NICHOLAS JACKSON

MA267 GROUPS AND RINGS

DRAFT: 07/12/2023

MATHEMATICS INSTITUTE, UNIVERSITY OF WARWICK

Contents

Introduction

Still less is our essay intended as a textbook of the Glass Bead Game; indeed, no such thing will ever be written. The only way to learn the rules of this Game of games is to take the usual prescribed course, which requires many years; and none of the initiates could ever possibly have any interest in making these rules easier to learn.

— Hermann Hesse (1877–1962), *The Glass Bead Game* (1943)

These are the lecture notes for *MA267 Groups and Rings*, an introductory abstract algebra module primarily for second-year undergraduate students on joint mathematical degree programmes at the University of Warwick.

We may always depend upon it that algebra which cannot be translated into good English and sound common sense is bad algebra.

— William Kingdon Clifford (1845–1879), *The Common Sense of the Exact Sciences* (1886) 21

Organisation

Module leader	Dr Nicholas Jackson <Nicholas.Jackson@warwick.ac.uk> Zeeman B0.09, Economics S0.84 Pudding (honorary assistant module leader)
Assistants	Edison Au-Yeung, Alexandros Groutides
Credit	10 CATS
Assessment	One 2–hour examination in April (85%) Best three of four written assignments (15%)
Lectures	Monday 3pm–4pm: L5 (weeks 1–10) Tuesday 3pm–4pm: Chancellors 1 (weeks 1, 4–10), GLT2 (weeks 2, 3) Thursday 2pm-3pm: L5 (weeks 1–10)
Classes	Monday 2pm–3pm: MB0.08 (weeks 2–10) Tuesday 5pm–6pm: Zeeman B3.02 (weeks 2–10)

Content and learning objectives

This is an introductory abstract algebra module. As the title suggests, the two main objects of study are groups and rings. A group is a set with one binary operation; examples include the additive group of integers, groups of permutations, and groups of nonsingular matrices. Rings are sets with two binary operations, analogous to addition and multiplication. The most familiar example is the ring of integers with the usual addition and multiplication operations, and others include rings of polynomials, and rings of square matrices.

This module will assume no prior knowledge of group or ring theory, but students who have previously taken *MA151 Algebra 1* or similar will have met some of the basic concepts already.

We will assume some basic knowledge of the following topics (from *MA138 Sets and Numbers* or elsewhere):

Number theory congruence modulo–n, prime factorisation, the Euclidean algorithm, greatest common divisors (gcd) and least common multiples (lcm).

Sets and functions basic set theory and notation, injective and surjective functions, equivalence relations.

Polynomials multiplication and division, the Euclidean algorithm, the Remainder Theorem.

The topics we will cover include:

Group theory Basic definitions and properties of groups, subgroups and homomorphisms. Cosets and Lagrange's Theorem. Normal subgroups and quotient groups. Cyclic groups, permutation groups, dihedral groups. Isomorphism theorems. Group actions, orbits and stabilisers, conjugacy classes, simple groups. Classification of finitely-generated abelian groups.

Ring theory Basic definitions and properties of rings, subrings and homomorphisms. Ideals and quotient rings. Integral domains, Euclidean domains, Principal Ideal Domains (PIDs), Unique Factorisation Domains (UFDs). Prime and irreducible elements. Fields. Polynomial rings.

By the end of the module, the student should have a good working knowledge of the basic concepts of group theory and ring theory, and be familiar with a number of standard theorems and techniques.

Assessment

The assessment for this module consists of the following:

Assignments Four assignments, with deadlines in weeks 3, 5, 7 and 9 of term 1. The best three marks will together comprise 15% of the overall mark for the module.

Exam One two-hour exam, early in term 3, consisting of one compulsory question (worth 40 marks) and two optional questions (worth 20 marks) from a choice of three, giving a total mark out of 80. The exam will comprise the remaining 85% of the overall mark for the module.

Synergies and further study

This module works well alongside the following other modules:

- *MA243 Geometry*
- *MA257 Introduction to Number Theory*
- *MA266 Multilinear Algebra*

This module provides useful background or assumed knowledge for the following modules:

- *MA257 Introduction to Number Theory*
- *MA3E1 Groups and Representations*
- *MA3G6 Commutative Algebra*
- *MA3J9 Historical Challenges in Mathematics*
- *MA377 Rings and Modules*

- *MA3F1 Introduction to Topology*
- *MA3K4 Introduction to Group Theory*
- *MA3J3 Bifurcations, Catastrophes and Symmetry*
- *MA3D5 Galois Theory*
- *MA3H6 Algebraic Topology*
- *MA3J2 Combinatorics II*
- *MA3A6 Algebraic Number Theory*
- *MA4L6 Analytic Number Theory*
- *MA4H4 Geometric Group Theory*
- *MA426 Elliptic Curves*
- *MA473 Reflection Groups*
- *MA453 Lie Algebras*
- *MA4J8 Commutative Algebra II*
- *MA4M6 Category Theory*

Further reading

These notes contain all the material covered in this module, but you may find it helpful to consult one or more of the following books:

- Lara Alcock, *How to Think About Abstract Algebra*, Oxford University Press (2021)
- M. A. Armstrong, *Groups and Symmetry*, Undergraduate Texts in Mathematics, Springer (1988)
- John B. Fraleigh, *A First Course in Abstract Algebra*, 8th edition, Pearson (2020)
- Joseph Gallian, *Contemporary Abstract Algebra*, 10th edition, CRC Press (2021)
- Nicholas Jackson, *A Course in Abstract Algebra*, Oxford University Press (forthcoming)

However, there is a pleasure in recognizing old things from a new point of view. Also, there are problems for which the new point of view offers a distinct advantage.

— Richard Feynman (1918–1988), *Space-time approach to non-relativistic quantum mechanics*, Reviews of Modern Physics 20 (1948) 367–387

1 Groups

Much of modern mathematics concerns the study of different kinds of structures attached to sets. For example, in linear algebra we study **vector spaces**: sets of "vectors" equipped with "vector addition" and "scalar multiplication" operations. In topology we study **metric spaces** (sets with some well-defined notion of distance between given elements) and **topological spaces** (sets with designated families of "open subsets").[1]

In abstract algebra we are concerned with sets which have some kind of generalised multiplication and/or addition operations. In this module we will investigate two very important objects: **groups** and **rings**.

[1] Considering such objects in generality leads to a branch of mathematics called **category theory**, which is beyond the scope of this module. Sometimes gently derided as "generalised abstract nonsense", it has proved to be a very powerful approach that has enabled major developments not just in mathematics, but also in other fields such as theoretical computer science. The basic idea is that a **category** consists of a class of **objects** and, for each ordered pair of objects, a set of **morphisms**, satisfying a few simple axioms. So the category Set comprises sets and functions mapping between them, while the category $\mathsf{Vect}_{\mathbb{K}}$ consists of vector spaces over a field $\mathbb{K}$ with linear maps between them. Where this becomes particularly useful is when we start looking at **functors**: structure-preserving maps from one category to another. This enables us to translate one sort of mathematical problem into another, which might be easier to solve. If you take MA3F1 Introduction to Topology next year, you'll meet something called the **fundamental group** of a topological space, which is a functor $\pi_1 \colon \mathsf{Top}_* \to \mathsf{Group}$ from the category of based topological spaces to the category of groups. It's essentially a machine for turning a possibly difficult problem in topology into an analogous problem in group theory that's hopefully easier to solve.

1.1 Definitions and elementary properties

We'll use the integers $\mathbb{Z}$ as our motivating example for both of these concepts. From a very early age we learn how to add and multiply integers together. Leaving multiplication aside for the moment, we notice that there are five basic properties that integer addition satisfies, for any integers $a, b, c \in \mathbb{Z}$.

(i) If we add two integers together, we get an integer. That is, $a+b \in \mathbb{Z}$. We say $\mathbb{Z}$ is **closed** under addition.

(ii) The order doesn't matter: we get the same answer either way round. That is, $a+b = b+a$. We say integer addition is **commutative**.

(iii) Parentheses don't matter when we're adding three or more integers together. That is, $(a+b)+c = a+(b+c)$. We say integer addition is **associative**.

(iv) There is a special integer, zero, that doesn't change anything we add it to. That is, $a+0 = a = 0+a$. We call this the **(additive) identity**.

(v) Every integer has a corresponding negative partner. That is, for any $a \in \mathbb{Z}$ there exists $-a \in \mathbb{Z}$ such that $a+(-a) = 0 = (-a)+a$. We call $-a$ the **(additive) inverse** of a.

We want to generalise this idea to an arbitrary (finite or infinite) set, and see what other familiar situations have a similar structure.

What does addition do? It's an operation that takes an ordered pair of elements of our chosen set $\mathbb{Z}$ and gives us a single element of $\mathbb{Z}$ in return. It's effectively a function $f \colon \mathbb{Z} \to \mathbb{Z}$ defined by $f(a, b) = a+b$. We give a function of this sort a special name:

Definition 1.1 Let S be a set. Then a **binary operation** on S is a function $f : S \times S \to S$.

For notational reasons, we will usually write a binary operation not as a function, but as an operator placed between the two arguments: so, for example $a*b$ instead of $f(a,b)$. In particular, this makes the commutativity and associativity conditions neater:

$$a * b = b * a \qquad \text{and} \qquad (a * b) * c = a * (b * c)$$

Although many of the objects we want to study will satisfy the commutativity condition, it turns out that many interesting examples don't, so we'll leave that one as optional for now. But we'll require associativity. We'll also require the existence of inverse elements, and an identity element.[2] This leads us to the following definition:[3]

Definition 1.2 A **group** $G = (G, *)$ comprises a set G together with a binary operation $* : G \times G \to G$, such that:

(G1) The binary operation $*$ is **associative**; that is,

$$g * (h * k) = (g * h) * k$$

for all $g, h, k \in G$.

(G2) There exists an element $e \in G$, the **identity** (or **neutral element**), such that

$$g * e = g = e * g$$

for all $g \in G$.

(G3) For every $g \in G$ there exists an element $g^{-1} \in G$, the **inverse** of g, such that

$$g * g^{-1} = e = g^{-1} * g.$$

Although we decided not to include the commutativity requirement by default, groups which do satisfy this property form an important subclass which we will study in depth. They are named after the early 19th century Norwegian mathematician Niels Henrik Abel, one of the pioneers of group theory.

Niels Henrik Abel (1802–1829)

Definition 1.3 A group $G = (G, *)$ is said to be **abelian** if:

(G4) The operation $*$ is commutative; that is, $g * h = h * g$ for all $g, h \in G$.

Sometimes we will want to discuss the size of a group:[4]

Definition 1.4 Let G be a group. Then the **order** of G, denoted $|G|$ is the number of elements in G. This may be finite or infinite.

Time for some examples. The first is our original example:

Example 1.5 The set $\mathbb{Z}$ of integers forms an abelian group under the usual addition operation.

Similarly, the sets $\mathbb{Q}$ of rational numbers, $\mathbb{R}$ of real numbers and $\mathbb{C}$ of complex numbers also form abelian groups under the corresponding addition operations.

The nonzero elements of the latter three examples above also form groups under multiplication.[5]

[2] There are less restrictive variants of this structure, which we won't cover in this module. A set S with a binary operation $*$ is called a **magma**. A magma whose operation is associative is called a **semigroup**, and a semigroup with an identity is called a **monoid**. A **group** can thus be thought of as a monoid where every element is invertible.

[3] Some books list the closure requirement as an additional criterion in the definition. For example:

(G0) The set G is **closed** under the operation $*$; that is,

$$g * h \in G$$

for all $g, h \in G$

but in our case this is automatically satisfied as part of the way we've defined a binary operation.

[4] Infinity is a tricky concept, and sometimes we might want to make a distinction between **countable** infinite sets, such as $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$, and **uncountable** infinite sets, such as $\mathbb{R}$ and $\mathbb{C}$.

[5] But $\mathbb{Z}$ doesn't. Why not?

Example 1.6 Let $\mathbb{K}$ be one of the sets $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$, and let $\mathbb{K}^* = \mathbb{K} \setminus \{0\}$. Then $\mathbb{K}^*$ forms a group under multiplcation.

The next example comes from modulo–n arithmetic, and yields an important class of groups that we'll study further later.

Example 1.7 Let $\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$ be the set consisting of the first n non-negative integers, for some positive integer n. Then let $+ : \mathbb{Z}_n \times \mathbb{Z}_n \to \mathbb{Z}_n$ be the operation of **addition modulo** n; that is, given any $a, b \in \mathbb{Z}_n$, define $a+b$ to be the remainder of the sum $a+b \in \mathbb{Z}$ after division by n.

Then $\mathbb{Z}_n = (\mathbb{Z}_n, +)$ is the **cyclic group** of order n.

Can we form a group from $\mathbb{Z}_n$ using modulo–n multiplication instead of addition? Well, the problem we run into is that a given element of $\mathbb{Z}_n$ isn't guaranteed to have a modulo–n multiplicative inverse. And zero certainly doesn't. But we can define a multiplicative group structure on some appropriate subset of $\mathbb{Z}_n$:

Example 1.8 Let

$$U_n = \{m \in \mathbb{Z}_n : \gcd(m, n) = 1\}.$$

This set forms a group under multiplication modulo n, and we call it the **group of units modulo** n.

As an exercise, prove that U_n does form a group. Use the extended Euclidean Algorithm, from *MA132 Foundations* or *MA138 Sets and Numbers*.

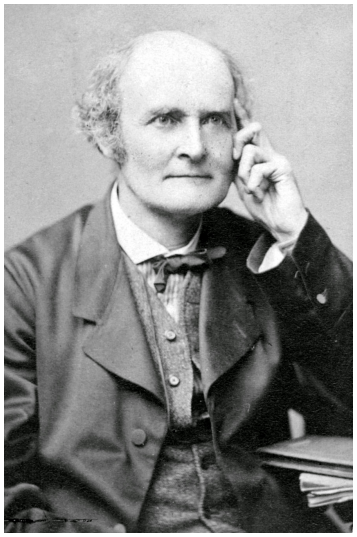
These examples are groups of numbers, but we can form groups from other mathematical objects as well. In a little while, we will meet classes of groups formed from symmetry operations on geometric objects, and groups formed from permutations on sets. For the moment, however, we will look at some important groups formed from matrices:

Example 1.9 For any integers $m, n > 0$ we can define $M_{m\times n}(\mathbb{R})$ or $\mathbb{R}^{m\times n}$ to be the set of $m\times n$ matrices with entries in K. This set forms a group under the usual matrix addition operation: matrix addition is associative, the zero matrix serves as the required identity element, and for any $m\times n$ matrix A there is an additive inverse $-A$.

All of these examples so far are abelian: their operation is commutative. But matrix multiplication isn't commutative, and that yields a number of other interesting matrix groups:

Example 1.10 Denote by $GL_n(\mathbb{R})$ or $GL(n, \mathbb{R})$ the set of $n\times n$ invertible matrices with real entries. Equivalently, this is the set of $n\times n$ real matrices with nonzero determinant. This set forms a group (the **general linear group**) under matrix multiplication.

Example 1.11 Let $SL_n(\mathbb{R})$ or $SL(n, \mathbb{R})$ denote the set of $n\times n$ invertible real matrices with determinant equal to 1. This set also forms a group (the **special linear group**) under ordinary matrix multiplication.



Arthur Cayley (1821–1895)

$+_4$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

Table 1.1: Cayley table for the cyclic group $\mathbb{Z}_4$

$*$	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

Table 1.2: Cayley table for the Klein group V_4

Felix Klein (1849–1925)

Example 1.12 Let $O_n(\mathbb{R})$ or $O(n,\mathbb{R})$ be the set of $n\times n$ real orthogonal matrices. (That is, matrices m such that $M^TM = I = MM^T$, or equivalently that $M^{-1} = M^T$.) This set forms the **orthogonal group** under matrix multiplication.

Similarly, let $SO_n(\mathbb{R})$ or $SO(n,\mathbb{R})$ be the set of $n\times n$ orthogonal matrices with determinant equal to 1. This forms the **special orthogonal group** under ordinary matrix multiplication.

We can generalise all of these matrix groups to other scalar fields such as $\mathbb{Q}$ or $\mathbb{C}$. Except when $n = 1$, the multiplicative groups $GL_n(\mathbb{R})$, $SL_n(\mathbb{R})$, $O_n(\mathbb{R})$ and $SO_n(\mathbb{R})$ are not abelian.

But what do these groups actually look like? Well, one way of displaying the group structure, at least of relatively small groups, is to write down the **multiplication table** or **Cayley table**. This is the same idea as an ordinary multiplication table that we learn about in primary school, except that we use the given group operation instead of ordinary multiplication. Table 1.1 shows the multiplication table (or in this case, perhaps the **addition table**) for the cyclic group $\mathbb{Z}_4$. We can't do this for larger groups or infinite groups, so much of the rest of this module will be concerned with finding different methods to understand these structures.

Another simple but important example is named after the German mathematician Felix Klein:

Example 1.13 Let

$$V_4 = \{e, a, b, c\}$$

and define a group structure on V_4 as follows:

(i) Let e be the identity element.
(ii) Let $a * a = b * b = c * c = e$.
(iii) Let $a * b = b * a = c$.

There is a unique group structure determined by these conditions, and its multiplication table is shown in Table 1.2.

This group is called the **Klein group**, the **Klein 4–group**, or the **Viergruppe**. As an exercise convince yourself that it is indeed an abelian group.

Something that we should really address before we go any further is the question of notation. In almost all the examples we've seen so far, the group operation was either "addition" or "multiplication". In fact, we will almost always write group operations using either additive or multiplicative notation (rather than using a symbol like $*$ as in Definition 1.2), even if the operation isn't called addition or multiplication.

The two notations we will mostly use are:

Multiplicative groups where we omit the sign representing the operation (so $g*h$ becomes gh), we denote the identity element by 1, and the inverse of an element g by g^{-1}.

Additive groups where we represent the operation by $+$, the identity element by 0, and the inverse of an element g by $-g$.

Sometimes we might find ourselves discussing more than one group at a time, and may need to distinguish between the identity elements

of two different groups G and H, say. If so, we will denote them by 1_G and 1_H (or 0_G and 0_H).

We will adopt the convention that additive notation will only be used for abelian groups; that is, any operation we denote by $+$ may be assumed to be commutative. Multiplicative notation, however, may be used both for abelian and for nonabelian groups. By default we will use multiplicative notation.

Now it's time to prove some basic properties of groups. These all follow from Definition 1.2.

First, we will prove the Cancellation Law:

Proposition 1.14 (Cancellation Law) *Let G be a group, and suppose that $g, h, k \in G$.*

(i) *If $gh = gk$ then $h = k$, and*
(ii) *if $hg = kg$ then $h = k$.*

Proof To prove part (i), suppose that $gh = gk$. Then multiplying on the left by g^{-1} we have $g^{-1}(gh) = g^{-1}(gk)$, and by the associativity condition this is equivalent to $(g^{-1}g)h = (g^{-1}g)k$, so $1h = 1k$ and hence $h = k$ as claimed.

Part (ii) can be proved by multiplying on the right by g^{-1}. □

This works precisely because every element in a group has an inverse.

So far, we've talked about *the* identity element of a group, and *the* inverse of a given element, quietly glossing over the possibility that these might not always be unique. We'll justify ourselves now:

Lemma 1.15 *Let G be a group. Then G has a unique identity element 1, and each g in G has a unique inverse g^{-1}. That is:*

(i) *Suppose that $e \in G$ such that $eg = g$ for all $g \in G$ (that is, e is a **left identity element**), then $e = 1$.*
(ii) *Given $g \in G$, if there exists some element $h \in G$ such that $hg = 1$ (that is, h is a **left inverse** of g) then $h = g^{-1}$.*

The group axioms say that a group G has an identity element 1, and every element $g \in G$ has an inverse g^{-1}. This lemma says that any other element that behaves like an identity is actually equal to 1 itself, and any other element that behaves like the inverse of g must actually be equal to g^{-1} itself.

Proof

(i) Because e is a left identity element, we have $e1 = 1$. And because 1 is a two-sided identity element, we also have $e1 = e$. Putting these together, we see that $e = e1 = 1$.
(ii) Because h is a left inverse of g, we have $hg = 1$. And because g^{-1} is a two-sided inverse of g, we have $gg^{-1} = 1$. Combining these, we have

$$h = h1 = h(gg^{-1}) = (hg)g^{-1} = 1g^{-1} = g^{-1}.$$

Hence every group has a unique identity element, and every element has a unique inverse. □

Lemma 1.16 *Suppose that g and h are arbitrary elements of some group G. Then*

$$(gh)^{-1} = h^{-1}g^{-1}.$$

Proof We can check this by means of the following calculation:

$$(h^{-1}g^{-1})(gh) = h^{-1}(g^{-1}(gh)) = h^{-1}(g^{-1}g)h^{-1} \\ = h1h^{-1} = hh^{-1} = 1.$$

Thus $h^{-1}g^{-1}$ is a left inverse of gh. We can either prove it is also a right inverse by a similar calculation, or we can appeal to part (ii) of Lemma 1.15. Either way, the result follows. □

In a multiplicative group, we'll define $g^2 = gg$, $g^3 = gg^2 = ggg$, and so on. Formally, for $n \in \mathbb{N}$ we define g^n inductively by

$$g^1 = g \qquad \text{and} \qquad g^{n+1} = gg^n.$$

We also define g^0 to be the identity element 1, and $g^{-n} = (g^n)^{-1} = (g^{-1})^n$ to be the inverse of g^n. Then we have

$$g^{m+n} = g^m g^n$$

for all $m, n \in \mathbb{Z}$.

In an additive group, we replace g^n with $ng = g + \cdots + g$, and g^{-n} with $(-n)g = -(ng) = n(-g)$.

Looking at the Klein group in Example 1.13, we see that multiplying any of the elements a, b or c by itself yields the identity element e. In other words, $a^2 = b^2 = c^2 = e$. One consequence of this is that each of these elements is equal to its own inverse.[6]

[6] Elements with this self-inverse property are sometimes called **involutory**.

In the group $\mathbb{Z}_4$, whose Cayley table is shown in Table 1.1, we can combine the element 2 with itself to get $2 +_4 2 = 0$. But we have to add four copies of the element 1 together to get 0, and the same goes for 3. And in the additive group $\mathbb{Z}$, there are no nonzero elements which can be added to each other a finite number of times to get the identity element 0.

Let's formalise all this with a definition:[7]

[7] As sometimes happens in mathematics, we've used the same word to mean two different things. In this case, recall that the **order** of a group G is the cardinality (or, if finite, the number of elements) of G. But the **order** of an *element* of G means something different.

Definition 1.17 Let $g \in G$ be an element of some group G. The **order** of g, denoted $|g|$, is the smallest positive integer n such that

$$g^n = g \cdots g = 1$$

(if G is a multiplicative group) or

$$ng = g + \cdots + g = 0$$

(if G is an additive group).

If no such finite integer exists, then we say g has infinite order.

So, for example, in the Klein group V_4 we have $|e| = 1$ and $|a| = |b| = |c| = 2$. And in the cyclic group $\mathbb{Z}_4$ we have $|0| = 1$, while $|2| = 2$ and $|1| = |3| = 4$.

This gives us a big clue that although V_4 and $\mathbb{Z}_4$ have the same number of elements (that is, $|V_4| = |\mathbb{Z}_4| = 4$) they have different

internal structures. We'll look into this idea further in a little while, but first we'll prove a couple of very basic results about the order of group elements:

Lemma 1.18 *Let g be an element of a group G. Then $|g| = 1$ if and only if $g = 1$.*

Proof If $g = 1$ then we have $g^1 = g = 1$, and there is no smaller $k \in \mathbb{N}$ such that $g^k = 1$.

Conversely, if $|g| = 1$ then $g^1 = 1$, but $g^1 = g$ by definition, and hence $g = 1$. □

Lemma 1.19 *Let g be an element of a group G with $|g| = n$. Then $|g| = |g^{-1}|$.*

Proof Since $|g| = n$, we have $g^n = 1$, and n is the smallest positive integer with this property. Then $(g^{-1})^n = (g^n)^{-1} = 1^{-1} = 1$. Furthermore if $0 < k < n$ such that $(g^{-1})^k = 1$, then this means that $(g^k)^{-1} = 1$, and hence $g^k = 1$. This contradicts the hypothesis that $|g| = n$, so it must be the case that $|g^{-1}| = n$. □

Lemma 1.20 *Let g be an element of a group G with $|g| = n$. Then $g^k = 1$ if and only if $n|k$.*

Proof If $n|k$ then there exists some $m \in \mathbb{N}$ such that $k = mn$. Then

$$g^k = g^{mn} = (g^n)^m = 1^m = 1.$$

Conversely, suppose that $g^k = 1$. By Euclid's Division Theorem,[8] we know that if $|k| > n$ then there exist integers q, r such that

$$k = qn + r \qquad \text{and} \qquad 0 \leqslant r < n.$$

Then we have

$$g^k = g^{qn+r} = g^{qn}g^r = (g^n)^q g^r = 1^q g^r = g^r$$

But $g^k = 1$, so this implies that $g^r = 1$, and since $r < n$, which is the smallest positive integer such that $g^n = 1$, it must be the case that $r = 0$. Hence $k = qn$, and so $n|k$ as claimed. □

1.2 *Structural equivalence*

A little while ago we remarked that the Klein group V_4 and the cyclic group $\mathbb{Z}_4$ have the same number of elements, but there were differences in the internal structure of each group. In particular, $\mathbb{Z}_4$ has two elements of order 4, but every element of V_4 (apart from the identity e) has order 2.

There certainly exist bijections between V_4 and $\mathbb{Z}_4$, because they both have the same number of elements. But that's not enough for us: we want to compare the structures as well. And the order of individual elements is a fundamental aspect of that structure. Ultimately, we're looking for bijections between groups that in some way preserve this structure. We'll get to that soon, but first we'll look at another example.

[8] This was covered in *MA138 Sets and Numbers* and *MA132 Foundations*, but if you've not seen it before, or need a reminder, here it is:

Theorem 1.21 (Division Theorem) *Let $a, b \in \mathbb{N}$. Then there exist unique integers $q, r \in \mathbb{Z}$ such that*

$$a = qb + r$$

with $0 \leqslant r < b$.

Euclid of Alexandria (fl. 300 BC) detail from *The School of Athens* by Raphael (1483–1520)

	1	ω	ω^2
1	1	ω	ω^2
ω	ω	ω^2	1
ω^2	ω^2	1	ω

Table 1.3: Cayley table for the group G in Example 1.22

$\cdot$	I	A	B
I	I	A	B
A	A	B	I
B	B	I	A

Table 1.4: Cayley table for the group H in Example 1.22

$+_3$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

Table 1.5: Cayley table for the cyclic group $\mathbb{Z}_3$

Example 1.22 Let $G = \{1, \omega, \omega^2\}$, where $\omega = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$. This forms a group under multiplication, which has Cayley table shown in Table 1.3.

Now let $H = \{I, A, B\}$ where

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A = \frac{1}{2}\begin{bmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix}, \quad B = \frac{1}{2}\begin{bmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{bmatrix}.$$

This forms a group under matrix multiplication, with the Cayley table shown in Table 1.4

Compare this with the Cayley table for $\mathbb{Z}_3$ in Table 1.5. We can see that apart from some simple relabelling, these groups all have essentially the same structure, and on some level we can view them all as different representatives of the "same" group.

We want some notion of equivalence between groups that recognises and preserves the actual structure, while being pretty much agnostic about the specific form of the groups in question. As in Example 1.22 we want to recognise the way the elements of the groups interact, while not really caring particularly whether those elements are integers, complex numbers or matrices.

The key idea, as remarked earlier, is that we want a bijective function[9] between groups that in some way respects the group structure.

Definition 1.24 Two groups G and H are said to be **isomorphic** if there exists a bijective function (an **isomorphism**) $f\colon G \to H$ such that

$$f(ab) = f(a)f(b)$$

for all $a, b \in G$. We denote this by $G \cong H$.

We'll now prove a couple of basic facts about isomorphisms:

Lemma 1.25 *Let $f\colon G \to H$ be an isomorphism of groups. Then*

(i) $f(1_G) = 1_H$, *and*

(ii) $f(g^{-1}) = f(g)^{-1}$ *for all* $g \in G$.

Proof

(i) Since f is a bijection (and hence surjective) for any $h \in H$ there exists some $g \in G$ such that $f(g) = h$. Then

$$f(1_G)h = f(1_G)f(g) = f(1_G g) = f(g) = h.$$

Hence $f(1_G)$ is a left identity in H, and by Lemma 1.15 (i) it must be *the* identity 1_H in H.

(ii) For any $g \in G$ we have

$$f(g^{-1})f(g) = f(g^{-1}g) = f(1_G) = 1_H,$$

hence $f(g^{-1})$ is a left inverse of $f(g)$, and so by Lemma 1.15 (ii) it must be *the* inverse of $f(g)$, namely $f(g)^{-1}$.

Thus $f(1_G) = 1_H$ and $f(g^{-1}) = f(g)^{-1}$ for all $g \in G$, as claimed. □

The next proposition relates to our discussion about the different orders of elements in V_4 and $\mathbb{Z}_4$. Isomorphisms preserve the orders of individual elements:

[9] You should hopefully have met injective, surjective and bijective functions before, but in case you haven't (or need a reminder), here's the definition:

Definition 1.23 Let $f\colon A \to B$ be a function mapping from a set A (the **domain**) to a set B (the **codomain**).

We say that f is **injective** or **one-one** if, for any elements $x, y \in A$, we have $f(x) = f(y)$ *only* when $x = y$.

We say that f is **surjective** or **onto** if, for any element $b \in B$ there exists an element $a \in A$ such that $f(a) = b$.

And we say that f is **bijective** (or a **bijection**) if it is both injective and surjective.

Equivalently, if $f\colon A \to B$ is injective, then every element of B is mapped to by *at most* one element of A. Distinct elements of A are mapped to distinct elements of B.

And if $f\colon A \to B$ is surjective, then every element of B is mapped to by *at least* one element of A.

Proposition 1.26 *Let $f\colon G \to H$ be an isomorphism. Then $|g| = |f(g)|$ for all $g \in G$.*

To prove this, we need to consider the finite-order and infinite-order cases separately.

Proof Suppose first that $|g| = n$ is finite. Then

$$f(g)^n = f(g^n) = f(1_G) = 1_H$$

and hence $|f(g)| \leqslant n = |g|$.

Now let $m = |f(g)|$. Then

$$f(g^m) = f(g)^m = 1_H = f(1_G).$$

Since f is a bijection, and hence injective, we must have $g^m = 1_G$, so $|g| \leqslant m$. Hence

$$|f(g)| \leqslant |g| \leqslant |f(g)|$$

and so $|f(g)| = |g|$ as claimed.

