MT2505
Abstract Algebra

Martyn Quick

Version 1.1, 9th January 2024

University of St Andrews

Contents

Motivation: What is an algebraic structure? **1**

1 Rings and Fields **3**
Basic consequences of the axioms . . . 6
Matrix rings . . . 7
Polynomial rings . . . 10
Fields . . . 14
Subrings . . . 15

2 Greatest Common Divisors and the Euclidean Algorithm **17**
The Extended Euclidean Algorithm . . . 20

3 Equivalence Relations **23**

4 Congruences and Modular Arithmetic **29**

5 Groups **37**
Cayley tables . . . 40
Non-examples . . . 42
Basic properties of groups . . . 43
Powers of elements . . . 46
Use of additive notation . . . 47

6 Permutations and Symmetric Groups **49**
Functions . . . 49
Symmetric groups . . . 51
Symmetric groups of finite degree . . . 53
Cycle notation . . . 55

7 Isometries **61**
Isometries of the plane and of subsets of the plane . . . 61
Dihedral groups . . . 64

8 Subgroups **67**
Cyclic subgroups . . . 70
Alternating groups . . . 74

9 Lagrange's Theorem **79**
Cosets . . . 79

10 Homomorphisms, Normal Subgroups and Quotients Groups **83**

Homomorphisms 83
Isomorphisms 85
Kernels and images 86
Normal subgroups 88
Quotient groups 90
The First Isomorphism Theorem 92

Bibliography **95**

Versions **97**

Motivation: What is an algebraic structure?

The purpose of this introductory chapter is to motivate the content of the module and specifically to consider the question "What is algebra?" To attempt to answer this question, we shall list various examples that will be typical algebraic structures. Algebra could then be viewed as the study of mathematical structures of the type we present here.

Perhaps previously you have seen the word "algebra" most commonly used in the context of polynomials. Indeed, you may have done a number of things with polynomials, for example:

- We might view a polynomial as a function $f\colon \mathbb{R} \to \mathbb{R}$ and then perform differentiation, integration and perhaps sketch the graph.
- We might solve polynomial equations. This could be achieved by factorizing the polynomial.
- We perform algebraic operations where we add and/or multiply polynomials.

This first of these tasks is not typical of the study of algebra. Although calculus can sometimes turn up in algebra, it is not central to the topic. The other two types of task, however, are directly related to the sort of thing that we do in algebra. Indeed, these two tasks are linked since to factorize a polynomial is to express it as a product of polynomials of smaller degree. A general description of what we study in algebra is *algebraic operations* such as the addition and multiplication that can be performed with polynomials.

Other examples of structures where such algebraic operations occur include:

- The set of $n \times n$ matrices with entries that are real numbers. We can add and multiply such matrices to produce further matrices of the same size.
- The set $\mathcal{T}_X$ of functions $f\colon X \to X$ from a fixed set X to itself. We can compose two such functions to produce another function in $\mathcal{T}_X$:

  $$(f \circ g)(x) = f(g(x)).$$

  Composition is consequently an operation on the set $\mathcal{T}_X$ of all functions $X \to X$.
- We can add, multiply and subtract integers.
- We can add, multiply and subtract real numbers. We can divide two real numbers *only* when we divide by a non-zero real number.

Some of these examples might not seem the sort of thing that you were expecting to appear in the topic of "algebra" but they give some idea what the study of abstract algebra is concerned with. We wish to study structures that possess operations that behave a bit like addition or multiplication and then understand common themes that arise. By doing this in reasonable generality, the hope is that these common themes will become transparent (and this course — and those that follow — is intended to reveal these themes).

We will finish this introductory chapter by making the core definition that will enable us to formulate the key ideas of the course.

Definition 0.1 Let A be a set. A *binary operation* on A is a function

$$A \times A \to A$$

defined on the set $A \times A = \{ (a, b) \mid a, b \in A \}$ of pairs of points in A and that takes values in A.

Although we have defined a binary operation as a function, we will usually use a notation that suggests a similarity to addition or multiplication when denoting a binary operation. Thus, one example is that we might write $a * b$ to denote the image of a pair (a, b) under a binary operation. Other common notations are to use *multiplicative notation* where we write ab for the effect of combining a and b under the given binary operation and *additive notation* where we use the notation $a + b$. We shall view a binary operation as a way to combine two elements of the set A so as to result in another element of A (possibly one of the original two back again).

Example 0.2 The following defines a binary operation on the set $\mathbb{R}$ of real numbers:

$$a * b = 0 \qquad \text{for all } a, b \in \mathbb{R}.$$

This is a binary operation on $\mathbb{R}$: It is not a very interesting binary operation.

We shall actually be interested in binary operations that have natural properties and which arise in interesting examples. We shall meet examples of such operations and their properties in our first main chapter.

