h \in H} hN = HN$, where we use Lemma 3.11 in the second equality to translate between right and left coset,

(ii) It is first clear that $NH \subseteq G$, because both N and H are (at least) subsets. This then allows for the use of the Subgroup Criterion. Indeed, let $nh, mk \in NH$. Now, we can use (i) above to re-write $nhmk = nm'h'k$ for some $m' \in N$ and $h' \in H$; this is discussed in Remark 3.98.

Basically, because $h \in H$ and $m \in N$, we have $hm \in HN$, but we can always re-write this so that something in N appears on the left (this is what we call m') and something in H appears on the right (this is what we call h'). Thus, $nhmk = nm'h'k \in NH$ because each of N and H are closed under the operation ($nm' \in N$ and $h'k \in H$); this shows closure under the operation.

Next, $1_G \in N$ and $1_G \in H$ since they are subgroups, so $1_G \in NH$.

Finally, for $nh \in NH$, we see that $(nh)^{-1} = h^{-1}n^{-1}$. Again, we can re-write this using (i) above so that it is of the form $(nh)^{-1} = h^{-1}n^{-1} = n'h'$ for $n' \in N$ and $h' \in H$. Thus, $(nh)^{-1} \in NH$; it is closed under forming inverses.

(iii) Let $n \in N$ and $x \in NH$. We must show that $x^{-1}nx \in N$ for normality to hold. First, we write $x = mh$ for some $m \in N$ and $h \in H$. Clearly, $x \in G$. Because of the usual normality $N \trianglelefteq G$ within G, it immediately follows that $x^{-1}nx \in N$. □

Exercise 27 Prove the Third Isomorphism Theorem.

[**Hint:** Apply the First Isomorphism Theorem to $f : G/M \to N/M$ where $f(Mg) = Ng$.]

Solution: □

Exercise 28 Give an example of some $\mathbb{Z}_n$ which is **not** a field and explain why.

[**Hint:** By Example 4.2(iii), it must be that n isn't a prime number.]

Solution: □

Exercise 29 Give an example of a field K and a set V where V is **not** a K-vector space.

Solution: □