Suppose instead that g has infinite order. Then the elements g^k are distinct for all $k \in \mathbb{Z}$. Since f is a bijection, and hence injective, it follows that the elements $f(g^k) = f(g)^k$ are also distinct for all $k \in \mathbb{Z}$. Therefore $|f(g)| = |g| = \infty$. □

1.3 *Cyclic groups*

We met the cyclic groups $\mathbb{Z}_n$ in Example 1.7, and we want to look at them in a bit more detail now. First of all, we observe that we can construct the entirety of $\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$ using just the group operation $+_n$ and the element 1:

$$\begin{aligned} 0 &= 0 \\ 1 &= 1 \\ 2 &= 1 +_n 1 \\ 3 &= 1 +_n 1 +_n 1 \\ &\vdots \\ n-1 &= 1 +_n \cdots +_n 1 \end{aligned}$$

Formalising this idea we get the following definition:

Definition 1.27 A group G is **cyclic** if it consists of all the integral powers of a single given element. That is, G is cyclic if there exists some element $g \in G$ such that for any $h \in G$ there exists $k \in \mathbb{Z}$ such that $g^k = h$. Or, equivalently,

$$G = \{g^k : k \in \mathbb{Z}\}$$

for some $g \in G$.

The element g is called a **generator** of G.

The cyclic groups we've met so far have been additive rather than multiplicative, but that's really just a matter of notation. We could just as easily have defined

$$\mathbb{Z}_n = \{t^k : 0 \leqslant k < n\}$$

$+_4$	1	t	t^2	t^3
1	1	t	t^2	t^3
t	t	t^2	t^3	1
t^2	t^2	t^3	1	t
t^3	t^3	1	t	t^2

Table 1.6: Cayley table for the cyclic group $\mathbb{Z}_4$ in multiplicative form

for some symbol t and used multiplication such that $t^n = 1$. See Table 1.6 and compare it with Table 1.1.

The additive group $\mathbb{Z}$ of integers is also cyclic: it can be generated additively by the element $1 \in \mathbb{Z}$, since every integer k is of the form $k1$. And as noted above, the finite cyclic groups $\mathbb{Z}_n$ can all be generated by the element $1 \in \mathbb{Z}_n$ using modulo–n addition.[10]

We now provide a complete classification (up to isomorphism) of cyclic groups:

Proposition 1.28 *Any two infinite cyclic groups are isomorphic to $\mathbb{Z}$, and any two finite cyclic groups of order n are isomorphic to $\mathbb{Z}_n$.*

Proof Suppose that G and H are infinite cyclic groups, such that G is generated by some element $g \in G$, and H is generated by some element $h \in H$. Then

$$G = \{g^k : k \in \mathbb{Z}\} \qquad \text{and} \qquad \{h^k : k \in \mathbb{Z}\}.$$

We observed earlier that the elements $g^k \in G$ are all distinct, and so the map $f\colon G \to H$ defined by $f(g^k) = h^k$ for all $k \in \mathbb{Z}$ is a bijection. It also satisfies the structural property

$$f(g^k g^l) = f(g^{k+l}) = h^{k+l} = h^k h^l = f(g^k) f(g^l)$$

and is hence the required isomorphism. Since $\mathbb{Z}$ is also an infinite cyclic group, it follows that any two infinite cyclic groups are isomorphic to each other, and to $\mathbb{Z}$.

Now suppose that G and H are finite cyclic groups of order n. Then

$$G = \{g^k : k \in \mathbb{Z}_n\} \qquad \text{and} \qquad H = \{h^k : k \in \mathbb{Z}_n\}.$$

Again, we define the map $f\colon G \to H$ with $f(g^k) = h^k$. This is also a bijection and satisfies the structural property, and is hence an isomorphism. Since $\mathbb{Z}_n$ is also a finite cyclic group of order n, it follows that any two finite cyclic groups of order n are isomorphic to each other, and to $\mathbb{Z}_n$. □

[10] Is 1 the only generator of $\mathbb{Z}_n$? If not, which other elements of $\mathbb{Z}_n$ are generators?

1.4 Symmetry groups

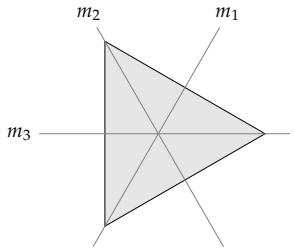


Figure 1.1: Axes of symmetry of an equilateral triangle

Another rich source of groups, and one of the original motivations for the subject, is geometry. This has been a particularly important line of inquiry in particle physics and molecular chemistry.

Example 1.29 Consider an equilateral triangle (see Figure 1.1). There are six different symmetry operations we can perform on this:

- The identity operation, which just maps the triangle to itself.
- Reflections in each of the three axes of symmetry.
- Rotations through $\pm\frac{1}{3}$ full turn.

We can compose these operations as if they were functions, by doing one and then another. In each case, we get one of the six operations on the list.

So, here we have a set of objects (in this case, symmetry operations or **isometries**) that is closed under a binary operation (composition). We can turn this into a group:

Example 1.30 Let

$$D_3 = \{e, r, r^2, m_1, m_2, m_3\}$$

be the set of symmetry operations of an equilateral triangle. Here,

- e is the identity operation,
- r is an anticlockwise rotation through $\frac{2\pi}{3}$,
- r^2 is an anticlockwise rotation through $\frac{4\pi}{3}$ (or equivalently a clockwise rotation through $\frac{2\pi}{3}$), and
- m_1, m_2 and m_3 are reflections in the axes shown in Figure 1.1.

Composing each of these yields the group structure shown in Table 1.7. We call this the **dihedral group** of the triangle.

$\cdot$	e	r	r^2	m_1	m_2	m_3
e	e	r	r^2	m_1	m_2	m_3
r	r	r^2	e	m_3	m_1	m_2
r^2	r^2	e	r	m_2	m_3	m_1
m_1	m_1	m_2	m_3	e	r	r^2
m_2	m_2	m_3	m_1	r^2	e	r
m_3	m_3	m_1	m_2	r	r^2	e

Table 1.7: The Cayley table for the dihedral group D_3

More generally:[11]

Definition 1.31 Let $n \in \mathbb{N}$ with $n \geqslant 3$, and denote by P_n the regular n–sided polygon in the plane with vertices at the points $\left(\cos\left(\frac{2\pi k}{n}\right), \sin\left(\frac{2\pi k}{n}\right)\right)$ for $0 \leqslant k < n$. Then

$$D_n = \{e, r, r^2, \ldots, r^{n-1}, m_1, \ldots, m_n\}$$

is the **dihedral group** of P_n.

Here, r is an anticlockwise rotation of P_n through an angle $\frac{2\pi}{n}$. Then for $0 \leqslant k < n$, the power r^k denotes a $\frac{2\pi k}{n}$ anticlockwise rotation, with $e = r^0$ the identity map.

Furthermore, m_k denotes a reflection in the line through the origin that makes an angle $\frac{k\pi}{n}$ with the positive horizontal axis. If n is odd, then these lines will pass through a vertex and the midpoint of its opposite side. If n is even, then half of these lines will pass through opposite vertices, and the other half will pass through the midpoints of opposite sides.

See Figure 1.2 and Table 1.8 for the diagram and Cayley table for the dihedral group D_4.

[11] In these notes, we denote the dihedral group of the n–sided polygon by D_n. However, you should be aware that since $|D_n| = 2n$, some books denote it D_{2n}. It should be clear from context, but it's something to watch out for.

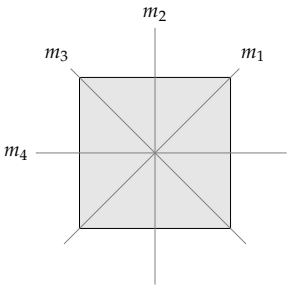


Figure 1.2: Axes of symmetry of the square

	e	r	r^2	r^3	m_1	m_2	m_3	m_4
e	e	r	r^2	r^3	m_1	m_2	m_3	m_4
r	r	r^2	r^3	e	m_4	m_1	m_2	m_3
r^2	r^2	r^3	e	r	m_3	m_4	m_1	m_2
r^3	r^3	e	r	r^2	m_2	m_3	m_4	m_1
m_1	m_1	m_2	m_3	m_4	e	r	r^2	r^3
m_2	m_2	m_3	m_4	m_1	r^3	e	r	r^2
m_3	m_3	m_4	m_1	m_2	r^2	r^3	e	r
m_4	m_4	m_1	m_2	m_3	r	r^2	r^3	e

Table 1.8: The multiplication table for the dihedral group D_4

The dihedral groups are not in general abelian. We can see this from examining Tables 1.7 and 1.8: neither is symmetric in the leading (top-left to bottom right) diagonal. We will look in more detail at the dihedral groups later.

1.5 *Permutation groups*

Now we will look at an important class of groups. Historically, these were some of the first groups studied in generality: the abstract concept of a group as we understand it now didn't really evolve until the late 19th century.

Definition 1.32 A **permutation** on a set X is a bijection $\sigma\colon X \to X$. We denote by $\mathrm{Sym}(X)$ the set of all permutations on X.

Choose a (finite or infinite) set X and consider two permutations σ and τ. Since these are functions, we can compose them to get new functions $\sigma\circ\tau$ and $\tau\circ\sigma$ defined by

$$(\sigma\circ\tau)(x) = \sigma(\tau(x)) \qquad \text{and} \qquad (\tau\circ\sigma)(x) = \tau(\sigma(x)).$$

These are both well-defined functions, and because the composite of two bijections is also a bijection, they are also permutations on the set X.

So we have a binary operation on $\mathrm{Sym}(X)$. This operation is associative, because function composition is associative. We have an identity permutation: the identity map $\iota\colon X \to X$, where $\iota(x) = x$ for all $x \in X$. And for any permutation $\sigma \in \mathrm{Sym}(X)$ there is a well-defined inverse permutation σ^{-1} which we can regard either as the inverse of the map σ, or as the permutation that puts every element of X back to where it was before σ shuffled everything around. Hence $\mathrm{Sym}(X)$ forms a group under composition:

Definition 1.33 Let X be a (finite or infinite) set. The group $\mathrm{Sym}(X)$, of all permutations $\sigma\colon X \to X$, is the **symmetric group** on X.

If $X = \{1, \ldots, n\}$ is a finite set consisting of n elements, we call $\mathrm{Sym}(X)$ the **symmetric group** on n objects, and denote it S_n.

If X is an infinite set, then $\mathrm{Sym}(X)$ will also be infinite. But if X is finite, then there are only finitely many distinct ways of rearranging the elements of X, and so S_n has finite order:

Proposition 1.34 *The finite symmetric group S_n has order $n!$.*

Proof A permutation of the set $X = \{1, \ldots, n\}$ is completely determined by how it maps the numbers amongst themselves. There are n choices for where 1 maps to, then $(n-1)$ choices for where 2 goes (since it can map to any of the remaining numbers except for the one we mapped 1 to), then $(n-2)$ possible choices for where 3 maps to, and so on. So $|S_n| = n(n-1)(n-2)\ldots 1 = n!$ as claimed. □

More generally, we have the following fact:

Proposition 1.35 *Let X and Y be two sets with $|X| = |Y|$. Then* $\mathrm{Sym}(X) \cong \mathrm{Sym}(Y)$.

To help us work with permutations, we would like a consistent notation. We have a couple of options.

One approach is that since $\sigma\colon X \to X$ is determined completely by its action on the elements of X, we can represent it as an array:

$$\begin{bmatrix} 1 & 2 & \ldots & n \\ \sigma(1) & \sigma(2) & \ldots & \sigma(n) \end{bmatrix}$$

The first row lists the elements of X and the second lists their images under the action of σ. So, suppose that $\sigma \in S_5$ maps

$$1 \mapsto 1, \qquad 2 \mapsto 3, \qquad 3 \mapsto 5, \qquad 4 \mapsto 4, \qquad 5 \mapsto 2.$$

Then we can represent σ by the array

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 5 & 4 & 2 \end{bmatrix}.$$

Suppose that we have another permutation $\tau \in S_5$ such that

$$1 \mapsto 2, \qquad 2 \mapsto 4, \qquad 3 \mapsto 5, \qquad 4 \mapsto 1, \qquad 5 \mapsto 3.$$

Then τ can be represented by the array

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 5 & 1 & 3 \end{bmatrix}.$$

We can represent composition quite easily by stacking the arrays on top of each other:

$$\tau\sigma = \begin{matrix} \sigma \\ \tau \end{matrix} \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 5 & 4 & 2 \\ 2 & 4 & 5 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 5 & 3 & 1 & 4 \end{bmatrix}$$

In general, composition isn't commutative, so we have to be careful of the order.[12] For example,

$$\sigma\tau = \begin{matrix} \tau \\ \sigma \end{matrix} \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 5 & 1 & 3 \\ 1 & 3 & 5 & 4 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 4 & 2 & 1 & 5 \end{bmatrix} \neq \tau\sigma$$

This notation is quite clear, and makes it easy to work out the composite of two permutations, but it becomes unwieldy with larger numbers of permuting objects. It also doesn't really tell us much about the internal structure of the permutation.

For example, σ leaves 1 and 4 unchanged, but maps $2 \mapsto 3$, $3 \mapsto 5$ and $5 \mapsto 2$. So repeated applications of σ leave 1 and 4 where they are, while 2, 3 and 5 cycle amongst themselves. We can depict all this graphically (see Figure 1.3).

Ideally, however, we want a more compact notation that will enable us to see the permutation's internal structure. The key is to split the permutation into disjoint cyclic subpermutations. For example, the three-element cycle in σ can be written as $(2, 3, 5)$, because each element in the list maps to the next one along, wrapping back round to the beginning. The fixed elements 1 and 4 could be written as single-element cycles (1) and (4), but by convention we usually just omit these for conciseness. So the ordered list $(2, 3, 5)$ encodes everything we need to know about the permutation σ.

Definition 1.36 Let X be a set, and suppose that $x_1, \ldots, x_k$ are distinct elements of X. The **cycle** $(x_1, \ldots, x_k)$ denotes the permutation $\phi \in \mathrm{Sym}(X)$ such that:

(i) $\phi(x_i) = x_{i+1}$ for $1 \leqslant i < k$,
(ii) $\phi(x_k) = x_1$, and
(iii) $\phi(y) = y$ for all $y \in X \setminus \{x_1, \ldots, x_k\}$.

Similarly, we can write

$$\tau = (1, 2, 4)(3, 5), \qquad \tau\sigma = (1, 2, 5, 4), \qquad \sigma\tau = (1, 3, 2, 4).$$

The permutation τ consists of two nontrivial cycles $(1, 2, 4)$ and $(3, 5)$, which don't interact with each other: they act on disjoint subsets of elements. In fact, we can write any permutation as a product of disjoint cycles:

[12] Because we regard permutations as bijections, and the product operation as being composition, we write products from right to left, rather than left to right. So $\tau\sigma$ means σ followed by τ. Stacking the arrays vertically, we read down the page, so $\begin{smallmatrix} \sigma \\ \tau \end{smallmatrix}$ means σ followed by τ.

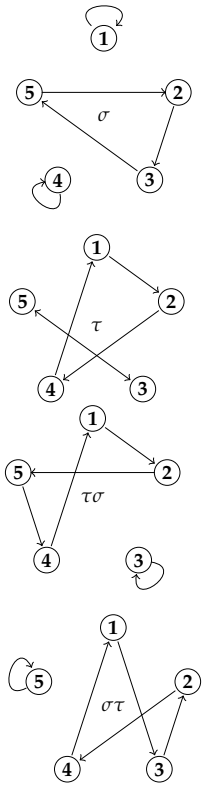


Figure 1.3: Graphical depictions of permutations σ, τ, $\tau\sigma$ and $\sigma\tau$ in S_5

Proposition 1.37 *Any permutation $\sigma \in S_n$ can be written as a product of disjoint cycles.*

Proof We will prove this by giving a well-defined procedure for decomposing a given permutation into disjoint cycles:

- Open parentheses $($.
- Write down the first element 1.
- Write down $\sigma(1)$.
- Write down $\sigma^2(1) = \sigma(\sigma(1))$.

  $\vdots$
- When we get back to 1, close parentheses $)$.

Now repeat this process, but instead of starting with 1, start with the smallest integer not yet seen. Continue until all integers $1, \ldots, n$ have been written down. Then delete all single-element cycles. What remains is a product of all the disjoint cycles in σ. □

It isn't immediately obvious how to multiply permutations together, but with practice it turns out to be easier than it might appear.

Example 1.38 Let $\sigma = (2,3,5)$ and $\tau = (1,2,4)(3,5)$. We calculate $\sigma\tau$ as follows:

$$\sigma\tau = (2,3,5)(1,2,4)(3,5)$$

Start with 1, and read through the list of cycles from right to left, applying each one in turn until you've done them all:

$$1 \overset{(3,5)}{\longmapsto} 1 \overset{(1,2,4)}{\longmapsto} 2 \overset{(2,3,5)}{\longmapsto} 3$$

Now do the same process, but starting with the number (in this case 3) that you ended up with last time:

$$3 \overset{(3,5)}{\longmapsto} 5 \overset{(1,2,4)}{\longmapsto} 5 \overset{(2,3,5)}{\longmapsto} 2$$

Now do it again, and again, until you end up back at 1:

$$2 \overset{(3,5)}{\longmapsto} 2 \overset{(1,2,4)}{\longmapsto} 4 \overset{(2,3,5)}{\longmapsto} 4$$
$$4 \overset{(3,5)}{\longmapsto} 4 \overset{(1,2,4)}{\longmapsto} 1 \overset{(2,3,5)}{\longmapsto} 1$$

This gives the first disjoint cycle $(1,3,2,4)$. Now repeat this process with the smallest number (in this case 5) not yet seen:

$$5 \overset{(3,5)}{\longmapsto} 3 \overset{(1,2,4)}{\longmapsto} 3 \overset{(2,3,5)}{\longmapsto} 5$$

Thus 5 is unchanged by $\sigma\tau$, so our next cycle is (5), except that by convention we omit length–1 cycles. Since all of the numbers $1, \ldots, 5$ are now accounted for, we are done, and $\sigma\tau = (1,3,2,4)$.

Now we introduce a couple of definitions that will be useful later.

Definition 1.39 Let $\sigma = (x_1, \ldots, x_k)$ be a finite permutation in some (possibly infinite) symmetric group $\mathrm{Sym}(X)$. Then σ has **length** or **periodicity** $l(\sigma) = k$. This is equal to the order of the element σ in $\mathrm{Sym}(X)$. A cycle of length 2 is called a **transposition**.

It's probably about time we looked at a few concrete examples.

Example 1.40 The symmetric group S_3 has $3! = 6$ elements. These are:

- The identity permutation $\iota = (\ )$.
- Three transpositions $(1,2)$, $(1,3)$ and $(2,3)$.
- Two 3–cycles $(1,2,3)$ and $(1,3,2)$.

Example 1.41 The symmetric group S_4 has $4! = 24$ elements. These are:

- The identity permutation $\iota = (\ )$.
- Six transpositions:

$$(1,2) \quad (1,3) \quad (1,4) \quad (2,3) \quad (2,4) \quad (3,4)$$

- Eight 3–cycles

$$\begin{array}{cccc} (1,2,3) & (1,2,4) & (1,3,4) & (2,3,4) \\ (1,3,2) & (1,4,2) & (1,4,3) & (2,4,3) \end{array}$$

- Three double transpositions:

$$(1,2)(3,4) \qquad (1,3)(2,4) \qquad (1,4)(2,3)$$

- Six 4–cycles:

$$\begin{array}{ccc} (1,2,3,4) & (1,3,2,4) & (1,4,2,3) \\ (1,2,4,3) & (1,3,4,2) & (1,4,3,2) \end{array}$$

It turns out that every finite permutation can be written as a product of transpositions, although these transpositions need not be disjoint.

Proposition 1.42 *Let $\sigma \in \mathrm{Sym}(X)$ be a finite permutation in some (possibly infinite) symmetric group $\mathrm{Sym}(X)$. Then σ can be written as a product of (not necessarily disjoint) transpositions.*

Proof We know from Proposition 1.37 that any finite permutation $\sigma \in \mathrm{Sym}(X)$ can be written as a product of disjoint cycles. We will now show that any finite-length cycle can be written as a product of transpositions (that is, length–2 cycles).

Consider a cycle $(x_1, \ldots, x_k)$. Then this can be written as

$$(x_1, \ldots, x_k) = (x_1, x_k)(x_1, x_{k-1}) \ldots (x_1, x_3)(x_1, x_2).$$

Doing this for each of the (disjoint) cycles in σ yields a product of (not necessarily disjoint) transpositions. □

Corollary 1.43 *The transpositions in S_n generate S_n.*

So, we can decompose a finite permutation into a product of disjoint cycles, and this decomposition is unique up to a certain amount of reordering. But the decomposition into transpositions isn't necessarily going to be unique. It's not even the case that two different decompositions will have the same number of transpositions. But we can at least talk about the parity of the number of transpositions:

Definition 1.44 Let $\sigma \in \mathrm{Sym}(X)$ be a finite permutation in a (possibly infinite) symmetric group $\mathrm{Sym}(X)$. We say σ is **even** if it can be written as a product of an even number of transpositions, and **odd** if it can be written as a product of an odd number of transpositions.

In order for this to make sense, we need the following:

Proposition 1.45 *Let $\sigma \in \mathrm{Sym}(X)$ be a finite permutation on some set X. Then σ is either even or odd, but not both.*

What this says is that we can separate finite permutations into two types, depending on whether they decompose as an odd or even number of transpositions. We'll omit the proof because it's not very illuminating and would be a bit of a digression at this point.

But now consider two permutations σ and τ. If both are even, then their product will also be even. If one is even and one is odd, then their product will be odd. And if both are odd, then their product will be even.[13] What this means is that the set of even permutations in $\mathrm{Sym}(X)$ is closed under composition. This gives us another important subfamily of permutation groups:

[13] Try to convince yourself that this is true, perhaps by looking at a few examples.

Definition 1.46 Let X be a (possibly infinite) set. Denote by $\mathrm{Alt}(X)$ the group of even permutations of X. This is the **alternating group** on X.

If $X = \{1, \ldots, n\}$ then we will usually denote $\mathrm{Alt}(X)$ by A_n.

We end this chapter with the observation that a cycle of length k, for $k \geqslant 2$, is an even permutation if k is odd, and an odd permutation if k is even. So, in particular, 3–cycles are even permutations.

Example 1.47 The alternating group A_3 consists of the identity permutation $\iota = (\ )$ and the two 3–cycles $(1,2,3)$ and $(1,3,2)$.

Example 1.48 The alternating group A_4 consists of the identity permutation $\iota = (\ )$, the eight 3–cycles

$$\begin{array}{cccc} (1,2,3) & (1,2,4) & (1,3,4) & (2,3,4) \\ (1,3,2) & (1,4,2) & (1,4,3) & (2,4,3) \end{array}$$

and the three double transpositions

$$(1,2)(3,4) \qquad (1,3)(2,4) \qquad (1,4)(2,3)$$

Notice that in both of these cases, $|A_n| = \frac{1}{2}|S_n|$. This happens to be true for all $n \in \mathbb{N}$, and the proof is left as an exercise.

The reader will find no figures in this work. The methods which I set forth do not require either constructions or geometrical or mechanical reasonings: but only algebraic operations, subject to a regular and uniform rule of procedure.

— Joseph-Louis Lagrange
(1736–1813),
preface to *Mécanique Analytique*
(1788)

2 *Subgroups*

If we look carefully at the Cayley tables of some of the groups we met in the last chapter, we can see some interesting internal structure. The dihedral group D_3, for example, has an obvious block consisting of just the identity and the two rotations (see Table 2.1). We could throw away the three reflections and still be left with a perfectly respectable group (which in this case happens to be isomorphic to $\mathbb{Z}_3$). Or, we could throw away everything except the identity e and one reflection (m_1, say) and still have an order–2 group isomorphic to $\mathbb{Z}_2$ (see Table 2.2).

$\cdot$	e	r	r^2	m_1	m_2	m_3
e	e	r	r^2	m_1	m_2	m_3
r	r	r^2	e	m_3	m_1	m_2
r^2	r^2	e	r	m_2	m_3	m_1
m_1	m_1	m_2	m_3	e	r	r^2
m_2	m_2	m_3	m_1	r^2	e	r
m_3	m_3	m_1	m_2	r	r^2	e

Table 2.1: The multiplication table for the dihedral group D_3 with the subgroup $\{e, r, r^2\}$ highlighted

$\cdot$	e	r	r^2	m_1	m_2	m_3
e	e	r	r^2	m_1	m_2	m_3
r	r	r^2	e	m_3	m_1	m_2
r^2	r^2	e	r	m_2	m_3	m_1
m_1	m_1	m_2	m_3	e	r	r^2
m_2	m_2	m_3	m_1	r^2	e	r
m_3	m_3	m_1	m_2	r	r^2	e

Table 2.2: The multiplication table for the dihedral group D_3 with the subgroup $\{e, m_1\}$ highlighted

2.1 *Definitions, examples and elementary properties*

We'll start with a definition.

Definition 2.1 Let G be a group. A subset $S \subset G$ is a **subgroup** of G if it forms a group under the same operation as G. We denote this by $H \leqslant G$.

Subgroups inherit their identity elements from their parent groups:

Lemma 2.2 *If H is a subgroup of a group G, then the identity element 1_H is equal to the identity element 1_G of G.*

Proof For every $h \in H$ we have $1_G h = h$ by the definition of the identity in G. Applying Lemma 1.15 (i) to the group H, it follows that $1_H = 1_G$. □

We can use the following proposition as a method for checking whether a given subset is actually a subgroup.

Proposition 2.3 *Let H be a nonempty subset of a group G. Then $H \leqslant G$ if and only if:*

(i) *H is closed under the group operation, that is, $h_1 h_2 \in H$ for all $h_1, h_2 \in H$; and*

(ii) *H contains all required inverses, that is $h^{-1} \in H$ for all $h \in H$.*

Proof The subset H is a subgroup of G if and only if the four group axioms (G0)–(G3) hold.[1] Two of these, the closure axiom (G0) and the inverse axiom (G3) are conditions (i) and (ii) of the lemma, and so if H is a subgroup then conditions (i) and (ii) must hold.

Conversely, suppose that (i) and (ii) hold. Then we need to show that the associativity axiom (G1) and the identity axiom (G2) also hold in H. The first of these is simple: since the inherited operation is associative in G, it must also be associative in H because H is

[1] Definition 1.2, page 2.

a subset of H. And since H is nonempty, it must contain at least one element h. By condition (ii) it must also contain the inverse h^{-1} of h, and then $hh^{-1} = 1$, which must also belong to H by the closure condition (i). Hence the identity condition (G2) holds in H and therefore H is a subgroup. □

Having introduced the concept and proved a couple of basic facts, it's time to look at some examples.

Example 2.4 Let G be any group. Then G is a subgroup of itself. Also, the **trivial subgroup** $\{1\}$, consisting just of the identity, is also a subgroup of G. Subgroups other than G are called **proper** subgroups, and subgroups other than $\{1\}$ are said to be **nontrivial**.

Example 2.5 The nonzero real numbers $\mathbb{R}^*$ form a subgroup of the multiplicative group $\mathbb{C}^*$ of nonzero complex numbers. Another subgroup of $\mathbb{C}^*$ is the one consisting of all the complex numbers $z \in \mathbb{C}$ such that $|z| = 1$.

Example 2.6 The matrix groups $SL_n(\mathbb{R})$, $O_n(\mathbb{R})$ and $SO_n(\mathbb{R})$ are all subgroups of $GL_n(\mathbb{R})$, and furthermore $SO_n(\mathbb{R})$ is a subgroup of $O_n(\mathbb{R})$.

Example 2.7 If g is any element of a group G, then we define the **cyclic subgroup generated by** g to be

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

Let's look at this example in more detail. For example, if $G = \mathbb{Z}$, then the subset

$$5\mathbb{Z} = \{5n : n \in \mathbb{Z}\},$$

consisting of all integer multiples of 5, is the cyclic subgroup generated by 5.

If $G = \langle g \rangle$ is a finite cyclic group of order n, and m is a positive integer dividing n, then the cyclic subgroup $\langle g^m \rangle$ has order n/m and consists of the elements g^{mk} for $0 \leqslant k < n/m$.

Example 2.8 Let X be a set. The alternating group $\mathrm{Alt}(X)$ is a subgroup of the symmetric group $\mathrm{Sym}(X)$. And if X is a finite set with n elements, we have $A_n \leqslant S_n$.

We end this section with another basic fact: intersections of subgroups are subgroups.

Proposition 2.9 *Let G be a group, and suppose that H and K are both subgroups of G. Then their intersection $H \cap K$ is also a subgroup of G.*

Proof Since H and K are subgroups of G, they must both contain the identity 1_G. Hence $1_G \in H \cap K \neq \emptyset$.

Having shown that $H \cap K$ is nonempty, we can apply Proposition 2.3.

Let $a, b \in H \cap K$ be arbitrary elements. Since $a, b \in H$ and H is a subgroup, their product ab must also lie in H. Similarly, since $a, b \in K$, by closure $ab \in K$. Hence $ab \in H \cap K$ and so $H \cap K$ is closed under the group operation.

Now consider $a \in H \cap K$. Since $a \in H$ and H is a subgroup, the

inverse a^{-1} also lies in H. And since $a \in K$ it follows that $a^{-1} \in K$. Therefore $a^{-1} \in H \cap K$, and hence $H \cap K \leqslant G$. □

In the previous chapter, we spent some time studying permutations. One reason for this is the following important result, which says that every group can be regarded as a permutation group.

Theorem 2.10 (Cayley's Theorem) *Any group G is isomorphic to a subgroup of $\mathrm{Sym}(X)$ for some set X.*

Proof For every element $g \in G$ we define the function $\lambda_g \colon G \to G$ by $\lambda_g(h) = gh$ for all $h \in G$. In the special case where $g = 1$, the function λ_1 is just the identity map on G.

More generally, λ_g is a bijection $G \to G$ and hence a permutation on G. In other words, $\lambda_g \in \mathrm{Sym}(G)$.

The injectivity of λ_g follows from the left cancellation law:[2] if $\lambda_g(h) = \lambda_g(k)$ for some $h, k \in G$, then this means that $gh = gk$, whence $h = k$.

[2] Proposition 1.14, page 5.

To prove surjectivity of λ_g, given some $h \in G$ we want to find some $k \in G$ such that $h = \lambda_g(k)$. But this is the same as saying $h = gk$, from which we see that $k = g^{-1}h$, and so $\lambda_g(k) = gg^{-1}h = h$. Therefore λ_g is a bijection, and thus a permutation of G.