Chapter 1

Rings and Fields

In this chapter, we shall introduce a type of algebraic structure that is studied in mathematics, namely the *ring*. This type of structure is intended to reflect the behaviour that one observes with numbers, polynomials, and matrices. These three types of mathematical object have the following common properties:

- one can add objects of the same type, and
- one can multiply objects of the same type.

Thus in the definition of "ring" we shall refer to two binary operations and they will be denoted to look like addition and multiplication.

As we observed in Example 0.2, it is possible to define binary operations that have strange but not very useful behaviours. In the case of numbers, polynomials and matrices, this bad behaviour is actually not what we observe with their addition and multiplication operations. These operations all behave quite naturally. For example, we observe that

$$(a+b)+c = a+(b+c)$$
$$(ab)c = a(bc)$$
$$a(b+c) = ab+ac$$

when a, b and c are real numbers. Moreover, similar formulae hold when we perform addition and multiplication with polynomials or matrices.

The definition we now formulate reflects these observations:

Definition 1.1 A *ring* is a set R together with two binary operations

$$(a,b) \mapsto a+b \qquad \text{and} \qquad (a,b) \mapsto ab,$$

that we shall call *addition* and *multiplication*, respectively, such that the following conditions all hold:

A1: $a+b = b+a$ for all $a, b \in R$,

A2: $(a+b)+c = a+(b+c)$ for all $a, b, c \in R$,

A3: there is some element $0 \in R$, called *zero*, such that $0+a = a+0 = a$ for all $a \in R$,

A4: for every $a \in R$, there is some element $-a \in R$, called the *negative* of a, such that $a+(-a) = (-a)+a = 0$,

M2: $(ab)c = a(bc)$ for all $a, b, c \in R$,

D: $a(b+c) = ab+ac$ and $(a+b)c = ac+bc$ for all $a, b, c \in R$.

Comments:

(i) The conditions listed above are often called the "axioms" of a ring. The labelling that we give for them is not universally adhered to. It has been chosen to coincide with that used in the module *MT3505 Algebra: Rings and Fields*. Note that A1–A4 are conditions relating only to the addition in a ring, while M2 refers only to the multiplication. Condition D is consequently significant as it says something about how addition and multiplication interact. If there were no condition involving both operations then this type of algebraic structure would not be so interesting: We would have two binary operations but they would not have any interaction with each other.

(ii) Note that Condition A3 ensures that every ring contains at least one element, namely the zero. Hence the underlying set R *must* be non-empty.

(iii) Usually we do not distinguish explicitly between an algebraic structure and its underlying set R. Thus we shall say

"R is a ring"

to mean: R is a set upon which addition and multiplication operations are defined such that Conditions A1–A4, M2 and D all hold.

(iv) Some textbooks (and some lecture notes) explicitly state two additional conditions

$$a + b \in R \qquad \text{for all } a, b \in R$$

and

$$ab \in R \qquad \text{for all } a, b \in R.$$

Note, however, that to say "+ is a binary operation" is to say that

$$\begin{aligned} R \times R &\to R \\ (a, b) &\mapsto a + b \end{aligned}$$

is a function taking values in R. Thus the above two conditions are built into our requirement that addition and multiplication are binary operations on the given set R in Definition 1.1.

We have additional names that refer to the conditions appearing in Definition 1.1. These are the following terms:

Definition 1.2 Let A be any set and $*$ be a binary operation on A.

(i) We say that the binary operation $*$ is *commutative* if

$$a * b = b * a \qquad \text{for all } a, b \in A.$$

(ii) We say that the binary operation $*$ is *associative* if

$$(a * b) * c = a * (b * c) \qquad \text{for all } a, b, c \in A.$$

(iii) An *identity* for the binary operation $*$ is an element $e \in A$ such that

$$e * a = a * e = a \qquad \text{for all } a \in A.$$

(iv) If there is an identity e for the binary operation $*$, then an *inverse* for $a \in A$ is an element $b \in A$ such that

$$a * b = b * a = e.$$

Thus, a ring is a set endowed with two binary operations, addition and multiplication, satisfying the following properties:

- addition is commutative and associative,
- there is an additive identity (called zero),
- every element has an additive inverse (its negative),
- multiplication is associative,
- the *distributive laws* (D) hold.

How to verify that a mathematical object is a ring:

- Say what the underlying set R is.
- Say what the addition and multiplication operations are *and* verify these actually are binary operations: that is, check $a + b \in R$ and $ab \in R$ for all $a, b \in R$.
- Verify Conditions A1–A4, M2 and D. In the case of A3 and A4, this will usually involving stating explicitly what the zero and what the negatives are.