Let $S = \{\lambda_g : g \in G\}$. We now want to show that this subset $S \subseteq \mathrm{Sym}(G)$ is actually a subgroup of $\mathrm{Sym}(G)$, and furthermore that it's isomorphic to G itself. The group operation in S is just the usual composition operation inherited from $\mathrm{Sym}(G)$.

This set S is nonempty, so we can apply Proposition 2.3. First we check the closure condition. Given $g, h, k \in G$, we have

$$(\lambda_g \circ \lambda_h)(k) = \lambda_g(\lambda_h(k)) = \lambda_g(hk) = g(hk) = (gh)k = \lambda_{gh}(k)$$

and since for any $g, h \in G$ the product $gh \in G$ so the permutation λ_{gh} also lies in S. So S is closed under composition.

Now we need to show that S contains all required inverses. The function λ_1 is just the identity map on G, and hence the identity permutation $\iota \in \mathrm{Sym}(G)$. So the inverse λ_g^{-1} is the permutation λ_h such that

$$\lambda_h \circ \lambda_g = \lambda_1 = \lambda_g \circ \lambda_h$$

But we know from the previous paragraph that $\lambda_g \circ \lambda_h = \lambda_{gh}$, so what we're really looking for is $h \in G$ such that $hg = 1$. That is, $h = g^{-1}$, and hence $\lambda_g^{-1} = \lambda_{g^{-1}}$.

So S is nonempty, closed under composition, and contains both an identity element and a full set of inverses, and is therefore a subgroup of $\mathrm{Sym}(G)$. All we need to do now is show that $S \cong G$, which requires us to find a bijection $f \colon G \to S$ satisfying the structural condition in Definition 1.24.

The obvious candidate for this isomorphism f is the map that takes an element g to its corresponding permutation $\lambda_g \in S$. So define $f(g) = \lambda_g$ for all $g \in G$.

This map f is injective: suppose that $f(g) = \lambda_g = \lambda_h = f(h)$ for some $g, h \in G$. Then it follows that $\lambda_g(k) = gk = hk = \lambda_h(k)$ for all $k \in G$, and by the left cancellation law it follows that $g = h$.

Surjectivity follows from the definition of S: for any $\lambda_g \in S$ there exists $g \in G$ such that $f(g) = \lambda_g$.

ThusHence f is a bijection, and all that remains is to check the structure condition:

$$f(g) \circ f(h) = \lambda_g \circ \lambda_h = \lambda_{gh} = f(gh).$$

Hence f is the required isomorphism $G \cong S \leqslant \mathrm{Sym}(G)$. □

2.2 *Cosets and Lagrange's Theorem*

In the proof of Cayley's Theorem, we introduced bijections $\lambda_g \colon G \to G$ for all $g \in G$. Since they are bijections, we have $\lambda_g(G) = G$ in all cases. But what happens if we choose a subgroup $H \leqslant G$ and look at the various images $\lambda_g(H)$ as g varies throughout G?

Well, it turns out that

$$\lambda_g(H) = \{gh : h \in H\} \subseteq G.$$

Let's look at a concrete example.

Example 2.11 Let $R_3 = \{e, r, r^2\} \subset D_3$. Then we have:

$$\begin{aligned} \lambda_e(R_3) &= \{e, r, r^2\} = R_3, & \lambda_{m_1}(R_3) &= \{m_1, m_2, m_3\} = D_3 \setminus R_3, \\ \lambda_r(R_3) &= \{r, r^2, e\} = R_3, & \lambda_{m_2}(R_3) &= \{m_2, m_3, m_1\} = D_3 \setminus R_3, \\ \lambda_{r^2}(R_3) &= \{r^2, e, r\} = R_3, & \lambda_{m_3}(R_3) &= \{m_3, m_1, m_2\} = D_3 \setminus R_3. \end{aligned}$$

So the action of the permutations on the subgroup R_3 partitions D_3 into two distinct subsets: R_3 itself and its complement $D_3 \setminus R_3$.

These subsets are going to be important, so we'll give them a special name:

Definition 2.12 Let G be a group, let $H \leqslant G$ be a subgroup of G, and let $g \in G$. Then the **left coset** of H determined by g is

$$gH = \{gh : h \in H\}$$

and the corresponding **right coset** is

$$Hg = \{hg : h \in H\}.$$

In the case of additive groups, we usually denote the cosets by $g+H$. Note that it is always the case that g lies in gH and Hg, because since $1 \in H$ we have $g = g1 \in gH$ and $g = g1 \in Hg$.

Proposition 2.13 *Let G be a group, let $H \leqslant G$, and suppose that $g, k \in G$. Then the following statements are equivalent:*

(i) $k \in gH$,
(ii) $gH = kH$,
(iii) $g^{-1}k \in H$.

Proof First we show that (i) $\Longrightarrow$ (ii). Suppose that $k \in gH$. Then $k = gh$ for some fixed $h \in H$. Multiplying on the right by h^{-1} gives $g = kh^{-1}$. Let $x \in gH$. Then for some $h_1 \in H$ we have $x = gh_1 = kh^{-1}h_1 \in kH$, so $gH \subseteq kH$. Similarly, if $x \in kH$ then

for some $h_2 \in H$ we have $x = kh_2 = ghh_2 \in gH$, so $kH \subseteq gH$, and hence $kH = gH$.

Showing that (ii) $\implies$ (i) is fairly straightforward. Suppose that $gH = kH$. Then $k = k1 \in kH = gH$, so $k \in gH$.

Now suppose that $k \in gH$. Then again, $k = gh$ for some fixed $h \in H$, and multiplying on the left by g^{-1} gives $g^{-1}k = h \in H$, so (i) $\implies$ (iii).

Finally, if $g^{-1}k \in H$, then putting $h = g^{-1}k$ we have $gh = k$, so $k \in gH$. Hence (iii) $\implies$ (i), and all three conditions are equivalent. □

We saw in the example above that the rotation subgroup $R_3 = \{e, r, r^2\}$ has two cosets in D_3, namely R_3 itself and its complement. What about another subgroup?

Example 2.14 Let $H = \{e, m_1\} \leqslant D_3$. Then the left and right cosets of H in D_3 are as follows:

$$\begin{aligned} eH = m_1H = H, &\qquad He = Hm_1 = H, \\ rH = m_3H = \{r, m_3\}, &\qquad Hr = Hm_2 = \{r, m_2\}, \\ r^2H = m_2H = \{r^2, m_2\}, &\qquad Hr^2 = Hm_3 = \{r^2, m_3\}. \end{aligned}$$

There are two things to notice here. The first is that the left cosets neatly partition D_3 into three disjoint subsets, and so do the right cosets. We'll look at this behaviour now. The other thing to notice is that the left cosets of H aren't necessarily the same as the right cosets of H, although this was the case when we looked at R_3 a little while ago. We'll investigate this in the next chapter.

Corollary 2.15 *Two left cosets g_1H and g_2H of H in G are either equal or disjoint.*

Proof If g_1H and g_2H aren't disjoint, then there exists some element $k \in g_1H \cap g_2H$. But then $k \in g_1H$ so $kH = g_1H$, and also $k \in g_2H$, so $kH = g_2H$. Hence $g_1H = kH = g_2H$. □

This immediately implies the following:

Corollary 2.16 *The left cosets of H in G partition G.*

Proposition 2.17 *Let G be a group, and H a finite subgroup of G. All left cosets of H have exactly $|H|$ elements.*

Proof Let $g \in G$ and define $f\colon H \to gH$ by $f(h) = gh$. This is injective, since for any $h_1, h_2 \in H$ with $f(h_1) = f(h_2)$ we have $gh_1 = gh_2$, and the cancellation law[3] then implies that $h_1 = h_2$. And f is surjective, since for any $k \in gH$ there must be some $h \in H$ such that $k = gh$, and then $f(h) = k$. Therefore f is a bijection, and $|H| = |gH|$. □

[3] Proposition 1.14, page 5.

We now have all the ingredients to prove the following important theorem:

Theorem 2.18 (Lagrange's Theorem) *Let G be a finite group and H a subgroup of G. Then the order or H divides the order of G.*

Proof By Corollary 2.16, the left cosets of H partition G. That is, $G = \bigcup_{g \in G} gH$, and either $g_1H = g_2H$ or $g_1H \cap g_2H = \emptyset$ for

all $g_1, g_2 \in G$. And Proposition 2.17 tells us that all the left cosets (including $1H = H$ itself) have the same number of elements. Hence the order of G must be a multiple of the order of H. □

Joseph-Louis Lagrange (1736–1813)

Definition 2.19 The number of distinct left cosets of H in G is called the **index** of H in G, and is denoted $|G : H|$.

Another version of Lagrange's Theorem is as follows:

Theorem 2.20 *Let G be a finite group and H a subgroup of G. Then*

$$|G| = |H| \cdot |G : H|.$$

Proposition 2.21 *Let G be a finite group. Then for any $g \in G$, the order $|g|$ of g divides the order $|G|$ of G.*

Proof Suppose that $|g| = n$. Then recall from Example 2.7 that the powers $\{g^k : k \in \mathbb{Z}\}$ of g form a finite subgroup of G; this is the cyclic subgroup $\langle g \rangle$ generated by g. But $\langle g \rangle = \{g^k : 0 \leqslant i < n\}$ and so $|\langle g \rangle| = |g| = n$. By Lagrange's Theorem, the order of any subgroup of a finite group G must divide the order of G. Hence n must be a factor of $|G|$. □

Lagrange's Theorem provides a necessary condition on the order of a subgroup of a finite group. It says that if H is a subgroup of a finite group G, then $|H|$ must divide $|G|$. But this condition is not sufficient. Even if some integer k divides $|G|$, it doesn't mean that G actually has a subgroup with that many elements. The smallest counterexample concerns the alternating group A_4, which has order 12 but doesn't have a subgroup of order 6, even though $6|12$.[4]

[4] Proposition 6.33, page 60.

The following theorem provides a partial converse, however:[5]

Theorem 2.22 (Cauchy's Theorem) *Let G be a finite group of order $|G| = n$. If p is a prime factor of n, then G contains a nontrivial element (and hence a cyclic subgroup) of order p.*

[5] This proof is a slightly expanded version of the elegant ten-line proof in the following article:

James H. McKay, *Another Proof of Cauchy's Group Theorem*, The American Mathematical Monthly 66.2 (1959) 119.

Augustin-Louis Cauchy (1789–1857)

Proof Let

$$S = \{(g_1, \ldots, g_p) : g_1, \ldots, g_p \in G \text{ and } g_1 \ldots g_p = 1\}.$$

This set has n^{p-1} members, since we have n choices for each of the elements $g_1, \ldots, g_{p-1}$, and then g_p has to be $(g_1 \ldots g_{p-1})^{-1}$.

We define an equivalence relation on S as follows: consider two ordered p–tuples in S to be equivalent if one is a cyclic permutation of the other.

If all of the elements in a given p–tuple are the same (that is, if it is of the form $(g, \ldots, g)$ for some $g \in G$) then it is the only element in its equivalence class. And if a given p–tuple has at least two distinct elements then that equivalence class contains exactly p members.

So each of the equivalence classes determined by this relation contain either a single element, or p elements.

Let r denote the number of elements g such that $g^p = 1$. Then r is the number of equivalence classes with exactly one member. Let s be the number of equivalence classes with exactly p members. Then

$r + sp = n^{p-1} = |S|$. We know that $p|n$, so $p|n^{p-1}$, and obviously $p|sp$. So it must be the case that $p|r$ as well.

Finally, we know that $r > 0$ since at the very least $(1, \ldots, 1) \in S$, as $1^p = 1$. And since $p|r$ there must be at least one other single-element equivalence class comprising a p–tuple $(g, \ldots, g)$ for $g \neq 1$. Then $g^p = 1$ and so $|g| = p$. The cyclic subgroup $\langle g \rangle$ generated by this element has exactly p distinct elements. □

We'll finish with a nice application of Lagrange's Theorem:

Proposition 2.23 *For any $n, r \in \mathbb{Z}$ such that $0 \leqslant r \leqslant n$, the binomial coefficient $\binom{n}{r} = \frac{n!}{(n-r)!r!}$ is an integer.*

Proof Recall from Proposition 1.34 that $|S_n| = n!$. Now let H be the subgroup of S_n consisting of permutations in which the first r elements are permuted amongst themselves, and the remaining $(n-r)$ elements are permuted amongst themselves. This is a subgroup of S_n: it is nonempty, closed under composition, and contains all required inverses. And $|H| = r!(n-r)!$ because there are $r!$ possible permutations for the first r elements, and $(n-r)!$ possible permutations of the remaining $(n-r)$ elements. By Lagrange's Theorem, $|H|$ must divide $|S_n|$, and hence $\binom{n}{r}$ must be an integer. □

All parts should go together without forcing. You must remember that the parts you are reassembling were disassembled by you. Therefore, if you can't get them together again, there must be a reason. By all means, do not use a hammer.

— IBM maintenance manual (c.1925)

3 Normal Subgroups and Quotients

If a group G is abelian, then clearly for any subgroup $H \leqslant G$ and element $g \in G$, the left coset gH will be equal to the right coset Hg, because all the elements of G commute with each other, so

$$gH = \{gh : h \in H\} = \{hg : h \in H\} = Hg.$$

This will sometimes happen for nonabelian groups, as we saw in Example 2.11. But sometimes it won't, as we saw in Example 2.14.

3.1 Normal subgroups

Not every subgroup has this property, so we'll give a special name to those that do:

Definition 3.1 Let G be a group. A subgroup $H \leqslant G$ is said to be **normal** in G, or a **normal subgroup** of G, if the left coset gH is equal to the right coset Hg for all $g \in G$. We will denote this by $H \trianglelefteq G$. (We may also use the notation $H \lhd G$ if H is a proper subgroup of G.)

Some examples:

Example 3.2 For any group G, the trivial subgroup $\{1\}$ and the group G itself are both normal subgroups of G.

Example 3.3 If G is an abelian group, then all of its subgroups are normal.

Example 3.4 As we saw in Example 2.11, the rotation subgroup $R_3 = \{e, r, r^2\}$ is a normal subgroup of the dihedral group D_3. More generally, the rotation subgroup

$$R_n = \{e, r, r^2, \ldots, r^{n-1}\} < D_n$$

is normal in D_n.

The rotation subgroup R_n is an index–2 subgroup of the dihedral group D_n; that is, $|D_n : R_n| = 2$. It turns out that this is relevant to its normality:

Proposition 3.5 *Let G be a group. If H is a subgroup of G with index $|G : H| = 2$ then H is a normal subgroup of G.*

Proof Recall that we defined the index $|G : H|$ to be the number of distinct left cosets of H in G.[1] If $|G : H| = 2$, then one of these cosets must be H itself, and since the cosets partition G,[2] the other

[1] Definition 2.19, page 22.

[2] Corollary 2.16, page 21.

must be its complement $G \setminus H$. The same applies to right cosets. Hence, for $g \in G$, if $g \in H$ then $gH = H = Hg$. And if $g \notin H$, then $gH = G \setminus H = Hg$. In either case, $gH = Hg$, so $H \lhd G$. □

Example 3.6 As remarked at the end of Chapter 1, $|A_n| = \frac{1}{2}|S_n|$, and hence $|S_n : A_n| = 2$. By the above proposition, then, $A_n \lhd S_n$. More generally, $\mathrm{Alt}(X) \lhd \mathrm{Sym}(X)$.

The following proposition gives an alternative (and sometimes more useful) characterisation of normal subgroups.

Proposition 3.7 *Let H be a subgroup of a group G. Then H is normal in G if and only if $ghg^{-1} \in H$ for all $g \in G$ and $h \in H$.*

Proof Suppose that $H \trianglelefteqslant G$, and suppose that $g \in G$ and $h \in H$. Then $gh \in gH$, and since H is normal, $gH = Hg$, so $gh \in Hg$ as well. This means that there exists some $h_1 \in H$ such that $gh = h_1 g$. Hence $ghg^{-1} = h_1 \in H$.

Conversely, suppose that $ghg^{-1} \in H$ for all $g \in G$ and $h \in H$.

Then for $gh \in gH$ we have $ghg^{-1} \in H$ and so $ghg^{-1} = h_2$ for some $h_2 \in H$. Then $gh = h_2 g \in Hg$, so $gH \subseteq Hg$.

Now suppose that $hg \in Hg$. Since $g^{-1} \in G$ we also have $g^{-1}hg \in H$, and so if we set $g^{-1}hg = h_3$, it follows that $hg = gh_3 \in gH$. So $Hg \subseteq gH$. Thus $gH = Hg$ and so $H \trianglelefteqslant G$. □

Now is a good time to introduce the following terminology:

Definition 3.8 Let G be a group. An element $g \in G$ is **conjugate** to an element $h \in G$ if there exists some $k \in G$ such that $kgk^{-1} = h$. We say also that h is the result of **conjugating** g with k.

So Proposition 3.7 says that $H \leqslant G$ is a normal subgroup of G exactly when it is closed under conjugation by all elements of G.

3.2 *Quotient groups*

As we saw earlier, given any subgroup H of a group G, we can neatly split G into a collection of subsets (called cosets) all of which are the same size. There are two ways of doing this, depending on whether we look at the left or right cosets, but if H is a normal subgroup of G these are the same. It turns out that (as long as H is a normal subgroup) we can define a group structure on these left (or right) cosets.

First, we need some notation:

Definition 3.9 Let A and B be subsets of a group G. We define their product

$$AB = \{ab : a \in A, b \in B\}$$

We'll use this notation a few times in the rest of these notes, but right now we're going to look at what happens if we take products of cosets.

Proposition 3.10 *Let N be a normal subgroup of a group G, and suppose that $g, h \in G$. Then $(gN)(hN) = (gh)N$.*

Proof Let $gn_1 \in gN$ and $hn_2 \in hN$. Then since N is a normal subgroup of G, we have $gN = Ng$, and so $n_1 h \in Nh$ is equal to some element $hn_3 \in hN$. Hence

$$(gn_1)(hn_2) = g(n_1 h)n_2 = g(hn_3)n_2 = (gh)(n_3 n_2) \in (gh)N$$

and so $(gN)(hN) \subseteq (gh)N$.

Now suppose that $ghn \in (gh)N$. Then $ghn = (g1)(hn) \in (gN)(hN)$ and hence $(gh)N \subseteq (gN)(hN)$. Therefore $(gN)(hN) = (gh)N$. □

This gives us a well-defined binary operation on the set of (left) cosets of N in G. In fact, we can turn this into a group:

Proposition 3.11 *If N is a normal subgroup of a group G, the set*

$$G/N = \{gN : g \in G\}$$

of all left cosets of N in G forms a group under the multiplication operation

$$(gN)(hN) = (gh)N$$

for all $g, h \in G$.

Proof We've just seen that $(gN)(hN) = (gh)N$, so the closure requirement is satisfied, and this is a binary operation on G/N. Associativity follows almost immediately from the associativity in G:

$$\begin{aligned}((gN)(hN))(kN) &= ((gh)N)(kN) = ((gh)k)N \\ &= (g(hk))N = (gN)((hk)N) = (gN)((hN)(kN))\end{aligned}$$

Since

$$\begin{aligned}N(gN) = (1N)(gN) = (1g)N = gN \\ = (g1)N = (gN)(1N) = (gN)N,\end{aligned}$$

the subgroup N itself serves as an identity element.

And since

$$(g^{-1}N)(gN) = (g^{-1}g)N = N = (gg^{-1})N = (gN)(g^{-1}N),$$

the coset $g^{-1}N$ is the inverse of the coset gN for all $g \in G$.

Therefore G/N is a group. □

Definition 3.12 Let G be a group, and N a normal subgroup of G. The group G/N constructed in Proposition 3.11 is called the **quotient group** or **factor group** of G by N.

Why do we need N to be a normal subgroup of G? Well, the reason is that if N isn't normal, the left and right cosets partition G differently and the set G/N doesn't have a group structure. Table 3.1 shows the group table for D_3/R_3, and we can see that we get a nice block structure (that looks very much like the group table for $\mathbb{Z}_2$). However, Table 3.2 shows the group table if we try to factor by the non-normal subgroup $H = \{e, m_1\}$, and we can see that this doesn't yield such a neat structure, and certainly not one that looks like the group table for $\mathbb{Z}_3$.

Time for some examples.

$*$	e	r	r^2	m_1	m_2	m_3
e	e	r	r^2	m_1	m_2	m_3
r	r	r^2	e	m_3	m_1	m_2
r^2	r^2	e	r	m_2	m_3	m_1
m_1	m_1	m_2	m_3	e	r	r^2
m_2	m_2	m_3	m_1	r^2	e	r
m_3	m_3	m_1	m_2	r	r^2	e

Table 3.1: Cosets of R_3 in D_3

$*$	e	m_1	r	m_3	r^2	m_2
e	e	m_1	r	m_3	r^2	m_2
m_1	m_1	e	m_2	r^2	m_3	r
r	r	m_3	r^2	m_2	e	m_1
m_3	m_3	r	m_1	e	m_2	r^2
r^2	r^2	m_2	e	m_1	r	m_3
m_2	m_2	r^2	m_3	r	m_1	e

Table 3.2: Left cosets of $\{e, m_1\}$ in D_3

Example 3.13 If we factor a group G by its trivial subgroup $\{1\}$ then we get a group isomorphic to G itself. That is, $G/\{1\} \cong G$.

And if we factor G by itself, we get a single coset consisting of the entirety of G, and the resulting quotient group is trivial. That is, $G/G \cong \{1\}$.

Example 3.14 Let G be the infinite cyclic group $\mathbb{Z}$ and let $N = n\mathbb{Z} = \{nk : k \in \mathbb{Z}\}$ be the subgroup generated by a fixed positive integer n. By Proposition 2.13 (using additive notation) we see that the cosets $i+n\mathbb{Z}$ and $j+n\mathbb{Z}$ are equal if and only if $i \cong j \pmod n$. So there are n distinct cosets:

$$n\mathbb{Z} = 0+n\mathbb{Z}, \quad 1+n\mathbb{Z}, \quad \ldots, \quad (n-1)+n\mathbb{Z}$$

The quotient group $\mathbb{Z}/n\mathbb{Z}$ is isomorphic to $\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$ via the isomorphism $f : i+n\mathbb{Z} \mapsto i$ for $0 \leqslant i < n$. We can visualise this as wrapping the infinite set $\mathbb{Z}$ of integers into a circle of n numbers: Figure 3.1 shows this for the case $\mathbb{Z}/12\mathbb{Z} \cong \mathbb{Z}_{12}$.

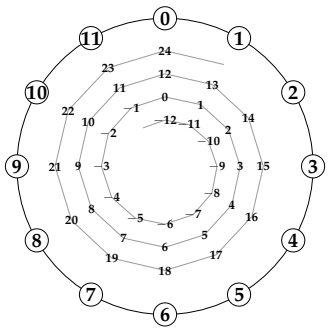


Figure 3.1: Wrapping $\mathbb{Z}$ into $\mathbb{Z}_{12}$.

Example 3.15 Suppose $G = \langle g \rangle \cong \mathbb{Z}_n$ is a finite cyclic group, with $|g| = n = lm$ composite. Let $N = \langle g^m \rangle$ be the cyclic subgroup generated by g^m. This subgroup has order $l = n/m$ and consists of the elements $\{g^{mk} : 0 \leqslant k < l\}$. Since all cosets have the form $g^k N$ for some $k \in \mathbb{Z}$, it follows that G/N is cyclic and is generated by gN. Its order is $|G/N| = |G|/|N| = n/l = m$. To see this, note that $g^i N = g^j N$ if and only if $g^{i-j} \in N$, if and only if $m|(i-j)$, and so the distinct cosets of N in G are $g^k N$ for $0 \leqslant k < m$. In particular, $(gN)^m = g^m N = N$ is the identity element of G/N, and the order of gN is m.

3.3 *Direct products*

Given two groups G and H we can combine them in the following way to form a new group.[3]

[3] There is a technical difference between the direct sum and the direct product, but this only becomes relevant when taking the direct sum or product of infinitely many groups, which we won't be doing. For finite collections of groups, the two concepts are equivalent, so we will use them interchangeably, largely dependent on whether the groups in question are using multiplicative or additive notation.

Definition 3.16 Given two groups G and H, their **direct product** is formed from the cartesian product of their underlying sets

$$G \times H = \{(g, h) : g \in G, h \in H\}$$

with the group operation

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

If G and H are additive groups, we might instead refer to the **direct sum**

$$G \oplus H = \{(g, h) : g \in G, h \in H\}$$

with group operation

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

As an exercise, prove that this is a group.

The two component groups of a direct product form normal sub-

groups:

Proposition 3.17 *Let G and H be groups. Their direct product $G\times H$ has a normal subgroup isomorphic to G and a normal subgroup isomorphic to H. Furthermore, $G\times H/G \cong H$ and $G\times H/H \cong G$.*

Proof First, consider the subset $S = \{(g,1) : g \in G\} \subseteq G\times H$. This is nonempty, since $1_{G\times H} = (1,1) \in S$. It is closed under multiplication, since $(g_1,1)(g_2,1) = (g_1g_2,1) \in S$. And it contains all necessary inverses, since if $(g,1) \in S$ then $(g^{-1},1) \in S$ as well, and $(g,1)(g^{-1},1) = (gg^{-1},1) = (1,1)$.

Hence $S \leqslant G\times H$, and $S \cong G$ by the map $f\colon S \to G$ given by $f(g,1) = g$.

For any $(g,1) \in S$ and $(k,h) \in G\times H$ we have

$$\begin{aligned}(k,h)(g,1)(k,h)^{-1} &= (k,h)(g,1)(k^{-1},h^{-1})\\ &= (kgk^{-1},h1h^{-1}) = (kgk^{-1},1) \in S,\end{aligned}$$

so $S \trianglelefteq G\times H$.

Consider the quotient $G\times H/S$. This comprises cosets of the form

$$(k,h)S = \{(k,h)(g,1) : g \in G\} = \{(kg,h) : g \in G\}.$$

We can define a map $f\colon G\times H/S \to H$ by $(k,h)S \mapsto h$. This is injective, since if $f((k_1,h_1)S) = f((k_2,h_2)S)$ then $h_1 = h_2$, and so $(k_1,h_1)S = (k_2,h_2)S$. And it is surjective since for any $h \in H$ we have $f((1,h)S) = h$. It satisfies the structural condition since for any $(k_1,h_1)S$ and $(k_2,h_2)S$ we have

$$\begin{aligned}f((k_1,h_1)S(k_2,h_2)S) &= f((k_1k_2,h_1h_2)S)\\ &= h_1h_2 = f((k_1,h_1)S)f((k_2,h_2)S).\end{aligned}$$

Thus f is an isomorphism $G\times H/S \cong H$. We might write this as $G\times H/H \cong H$ without ambiguity.

By a very similar argument, $G\times H$ has a normal subgroup isomorphic to H, and the corresponding quotient is isomorphic to G. □

The following proposition gives us sufficient conditions for a group to be isomorphic to the direct product of two of its subgroups.

Proposition 3.18 *Suppose $H,K \leqslant G$ are two subgroups of a group G, such that $H\cap K = \{1\}$, every element of H commutes with every element of K (that is, $hk = kh$ for all $h \in H$ and $k \in K$) and $HK = G$. Then $G \cong H\times K$.*

Proof Let $f\colon H\times K \to G$ with $f(h,k) = hk$ for all $h \in H$ and $k \in K$. Then if $h_1,h_2 \in H$ and $k_1,k_2 \in K$, we have

$$\begin{aligned}f((h_1,k_1)(h_2,k_2)) &= f(h_1h_2,k_1k_2) = h_1h_2k_1k_2\\ &= h_1k_1h_2k_2 = f(h_1,k_1)f(h_2,k_2).\end{aligned}$$

So f satisfies the required structure condition. It is surjective, since $G = HK$, and so any element of G can be expressed as a product $hk = f(h,k)$, which is the image of an ordered pair in $H\times K$.

All we need to show now is that f is injective. Suppose that $f(h_1,k_1) = f(h_2,k_2)$. Then $h_1k_1 = h_2k_2$, and so $h_2^{-1}h_1 = k_2k_1^{-1}$.

But the terms on the left hand side of this expression belong to H, while the terms on the right hand side belong to K. So both $h_2^{-1}h_1$ and $k_2k_1^{-1}$ belong to the intersection $H \cap K = \{1\}$, and so $h_2^{-1}h_1 = k_2k_1^{-1} = 1$. Then $h_1 = h_2$ and $k_1 = k_2$, so f is injective, and hence the required isomorphism. Thus $G \cong H \times K$. □

This is sometimes called the **internal direct product** of H and K. We can use this result to provide an alternative viewpoint on the Klein group V_4:

Proposition 3.19 *The Klein group V_4 is isomorphic to the direct product $\mathbb{Z}_2 \times \mathbb{Z}_2$.*

Proof Let $G = V_4 = \{e, a, b, c\}$ be the Klein group. Let $H = \{e, a\}$ and $K = \{e, b\}$ be two subgroups, both clearly isomorphic to $\mathbb{Z}_2$. Then $H \cap K = \{e\}$, and since G is abelian, every element of H commutes with every element of K. Furthermore, $HK = \{e, a, b, c\} = G$, and thus $V_4 \cong \mathbb{Z}_2 \times \mathbb{Z}_2$ as claimed. □

We will end this chapter with an important fact about direct products of finite cyclic groups.

Proposition 3.20 *$\mathbb{Z}_m \times \mathbb{Z}_n \cong \mathbb{Z}_{mn}$ if and only if $\gcd(m, n) = 1$; that is, if m and n are coprime.*

Proof The key is to examine the subgroup $\langle (1, 1) \rangle$ of $\mathbb{Z}_m \times \mathbb{Z}_n$ generated by the element $(1, 1,)$. Since both m and n are finite, the first coordinate will cycle back round to 0 after m additions, while the second will do so after n additions.

Suppose that m and n are coprime. Then after m additions, the first coordinate will be 0 but the second won't. Similarly, after n additions the second coordinate will be 0 but the first won't. To get back to $(0, 0)$ we need to add $(1, 1)$ to itself a number of times that contains both m and n as factors. The smallest such number is the least common multiple $\operatorname{lcm}(m, n)$. Since m and n are coprime, $\operatorname{lcm}(m, n) = mn$ and so this subgroup $\langle (1, 1) \rangle$ has mn distinct elements. But $|\mathbb{Z}_m \times \mathbb{Z}_n| = mn$ too, so $(1, 1)$ generates the entirety of $\mathbb{Z}_m \times \mathbb{Z}_n$. So $\mathbb{Z}_m \times \mathbb{Z}_n$ is cyclic, and hence isomorphic to $\mathbb{Z}_{mn}$ by Proposition 1.28.

To prove the converse, suppose that $\gcd(m, n) = d > 1$. Then both m and n are divisible by d, and so $\operatorname{lcm}(m, n) = \frac{mn}{d}$. This means that the subset $\langle (1, 1) \rangle$ has $\frac{mn}{d}$ elements and hence can't be isomorphic to $\mathbb{Z}_{mn}$, but instead is isomorphic to $\mathbb{Z}_{mn/d}$.

In fact, no element $(a, b) \in \mathbb{Z}_m \times \mathbb{Z}_n$ can generate the entire group, since

$$(a, b) + \cdots + (a, b) = \left(\tfrac{mn}{d}a, \tfrac{mn}{d}b\right) = \left(\tfrac{n}{d}ma, \tfrac{m}{d}nb\right) = (0, 0).$$

So $\mathbb{Z}_m \times \mathbb{Z}_n$ isn't cyclic and can't be isomorphic to $\mathbb{Z}_{mn}$. □

4 Homomorphisms

In the judgment of the most competent living mathematicians, Fräulein Noether was the most significant creative mathematical genius thus far produced since the higher education of women began. In the realm of algebra, in which the most gifted mathematicians have been busy for centuries, she discovered methods which have proved of enormous importance in the development of the present-day younger generation of mathematicians.

— Albert Einstein (1879–1955), letter to the New York Times, 5 May 1935

EARLY IN THE FIRST CHAPTER, we introduced the concept of an **isomorphism**, a structure-preserving bijection between groups. Now we will generalise this notion, dropping the bijectivity requirement.

4.1 Structure-preserving maps

Structures are the weapons of the mathematician.

— attributed to Nicolas Bourbaki (1934–)

We start with the following definitions:

Definition 4.1 Let G and H be groups. A function $f\colon G \to H$ is a **homomorphism** if, for all $g_1, g_2 \in G$ we have

$$f(g_1 g_2) = f(g_1) f(g_2).$$

In addition, we have the following types of homomorphism:

- An injective homomorphism is called a **monomorphism**; that is, if $f(g_1) = f(g_2)$ only when $g_1 = g_2$.
- A surjective homomorphism is called an **epimorphism**; that is, if for all $h \in H$ there exists some $g \in G$ with $f(g) = h$.
- A homomorphism $f\colon G \to G$ from a group to itself is called an **endomorphism**.
- An **isomorphism** is a bijective homomorphism.
- A bijective endomorphism (that is, an isomorphism from a group to itself) is called an **automorphism**.

Homomorphisms behave properly with respect to the group operation, and as a consequence they also respect identities and inverses:

Proposition 4.2 *Let G and H be groups, and suppose that $f\colon G \to H$ is a homomorphism. Then $f(1_G) = 1_H$, and $f(g^{-1}) = f(g)^{-1}$ for all $g \in G$.*

Proof First, suppose that $f(1_G) = h \in H$. Then

$$1_H h = h = f(1_G) = f(1_G 1_G) = f(1_G) f(1_G) = hh,$$

and so $h = 1_H$ by the cancellation law.[1]

Secondly, if $g \in G$ and $f(g) = h$, then

$$f(g^{-1}) f(g) = f(g^{-1} g) = f(1_G) = 1_H = h^{-1} h = f(g)^{-1} f(g),$$

so $f(g^{-1}) = f(g)^{-1}$, again by the cancellation law. □

Now for some examples of homomorphisms.

[1] Proposition 1.14, page 5.

Example 4.3 For any group G there exists an **identity homomorphism** $\mathrm{id}\colon G \to G$ given by $\mathrm{id}(g) = g$ for all $g \in G$.

Example 4.4 For any groups G and H there exists a **trivial** or **zero homomorphism** $z\colon G \to H$ given by $z(g) = 1_H$ for all $g \in G$.

Example 4.5 Let $k \in G$ be some fixed element of a group G. Then for any $g, h \in G$ we have

$$(kgk^{-1})(khk^{-1}) = kg(k^{-1}k)hk^{-1} = k(gh)k^{-1},$$

so the map $f_k\colon G \to G$ given by $f_k(g) = kgk^{-1}$ is a homomorphism from G to itself.

In fact, it is an isomorphism. If $f_k(g) = f_k(h)$, then $kgk^{-1} = khk^{-1}$ and the cancellation laws imply that $g = h$, so it is injective. And it is surjective, since for any $h \in G$ we have $f_k(k^{-1}hk) = kk^{-1}hkk^{-1} = h$.

If G is abelian, then f_k is the identity homomorphism, since $f_k(g) = kgk^{-1} = kk^{-1}g = g$ for all $g \in G$. So these **conjugation homomorphisms** (or **conjugation automorphisms**) are only interesting when G is nonabelian.

Example 4.6 Let $\mathbb{R}^*$ be the multiplicative group of nonzero real numbers, and consider the map $f\colon GL_n(\mathbb{R}) \to \mathbb{R}^*$ defined by $f(A) = \det(A)$ for all nonsingular $n \times n$ real matrices $A \in GL_n(\mathbb{R})$. Recall from linear algebra that the determinant has a multiplicative property: $\det(AB) = \det(A)\det(B)$ for any $A, B \in GL_n(\mathbb{R})$. It is therefore a homomorphism.

Inclusion of a subgroup into its parent group is a homomorphism:

Example 4.7 Let H be a subgroup of a group G. Then the **inclusion homomorphism** $i\colon H \to G$ is defined by $i(h) = h \in G$. Sometimes we denote this with a hooked arrow: $i\colon H \hookrightarrow G$.

More generally, suppose that $G = G_1 \times \cdots \times G_n$ is a direct product of n groups. Then the canonical inclusion homomorphism $i_k\colon G_k \hookrightarrow G$ is defined by $i_k(g) = (1_{G_1}, \ldots, 1_{G_{k-1}}, g, 1_{G_{k+1}}, \ldots, 1_{G_n})$ for $1 \leqslant k \leqslant n$.

These homomorphisms are injective.

And the projection of a direct product onto one of its factors is also a homomorphism:

Example 4.8 Let $G = G_1 \times \cdots \times G_n$ be a direct product of n groups. The canonical **projection homomorphism** $p_k\colon G \to G_k$ given by $p_k(g_1, \ldots, g_k, \ldots, g_n) = g_k$ is a homomorphism for all $g_1 \in G_1, \ldots, g_n \in G_n$, and $1 \leqslant k \leqslant n$.

These homomorphisms are surjective.

Composites of homomorphisms are homomorphisms:

Proposition 4.9 *Let G, H and K be groups, and suppose that $\phi\colon G \to H$ and $\psi\colon H \to K$ are homomorphisms. Then the composite map $\psi\circ\phi\colon G \to K$ is also a homomorphism.*

Proof Suppose that $g_1, g_2 \in G$. Then

$$\begin{aligned}(\psi\circ\phi)(g_1g_2) &= \psi(\phi(g_1g_2))\\ &= \psi(\phi(g_1)\phi(g_2))\\ &= \psi(\phi(g_1))\psi(\phi(g_2))\\ &= (\psi\circ\phi)(g_1)(\psi\circ\phi)(g_2)\end{aligned}$$

and so $\psi\circ\phi$ is also a homomorphism. □

4.2 *Kernels and images*

In linear algebra we meet the concepts of the **kernel** or **null space** of a linear map, and also the **image** or **column space**. Now we formulate and study the analogous concepts for groups.

Definition 4.10 Given a homomorphism $f\colon G \to H$, we define the **image** of f to be the set

$$\mathrm{im}(f) = \{f(g) : g \in G\} \subseteq H.$$

The image of a linear map is a subspace of the codomain, and we have a similar situation for the image of a group homomorphism:

Proposition 4.11 *Let $f\colon G \to H$ be a homomorphism. Then* $\mathrm{im}(f)$ *is a subgroup of H.*

Proof First, recall from Proposition 4.2 that $f(1_G) = 1_H$, so $1_H \in \mathrm{im}(f)$ and hence $\mathrm{im}(f)$ is nonempty.

Now suppose that $h_1, h_2 \in \mathrm{im}(f) \subseteq H$. Then there exist $g_1, g_2 \in G$ such that $h_1 = f(g_1)$ and $h_2 = f(g_2)$. And

$$h_1h_2 = f(g_1)f(g_2) = f(g_1g_2) \in \mathrm{im}(f).$$

By Proposition 4.2 again, we have $f(g)^{-1} = f(g^{-1}) \in \mathrm{im}(f)$. Hence by Proposition 2.3, $\mathrm{im}(f)$ is a subgroup of H. □

Definition 4.12 Given a homomorphism $f\colon G \to H$, we define the **kernel** of f to be the set

$$\ker(f) = \{g \in G : f(g) = 1_H\} \subseteq G.$$

There is an important link between kernels and injective homomorphisms: a homomorphism is injective exactly when it has trivial kernel.

Proposition 4.13 *Let $f\colon G \to H$ be a homomorphism. Then f is injective if and only if* $\ker(f) = \{1\}$.

Proof Since $1_G \in \ker(f)$, if f is injective then it must be the case that $\ker(f) = \{1_G\}$. If $\ker(f)$ contains anything else, then that would mean there is more than one distinct element that maps to 1_H, contradicting the injectivity of f.

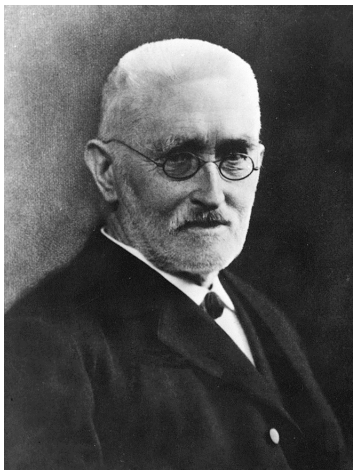
Conversely, suppose that $\ker(f) = \{1_G\}$, and let $g_1, g_2 \in G$ with $f(g_1) = f(g_2)$. Then $1_H = f(g_1)^{-1}f(g_2) = f(g_1^{-1}g_2)$. Hence $g_1^{-1}g_2 \in \ker(f)$ and so $g_1^{-1}g_2 = 1_G$, which implies that $g_1 = g_2$. Hence f is injective. □

The image of a homomorphism is a subgroup of the codomain. What about the kernel? The following result answers this question: kernels are normal subgroups, and all normal subgroups are kernels.

Proposition 4.14

(i) *Let $f: G \to H$ be a homomorphism. Then* $\ker(f)$ *is a normal subgroup of G.*

(ii) *Let N be a normal subgroup of a group G. Then the map $p: G \to G/N$ defined by $p(g) = gN$ is a surjective homomorphism with kernel N.*



Richard Dedekind (1831–1916)

Proof

(i) Clearly $\ker(f)$ is a non-empty subset of G, as $1_G \in \ker(f)$. Now consider any $g_1, g_2 \in \ker(f)$. Then

$$f(g_1g_2) = f(g_1)f(g_2) = 1_H1_H = 1_H$$

so $g_1g_2 \in \ker(f)$, and

$$f(g_1^{-1}) = f(g_1)^{-1} = 1_H^{-1} = 1_H$$

so $g^{-1} \in \ker(f)$. Hence $\ker(f)$ is a subgroup of G by Proposition 2.3.
If $g \in G$ and $k \in \ker(f)$, then

$$f(gkg^{-1}) = f(g)f(k)f(g^{-1}) = f(g)1_Hf(g)^{-1} = 1_H,$$

so $gkg^{-1} \in \ker(f)$ and therefore $\ker(f) \trianglelefteq G$ by Proposition 3.7.

(ii) For any $a, b \in G$ we have

$$p(ab) = (ab)N = (aN)(bN) = p(a)p(b),$$

so $p: G \to G/N$ is a homomorphism. Now consider any $gN \in G/N$. Then $gN = p(g)$, and so p is surjective. Finally, $p(g) = 1_{G/N}$ implies that $gN = 1_GN = N$, so $g \in N$ and thus $\ker(p) = N$.

This completes the proof. □

Emmy Noether (1882–1935)

4.3 *The Isomorphism Theorems*

In the last section of this chapter we will state and prove three theorems relating quotients, homomorphisms and subgroups, due to the German mathematicians Richard Dedekind (1831–1916) and Emmy Noether (1882–1935).

The first of these is the most important. It is sometimes called the *Fundamental Theorem of Homomorphisms*.

Theorem 4.15 (First Isomorphism Theorem) *Let $f: G \to H$ be a homomorphism with $K = \ker(f)$. Then $G/\ker(f) \cong \mathrm{im}(f)$. More precisely, there is an isomorphism $\phi: G/K \to \mathrm{im}(f)$ defined by $\phi(gK) = f(g)$ for all $g \in G$.*

Proof We need to show three things: that the function ϕ is well-defined, that it is bijective, and that it is a homomorphism.

The first of these is important: we need to show that $\phi(gK) = f(g)$ really does define a function from G/K to $\mathrm{im}(f)$. This isn't immediately obvious, because we can have $g_1K = g_2K$ when $g_1 \neq g_2$, and when that happens we need to make sure that $f(g_1) = f(g_2)$. This is what we mean by checking that ϕ is well-defined.

Suppose, then, that $g_1, g_2 \in G$ and that $g_1K = g_2K$. Then by Proposition 2.13 it follows that $g_1 = g_2k$ for some $k \in K$, and hence $f(g_1) = f(g_2k) = f(g_2)f(k) = f(g_2)$. So ϕ is indeed well-defined.

It is clear that ϕ is surjective: any element of $\mathrm{im}(f)$ is of the form $f(g)$ for some $g \in G$, and by the definition of ϕ we have $f(g) = \phi(gK)$. Hence any element of $\mathrm{im}(f)$ is of the form $\phi(gK)$ for some coset $gK \in G/K$.

To show injectivity, let $g \in G$ and suppose that $gK \in \ker(\phi)$. That is, $\phi(gK) = 1_H$. Then $f(g) = 1_H$, and so $g \in K$. Thus $gK = K = 1_{G/H}$, and hence $\ker(\phi) = \{K\}$. Then by Proposition 4.13, ϕ is injective, and hence bijective.

Finally, we need to show that ϕ is a homomorphism. Suppose that $g_1, g_2 \in G$. Then

$$\begin{aligned}\phi((g_1K)(g_2K)) = \phi((g_1g_2)K) &= f(g_1g_2) \\ &= f(g_1)f(g_2) = \phi(g_1K)\phi(g_2K)\end{aligned}$$

and hence ϕ is a homomorphism. This completes the proof. □

Let's illustrate this by looking at an example.

Example 4.16 Let G be the Klein group $V_4 = \{e, a, b, c\}$ and H be the cyclic group $\mathbb{Z}_2$. Define a homomorphism $f\colon V_4 \to \mathbb{Z}_2$ by $f(e) = f(a) = 0$ and $f(b) = f(c) = 1$. Then $\ker(f) = \{e, a\}$ and $\mathrm{im}(f) = \mathbb{Z}_2$. Applying the First Isomorphism Theorem to f we see that

$$V_4/\{e, a\} = V_4/\ker(f) \cong \mathrm{im}(f) = \mathbb{Z}_2.$$

We can also use the First Isomorphism Theorem together with Lagrange's Theorem[2] to disprove the existence of surjective homomorphisms in some circumstances:

[2] Theorem 2.18, page 21.

Example 4.17 There is no surjective homomorphism $f\colon D_9 \to \mathbb{Z}_4$. If such a homomorphism existed then $|\mathrm{im}(f)| = |\mathbb{Z}_4| = 4$. By the First Isomorphism Theorem, it would be the case that $G/\ker(f) \cong \mathrm{im}(f) = \mathbb{Z}_4$, and hence that $|\ker(f)| = |G|/|\mathrm{im}(f)| = 18/4 = \frac{9}{2}$. This is not an integer, so no such homomorphism can exist.

Less useful for our purposes, but still worth looking at, are the Second and Third Isomorphism Theorems. First, we state and prove the following lemma.

Lemma 4.18 *Let G be a group, let H be a subgroup of G, and let N be a normal subgroup of G. Then the following conditions hold:*

(i) *$HN = NH$ is a subgroup of G,*
(ii) *N is a normal subgroup of HN, and*
(iii) *$H \cap N$ is a normal subgroup of H.*

Proof

(i) Suppose that $h \in H$ and $n \in N$, so $hn \in HN$. Then since N is normal in G we have $hn \in hN = Nh \subseteq NH$, so $HN \subseteq NH$. Similarly, $nh \in NH$, but also $nh \in Nh = hN \subseteq HN$, so $NH \subseteq HN$ and thus $HN = NH$.
To see that it is a subgroup, we use Proposition 2.3. Since H and N are both subgroups of G, they are nonempty (each must at the very least contain 1_G) and hence HN is also nonempty. Now suppose that $h_1, h_2 \in H$ and $n_1, n_2 \in N$. Then the products h_1n_1 and h_2n_2 both belong to HN. Since $n_1h_2 \in NH = HN$, there exists some $n_3 \in N$ such that $n_1h_2 = h_2n_3$. Then

$$(h_1n_1)(h_2n_2) = h_1(n_1h_2)n_2 = h_1(h_2n_3)n_2 = (h_1h_2)(n_3n_2)$$

which lies in HN, and so HN is closed. And for any $hn \in HN$ we have $(hn)^{-1} = n^{-1}h^{-1} \in NH = HN$, so HN contains all required inverses. Therefore $HN \leqslant G$.

(ii) It is fairly clear that N is a subgroup of HN: it is a subset of HN and a subgroup of G, so it must therefore be a subgroup of HN as well. Normality also follows almost immediately: N is a normal subgroup of G and is therefore closed under conjugation by elements of G. But $HN \subseteq G$, so N must also be closed under conjugation by any elements of HN. Hence $N \trianglelefteq HN$.

(iii) By Proposition 2.9, we know that $H \cap N$ is a subgroup of G, and since it is clearly a subset of H, it must also be a subgroup of H. We can prove normality by showing closure under conjugation: suppose $h \in H$ and $k \in H \cap N$. Then hkh^{-1} belongs to H, since H is closed under the group operation, and $k \in H \cap N \subseteq H$. But it must also belong to N, since N is normal in G and hence closed under conjugation by any element of G, including any element of H. Thus hkh^{-1} must belong to the intersection $H \cap N$, and hence $H \cap N \trianglelefteq H$.

This completes the proof. □

Theorem 4.19 (Second Isomorphism Theorem) *Let G be a group, let H be a subgroup of G, and let N be a normal subgroup of G. Then*

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

Proof Let $p\colon G \to G/N$ be the surjective quotient homomorphism from Proposition 4.14 (ii), such that $p(g) = gN$ for all $g \in G$. Let $q\colon H \to G/N$ be the restriction of this homomorphism to the subgroup $H \leqslant G$, and we find that $\mathrm{im}(q)$ is the set of cosets hN where $h \in H$. Together, these cosets form the subgroup HN/N of G/N, and hence $\mathrm{im}(q) = HN/N$. Also, $\ker(q) = H \cap \ker(p) = H \cap N$. Applying the First Isomorphism Theorem to q we see that

$$H/(H \cap N) = H/\ker(q) \cong \mathrm{im}(q) = HN/N$$

as claimed □

For the Third Isomorphism Theorem, we need the following short lemma:

Lemma 4.20 *Let G be a group, and let $K \subseteq H \subseteq G$, where H and K are both normal subgroups of G. Then:*

(i) *K is a normal subgroup of H,*
(ii) *H/K is a normal subgroup of G/K, and*

Proof

(i) This is straightforward. Firstly, K is a subgroup of G and a subset of H, so it must be a subgroup of H. And since K is a normal subgroup of G, it is closed under conjugation by any element of G, including all the elements of H, so K must also be a normal subgroup of H.

(ii) The quotient H/K is nonempty, since at the very least it contains $K = 1_{H/K}$. It is closed under the product operation on cosets, since for any $h_1, h_2 \in H$ we have

$$(h_1K)(h_2K) = (h_1h_2)K \in H/K.$$

And for any $h \in H$ we have

$$(hK)^{-1} = (h^{-1})K \in H/K,$$

so by Proposition 2.3 H/K is a subgroup of G/K.
Now suppose that $g \in G$ and $h \in H$. Then $gK \in G/K$ and $hK \in H/K$, and

$$(gK)(hK)(gK)^{-1} = (gK)(hK)(g^{-1}K) = (ghg^{-1})K.$$

Since H is a normal subgroup of G, it follows that $ghg^{-1} \in H$, so $(ghg^{-1})K \in H/K$, hence H/K is closed under conjugation by elements of G/K and is therefore a normal subgroup of G/K.

This completes the proof. □

Theorem 4.21 (Third Isomorphism Theorem) *Let G be a group, and let $K \subseteq H \subseteq G$, where H and K are both normal subgroups of G. Then*

$$(G/K)/(H/K) \cong G/H.$$

Proof Define $f\colon G/K \to G/H$ by $f(gK) = gH$ for all $g \in G$. We need to check that this is well-defined; that is, if $g_1K = g_2K$ then $f(g_1) = f(g_2)$. This is straightforward, because $K \subseteq H$, so if $g_1K = g_2K$ then $g_1 = g_2k$ for some $k \in K$, and so $g_1H = (g_2k)H = g_2(kH)$. As $k \in K \subseteq H$, it follows that $kH = H$, hence $g_1H = g_2H$, and thus $f(g_1) = f(g_2)$.

We observe that $\mathrm{im}(f) = G/H$ and $\ker(f) = H/K$, then apply the First Isomorphism Theorem to f, to get

$$(G/K)/(H/K) = (G/K)/\ker(f) \cong \mathrm{im}(f) = G/H$$

as claimed. □

I tried to make out the names of plants, and collected all sorts of things, shells, seals, franks, coins and minerals. The passion for collecting, which leads a man to be a systematic naturalist, a virtuoso, or a miser, was very strong in me, and was clearly innate, as none of my sisters or brother ever had this taste.

— Charles Darwin (1809–1882), in: Francis Darwin, *The Life and Letters of Charles Darwin* (1887) I 27–28

5 *Classification of Groups*

So far, we have studied a number of properties of groups and seen various examples. It would be useful to have a complete list of small finite groups. In this chapter we will compile such a list, up to isomorphism, of groups with eight or fewer elements. We will then state and study a classification theorem for finitely-generated abelian groups.

5.1 *Generators and relations*

Recall that a group G is a **cyclic group** if it is generated a single element; that is, there exists some element $g \in G$ such that $G = \langle g \rangle = \{g^i : i \in \mathbb{Z}\}$. Furthermore, a **cyclic subgroup** is a subgroup consisting of all powers of a single element. We want to extend this idea now to cover groups generated by more than one element. The details of this are complicated, and mostly beyond the scope of this module, but the basic concepts will be sufficient for our purposes.

Definition 5.1 Let G be a group. A collection of elements $g_1, \ldots, g_r$ of G are said to **generate** G, or be **generators** for G, if every element of G can be written as a product of some or all of the elements $g_1, \ldots, g_r$ and their inverses.

That is, every element of the group G can be written as an expression like $g_2^2 g_3^{-1} g_1 g_4 g_3^{-1} g_1 g_2^{-2}$, which may be of arbitrary length. Such an expression is called a **word** in the generators $g_1, \ldots, g_r$ and their inverses.

Example 5.2 Recall from Definition 1.27 that a group G is **cyclic** if and only if it is generated by a single element.

Example 5.3 As noted in Corollary 1.43, the symmetric group S_n is generated by the transitions in S_n. (However, there are smaller generating sets for S_n.)

Example 5.4 The dihedral group D_n can be generated by the rotation r and the reflection m_1.

One way of defining a group is by giving a list of generators and **relations** between those generators: equations saying that certain words are equivalent to other words. If you write down the order of the group, a list of generators, and enough relations, you can fully describe the group up to isomorphism.

Definition 5.5 Let $X = \{x_1, \ldots\}$ be a (finite, infinite or empty) set of **generators**, and $R = \{w_1, \ldots\}$ be a (finite, infinite or empty) set of **relators**, that is, words in the generators and their formal inverses. Then a group G can be defined by a **presentation**:

$$G = \langle x_1, \ldots : w_1, \ldots \rangle = \langle X : R \rangle.$$

Here, we form the set of all possible words in the generators x_i and their formal inverses x_i^{-1}, and then simplify them by cancelling any adjacent generators and their corresponding inverses, and also replacing any instances of each word w_i with the empty word (that is, the identity).

This process determines a group, with the obvious concatenation and reduction operation. In general, it is not obvious whether this group will be finite or infinite.

Alternatively, we can replace the relators with **relations**: equations in the generators of the form $w_i = 1$ or equivalent.

There are lots of questions raised by this concept, including:

(i) Is there a way of deciding whether a given presentation determines a finite group, and if so what its order is?[1]

(ii) Can we tell whether two different presentations describe the same group?[2]

Full answers to these and other questions can be found in the notes for the module *MA467 Presentations of Groups* (if available), or the book by D L Johnson.[3]

[1] Sort of. The **Todd–Coxeter Algorithm** is a process for doing this, but if the group is infinite, the algorithm won't terminate. And there's no easy way of telling whether the algorithm will *never* terminate (because the group is infinite), or just that it hasn't terminated *yet* (because the group is finite but very large).

[2] Yes, in principle. Two presentations determine isomorphic groups if and only if we can change one presentation into the other via a finite sequence of **Tietze transformations**. But it's not always obvious what the required sequence of transformations is.

[3] D. L. Johnson, *Presentations of Groups*, second edition, London Mathematical Society Student Texts 15, Cambridge University Press (1997).

Proposition 5.6 *Let G be generated by two elements a and b subject to the relations $a^m = 1$, $b^n = 1$ and $ab = ba$. Then $G \cong \mathbb{Z}_m \times \mathbb{Z}_n$.*

Proof Since $a^m = 1$ and $b^n = 1$, we can always replace a^{-1} with a^{m-1}, and b^{-1} with b^{n-1}. We can also replace a^r and b^s with $a^{[r]_m}$ and $b^{[s]_n}$, where $[r]_m$ denotes the remainder of r modulo m, and $[s]_n$ denotes the remainder of s modulo n. The relation $ab = ba$ enables us to move all instances of the generator a in a word to the left, and all instances of the generator b to the right.

Hence any element of G determined by the presentation

$$G = \langle a, b : a^m, b^n, ab = ba \rangle$$

can be written in the form $a^k b^l$ where $0 \leqslant k < m$ and $0 \leqslant l < n$. There are mn distinct words of this form, so $|G| = mn$.

To show that $G \cong \mathbb{Z}_m \times \mathbb{Z}_n$, we define a function $f\colon G \to \mathbb{Z}_m \times \mathbb{Z}_n$ by $f(a^k b^l) = (k, l)$.

This function is injective: $(k_1, l_1) = (k_2, l_2)$ if and only if $a^{k_1} b^{l_1} = a^{k_2} b^{l_2}$. It is surjective, since any element $(k, l) \in \mathbb{Z}_m \times \mathbb{Z}_n$ is the image of an element $a^k b^l \in G$. Thus f is a bijection.

And for any elements $a^{k_1} b^{l_1}$ and $a^{k_2} b^{l_2}$ in G we have

$$\begin{aligned} f(a^{k_1} b^{l_1} a^{k_2} b^{l_2}) = f(a^{k_1+k_2} b^{l_1+l_2}) &= (k_1+k_2, l_1+l_2) \\ &= (k_1, l_1)(k_2, l_2) = f(a^{k_1} b^{l_1}) f(a^{k_2} b^{l_2}) \end{aligned}$$

so f is the required isomorphism. □

We want to prove a similar result for the dihedral groups D_n, but first we need to look more closely at the structure of these groups.

Let P_n denote the regular n–sided polygon with vertices labelled $1, \ldots, n$ where vertex i has coordinates $\left(\cos\left(\frac{2(i-1)\pi}{n}\right), \sin\left(\frac{2(i-1)\pi}{n}\right)\right)$. The element r represents an anticlockwise rotation through an angle $\frac{2\pi}{n}$ about the origin. The element m_i represents a reflection in a line passing through the origin, making an angle $\frac{2(i-1)\pi}{n}$ with the positive horizontal axis. See Figure 5.1 for illustrations of the cases $n = 5$ and 6.

We need to consider the odd and even cases separately:[4]

- When n is odd, the reflection m_{2i-1} passes through vertex i and the midpoint of the opposite edge, between vertices $(i+\frac{n-1}{2})$ and $(i+\frac{n+1}{2})$.
- When n is even, the reflection m_{2i-1} passes through vertex i and the opposite vertex $(i+\frac{n}{2})$, while the reflection m_{2i} passes through the midpoint of the edge between vertices i and $(i+1)$, and the midpoint of the opposite edge, between vertices $(i+\frac{n}{2})$ and $(i+\frac{n}{2}+1)$.

We can understand the action of these operations in terms of the way they permute the vertices of P_n.

The $\frac{2\pi}{n}$ anticlockwise rotation operation r shifts each vertex round by one place, so $1 \mapsto 2$, $2 \mapsto 3$, $\ldots$, $n \mapsto 1$. This can be represented by the permutation $\alpha = (1, 2, \ldots, n) \in S_n$.

The reflection m_1 passes through the vertex 1 and either the opposite vertex $\frac{n}{2}+1$ (if n is even) or the midpoint of the opposite edge (if n is odd). It transposes the vertices $2 \leftrightarrow n$, $3 \leftrightarrow (n-1)$, and so forth. If n is even, then the vertex $(\frac{n}{2}+1)$ is also fixed. We can represent m_1 by the permutation $\beta = (2, n)(3, n-1)\ldots \in S_n$.

We can deduce that $\alpha^n = 1$, and also that $\beta^2 = 1$: performing the $\frac{2\pi}{n}$ rotation n times yields the identity, while performing the m_1 reflection twice also leaves every vertex unchanged.

Furthermore, we can see, either geometrically or by considering the product of permutations, that $\alpha^k\beta$ represents the reflection operation m_k for $1 \leqslant k \leqslant n$. We can therefore see that there is a bijection between the elements of the group $D_n = \{1, r, \ldots, r^n, m_1, \ldots, m_n\}$ and the set

$$G = \{\alpha^k, \alpha^k\beta : 0 \leqslant k \leqslant n\}.$$

Again by considering the geometric operations, or composing the permutations, we can see that $\beta\alpha = \alpha^{-1}\beta = \alpha^{n-1}\beta$. These equations

$$\alpha^n = 1, \qquad \beta^2 = 1, \qquad \beta\alpha = \alpha^{-1}\beta$$

enable us to construct the entire Cayley table for G, which is isomorphic to D_n. We can calculate any of the four types of product

$$(\alpha^k)(\alpha^l) = \alpha^{k+l} \qquad (\alpha^k)(\alpha^l\beta) = \alpha^{k+l}\beta$$
$$(\alpha^k\beta)(\alpha^l) = \alpha^{k-l}\beta \qquad (\alpha^k\beta)(\alpha^l\beta) = \alpha^{k-l}$$

(where addition and subtraction of the exponents k and l is modulo n). This determines a presentation for the dihedral group D_n.

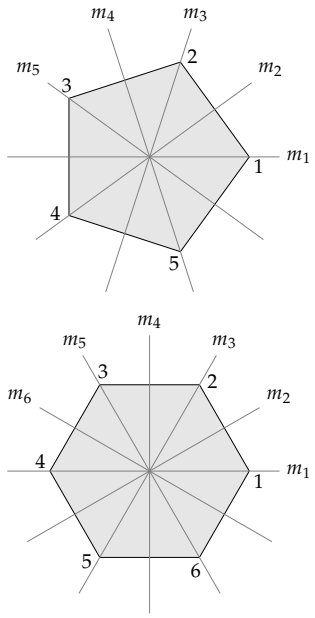


Figure 5.1: The regular pentagon P_5 and the regular hexagon P_6

[4] Here $1 \leqslant i \leqslant n$, and the indices of m_{2i} and m_{2i-1} are calculated modulo n. For example, when $n = 5$ and $i = 4$ we have $2i-1 = 7 \equiv 2 \pmod 5$, so m_2 is the reflection in the line passing through vertex $i = 4$ and the midpoint of the opposite edge.

Proposition 5.7 *Let G be a group generated by two elements a and b satisfying the relations $a^n = 1$, $b^2 = 1$ and $ba = a^{-1}b$. Then $G \cong D_n$.*

5.2 *Small finite groups*

We will now classify (that is, compile a complete list of) all groups of order up to 8.

The simplest case concerns groups of prime order:

Proposition 5.8 *Let G be a group with prime order $|G| = p$. Then G is isomorphic to the finite cyclic group $\mathbb{Z}_p$.*

Proof Let $g \neq 1$ be a nontrivial element of G. Then $|g| > 1$ by Lemma 1.18, and $|g|$ must divide $|G| = p$ by Proposition 2.21, so we have $|g| = p$. Then $1, g, g^2, \ldots, g^{p-1}$ are distinct elements of G, and there are p of them, so they comprise the entirety of G. That is, G consists entirely of powers of g, and so it is cyclic. By Proposition 1.28, G must be isomorphic to $\mathbb{Z}_p$. □

Next we classify the groups of order 4. Up to isomorphism, there are exactly two of them:

Proposition 5.9 *Let G be a group of order $|G| = 4$. Then G is isomorphic to either the cyclic group $\mathbb{Z}_4$ or the Klein group V_4.*

Proof First of all, we note that $\mathbb{Z}_4 \ncong V_4$. This follows from Proposition 1.26: any isomorphism $f : \mathbb{Z}_4 \to V_4$ must preserve the order of each element. That is, $|f(g)| = |g|$. But $\mathbb{Z}_4$ has two elements of order 4 while V_4 has no such elements, and hence no isomorphism can exist.

Now suppose that $G = \{1, a, b, c\}$. By Proposition 2.21, the order of each of these elements must be a factor of 4. The identity element has order 1 by Lemma 1.18, and is the only element that does. So the remaining elements a, b and c must have order either 2 or 4.

[5] Theorem 2.22, page 22.

Cauchy's Theorem[5] ensures the existence of at least one element of order 2. If one of the other elements, say a, has order 4, then G has a cyclic subgroup $\langle a \rangle = \{1, a, a^2, a^3\}$, which accounts for all the elements of G and hence $G \cong \mathbb{Z}_4$. Here, a^2 is the element of order 2 whose existence was guaranteed by Cauchy's Theorem.

[6] Proposition

If a, b and c all have order 2, then the cancellation laws[6]prop:cancellation show that the product of any two must be the third. For example, $ab = b$ implies that $a = 1$, which can't be true since we assumed that $a \neq 1$ (and also that $|a| = 2$). Similarly, $ab = a$ forces $b = 1$, which can't be true for the same reason. Neither can $ab = e$, since that would imply that $a = b^{-1} = b$. The only remaining possibility is that $ab = c$. This yields the Klein group V_4.

So G must be isomorphic to either $\mathbb{Z}_4$ (if it has an element of order 4) or V_4 (if it doesn't have such an element). □

Recall from Proposition 3.19 that $V_4 \cong \mathbb{Z}_2 \times \mathbb{Z}_2$. So every group of order 4 is isomorphic to either $\mathbb{Z}_4$ or $\mathbb{Z}_2 \times \mathbb{Z}_2$. We will explore this further in the next section when we look at the classification of finite and finitely-generated abelian groups.

We have now classified groups of order 1, 2, 3, 4, 5 and 7. The next on our list is order 6. We've seen two examples so far: the cyclic group $\mathbb{Z}_6$ and the dihedral group D_3 (which is isomorphic to the symmetric group S_3). In fact, any group of order 6 is isomorphic to one of these.

Proposition 5.10 *Let G be a group with six elements. Then G is isomorphic to either $\mathbb{Z}_6$ or D_3.*

Proof First, we note that $\mathbb{Z}_6 \ncong D_3$, since the former is abelian and the latter isn't.

Now suppose that G is a group of order 6. By Cauchy's Theorem[7] G has an element g of order 2 and an element h of order 3, since 2 and 3 are the prime factors of $|G| = 6$. These elements are distinct from each other, and neither is equal to the identity element 1. Each of them generates a cyclic subgroup

$$\langle g \rangle = \{1, g\} \cong \mathbb{Z}_2, \qquad \langle h \rangle = \{1, h, h^2\} \cong \mathbb{Z}_3.$$

There are therefore six possible elements:

$$G = \{1, g, h, gh, h^2, gh^2\}$$

Now consider the element hg. This element does not belong to the cyclic subgroup $\langle h \rangle$, and it isn't equal to g. So we have two possibilities: either $hg = gh$ or $hg = gh^2 = gh^{-1}$.

If $hg = gh$ then by Proposition 5.6 we have $G \cong \mathbb{Z}_2 \times \mathbb{Z}_3$. And since 2 and 3 are coprime, $\mathbb{Z}_2 \times \mathbb{Z}_3 \cong \mathbb{Z}_6$ by Proposition 3.20.

If $hg = gh^2 = gh^{-1}$ then by Proposition 5.7 we have $G \cong D_3$.

These are the only two possible cases, and so a group of order 6 must be isomorphic to either $\mathbb{Z}_6$ or D_3. □

Now we have classified groups of order up to 7. Time to look at groups of order 8. There are a few of these that we have met (at least in principle) already: the cyclic group $\mathbb{Z}_8$ and the dihedral group D_4. Other possibilities include $\mathbb{Z}_2 \times \mathbb{Z}_4$ and $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$, which by Proposition 3.20 aren't isomorphic to each other or to $\mathbb{Z}_8$. But there is a fifth group of order 8 that we haven't yet seen, and which we will define now. This is called the **quaternion group**, and we will denote it Q_8.

There several ways of defining this group. One way is as the subgroup of $GL_2(\mathbb{C})$ consisting of the following 8 matrices:[8]

$$E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad I = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} \qquad J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \qquad K = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}$$
$$-E = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \qquad -I = \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix} \qquad -J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \qquad -K = \begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix}$$

The multiplication table for this group is shown in Table 5.1. (Note that here we have used E rather than I for the 2×2 identity matrix.)

Proposition 5.11 *Let G be a group generated by two elements a and b that satisfy the equations $a^4 = 1$, $b^2 = a^2$ and $ba = a^{-1}b$. Then $G \cong Q_8$.*

Proof The first relation $a^4 = 1$ tells us that we need only consider positive powers of a, since we can replace any occurrence of a^{-1}

[7] Theorem 2.22, page 22.

	E	I	J	K	$-E$	$-I$	$-J$	$-K$
E	E	I	J	K	$-E$	$-I$	$-J$	$-K$
I	I	$-E$	K	$-J$	$-I$	E	$-K$	J
J	J	$-K$	$-E$	I	$-J$	K	E	$-I$
K	K	J	$-I$	$-E$	$-K$	$-J$	I	E
$-E$	$-E$	$-I$	$-J$	$-K$	E	I	J	K
$-I$	$-I$	E	$-K$	J	I	$-E$	K	$-J$
$-J$	$-J$	K	E	$-I$	J	$-K$	$-E$	I
$-K$	$-K$	$-J$	I	E	K	J	$-I$	$-E$

Table 5.1: Multiplication table for the quaternion group Q_8

[8] The **Pauli matrices**

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

are particularly relevant in particle physics, where they represent observables relating to the spin of spin–$\frac{1}{2}$ particles such as protons, neutrons and electrons, and in quantum computing, where they represent an important class of single-qubit operations.

These are related to our construction of Q_8 as follows:

$$I = i\sigma_z \qquad J = i\sigma_y \qquad K = i\sigma_x$$

with a^3. The first and second relations together imply that $b^4 = 1$, since

$$b^4 = (b^2)^2 = (a^2)^2 = a^4 = 1.$$

Hence we only have to consider positive powers of b, since $b^{-1} = b^3$. We can also rewrite the third relation $ba = a^{-1}b$ as $ba = a^3b$.

Putting all this together, any finite word composed from the generators a and b and their inverses can be rearranged in the form $a^k b^l$ where k and l are non-negative integers. Using the second relation we can rewrite b^n as either a^n (if n is even) or $a^{n-1}b$ (if n is odd) to get a word of the form a^k or $a^k b$. Since $a^4 = 1$ we know that $0 \leqslant k < 4$, so G has eight elements:

$$G = \{1, a, a^2, a^3, b, ab, a^2b, a^3b\}.$$

The function $f\colon G \to Q_8$ that maps

$$\begin{array}{llll} 1 \mapsto E & a \mapsto I & a^2 \mapsto -E & a^3 \mapsto -I \\ b \mapsto J & ab \mapsto K & a^2b \mapsto -J & a^3b \mapsto -K \end{array}$$

is an isomorphism, and so $G \cong Q_8$. □

Just before we classify the groups of order 8, we need the following simple lemma:

Lemma 5.12 *Let G be a group, and suppose that $g^2 = 1$ for all $g \in G$. Then G is abelian.*

Proof If $g^2 = 1$ then $g^{-1} = g$. Then for any $g, h \in G$ we have

$$gh = (gh)^{-1} = h^{-1}g^{-1} = hg.$$

Hence G is abelian. □

We are now ready to classify the groups of order 8.

Proposition 5.13 *Let G be a group of order 8. Then G is isomorphic to one of the following groups:*

$$\mathbb{Z}_8, \quad \mathbb{Z}_4 \times \mathbb{Z}_2, \quad \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2, \quad D_4, \quad Q_8.$$

Proof Let G be a group with eight elements. By Proposition 2.21, each of these elements must have order 1 (the identity), 2, 4 or 8. We have a number of cases to consider:

Case 1 If G has an element of order 8, then the cyclic subgroup generated by this element is

$$\langle g \rangle = \{1, g, g^2, \ldots, g^7\}.$$

This subgroup has eight distinct elements and must therefore be the entirety of G. In this case, G is a cyclic group of order 8, isomorphic to $\mathbb{Z}_8$ by Proposition 1.28.

Case 2 Suppose instead that G has an element g of order 4, but no elements of order 8. Then

$$\langle g \rangle = \{1, g, g^2, g^3\} \cong \mathbb{Z}_4.$$

Let h be some element of $G \setminus \langle g \rangle$ and consider the coset $\langle g \rangle h = \{h, gh, g^2h, g^3h\}$. Then

$$G = \langle g \rangle \cup \langle g \rangle h = \{1, g, g^2, g^3, h, gh, g^2h, g^3h\}.$$

Consider the element hg. This clearly isn't a power of g, and is thus not in $\langle g \rangle$. Also, $hg \neq h$, since the cancellation law[9]prop:cancellation would then imply that $g = 1$, which we've already decided is not the case. Furthermore, if $hg = g^2h$ then $g = h^{-1}g^2h$, which then implies that

$$g^2 = (h^{-1}g^2h)^2 = h^{-1}g^2hh^{-1}g^2h = h^{-1}g^4h = h^{-1}1h = h^{-1}h = 1.$$

But g has order 4, so $g^2 \neq e$.

So we're left with two possibilities. Either $hg = gh$, or $hg = g^3h$. Also, we have two possibilities for the order of h: either $|h| = 2$ or $|h| = 4$. Note that $h \in \langle g \rangle h$, since $h \notin \langle g \rangle$. Furthermore, $h^2 \neq g$, since this would require $|h| = 8$. And $h^2 \neq g^3$ for the same reason. So if h has order 2 then $h^2 = 1$, and if h has order 4 then the only possibility is that $h^2 = g^2$.

So if G has an element of order 4 then we have four possibilities:

(i) If $hg = gh$ and $|h| = 2$ then G is abelian, and $G \cong \mathbb{Z}_4 \times \mathbb{Z}_2$ by Proposition 5.6. The map $f\colon G \to \mathbb{Z}_4 \times \mathbb{Z}_2$ given by $g \mapsto (1,0)$ and $h \mapsto (0,1)$ is an isomorphism.

(ii) If $hg = g^3h$ and $|h| = 2$ then G is isomorphic to D_4 by Proposition 5.7. The map $f\colon G \to D_4$ with $f(g) = r$ and $f(g) = m_1$ is an isomorphism.

(iii) If $hg = gh$ and $|h| = 4$ then G is abelian. Also, $|gh^{-1}| = 2$ since

$$(gh^{-1})^2 = gh^{-1}gh^{-1} = g^2h^{-2} = g^2g^2 = g^4 = 1.$$

In this case, the function $f\colon G \to \mathbb{Z}_4 \times \mathbb{Z}_2$ given by $g \mapsto (1,0)$ and $gh^{-1} \mapsto (0,1)$ is an isomorphism.

(iv) If $hg = g^3h$ and $|h| = 4$ then $G \cong Q_8$ by Proposition 5.11. The map $f\colon G \to Q_8$ where $f(g) = I$ and $f(h) = J$ is an isomorphism.

Case 3 Now suppose that every element of G (apart from the identity) have order 2. In this case, G is abelian by Lemma 5.12. Choose $g, h, k \in G \setminus \{1\}$ such that $gh \neq k$. The subgroup $\{1, g, h, gh\}$ is isomorphic to the Klein group $V_4 \cong \mathbb{Z}_2 \times \mathbb{Z}_2$. Now let $K = \{1, k\} = \langle k \rangle \cong \mathbb{Z}_2$. Then $HK = G$, the intersection $H \cap K = \{1\}$, and every element of H commutes with each element of K since G is abelian.

Therefore, by Proposition 3.18 we have

$$G \cong HK \cong H \times K \cong \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2.$$

There are no further cases to consider, so the proof is complete. □

Table 5.2 summarises the results in this section.

Classifying groups of higher order starts to require more sophisticated techniques. In particular, the cases $|G| = 2^k$ for some k tend to be particularly tricky: there are 14 groups of order 16 and 51 groups of order 32. Up to isomorphism there are 49 910 529 484 groups of order $\leqslant 2000$, and of these 49 487 365 422 (just over 99.15%) have order 1024.

[9] Proposition

n	Groups of order n
1	$\{1\}$
2	$\mathbb{Z}_2$
3	$\mathbb{Z}_3$
4	$\mathbb{Z}_4$, $V_4 \cong \mathbb{Z}_2 \times \mathbb{Z}_2$
5	$\mathbb{Z}_5$
6	$\mathbb{Z}_6$, $D_3 \cong S_3$
7	$\mathbb{Z}_7$
8	$\mathbb{Z}_8$, $\mathbb{Z}_4 \times \mathbb{Z}_2$, $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$, D_4, Q_8

Table 5.2: Groups of order $\leqslant 8$

5.3 *Finitely-generated abelian groups*

Earlier, we saw that the cyclic group $\mathbb{Z}_4$ and the Klein group $V_4 \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2$ are not isomorphic. But by Proposition 3.20 we know that $\mathbb{Z}_m \oplus \mathbb{Z}_n \cong \mathbb{Z}_{mn}$ if and only if $\gcd(m, n) = 1$. Now we want to generalise this idea to classify all finitely-generated abelian groups. That is, abelian groups with finitely many generators.

We will state the classification theorem, but a full proof would involve too much of a digression, so we'll omit it. However, we will study a general method for finding which abelian group a given presentation determines.

First, we introduce a standard form for an abelian group.

Definition 5.14 Let G be an abelian group such that

$$G \cong \mathbb{Z}_{m_1} \oplus \cdots \oplus \mathbb{Z}_{m_k} \oplus \mathbb{Z}^r$$

where $m_1, \ldots, m_k, r \in \mathbb{Z}$, $r \geqslant 0$ and $m_i | m_{i+1}$ for $1 \leqslant i < k$. This is called an **invariant factor decomposition** of G, with **torsion coefficients** or **invariant factors** $m_1, \ldots, m_k$ and **rank** r.

Here, $\mathbb{Z}^r$ denotes a direct sum of r copies of $\mathbb{Z}$. That is, $\mathbb{Z}^r = \mathbb{Z} \oplus \cdots \oplus \mathbb{Z}$.

Any group of this form (that is, a direct sum of finite and/or infinite cyclic groups) is abelian. Less obviously, any abelian group has a unique decomposition of this form.

Theorem 5.15 *Let G be a finitely-generated abelian group. Then G has a unique invariant factor decomposition*

$$G \cong \mathbb{Z}_{m_1} \oplus \cdots \oplus \mathbb{Z}_{m_k} \oplus \mathbb{Z}^r$$

in the sense that if H is another finitely generated abelian group with invariant factor decomposition

$$H \cong \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_l} \oplus \mathbb{Z}^s,$$

then $G \cong H$ if and only if $k = l$, $r = s$ and $m_i = n_i$ for $1 \leqslant i < k$.

Finite abelian groups have a similar form except without any infinite summands (copies of $\mathbb{Z}$).

Corollary 5.16 *Let G be a finite abelian group. Then G has a unique invariant factor decomposition*

$$G \cong \mathbb{Z}_{m_1} \oplus \cdots \oplus \mathbb{Z}_{m_k}.$$

An element of a group is said to be a **torsion element** if it has finite order. And a group is called **torsion-free** if it has no elements of finite order. Torsion-free finitely-generated abelian groups are formed from finitely many copies of $\mathbb{Z}$:

Corollary 5.17 *Let G be a torsion-free abelian group. Then $G \cong \mathbb{Z}^r$ for some non-negative integer r.*

A finitely-generated abelian group has a presentation of the form

$$\langle x_1, \ldots, x_n : r_1, \ldots, r_m \rangle$$

where $x_1, \ldots, x_n$ are the generators, and $r_1, \ldots, r_m$ are the relators. We will assume further that every generator commutes with every other generator. We can either incorporate this into the presentation as $n(n-1)$ relators of the form $x_i x_j x_i^{-1} x_j^{-1}$ for $1 \leqslant i < j \leqslant n$,[10] or tacitly assume them by using additive notation for our relators.

[10] These can be rewritten as relations of the form $x_i x_j = x_j x_i$.

Using additive notation, our relators will be $\mathbb{Z}$–linear combinations of the generators, that is, expressions of the form

$$k_1 x_1 + k_2 x_2 + \cdots + k_n x_n$$

where $k_1, \ldots, k_n \in \mathbb{Z}$. Each of these relators is equal to the (additive) identity 0, and so effectively we have a homogeneous system of simultaneous equations with integer coefficients.

In linear algebra we learn a standard method of solving systems of linear equations with real coefficients: we write down the augmented matrix of the system and apply elementary operations to convert it to **reduced row echelon form**.

The approach we use here is similar but with a couple of differences: since our coefficients are in $\mathbb{Z}$ rather than $\mathbb{R}$, we can't divide a row (or column) by an arbitrary number, so that limits one of our elementary operations. And instead of reduced row echelon form, we want our matrix to be in a slightly different form:

Definition 5.18 An $m \times n$ matrix A of rank r is in **Smith normal form** if every element is zero, except possibly for the diagonal elements a_{ii} for $1 \leqslant i \leqslant r$, and furthermore that $a_{ii} | a_{(i+1)(i+1)}$ for $1 \leqslant i < r$.

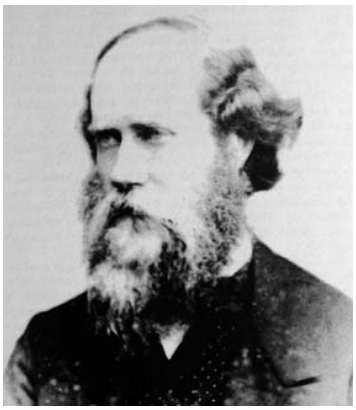
That is, A is in the form shown in Figure 5.2, where the last $(n-r)$ columns and $(m-r)$ rows are all zero.

$$\begin{bmatrix} a_{11} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & a_{rr} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

Figure 5.2: A matrix in Smith normal form

Example 5.19 The following matrices are in Smith normal form:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 12 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 7 & 0 & 0 & 0 \\ 0 & 14 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 4 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$



Henry Smith (1826–1883)

We need to define the appropriate elementary row operations:

Definition 5.20 Let A be an $m \times n$ matrix with integer entries. Then we may apply one or more of the following **elementary row** and **column operations** to A in order to obtain a similar matrix.

- **E1** Swap two rows (or columns).
- **E2** Multiply all entries of a row (or column) by -1.
- **E3** Add an integer multiple of one row (or column) to another row (or column).

The next proposition is the key to the whole problem:

Proposition 5.21 *Any $m \times n$ integer matrix A can be transformed into Smith normal form by a finite sequence of row operations of type E1, E2 and E3.*

Equivalently, there exists an $m \times m$ integer matrix P and an $n \times n$ integer matrix Q such that PAQ is an $m \times n$ matrix in Smith normal form.

The following algorithm provides a constructive proof of this fact:

Algorithm 5.22 Assume that A is not already in Smith normal form.

1 Use row operations of type E1 to arrange the matrix so that the nonzero entries in the first row and first column are in ascending order of absolute value, followed by any zero elements, so that $|a_{11}| \leqslant |a_{12}| \leqslant \cdots \leqslant |a_{1s}|$ and $|a_{11}| \leqslant |a_{21}| \leqslant \cdots \leqslant |a_{t1}|$.

2 Use row and column operations of type E2 to ensure that all of the entries $a_{11}, \ldots, a_{1s}$ in the first row are positive, and all of the entries $a_{11}, \ldots, a_{t1}$ in the first column are also positive.

3 If a_{11} divides all other nonzero entries in row 1, then go to step 4.
Otherwise, let a_{1j} be the first nonzero entry in the first row which isn't an integer multiple of a_{11}. Then we can find non-negative integers q and p such that $a_{1j} = qa_{11} + p$ where $0 \leqslant p < a_{11}$.
Apply a column operation of type E3, subtracting q times column 1 from column j.
Repeat this process for all the other columns for which the first element is not an integer multiple of a_{11}.
Go to step 1.

4 If a_{11} divides all of the other nonzero entries in the first column, then go to step 5.
Otherwise, let a_{k1} be the first nonzero entry in column 1 that isn't an integer multiple of a_{11}. Then we can find non-negative integers q and p such that $a_{k1} = qa_{11} + p$ where $0 \leqslant p < a_{11}$.
Apply an E3 row operation, subtracting q times row 1 from row k.
Repeat this process for all the other rows for which the first element is not an integer multiple of a_{11}.
Go to step 1.

5 Every entry in row 1 and column 1 is now a multiple of a_{11}.
Now apply column operations of type E3, subtracting multiples of column 1 from each of the other columns with nonzero first element, so that a_{11} is the only nonzero element on the first row.
Similarly, apply row operations of type E3, subtracting multiples of row 1 from all the other rows that have a nonzero first element, so that a_{11} is left as the only nonzero element in column 1.
The matrix is then in the form

$$\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

and we can apply steps 1–5 to the $(m-1) \times (n-1)$ submatrix

$$\begin{bmatrix} a_{22} & \cdots & a_{2n} \\ \vdots & & \vdots \\ a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

to get an $m \times n$ matrix where the only nonzero elements in the first two rows and columns are on the diagonal. We repeat this until we get a matrix whose only nonzero entries are on the diagonal.

These diagonal entries won't necessarily satisfy the divisibility criterion $a_{11}|a_{22}|\dots|a_{rr}$. If they don't, we proceed to the next step.

6 For $1 \leqslant i \leqslant r-1$, compare a_{ii} with each a_{jj} for $i < j \leqslant r$. Let i and j be the lowest integers for which a_{ii} doesn't divide a_{jj}. Use a row operation of type E3 to add row j to row i, and then reduce this new $m \times n$ matrix using steps 1–5.

This algorithm will eventually terminate, yielding a matrix in Smith normal form. Let's try a couple of examples.

Example 5.23 Let G be the abelian group with generators x and y, and relations

$$2x = 0, \qquad\qquad 3y = 0.$$

(This group is isomorphic to $\mathbb{Z}_2 \oplus \mathbb{Z}_3$.) This yields the coefficient matrix $A = \left[\begin{smallmatrix} 2 & 0 \\ 0 & 3 \end{smallmatrix}\right]$. Applying Algorithm 5.22 we see A is already diagonal, so we skip to step 6. Now $a_{11} = 2$, which doesn't divide $a_{22} = 3$, so we perform an operation of type E3, adding row 2 to row 1, to get the matrix $\left[\begin{smallmatrix} 2 & 3 \\ 0 & 3 \end{smallmatrix}\right]$. Starting again from step 1, the matrix evolves as follows:

$$\left[\begin{smallmatrix} 2 & 3 \\ 0 & 3 \end{smallmatrix}\right] \longmapsto \left[\begin{smallmatrix} 2 & 1 \\ 0 & 3 \end{smallmatrix}\right] \longrightarrow \left[\begin{smallmatrix} 1 & 2 \\ 3 & 0 \end{smallmatrix}\right] \longmapsto \left[\begin{smallmatrix} 1 & 0 \\ 3 & -6 \end{smallmatrix}\right] \longrightarrow \left[\begin{smallmatrix} 1 & 0 \\ 0 & -6 \end{smallmatrix}\right].$$

We can perform a final E2 move to get $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 6 \end{smallmatrix}\right]$, and then read off the new relations

$$x = 0, \qquad 6y = 0$$

which yields the group $\mathbb{Z}_6$, as expected from Proposition 3.20.

Example 5.24 Let G be the abelian group with four generators w, x, y and z, and three relations

$$\begin{aligned} 16w + 56x + 4y + 48z &= 0, \\ 4w + 16x + 4y - 8z &= 0, \\ 10w + 22x - 2y + 70z &= 0. \end{aligned}$$

Under the application of Algorithm 5.22, the coefficient matrix evolves as follows:

$$\left[\begin{smallmatrix} 16 & 56 & 4 & 48 \\ 4 & 16 & 4 & -8 \\ 10 & 22 & -2 & 70 \end{smallmatrix}\right] \xrightarrow{\text{E1}} \left[\begin{smallmatrix} -2 & 10 & 22 & 70 \\ 4 & 4 & 16 & -8 \\ 4 & 16 & 56 & 48 \end{smallmatrix}\right] \xrightarrow{\text{E2}} \left[\begin{smallmatrix} 2 & 10 & 22 & 70 \\ 4 & -4 & -16 & 8 \\ 4 & -16 & -56 & -48 \end{smallmatrix}\right]$$

$$\xrightarrow{\text{E3}} \left[\begin{smallmatrix} 2 & 0 & 0 & 0 \\ 4 & -24 & -60 & -132 \\ 4 & -36 & -100 & -188 \end{smallmatrix}\right] \xrightarrow{\text{E2}} \left[\begin{smallmatrix} 2 & 0 & 0 & 0 \\ 0 & 24 & 60 & 132 \\ 0 & 36 & 100 & 188 \end{smallmatrix}\right] \xrightarrow{\text{E3}} \left[\begin{smallmatrix} 2 & 0 & 0 & 0 \\ 0 & 24 & 12 & 12 \\ 0 & 36 & 28 & 8 \end{smallmatrix}\right]$$

$$\xrightarrow{\text{E1}} \left[\begin{smallmatrix} 2 & 0 & 0 & 0 \\ 0 & 8 & 28 & 36 \\ 0 & 12 & 12 & 24 \end{smallmatrix}\right] \xrightarrow{\text{E3}} \left[\begin{smallmatrix} 2 & 0 & 0 & 0 \\ 0 & 8 & 4 & 4 \\ 0 & 12 & -24 & -24 \end{smallmatrix}\right] \xrightarrow{\text{E1}} \left[\begin{smallmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 4 & 8 \\ 0 & -24 & -24 & 12 \end{smallmatrix}\right]$$

$$\xrightarrow{\text{E2}} \left[\begin{smallmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 4 & 8 \\ 0 & 24 & 24 & -12 \end{smallmatrix}\right] \xrightarrow{\text{E3}} \left[\begin{smallmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 24 & 0 & -60 \end{smallmatrix}\right] \xrightarrow{\text{E1}} \left[\begin{smallmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 24 & -60 & 0 \end{smallmatrix}\right]$$

$$\xrightarrow{\text{E3}} \left[\begin{smallmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & -60 & 0 \end{smallmatrix}\right] \xrightarrow{\text{E2}} \left[\begin{smallmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 60 & 0 \end{smallmatrix}\right]$$

The corresponding relations are therefore

$$2w = 0, \qquad\qquad 4x = 0, \qquad\qquad 60y = 0.$$

The fourth generator z is free, since it has no associated relation, and hence $G \cong \mathbb{Z}_2 \oplus \mathbb{Z}_4 \oplus \mathbb{Z}_{60} \oplus \mathbb{Z}$.

We can use Theorem 5.15 together with Proposition 3.20 to compile a complete list (up to isomorphism) of finite abelian groups of a given order.

Example 5.25 Let G be an abelian group of order 12. Then since G must be isomorphic to a direct sum of finite cyclic groups, we have the following possibilities:

$$\mathbb{Z}_{12}, \quad \mathbb{Z}_2 \oplus \mathbb{Z}_6, \quad \mathbb{Z}_3 \oplus \mathbb{Z}_4, \quad \text{and} \quad \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_3.$$

By Proposition 3.20 we know that

$$\mathbb{Z}_3 \oplus \mathbb{Z}_4 \cong \mathbb{Z}_{12} \quad \text{and} \quad \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_3 \cong \mathbb{Z}_2 \oplus \mathbb{Z}_6.$$

Only two of these four satisfy the invariant factor condition, namely $\mathbb{Z}_{12}$ and $\mathbb{Z}_2 \oplus \mathbb{Z}_6$. Hence, up to isomorphism there are two abelian groups of order 12.

There is another classification theorem for finitely-generated abelian groups, in which the torsion subgroups are grouped in a different way.

Definition 5.26 Suppose that

$$G \cong \mathbb{Z}_{p_1^{n_1}} \oplus \cdots \oplus \mathbb{Z}_{p_k^{n_k}} \oplus \mathbb{Z}^r$$

where $p_1, \ldots, p_k \in \mathbb{N}$ are (not necessarily distinct) prime integers, and $n_1, \ldots, n_k, r \in \mathbb{Z}$.

This is called a **primary decomposition** of G with **torsion coefficients** $p_1^{n_1}, \ldots, p_k^{n_k}$ and **rank** r.

Theorem 5.27 *A finitely-generated abelian group G has a unique primary decomposition*

$$G \cong \mathbb{Z}_{p_1^{n_1}} \oplus \cdots \oplus \mathbb{Z}_{p_k^{n_k}} \oplus \mathbb{Z}^r$$

in the sense that if H is another finitely-generated abelian group with primary decomposition

$$H \cong \mathbb{Z}_{q_1^{m_1}} \oplus \cdots \oplus \mathbb{Z}_{q_l^{m_l}} \oplus \mathbb{Z}^s$$

then $G \cong H$ if and only if $k = l$, $r = s$ and the primes $p_1, \ldots, p_k$ are equal to the primes $q_1, \ldots, q_k$, up to possible rearrangement.

The Earth is full of anger,
The seas are dark with wrath,
The Nations in their harness
Go up against our path:
Ere yet we loose the legions –
Ere yet we draw the blade,
Jehovah of the Thunders,
Lord God of Battles, aid!

— Rudyard Kipling (1865–1936),
Hymn Before Action (1896)

6 Group Actions

MANY of the groups we've met so far consist of functions mapping from a set to itself. For example, the group $\mathrm{Sym}(X)$ of permutations on a given set X, the general linear group $GL_n(\mathbb{R})$ of invertible linear transformations on $\mathbb{R}^n$, the dihedral group D_n of symmetry operations on the regular n–gon. In this chapter, we will study this situation in more detail.

6.1 Groups acting on sets

When considering groups of transformations on some set, we often say that the group "acts on" the set. In general, we expect the identity element to leave the set unchanged. We also expect the transformations to behave consistently with respect to the group operation. The following definition formalises this idea:

Definition 6.1 Let G be a group and X a set. An **action** of G on X is a map $\cdot\colon G\times X \to X$ satisfying the following two properties:

A1 $1_G \cdot x = x$ for all $x \in X$, and
A2 $(gh)\cdot x = g\cdot(h\cdot x)$ for all $g, h \in G$ and $x \in X$.

We denote the image of (g, x) under the map $\cdot$ by $g \cdot x$. Strictly speaking, this is the definition for a **left action** of G on X. The definition for a right action is very similar: we consider a map $\cdot\colon X\times G \to X$ satisfying analogous properties. In this module we will only study left actions, although the corresponding theory for right actions is equivalent.

Example 6.2 If $G = \mathrm{Sym}(X)$ (or any subgroup of $\mathrm{Sym}(X)$), then G acts on X by setting $\sigma \cdot x = \sigma(x)$ for all $\sigma \in \mathrm{Sym}(X)$ and $x \in X$.

Example 6.3 The general linear group $GL_n(\mathbb{R})$ and as its various subgroups such as $SL_n(\mathbb{R})$, $O_n(\mathbb{R})$ and $SO_n(\mathbb{R})$, all act on $\mathbb{R}^n$ by $A \cdot \mathbf{v} = A\mathbf{v}$ for all matrices $A \in GL_n(\mathbb{R})$ and vectors $\mathbf{v} \in \mathbb{R}^n$.

Example 6.4 The dihedral group D_n acts on the regular n–sided polygon P_n with vertices at $\left(\cos\left(\frac{2k\pi}{n}\right), \sin\left(\frac{2k\pi}{n}\right)\right)$ for $k \in \mathbb{Z}$ and $0 \leqslant k < n$. Similarly, we can regard D_n as acting merely on the set of vertices of P_n, rather than the entire polygon.

In each of these examples, every element of the group G determines a permutation of the set X. That is, an element of $\mathrm{Sym}(X)$. This is always the case:

Proposition 6.5 *Let $\cdot$ be an action of a group G on a set X. For all $g \in G$ define the map $f_g \colon X \to X$ by $f_g(x) = g \cdot x$ for all $x \in X$. Then $f_g \in \mathrm{Sym}(X)$, and the map $\phi \colon G \to \mathrm{Sym}(X)$ given by $\phi(g) = f_g$ is a homomorphism.*

Proof This follows from Definition 6.1. Property A1 says that $\phi(1_G) = f_1$ is the identity map $\mathrm{id}_X \colon X \to X$. And property A2 tells us that $f_g \circ f_h = f_{gh}$ for all $g, h \in G$, since

$$(f_g \circ f_h)(x) = f_g(f_h(x)) = g \cdot (h \cdot x) = (gh) \cdot x = f_{gh}(x)$$

for all $x \in X$. Hence

$$\phi(g)\phi(g^{-1}) = f_g \circ f_{g^{-1}} = f_{gg^{-1}} = f_1 = \mathrm{id}_X,$$
$$\text{and} \quad \phi(g^{-1})\phi(g) = f_{g^{-1}} \circ f_g = fg^{-1}g = f_1 = \mathrm{id}_X.$$

So $\phi(g) = f_g$ and $\phi(g^{-1}) = f_{g^{-1}}$ are inverse maps, which means that $\phi(g) = f_g$ is bijective, hence $\phi(g) = f_g \in \mathrm{Sym}(X)$.

To see that ϕ is a homomorphism, we use property A2 again. Then

$$\phi(g)\phi(h) = f_g \circ f_h = f_{gh} = \phi(gh)$$

for all $g, h \in G$. □

Definition 6.6 Let G be a group and X be a set. The **kernel** of an action $\cdot$ of G on X is defined to be the kernel $K = \ker(\phi)$ of the homomorphism $\phi \colon G \to \mathrm{Sym}(X)$ in Proposition 6.5:

$$K = \{g \in G : g \cdot x = x \text{ for all } x \in X\}.$$

The action is said to be **faithful** if $K = \{1\}$, or equivalently (by Proposition 4.13) if ϕ is injective.

The actions in Examples 6.2, 6.3 and 6.4 are all faithful.

We can define some other actions of D_6 on the regular hexagon P_6:

Example 6.7 Let $E = \{e_1, e_2, e_3, e_4, e_5, e_6\}$ be the set of edges of P_6, where e_1 is the edge joining vertices 1 and 2, e_2 is the edge joining vertices 2 and 3, and so on, with e_6 joining vertices 6 and 1.

We can then define an action of D_6 on E with the homomorphism $\phi \colon D_6 \to \mathrm{Sym}(E) \cong S_6$ given by

$$\phi(r) = (e_1, e_2, e_3, e_4, e_5, e_6) \quad \text{and} \quad \phi(m_1) = (e_1, e_6)(e_2, e_3)(e_4, e_5).$$

This action is faithful.

Example 6.8 Let $D = \{d_1, d_2, d_3\}$ be the set of diagonals of the hexagon P_6, where d_1 joins vertices 1 and 4, d_2 joins vertices 2 and 5, and d_3 joins vertices 3 and 6.

Then we can define an action of D_6 on D by defining $\phi \colon D_6 \to \mathrm{Sym}(D) \cong S_3$, such that

$$\phi(r) = (d_1, d_2, d_3) \quad \text{and} \quad \phi(m_1) = (d_2, d_3).$$

This action is not faithful: its kernel is $\{1, r^3\}$.

There are a couple of important actions of a group on itself. The first of these arises from left multiplication in the group:

Example 6.9 We define the **left regular action** of a group G on itself by setting $g \cdot x = gx$ for all $g \in G$ and $x \in X = G$. Conditions A1 and A2 of Definition 6.1 hold, so this is an action. If g is in the kernel of the action, then that means that $gx = x$ for all $x \in G$, which implies that $g = 1$ by the right cancellation law, so this action is faithful.

We have met this action before in the proof of Cayley's Theorem.[1] The permutations $\lambda_g \in \mathrm{Sym}(G)$ given by multiplying on the left by an element $g \in G$ are exactly the images $\phi(g) \in \mathrm{Sym}(G)$ determined by this action.

[1] Theorem 2.10, page 19.

Alternatively, we can use the fact that this action is faithful, so the kernel K is trivial. Then $G \cong G/K$, and by the First Isomorphism Theorem[2] this is isomorphic to $\mathrm{im}(\phi) \leqslant \mathrm{Sym}(G)$.

[2] Theorem 4.15, page 34.

Another important example is given by conjugation:

Example 6.10 The **conjugation action** of a group G on itself is given by $g \cdot x = gxg^{-1}$ for all $g \in G$ and $x \in X = G$.

Suppose that g is in the kernel of this action. Then this means that $gxg^{-1} = x$ for all $x \in G$, which is equivalent to saying that $gx = xg$ for all $x \in G$. So the kernel of this action consists of those elements of G that commute with every element of G. This is called the **centre** of the group:

$$Z(G) = \{g \in G : gh = hg \text{ for all } h \in G\}.$$

The conjugation action will therefore be faithful only when the group has trivial centre; for example when $G = D_3$. But if G is abelian, then $Z(G) = G$, and the action will not be faithful.

6.2 *Orbits and stabilisers*

Let $\sigma = (1,2,4)(3,5) \in S_5$. The cyclic subgroup $G = \langle\sigma\rangle \leqslant S_5$ generated by this permutation acts on the set $X_5 = \{1,2,3,4,5\}$. Because of the cycle decomposition of σ, all of the elements $\sigma^k \in \langle\sigma\rangle$ permute 1, 2 and 4 amongst themselves, and also permute 3 and 5 amongst themselves. So the action of G partitions the set X_5 into two disjoint subsets $\{1,2,4\}$ and $\{3,5\}$.

Recall from *MA138 Sets and Numbers*, *MA132 Foundations* or elsewhere, that a partition determines an **equivalence relation** and vice versa.[3] We want to generalise this idea and consider the equivalence classes of group actions, so we introduce the following definition:

Definition 6.11 Let $\cdot$ be an action of a group G on a set X. We define a relation $\sim$ on X, so that for any $x, y \in X$, we say $x \sim y$ if and only if there exists some element $g \in G$ such that $g \cdot x = y$. This is an equivalence relation (check this).

The equivalence classes of this relation are called **orbits**. The orbit of a given element $x \in X$ is defined to be

$$\mathrm{Orb}_G(x) = \{g \cdot x : g \in G\}.$$

[3] Let S be a set. A **relation** $\sim$ on S is determined by a subset $R \subseteq S \times S$. For any two elements $x, y \in S$, we write $x \sim y$ if $(x,y) \in R$, and $x \not\sim y$ if $(x,y) \notin R$.

Thus any two elements $x, y \in S$ are either **related** ($x \sim y$) or not ($x \not\sim y$).

A relation $\sim$ on a set S is said to be:

- **reflexive** if $x \sim x$ for all $x \in S$,
- **symmetric** if, whenever $x \sim y$, then $y \sim x$ for all $x, y \in S$, and
- **transitive** if, whenever $x \sim y$ and $y \sim z$, then $x \sim z$ for all $x, y, z \in S$.

A relation that is reflexive, symmetric and transitive is called an **equivalence relation**.

An equivalence relation partitions the set S into **equivalence classes**. In particular, for any $x \in S$ the corresponding equivalence class is

$$[x] = \{y \in S : x \sim y\}.$$

The orbit $\mathrm{Orb}_G(x)$ of a given element $x \in X$ is essentially everything in X that can be reached from x by means of the group action.

In the example concerning the group $G = \langle \sigma \rangle \leqslant S_5$, where $\sigma = (1,2,4)(3,5)$, then the orbits are:

$$\mathrm{Orb}_G(1) = \mathrm{Orb}_G(2) = \mathrm{Orb}_G(4) = \{1,2,4\}$$
$$\mathrm{Orb}_G(3) = \mathrm{Orb}_G(5) = \{3,5\}$$

Sometimes the action may yield a single orbit (that is, the entirety of X), and we give such actions a special name:

Definition 6.12 An action of a group G on a set X is **transitive** if it only has one orbit. Equivalently, if for any $x, y \in X$ there exists some $g \in G$ such that $g \cdot x = y$.

The actions in Examples 6.2, 6.4, 6.7 and 6.8 are transitive. But the action of $GL_n(\mathbb{R})$ on $\mathbb{R}^n$ is not transitive: it has two orbits: $\{\mathbf{0}\}$ and $\mathbb{R}^n \setminus \{\mathbf{0}\}$.

The permutation $\sigma = (1,2,4)(3,5)$ affects every element of the set $X_5 = \{1,2,3,4,5\}$, but the permutation $\tau = (3,4)$ only affects 3 and 4, leaving 1, 2 and 5 unchanged. And the identity permutation $\iota = (\ )$ leaves all of X_5 fixed. It is often helpful to look at which elements of a group G leave a particular element of X fixed:

Definition 6.13 Let a group G act on a set X, and suppose that x is some element of X. Then the **stabiliser** of x in G is the subset

$$\mathrm{Stab}_G(x) = \{g \in G : g \cdot x = x\} \subseteq G.$$

That is, $\mathrm{Stab}_G(x)$ consists of the group elements that leave x fixed.

The stabiliser $\mathrm{Stab}_G(x)$ is not just a subset of G, it is a subgroup:

Proposition 6.14 *Let a group G act on a set X, and suppose that x is some element of X. Then:*

(i) *the stabiliser $\mathrm{Stab}_G(x)$ is a subgroup of G, and*

(ii) *the intersection $\bigcap_{x \in X} \mathrm{Stab}_G(x)$ is the kernel of the action of G on X.*

Proof

(i) Since 1_G acts trivially on any element $x \in X$, it follows that $1_G \in \mathrm{Stab}_G(x)$, and hence $\mathrm{Stab}_G(x)$ is a nonempty subset of G. Now suppose that $g, h \in \mathrm{Stab}_G(x)$. Then

$$(gh) \cdot x = g \cdot (h \cdot x) = g \cdot x = x$$

so $gh \in \mathrm{Stab}_G(x)$. Finally,

$$g^{-1} \cdot x = g^{-1} \cdot (g \cdot x) = (g^{-1}g) \cdot x = 1_G \cdot x = x$$

and hence $g^{-1} \in \mathrm{Stab}_G(x)$. Therefore $\mathrm{Stab}_G(x) \leqslant G$ by Proposition 2.3.

(ii) For any $g \in G$, we have $g \in \bigcap_{x \in X} \mathrm{Stab}_G(x)$ if and only if $g \cdot x = x$ for all $x \in X$. This is equivalent to g lying in the kernel of the action.

This completes the proof. □

Some books refer to the stabiliser $\mathrm{Stab}_G(x)$ as the **isotropy subgroup**.

The following important theorem gives a strong connection between orbits and stabilisers of a group action:

Theorem 6.15 (The Orbit–Stabiliser Theorem) *Let G be a finite group acting on a set X, and let $x \in X$. Then*

$$|G| = |\mathrm{Orb}_G(x)| \times |\mathrm{Stab}_G(x)|.$$

Proof Let $y \in \mathrm{Orb}_G(x)$. Then there exists some element $g \in G$ with $g \cdot x = y$. Let $H = \mathrm{Stab}_G(x)$. For some element $k \in G$ we have $k \cdot x = y$ if and only if $k \cdot x = g \cdot x$, or equivalently $(g^{-1}k) \cdot x = x$, which is the same as saying that $g^{-1}k \in \mathrm{Stab}_G(x) = H$. And by Proposition 2.13 this means that $k \in gH$.

So the elements $k \in G$ with $k \cdot x = y$ are exactly the elements of the coset gH. And by Proposition 2.17, we have $|gH| = |H|$. That is, for each $y \in \mathrm{Orb}_G(x)$ there are exactly $|H|$ elements $k \in G$ such that $k \cdot x = y$. Hence the total number of such elements $y \in \mathrm{Orb}_G(x)$ must be $|G|/|H|$, so

$$|G| = |\mathrm{Stab}_G(x)| \times |\mathrm{Orb}_G(x)|$$

as claimed. □

6.3 *Conjugacy classes*

Now we want to look at the orbits of the conjugation action. This will enable us to prove a couple of important results about alternating groups.

Definition 6.16 Let G be a group, and consider the conjugation action of G on itself. The orbits of this action are called **conjugacy classes**, and we will denote the conjugacy class of a given element $g \in G$ by

$$\mathrm{Cl}_G(g) = \mathrm{Orb}_G(g) = \{hgh^{-1} : h \in G\}.$$

What about the stabiliser $\mathrm{Stab}_G(g)$ of a given element $g \in G$? By definition, this comprises the elements h in G for which $h \cdot g = g$. That is, all $h \in G$ such that $hgh^{-1} = g$, or equivalently $hg = gh$. So the stabiliser $\mathrm{Stab}_G(g)$ consists of everything in G that commutes with the chosen element g. As noted earlier, the kernel of this action is the centre $Z(G)$ of the group.

Definition 6.17 Let G be a group. The **centraliser** of an element $g \in G$ is the subgroup

$$C_G(g) = \mathrm{Stab}_G(g) = \{h \in G : hgh^{-1} = g\} = \{h \in G : hg = gh\}.$$

Applying the Orbit–Stabiliser Theorem, we get the following result:

Proposition 6.18 *Let G be a finite group and let $g \in G$. Then*

$$|\mathrm{Cl}_G(g)| = |\mathrm{Orb}_G(g)| = |G|/|\mathrm{Stab}_G(g)| = |G|/|C_G(g)|.$$

Let's look at some examples.

Example 6.19 Let G be an abelian group. Then $Z(G) = G$. Also, for any $g \in G$, the centraliser $C_G(g) = G$, and the conjugacy class $\mathrm{Cl}_G(g) = \{g\}$.

Example 6.20 Let $G = D_4 = \{1, r, r^2, r^3, m_1, m_1r, m_1r^2, m_1r^3\}$. Then $\mathrm{Cl}_G(1) = \{g1g^{-1} : g \in G\} = \{1\}$, and $C_G(1) = G$.

Since $r^ir^2 = r^2r^i$ and $(r^im_1)r^2 = r^2(r_im_1)$ for $1 \leqslant i < 4$ we have

$$\mathrm{Cl}_G(r^2) = \{r^2\} \quad \text{and} \quad C_G(r^2) = G.$$

Now $\langle r \rangle \leqslant C_G(r)$ while $rm_1 = m_2 \neq m_4 = m_1r$. Hence

$$4 = |r| \leqslant |C_G(r)| < |G| = 8.$$

Lagrange's Theorem implies that $|C_G(r)| = 4$, so $C_G(r) = \langle r \rangle = \{1, r, r^2, r^3\}$. By Proposition 6.18, it follows that

$$|\mathrm{Cl}_G(r)| = |G|/|C_G(r)| = 8/4 = 2.$$

And since $m_1rm_1^{-1} = r^3$ we have $\mathrm{Cl}_G(r) = \{r, r^3\}$.

Similarly, $\{1, m_1, r^2, r^2m_1\} \leqslant C_G(m_1)$, while $rm_1 = m_2 \neq m_4 = m_1r$. Hence

$$4 \leqslant |C_G(m_1)| < |G| = 8,$$

and by Lagrange's Theorem $|C_G(m_1)| = 4$. Thus

$$C_G(m_1) = \{1, m_1, r^2, r^2m_1\} = \{1, m_1, r^2, m_3\} \cong V_4 \cong \mathbb{Z}_2 \times \mathbb{Z}_2.$$

By Proposition 6.18, we have

$$|\mathrm{Cl}_G(m_1)| = |G|/|C_G(m_1)| = 8/4 = 2.$$

Since $rm_1r^{-1} = r^2m_1 = m_3$ it must be the case that

$$\mathrm{Cl}_G(m_1) = \{m_1, r^2m_1\} = \{m_1, m_3\}.$$

Finally, we can see that

$$\mathrm{Cl}_G(rm_1) = \{rm_1, r^3m_1\} = \{m_2, m_4\}$$
$$\text{and} \quad C_G(rm_1) = \{1, rm_1, r^2, r^3m_1\} = \{1, m_2, r^2, m_4\}.$$

Hence the conjugacy classes of D_4 are

$$\mathrm{Cl}_G(1) = \{1\}, \qquad \mathrm{Cl}_G(m_1) = \mathrm{Cl}_G(m_3) = \{m_1, m_3\},$$
$$\mathrm{Cl}_G(r^2) = \{r^2\} \qquad \mathrm{Cl}_G(m_2) = \mathrm{Cl}_G(m_4) = \{m_2, m_4\},$$
$$\mathrm{Cl}_G(r) = \mathrm{Cl}_G(r^3) = \{r, r^3\}.$$

The conjugacy class of a permutation in a symmetric group $\mathrm{Sym}(X)$ is determined entirely by the permutation's cycle structure. We will prove this in two parts.

Proposition 6.21 *Let $\sigma, \tau \in \mathrm{Sym}(X)$. Then if we write σ in cycle form, we can obtain the conjugate $\tau\sigma\tau^{-1}$ by replacing each $x \in X$ in the cycles of σ by $\tau(x)$.*

Proof Suppose that $(x_1, \ldots, x_r)$ is a cycle of σ. Then $\sigma(x_1) = x_2$, and

hence $\tau\sigma(x_1) = \tau(x_2)$. Furthermore, $\tau\sigma\tau^{-1}\tau(x_1) = \tau(x_2)$. Similarly, $\tau\sigma\tau^{-1}\tau(x_i) = \tau(x_{i+1})$ for $1 \leqslant i < r$, and $\tau\sigma\tau^{-1}\tau(x_r) = \tau(x_1)$. Hence the conjugate $\tau\sigma\tau^{-1}$ has a cycle $(\tau(x_1), \tau(x_2), \ldots, \tau(x_r))$. □

For example, suppose $\sigma, \tau \in S_7$ such that $\sigma = (1,5)(2,4,7,6)$ and $\tau = (1,3,5,7,2,4,6)$, then $\tau\sigma\tau^{-1} = (3,7)(4,6,2,1)$.

Definition 6.22 Let $\sigma \in \mathrm{Sym}(X)$. We say that σ has **cycle type** $2^{r_2}3^{r_3}4^{r_4}\ldots$ if it has exactly r_i cycles of length i, for $i \geqslant 2$.

So Proposition 6.21 says that conjugation in $\mathrm{Sym}(X)$ doesn't change the cycle type; that is, conjugate cycles in $\mathrm{Sym}(X)$ have the same cycle type.

The converse holds as well, giving the following proposition.

Proposition 6.23 *Two permutations in* $\mathrm{Sym}(X)$ *are conjugate if and only if they have the same cycle type.*

Proof By Proposition 6.21, two conjugate permutations in $\mathrm{Sym}(X)$ have the same cycle type.

Conversely, suppose that two permutations $\sigma, \tau \in \mathrm{Sym}(X)$ have the same cycle type. Then we can define a bijective correspondence between the cycles in σ and those in τ such that each cycle in σ is paired with one of the same length in τ.

Suppose a cycle $(x_1, \ldots, x_r)$ in σ is paired with a cycle $(y_1, \ldots, y_r)$ in τ. Then we can construct a permutation $\alpha \in \mathrm{Sym}(X)$ such that $\alpha(x_i) = y_i$ for $1 \leqslant i \leqslant r$. Then by Proposition 6.21,

$$\alpha(x_1, \ldots, x_r)\alpha^{-1} = (\alpha(x_1), \ldots, \alpha(x_r)) = (y_1, \ldots, y_r).$$

We can do this for all the constituent cycles of σ to obtain an element $\alpha \in \mathrm{Sym}(X)$ such that $\alpha\sigma\alpha^{-1} = \tau$. Hence permutations with the same cycle type are conjugate in $\mathrm{Sym}(X)$. □

We can use this result to find the conjugacy classes of symmetric groups S_n:

Example 6.24 The group S_3 has three conjugacy classes:

cycle type	conjugacy class
1	$\{\iota\}$
2^1	$\{(1,2),(1,3),(2,3)\}$
3^1	$\{(1,2,3),(1,3,2)\}$

Example 6.25 The symmetric group S_4 has five conjugacy classes:

cycle type	conjugacy class
1	$\{\iota\}$
2^1	$\{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\}$
2^2	$\{(1,2)(3,4),(1,3)(2,4),(1,4)(2,3)\}$
3^1	$\{(1,2,3),(1,2,4),(1,3,4),(2,3,4),$ $(1,3,2),(1,4,2),(1,4,3),(2,4,3)\}$
4^1	$\{(1,2,3,4),(1,2,4,3),(1,3,2,4),$ $(1,4,3,2),(1,3,4,2),(1,4,2,3)\}$

Things get slightly complicated with the alternating groups $\mathrm{Alt}(X)$. The reason for this is that just because two even permutations are

conjugate in $\mathrm{Sym}(X)$, it doesn't necessarily mean that they are conjugate in $\mathrm{Alt}(X)$.

For example, $\sigma = (1,2,3,4,5)$ and $\tau(1,3,5,2,4)$ are both conjugate in S_5; in particular if we set $\alpha = (2,3,5,4) \in S_5$, then $\alpha\sigma\alpha^{-1} = \tau$. But α is a 4–cycle, and hence an odd permutation, so it doesn't belong to A_5. In fact, there is no even permutation that conjugates σ to τ. So σ and τ are conjugate in S_5, but not in A_5.

Sometimes the conjugacy class of a permutation in A_n is the same as its conjugacy class in S_n, and sometimes it is half of the corresponding class in S_n.

Proposition 6.26 *Let $G = S_n$ and $H = A_n$. Let $\sigma \in H$. Then either $\mathrm{Cl}_H(\sigma) = \mathrm{Cl}_G(\sigma)$ or $|\mathrm{Cl}_H(\sigma)| = \frac{1}{2}|\mathrm{Cl}_G(\sigma)|$.*

Proof Since $H = A_n \leqslant S_n = G$, if $\alpha \in H$ then $\alpha \in G$ and so $\alpha\sigma\alpha^{-1} \in \mathrm{Cl}_G(\sigma)$. Hence $\mathrm{Cl}_H(\sigma) \subseteq \mathrm{Cl}_G(\sigma)$.

Also, if $\alpha \in C_H(\sigma)$ then $\alpha \in H \leqslant G$ and $\alpha\sigma = \sigma\alpha$, so $\alpha \in C_G(\sigma)$. Thus $C_H(\sigma) \subseteq C_G(\sigma)$.

By Proposition 6.18,

$$|\mathrm{Cl}_G(\sigma)||C_G(\sigma)| = |S_n| = 2|A_n| = 2|\mathrm{Cl}_H(\sigma)||C_H(\sigma)| \qquad (6.1)$$

[4] Theorem 2.18, page 21.

Since $C_H(\sigma)$ is a subgroup of $C_G(\sigma)$, by Lagrange's Theorem,[4] $|C_H(\sigma)|$ divides $|C_G(\sigma)|$. We thus have three possible cases:

Case 1 $C_H(\sigma) = C_G(\sigma)$. Then $|\mathrm{Cl}_H(\sigma)| = \frac{1}{2}|\mathrm{Cl}_G(\sigma)|$ by (6.1).

Case 2 $|C_H(\sigma)| = \frac{1}{2}|C_G(\sigma)|$. Then $\mathrm{Cl}_H(\sigma) = \mathrm{Cl}_G(\sigma)$ by (6.1).

Case 3 $|C_H(\sigma)| < \frac{1}{2}|C_G(\sigma)|$. Since $\mathrm{Cl}_H(\sigma) \leqslant \mathrm{Cl}_G(\sigma)$ it must be the case that $|\mathrm{Cl}_H(\sigma)| \leqslant |\mathrm{Cl}_G(\sigma)|$, which contradicts (6.1).

Only cases 1 and 2 can occur, while case 3 is impossible. This completes the proof. □

Example 6.27 The possible cycle types for elements of A_4 are 1 (the identity permutation), 2^2 (double transpositions) and 3^1 (3–cycles). The 3–cycles form two conjugacy classes.

cycle type	conjugacy class	size
1	$\{\iota\}$	1
2^2	$\{(1,2)(3,4),(1,3)(2,4),(1,4)(2,3)\}$	3
3^1	$\{(1,2,3),(4,2,1),(2,4,3),(3,4,1)\}$	4
3^1	$\{(1,3,2),(4,1,2),(2,3,4),(3,1,4)\}$	4

Example 6.28 We now calculate the conjugacy classes in A_5. Let $G = S_5$ and $H = A_5$.

Cycle type 1 The identity permutation $\iota = (\ )$ forms a conjugacy class on its own.

Cycle type 2^2 Let $\sigma \in A_5$ have cycle type 2^2. There are 15 permutations in S_5 of this cycle type, so by Proposition 6.26, $|\mathrm{Cl}_H(\sigma)| = 15$ or $\frac{15}{2}$. But $|\mathrm{Cl}_H(\sigma)|$ must be an integer, so $|\mathrm{Cl}_H(\sigma)| = |\mathrm{Cl}_G(\sigma) = 15|$, and hence the permutations of cycle type 2^2 form a single conjugacy class in A_5.

Cycle type 3^1 There are 20 permutations of cycle type 3^1 in S_5. This is even, so we have to do a bit more work than the previous

case.
Consider two such permutations

$$\sigma = (x_1, x_2, x_3) \quad \text{and} \quad \tau = (y_1, y_2, y_3)$$

in A_5, where

$$\{x_1, x_2, x_3, x_4, x_5\} = \{y_1, y_2, y_3, y_4, y_5\} = \{1, 2, 3, 4, 5\}.$$

Here, x_4 and x_5 are the two elements of $\{1, 2, 3, 4, 5\}$ fixed by σ, while y_4 and y_5 are the two elements fixed by τ. By Proposition 6.23, σ and τ are conjugate in S_5. To see this, define $\alpha \in S_5$ to be the permutation

$$\alpha \colon x_1 \mapsto y_1, x_2 \mapsto y_2, x_3 \mapsto y_3, x_4 \mapsto y_4, x_5 \mapsto y_5.$$

Then $\alpha\sigma\alpha^{-1} = \tau$.
The problem here is that α might not belong to A_5. To resolve this, let $\beta = \alpha(x_4, x_5)$. Then $\beta\sigma\beta^{-1} = \tau$. Furthermore, either α or β is an even permutation, because β is α multiplied by a transposition. Hence either α or β belongs to A_5, so σ and τ lie in the same conjugacy class, and thus the permutations of type 3^1 form a single conjugacy class in A_5.

Cycle type 5^1 Let σ be a 5–cycle in A_5. There are 24 permutations of cycle type 5^1 in S_5, so by Proposition 6.26, $|\mathrm{Cl}_H(\sigma)| = 12$ or 24. By Proposition 6.18, $|\mathrm{Cl}_H(\sigma)|$ divides $|G| = |A_5| = 60$. But $24 \not| 60$, so it must be the case that $|\mathrm{Cl}_H(\sigma)| = 12$. Thus the permutations of cycle type 5^1 split into two conjugacy classes in A_5, each of size 12.

To summarise, A_5 has five conjugacy classes:

cycle type	size
1	1
2^2	15
3^1	20
5^1	12
5^1	12

6.4 *Simple groups*

Now we will briefly study an important class of groups. A full discussion is very much beyond the scope of these notes, but we will look at some relatively straightforward cases.

Definition 6.29 A group G is **simple** if it has no proper, nontrivial normal subgroups. That is, if its only normal subgroups are $\{1\}$ and G itself.

Proposition 6.30 *A simple abelian group G is cyclic of prime order.*

Proof Let G be an abelian group. All subgroups of an abelian group are normal, so we just need to find a proper, nontrivial subgroup to ensure G is not simple.

Choose some element $g \in G$ such that $g \neq 1$. If $|g|$ is infinite,

then the cyclic subgroup $\langle g^2 \rangle$ is nontrivial and proper, so G is not simple. If $|g|$ is finite but not prime, for example $|g| = mn$ for some $m, n \in \mathbb{Z}$, then the cyclic subgroup $\langle g^m \rangle$ has order $1 < n < |G|$, and is thus proper and nontrivial. Hence $|g| = p$ is prime, and it must be the case that $\langle g \rangle = G$, otherwise $\langle g \rangle$ would be a proper nontrivial subgroup.

Hence $G = \langle g \rangle \cong \mathbb{Z}_p$. □

There are infinitely many finite nonabelian simple groups. The full classification theorem is as follows:

Theorem 6.31 *Let G be a finite simple group. Then G is one of the following four types:*

(i) *cyclic groups $\mathbb{Z}_p$ of prime order,*
(ii) *alternating groups A_n for $n \geqslant 4$,*
(iii) *finite groups of Lie type,*
(iv) *sporadic groups.*

We've met types (i) and (ii) already, and types (iii) and (iv) are beyond the scope of this module and we will not discuss them further,[5] except to note that the full proof of this classification theorem took hundreds of mathematicians a few decades to complete, and is spread over several thousand pages of journal articles.

[5] The **finite groups of Lie type** are certain classes of matrix groups over finite fields, while the **sporadic groups** are 26 special cases that don't fit into the other three categories. The smallest sporadic group is the **Mathieu group** M_{11}, which has 7920 elements, and the largest is the **Monster**, which has approximately 8×10^{53} elements. It's finite, but it's still very large.

To finish this chapter, and also the part of this module concerned with group theory, we will prove a couple of results about alternating groups. Both of these results require the following simple lemma:

Lemma 6.32 *A subgroup H of a group G is normal if and only if it is a union of conjugacy classes.*

Proof By Proposition 3.7, $H \trianglelefteq G$ if and only if $ghg^{-1} \in H$ for all $g \in G$ and $h \in H$. But this is the same as saying that $H \trianglelefteq G$ if and only if $\mathrm{Cl}_G(h) \subseteq H$ for all $h \in H$. □

The first result is one that we mentioned earlier. It is the smallest counterexample to the converse of Lagrange's Theorem.

Proposition 6.33 *The alternating group A_4 has no subgroup of order 6.*

Proof Suppose that $H < A_4$ such that $|H| = 6$. Since $|A_4| = 12$, then $|A_4 : H| = 2$ and by Proposition 3.5 H must be a normal subgroup of A_4. By Lemma 6.32, H must be a union of conjugacy classes, and in Example 6.27 we found that A_4 has one conjugacy class of size 1 (containing the identity element), one of size 3, and two of size 4. The subgroup H must certainly contain the one-element conjugacy class consisting of the identity, and then we must find some combination of the remaining three conjugacy classes to provide five other elements. But there is no such combination, so the desired subgroup H can't exist. □

The proof of the following fact is similar.

Proposition 6.34 *The alternating group A_5 is simple.*

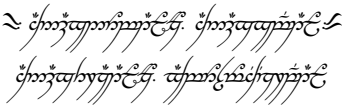
Proof Suppose that N is a proper, nontrivial normal subgroup of A_5. Then by Lemma 6.32, N must be a union of conjugacy classes. In Example 6.28 we found that A_5 has one conjugacy class of size 1,

15 and 20, and two of size 12. So $|N|$ must be the sum of some or all of the numbers 1, 12, 12, 15 and 20.

Furthermore, N must contain the identity, so 1 must be one of these numbers. By Lagrange's Theorem, $|N|$ must divide $|A_5| = 60$. But no such combination adds up to a divisor of 60 other than 1 or 60 itself.

To see this, we know that 1 must be included in the sum. And none of the numbers $1+12 = 13$, $1+15 = 16$ or $1+20 = 21$ divide 60. So any valid combination must include at least two of the numbers 12, 12, 15, 20. But then the sum is at least $1+12+12 = 25$, which doesn't divide 60. The next smallest possibility is $1+12+15 = 28$, which also doesn't divide 60. And the one after that is $1+12+20 = 33$ which is greater than 30, the largest proper divisor of 60.

So no such normal subgroup N exists, and hence A_5 is simple. □



One Ring to rule them all,
One Ring to find them,
One Ring to bring them all
and in the darkness bind them.

— J R R Tolkien (1892–1973),
The Fellowship of the Ring (1954)

7 Rings and Subrings

Now we begin the second part of this module. So far, we've been studying groups: sets equipped with a binary operation satisfying certain properties. Our original model for this structure was the additive structure of the integers. But there is another familiar operation on the set of integers: multiplication. In this section we introduce a new algebraic structure that has two operations, one analogous to addition and the other to multiplication.

7.1 Rings

The integers form an abelian group under addition, but the multiplicative structure is slightly weaker. In particular, no integer apart from 1 and -1 has a multiplicative inverse. And although integer multiplication is commutative, this isn't something we're going to insist on in general.

We'll start with the following definition.

Definition 7.1 A **ring** $R = (R, +, \cdot)$ is a set R together with two binary operations $+$ (called 'addition') and $\cdot$ (called 'multiplication') satisfying the following properties:

R1	$(R, +)$ is an abelian group.	**(additive group)**
R2	$(ab)c = a(bc)$ for all $a, b, c \in R$.	**(associativity)**
R3	$(a + b)c = ac + bc$ and $a(b + c) = ab + ac$ for all $a, b, c \in R$.	**(distributivity)**
R4	There exists an element $1 = 1_R \in R$ such that $1a = a1 = a$ for all $a \in R$.	**(identity)**

We will typically write ab instead of $a \cdot b$. The identity element of the group $(R, +)$ will be denoted 0_R, or usually just 0 (as with the usual additive notation for groups).

Definition 7.2 A ring R is **commutative** if it satisfies the following condition:

R5	$ab = ba$ for all $a, b \in R$.	**(commutativity)**

Some books omit the identity property R4 from the general definition of a ring, and give rings that do satisfy that requirement a special name, such as **unital rings**, **rings with unity**, or **rings with 1**. We will adopt the convention that rings do have multiplicative identities.

Rings satisfying only properties R1, R2 and R3 are sometimes called **nonunital rings**, **rings without 1**, or **rngs**.[1] These are interesting in

[1] Often pronounced "rung".

their own right, but beyond the scope of this module.

Time for some examples.

Example 7.3 The familiar number systems $\mathbb{Z}$ (integers), $\mathbb{Q}$ (rational numbers), $\mathbb{R}$ (real numbers) and $\mathbb{C}$ (complex numbers) all form commutative rings with the usual addition and multiplication operations.

Example 7.4 For any $n \in \mathbb{N}$, the set $\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$ forms a commutative ring with addition and multiplication modulo n.

Example 7.5 The set $\{0\}$ forms a ring with the (only possible) addition and multiplication operations $0 + 0 = 0$ and $0 \cdot 0 = 0$. This is sometimes called the **zero ring**.

Example 7.6 If R is a ring, then the set $M_n(R)$ or $M_{n\times n}(R)$ of all $n \times n$ matrices with entries in R forms a ring under the usual addition and multiplication operations.

Matrix rings are usually noncommutative: $M_n(R)$ is commutative if and only if R is the zero ring, or if R is commutative and $n = 1$.

Example 7.7 For a ring R, let $R[x]$ be the set of finite-degree polynomials with coefficients in R. Then $R[x]$ forms a ring under the usual addition and multiplication operations. In general, $R[x]$ is commutative if and only if R is.

Now for a few elementary properties.

Lemma 7.8 *Let R be a ring. Then $0a = 0 = a0$ and $(-1)a = -a = a(-1)$ for all $a \in R$.*

Proof First, note that

$$0a = (0 + 0)a = 0a + 0a$$

[2] Proposition 1.14, page 5.

by the distributivity condition. And by the cancellation law[2] in the additive group $(R, +)$ it follows that $0a = 0$. And $a0 = 0$ by a very similar argument.

Note that $-a$ denotes the additive inverse of a, and -1 denotes the additive inverse of 1. Then

$$(-1)a + 1a = ((-1) + 1)a = 0a = 0$$

[3] Lemma 1.15, page 5.

so $(-1)a = -a$ by the uniqueness of inverses in the group $(R, +)$.[3] And $a(-1) = -a$ by a very similar argument. □

Lemma 7.9 *Let R be a ring. Then R has a unique multiplicative identity element 1.*

Proof Let 1 and e be two identity elements of R. Then $1 = 1e = e$. So R has a unique identity element. □

Lemma 7.10 *Let R be a ring such that $0 = 1$. Then $R = \{0\}$.*

Proof For all $a \in R$ we have $a = a1 = a0 = 0$, so R must be the zero ring. □

7.2 *Subrings*

In Chapter 2 we studied the concept of a **subgroup**: a subset of a group that is a group in its own right. Now we will introduce and study the corresponding concept for rings.

Definition 7.11 A subset S of a ring R is a **subring** of R if it forms a ring under the same addition and multiplication operations as R, with the same identity element.

Just as with the definition of a ring, some books don't require a subring to have the same multiplicative identity element as R, or even an identity at all. Similarly to Proposition 2.3, the following result gives a method of checking whether a given subset is indeed a subring.

Proposition 7.12 *Let R be a ring, and let $S \subseteq R$ be a subset of R. Then S is a subring of R if and only if:*

(i) *$(S, +)$ is a subgroup of $(R, +)$,*
(ii) *$ab \in S$ for all $a, b \in S$, and*
(iii) *$1_R \in S$.*

Proof If S is a subring of R then all three conditions hold.

Now suppose $(S, +) \leqslant (R, +)$. Since $(R, +)$ is abelian, so is $(S, +)$ thus condition R1 in Definition 7.1 holds. Properties R2 (associativity) and R3 (distributivity) hold in S because they hold in R.

Conditions (i) and (ii) ensure that the restriction of the addition and multiplication operations of R to the subset S give valid addition and multiplication operations for S. And condition (iii) ensures that property R4 (identity) holds. Hence S is a ring in its own right, and thus a subring of R. □

Now we'll look at some examples.

Example 7.13 The set $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}$ is a subring of $\mathbb{C}$. This is the ring of **Gaussian integers**.

Example 7.14 The set $\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Z}\}$ is a subring of $\mathbb{R}$.

Example 7.15 The set

$$\left\{ \frac{a}{2^r} : a, r \in \mathbb{Z} \text{ and } r \geqslant 0 \right\}$$

is a subring of $\mathbb{Q}$.

Example 7.16 Let R be a ring, and let $\mathrm{UT}_n(R)$ and $\mathrm{LT}_n(R)$ be the sets of, respectively, upper and lower triangular $n \times n$ matrices with entries in R. These are both subrings of $M_n(R)$.

From these examples we can see that sometimes the easiest way of defining a ring is to present it as a subring of a ring we already know about. This way, we can avoid defining addition and multiplication, and prove associativity and distributivity: we just need to check closure under $+$, $-$ and $\cdot$, check that the set contains 1, and then

use Propositions 2.3 and 7.12.

The next result is a ring-theoretic analogue of Proposition 2.9.

Proposition 7.17 *Let R be a ring, and let S and T be subrings of R. Then the intersection $S \cap T$ is also a subring of R.*

Proof Since S and T are both subrings of R, then both $(S, +)$ and $(T, +)$ are additive subgroups of $(R, +)$, and by Proposition 2.9, $(S \cap T, +)$ is an additive subgroup of $(R, +)$.

Now suppose that $a, b \in S \cap T$. Then since S is a subring of R, the product $ab \in S$. And similarly $ab \in T$ since T is a subring of R. So $ab \in S \cap T$.

Finally, since S and T are both subrings of R, it follows that $1 \in S$ and $1 \in T$, so $1 \in S \cap T$. Therefore, by Proposition 7.12, $S \cap T$ is a subring of R. □

7.3 *Isomorphisms and direct products*

When we first started studying groups, we formulated and explored the concept of an **isomorphism**,[4] a structure-preserving bijection between groups. Now we want to formulate the corresponding notion for rings. Since rings are effectively additive abelian groups with some extra structure, we want an (additive abelian) group homomorphism that also respects the multiplicative structure.

[4] Definition 1.24, page 8.

Definition 7.18 Let R and S be rings. Then a function $f : R \to S$ is an **isomorphism** if:

(i) f is a bijection,
(ii) $f(r_1 + r_2) = f(r_1) + f(r_2)$ for all $r_1, r_2 \in R$, and
(iii) $f(r_1 r_2) = f(r_1) f(r_2)$ for all $r_1, r_2 \in R$.

We say that R and S are **isomorphic** if there is an isomorphism between them. If so, we write $R \cong S$.

Ring isomorphisms satisfy similar properties to group isomorphisms:

Lemma 7.19 *Let R and S be rings, and suppose that $f : R \to S$ is an isomorphism. Then:*

(i) $f(0_R) = 0_S$, *and*
(ii) $f(1_R) = 1_S$.

Proof The first of these follows from Proposition 4.2: since $f : R \to S$ is an isomorphism, we know that $f(0_R) = 0_S$.

To prove the second property, let $s \in S$ and suppose that $s = f(r)$ for some $r \in R$ (such an element r exists because f is surjective). Then

$$\begin{aligned} & sf(1_R) = f(r)f(1_R) = f(r1_R) = f(r) = s \\ \text{and} \qquad & f(1_R)s = f(1_R)f(r) = f(1_R r) = f(r) = s \end{aligned}$$

which means that $f(1_R)$ is an identity of S. It follows from Lemma 7.9 that $f(1_R) = 1_S$. □

Definition 7.20 Let R and S be rings. Their **direct product** $R\times S$ is the cartesian product

$$R\times S = \{(r,s) : r \in R, s \in S\}$$

of ordered pairs of elements from R and S, with the obvious component-wise addition and multiplication operations:

$$(r_1,s_1) + (r_2,s_2) = (r_1+r_2, s_1+s_2)$$
$$(r_1,s_1)(r_2,s_2) = (r_1r_2, s_1s_2)$$

for all $r_1, r_2 \in R$ and $s_1, s_2 \in S$.

It is straightforward to check that $R\times S$ is a ring under these operations. The multiplicative identity element is $1_{R\times S} = (1_R, 1_S)$.

Earlier,[5] we saw that the direct product $G\times H$ of two groups G and H contains a subgroup $G\times\{1_H\}$ isomorphic to G and a subgroup $\{1_G\}\times H$ isomorphic to H. The corresponding property doesn't hold in general for rings, however. The elements of the form $(r, 0_S)$ form a ring isomorphic to R, but the identity element of this ring is $(1_R, 0_S)$, whereas the identity element of $R\times S$ is $1_{R\times S} = (1_R, 1_S)$.

[5] Proposition 3.17, page 29.

The following important theorem is related to Proposition 3.20. It is usually called the *Chinese Remainder Theorem*, and its earliest known statement occurs in a mathematical treatise called *Sun Tzu Suan Ching* ('The Mathematical Classic of Sun Tzu') written sometime during the third to fifth centuries CE and attributed to a mathematician named Sun Tzu[6] (or Sunzi).[7,8] There have been suggestions that it be named after its original discoverer.

[6] 'Master Sun'.

[7] This Sun Tzu is a different person to the military strategist who wrote *The Art of War*, and lived during the sixth century BCE.

[8] Later discussions of the theorem occur in the work of the Indian mathematicians Aryabhata and Brahmagupta in the sixth and seventh centuries CE, and the *Liber Abaci* of Leonardo of Pisa (Fibonacci) in 1202 CE. The earliest known complete solution occurs in the *Shushu Chuichang* or *Mathematical Treatise in Nine Sections*, written by the Chinese mathematician Qin Jiushao in 1247 CE.

Theorem 7.21 (Chinese Remainder Theorem / Sun Tzu's Theorem) *The rings $\mathbb{Z}_m\times\mathbb{Z}_n$ and $\mathbb{Z}_{mn}$ are isomorphic if and only if m and n are coprime; that is, if* $\gcd(m,n) = 1$.

Proof If m and n are not coprime, then their least common multiple $l = \operatorname{lcm}(m,n) < mn$. The additive abelian groups $(\mathbb{Z}_m\times\mathbb{Z}_n, +)$ and $(\mathbb{Z}_{mn}, +)$ are not isomorphic, because the order of 1 in $(\mathbb{Z}_{mn}, +)$ is mn, but for any element $(a,b) \in \mathbb{Z}_m\times\mathbb{Z}_n$ we have $l(a,b) = (la, lb) = 0$, so no element of $\mathbb{Z}_m\times\mathbb{Z}_n$ has order mn, and so by Proposition 1.26 there is no isomorphism from $\mathbb{Z}_m\times\mathbb{Z}_n \to \mathbb{Z}_{mn}$.

Conversely, suppose that m and n are coprime, and let $[x]_m$ denote the residue of x modulo m. We now define $f\colon \mathbb{Z}_{mn} \to \mathbb{Z}_m\times\mathbb{Z}_n$ by $f(x) = ([x]_m, [x]_n)$.

It is clear that $f(x+y) = f(x) + f(y)$ and $f(xy) = f(x)f(y)$ for all $x, y \in \mathbb{Z}_{mn}$, so the structural conditions are satisfied.

We must now show that f is a bijection. If $f(x) = f(y)$ then $[x]_m = [y]_m$ and $[x]_n = [y]_n$, so m and n both divide $(x-y)$. But since m and n are coprime, this implies that mn divides $(x-y)$ and hence $[x]_{mn} = [y]_{mn}$. Hence f is injective. And since $|\mathbb{Z}_m\times\mathbb{Z}_n| = |\mathbb{Z}_{mn}| = mn$, f must be surjective too. Thus f is a bijective ring homomorphism, and hence an isomorphism. □

We can prove the following corollary by induction on k.

Corollary 7.22 *If $n = p_1^{n_1} \dots p_k^{n_k}$ is a decomposition of n into a product of distinct primes, then*

$$\mathbb{Z}_n \cong \mathbb{Z}_{p_1^{n_1}} \times \cdots \times \mathbb{Z}_{p_k^{n_k}}$$

as rings.

7.4 Integral domains and fields

For any two integers $a, b \in \mathbb{Z}$ we have $ab = 0$ only when either a or b is zero (or both). However, this property isn't shared by all rings: for example, the nonzero matrices $A = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ in $M_2(\mathbb{R})$ multiply to give $AB = \left[\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right] = BA$.

Definition 7.23 Let R be a ring. Then a nonzero element $a \in R$ is a **left zero divisor** if there exists a nonzero element $x \in R$ such that $ax = 0$. Similarly, a nonzero element $b \in R$ is a **right zero divisor** if there exists some nonzero element $y \in R$ such that $yb = 0$. An element that is both a left and right zero divisor is called a **two-sided zero divisor**.

In this terminology, we can say that $\mathbb{Z}$ has no zero divisors, while the matrix ring $M_2(\mathbb{R})$ does have zero divisors.

Rings that contain no zero divisors, and in which we can assume that $ab = 0$ implies that either $a = 0$ or $b = 0$ (or both), are particularly important, and we give them a special name:

Definition 7.24 An **integral domain** (or a **domain**) is a nontrivial commutative ring R that has no zero divisors. That is, if $ab = 0$ then either $a = 0$ or $b = 0$ (or both) for all $a, b \in R$.

Let's look at some examples.

Example 7.25 The rings $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are all integral domains.

Example 7.26 Let R and S be commutative rings. Their direct product $R \times S$ is not an integral domain. In particular, for any $r \in R$ and $s \in S$, the elements $(r, 0)$ and $(0, s)$ are zero divisors, because $(r, 0)(0, s) = (0, 0) = (0, s)(r, 0)$.

Example 7.27 Every subring of an integral domain is an integral domain. The reason for this is that if R is an integral domain, and S is a subring of R, then if R doesn't contain any zero divisors, neither does S.

For example, the rings $\mathbb{Z}[i]$ and $\mathbb{Z}[\sqrt{2}]$ are integral domains.

Proposition 7.28 *The ring $\mathbb{Z}_n$ is an integral domain if and only if n is prime.*

Proof If $n = 1$ then $\mathbb{Z}_n = \{0\}$ is the zero ring, which isn't a domain.

If $n = ab$ with $1 < a, b < n$, then $ab = 0$ with $a, b \neq 0$ in $\mathbb{Z}_n$, so $\mathbb{Z}_n$ is not a domain.

If n is prime, then n does not divide ab for any $0 < a, b < n$, so $ab \neq 0$ in $\mathbb{Z}_n$ and hence $\mathbb{Z}_n$ is an integral domain. □

Since a ring is an abelian group under addition, the additive cancellation laws[9] apply, but the same is not in general true for multiplication. However, the following proposition confirms that multiplicative cancellation laws hold in an integral domain:

[9] Proposition 1.14, page 5.

Proposition 7.29 *Let R be an integral domain. Suppose that $a, b, c \in R$ with $a \neq 0$ and either $ab = ac$ or $ba = ca$, then $b = c$.*

Proof Suppose that $ab = ac$. Then by the distributive law we have $ab - ac = a(b - c) = 0$. And since R is a domain with $a \neq 0$ that means that $b - c = 0$, and hence $b = c$.

The proof for $ba = ca$ is very similar. □

Every element in a group has a unique inverse, and every element in a ring has an additive inverse. But not every element in a ring need have a multiplicative inverse. Those that do are worthy of a special name:

Definition 7.30 Let R be a ring. An element $a \in R$ is a **unit** if it has a two-sided inverse under multiplication; that is, if there exists $b \in R$ such that $ab = 1 = ba$.

These invertible elements form a group:

Proposition 7.31 *Let R be a ring, and denote by R^* the set of all units in R. Then R^* forms a group under multiplication. This is called the* ***group of units*** *of the ring R.*

Proof If $a \in R^*$ is a unit, then it has an inverse a^{-1}. This inverse must also belong to R^*, because it is also invertible with inverse a. Hence every element of R^* has an inverse.

We need to check that R^* is closed under multiplication. To see this, consider two elements $a, b \in R^*$. Being units, they have inverses a^{-1} and b^{-1} in R^*. And by Lemma 1.16, ab is invertible, with inverse $b^{-1}a^{-1}$. So the restriction to R^* of the multiplication operation in R is a valid binary operation on R^*.

This operation is associative, since multiplication in R is associative.

And finally, the multiplicative identity element $1 \in R$ is its own inverse, and thus belongs to R^*. Hence R^* is a group. □

Example 7.32 The ring $\mathbb{Z}$ of integers has only two units, namely 1 and -1. So $\mathbb{Z}^* = \{1, -1\} \cong \mathbb{Z}_2$.

Example 7.33 The group $U_n = \{m \in \mathbb{Z}_n : \gcd(m, n) = 1\}$ introduced in Example 1.8 is the group of units of the ring $\mathbb{Z}_n$.

Example 7.34 In the rings $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$, every nonzero element is a unit, so we have

$$\mathbb{Q}^* = \mathbb{Q} \setminus \{0\}, \qquad \mathbb{R}^* = \mathbb{R} \setminus \{0\}, \qquad \mathbb{C}^* = \mathbb{C} \setminus \{0\}.$$

In the first and third of these examples we see the extremes of the possible outcomes. In $\mathbb{Z}$, the minimum possible number of elements (namely 1 and -1) are units, while in $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ everything but zero is a unit. This latter case is important, and we give it a couple of special names:

Definition 7.35 A nonzero ring R is said to be a **division ring** if $R^* = R \setminus \{0\}$; that is, if every nonzero element is a unit.

A commutative division ring is called a **field**.

There is a strong connection between fields and integral domains:

Proposition 7.36 *Every field is an integral domain.*

Proof Let F be a field. Suppose there exist nonzero elements $a, b \in F \setminus \{0\} = F^*$ such that $ab = 0$. Then a has a multiplicative inverse a^{-1}, so

$$b = 1b = a^{-1}ab = a^{-1}0 = 0,$$

which contradicts the assumption that $b \neq 0$. So F must be an integral domain. □

Proposition 7.37 *Every finite integral domain is a field.*

Proof Let $R = \{r_0, r_1, \ldots, r_n\}$ be a finite integral domain, with $r_0 = 0$. By the multiplicative cancellation laws in Proposition 7.29, for fixed $i > 0$, the n products $r_i r_j$ (where $1 \leqslant j \leqslant n$) are all distinct and nonzero. Since there are n possible values for these n products, they all occur exactly once. In particular, there is some j such that $r_i r_j = 1$, and hence r_i is a unit, with $r_i^{-1} = r_j$. Thus R is a field. □

Corollary 7.38 *The ring $\mathbb{Z}_n$ is a field if and only if n is prime.*

Proof This follows from Propositions 7.28 and 7.37. □

Definition 7.39 Let R be a ring. If there exists a positive integer n such that $na = 0$ for all $a \in R$, then we call the smallest such positive integer the **characteristic** of R, denoted $\mathrm{char}(R)$. If no such positive integer exists, we say that R has characteristic 0.

Example 7.40 The ring $\mathbb{Z}_n$ has characteristic n.

Example 7.41 The rings $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ have characteristic 0.

Example 7.42 The polynomial ring $R[x]$ has the same characteristic as R.

Mrs Erlynne: Ideals are dangerous things. Realities are better. They wound, but they're better.
Lady Windermere: If I lost my ideals, I should lose everything.

— Oscar Wilde (1854–1900), *Lady Windermere's Fan* (1893)

8 Ideals and Quotients

Next, we want to develop ring-theoretic concepts analogous to homomorphisms, normal subgroups and quotient groups.

8.1 Homomorphisms

First we begin with the concept of a ring homomorphism. This is very similar to the concept of a group homomorphism, and indeed we've met the bijective case already when we defined the notion of a ring isomorphism.[1] So we will define a ring homomorphism to be the more general case where we don't require bijectivity:

[1] Definition 7.18, page 66.

Definition 8.1 Let R and S be rings. A **ring homomorphism** is a function $f\colon R \to S$ that satisfies the following conditions:

(i) $f(r_1 + r_2) = f(r_1) + f(r_2)$ for all $r_1, r_2 \in R$,
(ii) $f(r_1 r_2) = f(r_1)f(r_2)$ for all $r_1, r_2 \in R$, and
(iii) $f(1_R) = 1_S$.

A (ring) **monomorphism** is an injective ring homomorphism, and a (ring) **epimorphism** is a surjective ring homomorphism.

From Proposition 4.2 it follows that $f(0_R) = 0_S$, and $f(-a) = -f(a)$ for all $a \in R$. It doesn't necessarily follow that $f(1_R) = 1_S$ from the additive and multiplicative conditions (i) and (ii) in Definition 8.1, so we incorporate it into the definition explicitly.[2]

[2] However, in some books (especially those that don't require rings to have multiplicative identity elements) ring homomorphisms aren't by default required to satisfy this condition, and ones which do are sometimes called **unital homomorphisms**.

Example 8.2 For any $n \in \mathbb{N}$, the "reduction modulo n" operation determines a ring homomorphism $f\colon \mathbb{Z} \to \mathbb{Z}_n$ given by $m \mapsto [m]_n$.

Example 8.3 Complex conjugation determines a ring homomorphism $f\colon \mathbb{C} \to \mathbb{C}$ given by $z \mapsto \bar{z}$.

More generally, a homomorphism $f\colon R \to R$ from a ring to itself is called a (ring) **endomorphism**, and an isomorphism from a ring to itself is called a (ring) **automorphism**.

Example 8.4 If $f\colon R \to S$ is a ring homomorphism, then there is an **induced homomorphism** $\overline{f}\colon R[x] \to S[x]$ defined by

$$\overline{f}(a_n x^n + \cdots + a_1 x + a_0) = f(a_n)x^n + \cdots + f(a_1)x + f(a_0).$$

We can define an induced homomorphism $\overline{f}\colon M_n(R) \to M_n(S)$ on matrix rings in a similar way.

Example 8.5 Let R be a ring, and let $R[x]$ be the ring of finite-degree polynomials with coefficients in R. Choose a fixed element $a \in R$. The map $\mathrm{ev}_a \colon R[x] \to R$ given by

$$\mathrm{ev}_a(p) = p(a)$$

for all polynomials $p \in R[x]$, is called the **evaluation map** or **evaluation homomorphism** at $a \in R$. (Important examples include the cases where $R = \mathbb{R}$ or $R = \mathbb{C}$.)

This is a ring homomorphism, since

$$\begin{aligned} ev_a(p+q) &= (p+q)(a) = p(a) + q(a) = \mathrm{ev}_a(p) + \mathrm{ev}_a(q), \\ \mathrm{ev}_a(pq) &= (pq)(a) = p(a)q(a) = \mathrm{ev}_a(p)\mathrm{ev}_a(q) \\ \text{and} \quad \mathrm{ev}_a(1_R) &= 1_R \end{aligned}$$

for all $p, q \in R[x]$.

Example 8.6 Let R be a commutative ring of prime characteristic p; that is, $pa = 0$ for all $a \in R$. Then the map $f \colon R \to R$ given by $f(a) = a^p$ is a homomorphism. The multiplicative condition

$$f(ab) = (ab)^p = a^p b^p = f(a)f(b)$$

is straightforward. The additive condition

$$f(a+b) = (a+b)^p = a^p + b^p = f(a) + f(b)$$

is less obvious, and holds because the commutativity in R implies

$$(a+b)^p = a^p + b^p + \sum_{k=1}^{p-1} \frac{p!}{k!(p-k)!} a^k b^{p-k}$$

and all the binomial coefficients in the sum are divisible by p.

Now we want to look at images and kernels of ring homomorphisms, as we did for group homomorphisms. The definition of the image of a ring homomorphism is straightforward:

Definition 8.7 Let $f \colon R \to S$ be a ring homomorphism. Then the **image** of f is

$$\mathrm{im}(f) = \{f(r) : f \in R\}.$$

The image of a group homomorphism is a subgroup of its codomain. Something very similar happens with ring homomorphisms:

Proposition 8.8 *Let $f \colon R \to S$ be a ring homomorphism. Then the image $\mathrm{im}(f)$ is a subring of the codomain S.*

Proof By Proposition 4.11, $(\mathrm{im}(f), +)$ is a subgroup of $(S, +)$. And since $f(1_R) = 1_S$, the identity $1_S \in \mathrm{im}(f)$.

Now suppose that $s_1, s_2 \in S$. Then there exist $r_1, r_2 \in R$ such that $f(r_1) = s_1$ and $f(r_2) = s_2$. And

$$s_1 s_2 = f(r_1)f(r_2) = f(r_1 r_2)$$

so $s_1 s_2 \in \mathrm{im}(f)$. Hence $\mathrm{im}(f)$ is a subring of S. □

We need to think a little bit about how to define the kernel of a

ring homomorphism $f: R \to S$. We want to consider the elements in R that map to the identity in S, but we have two identities to consider: the additive identity 0_S and the multiplicative identity 1_S. We will choose the additive identity 0_S: a ring is an abelian group with extra structure, so this way we ensure that the kernel of a ring homomorphism is effectively the kernel of an abelian group homomorphism with some extra structure.

Definition 8.9 Let $f: R \to S$ be a ring homomorphism. Then the **kernel** of f is

$$\ker(f) = \{r \in R : f(r) = 0_S\}.$$

By Proposition 4.13, a ring homomorphism is injective if and only if its kernel is trivial.

Now consider the "reduction modulo n" homomorphism $f: \mathbb{Z} \to \mathbb{Z}_n$ given by $f(m) = [m]_n$. The kernel of this map is the set $n\mathbb{Z}$ of multiples of n. This is not a subring of $\mathbb{Z}$, since it doesn't contain $1 \in \mathbb{Z}$.

So in general, the kernel of a ring homomorphism isn't necessarily a subring of its domain.[3]

[3] As noted earlier, some books don't require a ring or a subring to contain a multiplicative identity, and therefore do consider the kernel of a ring homomorphism to be a subring of its domain.

Proposition 8.10 *Let $f: R \to S$ be a ring homomorphism. Then $kr \in \ker(f)$ and $rk \in \ker(f)$ for all $r \in R$ and $k \in \ker(f)$.*

Proof If $r \in R$ and $k \in \ker(f)$, then

$$\begin{aligned} & f(kr) = f(k)f(r) = 0_S f(r) = 0_S \\ \text{and} \qquad & f(rk) = f(r)f(k) = f(r)0_S = 0_S \end{aligned}$$

so $kr \in \ker(f)$ and $rk \in \ker(f)$. □

8.2 *Ideals*

The next question we want to answer is: what is the ring-theoretic analogue of a normal subgroup? In Proposition 4.14 we found that kernels of group homomorphisms are normal subgroups, and every normal subgroup can be viewed as the kernel of some homomorphism. So we will use kernels of ring homomorphisms as the motivation for the following definition:

Definition 8.11 A subset I of a ring R is an **ideal** of R if:

I1 I is a subgroup of $(R, +)$, and
I2 $kr \in I$ and $rk \in I$ for all $r \in R$ and $k \in I$.

Condition I2 is sometimes called the **absorption condition**.

Proposition 8.10 tells us that for any ring homomorphism $f: R \to S$, the kernel $\ker(f)$ is an ideal of R.

The next proposition describes what happens if an ideal contains the multiplicative identity 1:

Proposition 8.12 *Let I be an ideal of a ring R. If $1_R \in I$ then $I = R$.*

Proof The absorption conditions I2 imply that $kr \in I$ and $rk \in I$ for all $r \in R$ and $k \in I$. In particular, setting $k = 1$, it follows that

$1r = r = r1 \in I$ for all $r \in R$, and hence $R \subseteq I$. And $I \subseteq R$ by definition, so $I = R$. □

In fact, the same thing happens if an ideal contains a unit:

Proposition 8.13 *Let I be an ideal of a ring R. If I contains a unit a, then $I = R$.*

Proof The absorption conditions I2 again imply that $ar \in I$ and $ra \in I$ for all $r \in R$. Setting $r = a^{-1}$ we see that $aa^{-1} = 1 = a^{-1}a \in I$, and then by Proposition 8.12 it follows that $I = R$. □

An important class of ideals are those generated by a single element.

Example 8.14 The kernel of the "reduction modulo n" homomorphism $f\colon \mathbb{Z} \to \mathbb{Z}_n$ is an ideal:

$$\ker(f) = n\mathbb{Z} = \{nm : m \in \mathbb{Z}\}$$

More generally, we have the following:

Definition 8.15 Let R be a ring, and suppose that $a \in R$. The **principal ideal** of R generated by a is the ideal

$$(a) = \left\{ \sum_{i=1}^{k} r_i a s_i : r_i, s_i \in R \right\}.$$

If R is commutative, then this simplifies to

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

We will only be concerned with principal ideals of commutative rings in this module.

Example 8.16 The principal ideals of $\mathbb{Z}$ are exactly those of the form $(n) = n\mathbb{Z}$ for $n \in \mathbb{N}$. We will see later that these are the only ideals of $\mathbb{Z}$.

Example 8.17 Let F be a field, and set $R = F[x]$, the polynomial ring over F. The principal ideal (x) consists of all polynomials with zero constant term; that is, those of the form

$$p = a_n x^n + \cdots + a_1 x.$$

8.3 *Quotient rings*

Possibly the most important aspect of normal subgroups is that we can use them to form quotient groups. The same is true for ideals. Since an ideal I of a ring R is a subgroup of the additive group $(R, +)$, we can consider its cosets

$$I{+}a = \{k + a : k \in R\}.$$

We know from Proposition 3.11 that these cosets form a group under addition, with the operation

$$(I{+}a) + (I{+}b) = I{+}(a{+}b)$$

for all $a, b \in R$. We just need to define a suitable multiplication operation and we get a ring:

Proposition 8.18 *Let I be an ideal of a ring R. The cosets of I form a ring under addition in the quotient group, and the multiplication operation*

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

*for all $a, b \in R$. This is the **quotient ring** R/I.*

Proof We know from Proposition 3.11 that R/I forms a group under addition, so axiom R1 in Definition 7.1 is satisfied.

We need to check that the multiplication in R/I is well-defined. To do this, suppose that $I+a = I+r$ and $I+b = I+s$. Then $(a-r)$ and $(b-s)$ belong to I, and so

$$ab = ab - as + as - rs + rs = a(b-s) + (a-r)s + rs.$$

Hence

$$ab - rs = a(b-s) + (a-r)s.$$

But by the absorption condition I2, it follows that $a(b-s) \in I$, since $(b-s) \in I$. And $(a-r)s \in I$, since $(a-r) \in I$. So $ab - rs \in I$, and hence $I+ab = I+rs$. The multiplication operation is therefore well-defined.

Properties R2 (associativity) and R3 (distributivity) follow automatically because they hold in R.

Finally, we set $1_{R/I} = I+1_R$. Then for any $I+a$ we have

$$\begin{aligned} & (I+a)(I+1_R) = I+a1_R = I+a \\ \text{and} \qquad & (I+1_R)(I+a) = I+1_Ra = I+a \end{aligned}$$

so property R4 (existence of an identity) holds, and R/I is a ring. □

Example 8.19 The quotient ring $\mathbb{Z}/(n)$ is isomorphic to $\mathbb{Z}_n$. The isomorphism $f\colon \mathbb{Z}_n \to \mathbb{Z}/(n)$ is given by $f(m) = m+(n)$.

8.4 The Isomorphism Theorems

There are ring-theoretic versions of the First,[4] Second[5] and Third Isomorphism Theorems.[6] We will state all three, but only prove the first:

[4] Theorem 4.15, page 34.
[5] Theorem 4.19, page 36.
[6] Theorem 4.21, page 37.

Theorem 8.20 (First Isomorphism Theorem) *Let $f\colon R \to S$ be a ring homomorphism with kernel I. Then $R/I \cong \mathrm{im}(f)$. More precisely, there is an isomorphism $\phi\colon R/I \to \mathrm{im}(f)$ defined by $\phi(I+a) = f(a)$ for all $a \in R$.*

Proof By the First Isomorphism Theorem for groups,[7] ϕ is a well-defined isomorphism of additive abelian groups. In particular, ϕ is a bijection. Furthermore,

[7] Theorem 4.15, page 34.

$$\phi((I+a)(I+b)) = \phi(I+ab) = f(ab) = f(a)f(b) = \phi(I+a)\phi(I+b)$$

for all $a, b \in R$. Hence ϕ is a ring isomorphism. □

We will apply this to our standard example:

Example 8.21 Let $f : \mathbb{Z} \to \mathbb{Z}_n$ be the "reduction modulo n" homomorphism. This is surjective and has kernel $\ker(f) = (n)$, so by the First Isomorphism Theorem we have $\mathbb{Z}/(n) = \mathbb{Z}/\ker(f) \cong \mathrm{im}(f) = \mathbb{Z}_n$.

The Second and Third Isomorphism Theorems are as follows:

Theorem 8.22 (Second Isomorphism Theorem) *Let R be a ring, let S be a subring of R, and let I be an ideal of R. Then*

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

Theorem 8.23 (Third Isomorphism Theorem) *Let R be a ring, and let I and J be ideals of R such that $I \subseteq J$. Then*

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

The first of these requires the following lemma:

Lemma 8.24 *Let R be a ring, let S be a subring of R, and I and ideal of R. Then*

(i) *$S+I$ is a subring of R,*
(ii) *I is an ideal of $S+I$, and*
(iii) *$S \cap I$ is an ideal of S.*

9 Domains

I never could do anything with figures, never had any talent for mathematics, never accomplished anything in my efforts at that rugged study, and to-day the only mathematics I know is multiplication, and the minute I get away up in that, as soon as I reach nine times seven . . . I've got it now. It's eighty-four. Well, I can get that far all right with a little hesitation. After that I am uncertain, and I can't manage a statistic.

— Mark Twain (Samuel Langhorne Clemens) (1835–1910), *In Aid of the Blind* (29 March 1906), *Mark Twain's Speeches* (1910) 322–332

In this last chapter, we will look at notions of divisibility in integral domains,[1] and generalise the notion of a prime or irreducible number to an arbitrary ring. We will then study three important classes of integral domains, each of which shares certain important properties with the ring $\mathbb{Z}$ of integers.

[1] Unless otherwise stated, we will be working with integral domains rather than more general rings in this chapter.

9.1 Divisibility

We'll start with the following definition:

Definition 9.1 Let $a, b \in R$ be elements of an integral domain R. We say that a **divides** b if there exists some $r \in R$ such that $b = ar$.

The following lemma draws a number of important connections between divisibility and principal ideals.

Lemma 9.2 *Let R be an integral domain. The following statements are equivalent for all $a, b \in R$:*

(i) $a|b$,
(ii) $b \in (a)$, *and*
(iii) $(b) \subseteq (a)$.

Proof To see that (i) implies (ii), suppose that $a|b$. Then $b = ar$ for some $r \in R$. Thus $b \in (a) = \{as : s \in R\}$.

To show that (ii) implies (iii), if $b \in (a)$ then $b = ar$ for some $r \in R$ and so

$$(b) = \{bt : t \in R\} = \{(ar)t : t \in R\} = \{a(rt) : t \in R\} \subseteq (a).$$

Finally, if $(b) \subseteq (a)$ then

$$(b) = \{bt : t \in R\} \subseteq \{as : s \in R\} = (a).$$

Hence $b \in (a)$ and so $b = ar$ for some $r \in R$, which confirms that (iii) implies (i). Thus all three statements are equivalent. □

We now want to consider the case where two elements are divisible by each other.

Definition 9.3 Let R be an integral domain. Two elements $a, b \in R$ are **associate** (written $a \sim b$) if both $a|b$ and $b|a$.

If we think about when this happens in our most familiar ring $\mathbb{Z}$, we can see that two integers m and n divide each other exactly when either $m = n$ or $m = -n$. That is, when one is equal to ± 1 times the

other. What is special about 1 and -1? They are the units (invertible elements) in $\mathbb{Z}$. This observation leads us to the following lemma.

Lemma 9.4 *The following statements are equivalent for any elements a and b in an integral domain R:*

(i) $a \sim b$,
(ii) $(b) = (a)$, *and*
(iii) *there exists a unit $q \in R$ with $a = qb$.*

Proof The equivalence of statements (i) and (ii) follows quickly from Lemma 9.2: if $a \sim b$ then $a|b$ and $b|a$, which is equivalent to saying that $(b) \subseteq (a)$ and $(a) \subseteq (b)$, which occurs if and only if $(b) = (a)$.

To show that (i) implies (iii), suppose first that $a = 0$. Then $a \sim b$ is equivalent to saying that $b = 0$. So now assume that $a, b \neq 0$. There exist $q, r \in R$ such that $a = qb$ and $b = ra$. Then $a = qb = q(ra) = (qr)a$, so $a(1 - qr) = 0$ and since $a \neq 0$ and R is an integral domain, it must be the case that $1 - qr = 0$, so $qr = 1$ and thus q is a unit.

Proving that (iii) implies (i) is straightforward. If $a = qb$ for some unit $q \in R$, then $a|b$. Furthermore $b = q^{-1}a$, and hence $b|a$, which means that $a \sim b$. □

Example 9.5 In $\mathbb{Z}$, the only units are ± 1 so $a \sim b$ if and only if $a = \pm b$; that is, $|a| = |b|$.

Example 9.6 Let F be a field, and consider the polynomial ring $F[x]$. The units in $F[x]$ are the nonzero constants, so $a \sim b$ if and only if $a = rb$ for some $r \in F \setminus \{0\}$.

Note also that every polynomial in $F[x]$ is associate to a unique **monic** polynomial (that is, one with leading coefficient 1). Given

$$f = a_n x^n + \cdots a_1 x + a_0 \in F[x],$$

with $a_0, \ldots, a_n \in F$ and $a_n \neq 0$, we can define a monic polynomial

$$g = x^n + \tfrac{a_{n-1}}{a_n} x^{n-1} + \cdots + \tfrac{a_1}{a_n} x + \tfrac{a_0}{a_n} \in F[x]$$

with $f = a_n g$. Hence $f \sim g$.

[2] Or **highest common factor**.

We can form the **greatest common divisor**[2] and **least common multiple** of two integers, and now we want to generalise these ideas to arbitrary domains.

Definition 9.7 Let R be an integral domain, and suppose that $a, b \in R$. A **greatest common divisor** $\gcd(a, b)$ (or **highest common factor** $\mathrm{hcf}(a, b)$) of a and b is an element d such that:

(i) $d|a$ and $d|b$, and
(ii) for any $c \in R$ with $c|a$ and $c|b$, then $c|d$.

Similarly, a **least common multiple** $\mathrm{lcm}(a, b)$ is an element $l \in R$ such that:

(i) $a|l$ and $b|l$, and
(ii) for any $m \in R$ with $a|m$ and $b|m$, then $l|m$.

We can generalise this definition in an obvious way to define the gcd or lcm of any set of elements of R.

Note also that $\gcd(0, a) = a$ and $\operatorname{lcm}(0, a) = 0$ for any $a \in R$.

Strictly speaking, greatest common divisors and least common multiples aren't unique: for example in $\mathbb{Z}$, both 2 and -2 are greatest common divisors of 4 and 6. In general, they are defined up to multiplication by a unit (in this case ± 1). However, we will sometimes abuse terminology and notation by speaking of *the* greatest common divisor of two elements and writing $\gcd(4, 6) = 2$ (or alternatively $\gcd(4, 6) = -2$).

Proposition 9.8 *Let R be an integral domain, and let $a, b \in R$. If $c, d \in R$ are greatest common divisors of a and b, then $c \sim d$. And if $l, m \in R$ are least common multiples of a and b, then $l \sim m$.*

This follows almost immediately from Definition 9.7.

Proof Since c and d are greatest common divisors of a and b, we have $c|d$ and $d|c$, so $c \sim d$.

And since l and m are least common multiples of a and b, we have $l|m$ and $m|l$, so $l \sim m$. □

Greatest common divisors and least common multiples don't necessarily exist in an arbitrary integral domain, but they do in certain types of domain (particularly Euclidean domains, unique factorisation domains, and principal ideal domains, which we will meet soon).

9.2 *Prime and irreducible elements*

Now we want to generalise the notion of a prime number to an arbitrary integral domain. There are two ways of defining prime elements, that are equivalent for integers, but not necessarily in an arbitrary domain.

One way is to say that a nonzero integer $p \neq \pm 1$ is prime if and only if whenever $p = mn$ then either m or n is equal to ± 1.

Definition 9.9 Let R be an integral domain, and suppose that $r \in R \setminus \{0\}$. Then r is **irreducible** if:

(i) r is not a unit, and
(ii) if $r = ab$ for some $a, b \in R$, then either a or b is a unit.

The other way is to say that a nonzero integer $p \neq \pm$ is prime if and only if whenever p divides mn then p divides either m or n.

Definition 9.10 Let R be an integral domain, and suppose that $r \in R \setminus \{0\}$. Then r is **prime** if:

(i) r is not a unit, and
(ii) if $r|ab$ for some $a, b \in R$, then $r|a$ or $r|b$.

In the ring $\mathbb{Z}$ these are equivalent, but this is not necessarily true in every integral domain. More precisely, in an integral domain, prime elements are irreducible, but not all irreducible elements are prime.

Proposition 9.11 *Let R be an integral domain, and suppose that $r \in R$ is prime. Then r is irreducible.*

Proof Let $r \in R$ be prime. Then by Definition 9.10, r is not a unit. Suppose that $r = ab$ for some $a, b \in R$. Then $r|r = ab$ and so we have $r|a$ or $r|b$. Without loss of generality, suppose that $r|a$. Now $a|r$ since $r = ab$, and so we have $r \sim a$, so $r = aq$ for some unit $q \in R$. Then $q = b$ by the cancellation laws[3], so b is a unit and r is irreducible. (If, on the other hand, $r|b$ then a similar argument shows that a must be a unit.) □

[3] Proposition 7.29, page 69.

The converse doesn't hold in general:

Example 9.12 Let $R = \mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} : a, b \in \mathbb{Z}\}$. Then

$$6 = 2\times3 = (1+\sqrt{-5})(1-\sqrt{-5})$$

in R. We will show that 2 is irreducible but not prime.

First, we note that 2 does not divide $1\pm\sqrt{-5}$, since if we set $2x = 1\pm\sqrt{-5}$ then this implies that $x = \frac{1}{2}\pm\frac{1}{2}\sqrt{-5}$, which doesn't belong to $R = \mathbb{Z}[\sqrt{-5}]$. Hence 2 is not prime.

We now show that 2 is irreducible. If $2 = ab$ with $a = x+y\sqrt{-5}$ and $b = s+t\sqrt{-5}$ then

$$4 = |ab|^2 = |a|^2|b|^2 = (x^2 + 5y^2)(s^2 + 5t^2).$$

Since $|a|^2, |b|^2 \in \mathbb{N}$, we have three cases to consider:

Case 1 If $|a|^2 = 1$ then $a^{-1} = x - y\sqrt{-5}$ and so a is a unit.

Case 2 If $|a|^2 = 4$ then $|b|^2 = 1$ and $b^{-1} = s - t\sqrt{-5}$ so b is a unit.

Case 3 If $|a|^2 = 2$ then we have a contradiction: there are no integers x and y with $|a|^2 = x^2 + 5y^2 = 2$. So this case can't happen.

Hence 2 is irreducible. We can show by similar arguments that 3, $(1+\sqrt{-5})$ and $(1-\sqrt{-5})$ are also irreducible in R.

9.3 *Euclidean domains*

In *MA132 Foundations* or *MA138 Sets and Numbers* you should have met the following result:

Proposition 9.13 *For any $a, b \in \mathbb{Z}$ with $b \neq 0$, there exist $q, r \in \mathbb{Z}$ such that $a = qb + r$ with $0 \leqslant r < |b|$.*

The **Euclidean Algorithm** in $\mathbb{Z}$ provides a constructive proof of this fact.

There is a similar result for polynomials over a field, using a polynomial version of the Euclidean Algorithm.

Proposition 9.14 *Let F be a field. For any polynomials $f, g \in F[x]$ with $g \neq 0$, there exist polynomials $q, r \in F[x]$ such that $f = qg + r$ and either $r = 0$ or $\deg(r) < \deg(g)$.*

Here, $\deg(f)$ is the **degree** of a nonzero polynomial f. Specifically,

for a nonzero polynomial

$$f = a_n x^n + \cdots + a_1 x + a_0$$

with $a_n \neq 0$, we define $\deg(f) = n$, with $\deg(f) = 0$ if f is a nonzero constant.[4]

[4] We will define $\deg(0) = -1$, although here we will only be concerned with nonzero polynomials.

Now we want to generalise this idea to arbitrary integral domains.

Definition 9.15 An integral domain R is a **Euclidean domain** if it admits a **norm** function $\nu\colon R \setminus \{0\} \to \mathbb{N} \cup \{0\}$ such that:

(i) $\nu(ab) \geqslant \nu(b)$ for all $a, b \in R \setminus \{0\}$, and

(ii) for all $a, b \in R$ with $b \neq 0$ there exist $q, r \in R$ such that $a = qb + r$ and either $r = 0$ or $\nu(r) < \nu(b)$.

Propositions 9.13 and 9.14 give the following two examples:

Example 9.16 The ring $\mathbb{Z}$ is a Euclidean domain with norm $\nu(a) = |a|$.

Example 9.17 Let F be a field. Then $F[x]$ is a Euclidean domain with norm $\nu(f) = \deg(f)$.

Another example concerns the Gaussian integers $\mathbb{Z}[i]$:

Example 9.18 As in Example 7.13, we set $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}$, the ring of **Gaussian integers**. We know that $\mathbb{Z}[i]$ is a subring of $\mathbb{C}$, so it is certainly an integral domain.

Let $z = x + yi$, and set $\nu(z) = |z|^2 = x^2 + y^2$. We want to show that this satisfies the requirements in Definition 9.15.

Firstly, condition (i) holds since $|zw| = |z||w|$ for all $z, w \in \mathbb{C}$.

To check condition (ii), let $a, b \in \mathbb{Z}[i]$ with $b \neq 0$. Then $a/b = x + yi$ for some $x, y \in \mathbb{Q}$. Choose $x_0, y_0 \in \mathbb{Z}$ with $|x - x_0| \leqslant \frac{1}{2}$ and $|y - y_0| \leqslant \frac{1}{2}$. Then

$$a = b(x + yi) = b(x_0 + y_0 i) + b((x - x_0) + (y - y_0)i) = qb + r$$

where $q = (x_0 + y_0 i) \in \mathbb{Z}[i]$ and $r = b((x - x_0) + (y - y_0)i)$. Since $r = a - qb$ we have $r \in \mathbb{Z}[i]$ and

$$\nu(r) = \nu(b)\nu((x - x_0) + (y - y_0)i) \leqslant \nu(b)(\tfrac{1}{4} + \tfrac{1}{4}) < \nu(b)$$

so condition (ii) holds, and hence $\mathbb{Z}[i]$ is a Euclidean domain.

9.4 *Principal ideal domains*

In Definition 8.15 we introduced the notion of a **principal ideal**: one generated by a single element of the ring. Some integral domains have no non-principal ideals:

Definition 9.19 An integral domain R is called a **principal ideal domain** (or **PID**) if every ideal of R is principal.

Proposition 9.20 *Every Euclidean domain is a PID.*

Proof Let R be a Euclidean domain and suppose that ν is a norm

for R. The trivial ideal $(0) = \{0\}$ and the full ring $R = (1)$ are both principal.

Suppose that I is a proper, nontrivial ideal of R and choose $b \in I \setminus \{0\}$ such that $\nu(b)$ is as small as possible. Then $(b) \subseteq I$.

Now let $a \in I$ be some arbitrary element of I. Because R is a Euclidean domain, we can find $q, r \in R$ such that $a = qb + r$ with either $r = 0$ or $\nu(r) < \nu(b)$. If $r \neq 0$ then $r = a - qb$, which belongs to I because $a \in I$, and $qb \in I$ by the absorption condition. But this is a contradiction because $r \in I$ and $\nu(r) < \nu(b)$, whereas we chose b such that $\nu(b)$ was minimal over I. So $r = 0$ and $a = qb$, which means that $a \in (b)$. Hence $I \subseteq (b)$, so $I = (b)$ is principal and R is therefore a PID. □

This proposition together with Example 9.16 yields the following corollary:

Corollary 9.21 *The ring $\mathbb{Z}$ is a PID.*

And Example 9.17 implies the following:

Corollary 9.22 *If F is a field, then the polynomial ring $F[x]$ is a PID.*

Not every PID is a Euclidean domain, but it is fairly difficult to find an example to demonstrate this. Probably the simplest example is $\mathbb{Z}[\alpha] = \{a + b\alpha : a, b \in \mathbb{Z}\}$ where $\alpha = \frac{1}{2}(1 + \sqrt{-19})$, but a proof of this fact is a little beyond the scope of this module.

Proposition 9.23 *If R is a PID then $\operatorname{lcm}(a, b)$ and $\gcd(a, b)$ exist for any $a, b \in R$. Furthermore, there exist $r, s \in R$ such that $\gcd(a, b) = ra + sb$.*

Proof By Lemma 8.24, $I = (a)+(b) = \{ra + sb : r, s \in R\}$ is an ideal of R, and since R is a PID, this ideal I is principal. Hence $I = (d)$ for some $d \in R$.

Similarly, $(a) \cap (b)$ is an ideal, so it must be equal to (l) for some $l \in R$.

We claim that d is a greatest common divisor and l is a least common multiple of a and b. Indeed, $(a) \subseteq (d) \supseteq (b)$, and whenever $(a) \subseteq (c) \supseteq (b)$ it follows that $(c) \supseteq (a)+(b) = (d)$.

Similarly, $(a) \supseteq (l) \subseteq (b)$, and whenever $(a) \supseteq (m) \subseteq (b)$ it follows that $(m) \subseteq (a) \cap (b) = (l)$. □

Proposition 9.24 *If R is a PID, then every irreducible element of R is prime.*

Proof Let $r \in R$ be irreducible. Then by definition r is not a unit. Suppose that $r|ab$ for some $a, b \in R$. Then by Proposition 9.23, an element $c = \gcd(a, b)$ exists.

Then $r = ct$ for some $t \in R$. Since r is irreducible, either c or t is a unit. We consider these cases separately.

Case 1 If t is a unit then $r \sim c$ and $c|a$, so $r|a$ and hence r is prime.

Case 2 If c is a unit, then by Proposition 9.23 we have $c = xa + yr$ for some $x, y \in R$. Multiplying both sides of this by b gives $cb = xab + yrb$. Now $r|ab$ and clearly $r|yrb$, so $r|cb$. This means that $ru = cb$ for some $u \in R$. But if c is a unit, then we can multiply by

c^{-1} to see that $r|b$, which implies again that r is prime. □

9.5 *Unique factorisation domains*

The Fundamental Theorem of Arithmetic says that every nonzero integer apart from ± 1 can be factorised as a product of primes, and that factorisation is unique up to the order of the factors and multiplication by ± 1. We want to generalise this.

Definition 9.25 An integral domain R is a **factorisation domain** (FD) if each non-unit element $a \in R \setminus \{0\}$ can be factorised as a product of irreducible elements $x = r_1 r_2 \dots r_n$.

Definition 9.26 A factorisation domain R is a **unique factorisation domain** (UFD) if, for each non-unit element $R \setminus \{0\}$ and any two factorisations

$$x = r_1 r_2 \dots r_n = s_1 s_2 \dots s_m,$$

where $r_1, \dots, r_n$ and $s_1, \dots, s_m$ are irreducible, then $m = n$ and there exists $\sigma \in S_n$ such that $r_i \sim s_{\sigma(i)}$ for $1 \leqslant i \leqslant n$.

Proposition 9.27 *Let R be a UFD. Then every irreducible element of R is prime.*

Proof Let $x \in R$ be irreducible. Then by definition x is not a unit. Now suppose that $x|ab$, and factorise $a = r_1 \dots r_k$ and $b = r_{k+1} \dots r_n$. Thus we have a factorisation $ab = r_1 \dots r_n$. On the other hand $ab = xy$ for some $y \in R$. Factorise $y = s_1 \dots s_m$, and then we have another factorisation $ab = x s_1 \dots s_m$. Since R is a UFD, x is associate to r_i for some i. If $i \leqslant k$ then $x|a$, and if $i > k$ then $x|b$. Hence x is prime. □

Proposition 9.28 *Every PID is a FD.*

Proof Let R be a PID, and suppose that $x \in R \setminus \{0\}$ is not a unit.

Suppose that x can't be factorised as a product of irreducible elements. Then

$$X = \{x \in R \setminus \{0\} : x \text{ is not a unit and can't be factorised}\}$$

is nonempty.

Let $x \in X$. Then x can't be irreducible, so we can write $x = yz$ for some $y, z \in R$, neither of which are units. If we could factorise both y and z into irreducibles, then we could do the same for x. So at least one of them, say y, can't be factorised, and hence $y \in X$. Since z is not a unit, $x \not| y$. By Lemma 9.2, $(x) \subset (y)$.

Since this is true for all $x \in X$, we can obtain an infinite sequence of elements $x_i \in X$ such that

$$(x_1) \subset (x_2) \subset \dots \subset (x_n) \subset (x_{n+1}) \subset \dots$$

where all the inclusions are proper.

Let $I = \bigcup_{i=1}^{\infty} (x_i)$. Then I is an ideal. To see this, let $r, s \in I$. Then for some $m, n \geqslant 0$ we have $r \in (x_m)$ and $s \in (x_n)$, and assuming

without loss of generality that $n \geqslant m$, we have $r, s \in (x_n)$, so $r+s \in (x_n) \subseteq I$. The other conditions can be checked similarly.

Since R is a PID, $I = (d)$ for some $d \in R$. Then $d \in (x_n)$ for some n. This implies that $I = (d) \subseteq (x_n)$. But this contradicts the assumption that (x_n) is properly contained in $(x_{n+1}) \subset I$.

Hence our initial assumption, that x can't be factorised into irreducible elements, was false. Therefore every nonzero, non-unit element of R can be factorised as a product of finitely many irreducible elements, and so R is a factorisation domain. □

Proposition 9.29 *Let R be a factorisation domain in which every irreducible element is prime. Ten R is a UFD.*

Corollary 9.30 *Every PID is a UFD.*

Example 9.31 The ring $\mathbb{Z}[\sqrt{-5}]$ is a factorisation domain but not a UFD.

Example 9.32 The ring $\mathbb{Z}[x]$ is a UFD but not a PID.

Proposition 9.33 *Let R be a UFD. Then for any $a, b \in R$, there exist elements $d = \gcd(a, b)$ and $l = \operatorname{lcm}(a, b)$.*