Example 1.3 The set $\mathbb{Z}$ of all integers forms a ring under the usual addition and multiplication operations. The verification is as follows:

- The sum of two integers is always an integer and the product of two integers is again always an integer.
- Conditions A1, A2, M2 and D are standard properties of the usual arithmetic operations.
- $0 \in \mathbb{Z}$ is an integer and it satisfies the requirement that $0 + a = a + 0 = a$ for all $a \in \mathbb{Z}$ appearing in Condition A3.
- If $a \in \mathbb{Z}$, then $-a$ is also an integer and this satisfies $a + (-a) = (-a) + a = 0$, so Condition A4 holds.

Hence the set $\mathbb{Z}$ of integers is indeed a ring under the usual addition and multiplication operations.

This example illustrates how one verifies many easy examples are rings: namely those with which we are essentially familiar from the arithmetic we learnt to perform when we were much younger. In these cases, Conditions A1–A4, M2 and D are usually things that we have relied upon for years and so it feels perhaps unusual to actually make explicit reference to them. In the same vein, we record some other easy examples:

Example 1.4 (i) The set $\mathbb{Q}$ of rational numbers is a ring under the usual addition and multiplication operations.

(ii) The set $\mathbb{R}$ of real numbers is a ring under the usual addition and multiplication operations.

(iii) The set $\mathbb{C}$ of complex numbers is a ring under the usual addition and multiplication operations.

The important point to note at this stage is that these are *not* the only examples. The reason to introduce the term "ring" is because it covers both the familiar examples of $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$, but also many other less familiar examples. Additional examples of rings are presented later in this chapter.

Basic consequences of the axioms

The fact that the addition in a ring is both associative and commutative means that we can manipulate sums of elements in the same way that we are familiar with. For example, if R is a ring and $a_1, a_2, \ldots, a_n$ are elements of R, then the sum

$$a_1 + a_2 + \cdots + a_n \tag{1.1}$$

should in theory be provided together with some bracketing since according to the definition $+$ can only be applied to a pair of elements at a time. However, the fact that addition is associative means that any two choices of bracketing actually yields the same answer as we can use the Associative Law A2 repeatedly to convert one to the other. This fact, in its general form, is given below as Theorem 1.5. A consequence is that we can safely omit the brackets in an expression such as (1.1) and we shall usually choose to do so.

Similarly the fact that addition is commutative means that we can rearrange the terms in the above sum (1.1). Consequently, it makes sense to write

$$\sum_{i=1}^{n} a_i$$

for the above sum as it does not matter the order in which we sum the terms nor how this sum is bracketed. We are permitted to manipulate these sums in the way with which we are familiar since Conditions A1 and A2 hold and in the examples that follow we shall do so.

For the record, we state the following theorem. The standard methods of proof is to proceed by induction on n. This is, however, omitted here since it is not the most enlightening of arguments and, apart from its use, does not give that much insight into algebra.

Theorem 1.5 (Generalized Associative Law) *Let A be a set and $*$ be an associative binary operation on A. If $a_1, a_2, \ldots, a_n$ are elements of A, then the element*

$$a_1 * a_2 * \cdots * a_n$$

is uniquely determined irrespective of how this expression is bracketed.

Matrix rings

Definition 1.6 Let R be any ring and n be a positive integer. An $n \times n$ *matrix* over R is an array consisting of n rows and n columns whose entries are selected from the ring R:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

where $a_{ij} \in R$ for all i and j. We shall abbreviate this by writing $A = [a_{ij}]$ to indicate that the (i, j)th entry of A is the element a_{ij}.

We define $\mathrm{M}_n(R)$ to be the set of all $n \times n$ matrices over R and define addition and multiplication of matrices as follows: If $A = [a_{ij}]$ and $B = [b_{ij}]$ are $n \times n$ matrices over R, then

$$A + B = [a_{ij} + b_{ij}]$$
$$AB = [c_{ij}]$$

where

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

We call $\mathrm{M}_n(R)$ a *matrix ring* over R.

So to add two $n \times n$ matrices A and B, we simply add the corresponding entries of each matrix. To multiply A and B, the (i, j)th entry is equal to the sum of the values obtained when we multiply each entry of the ith row of A by the corresponding jth column of B. Both operations will be familiar to students from their previous studies (arising, for example, in *MT1002 Mathematics* and *MT2501 Linear Mathematics*).

We have called $\mathrm{M}_n(R)$ a "matrix ring" and so one should actually verify that this is justified; that is, that the collection of $n \times n$ matrices with entries from R is indeed a ring in the sense of Definition 1.1. The full verification appears in these lecture notes, but some steps will be omitted in the lectures.

Theorem 1.7 *Let R be a ring and n be a positive integer. Then the matrix ring $\mathrm{M}_n(R)$ is indeed a ring with respect to the addition and multiplication given in Definition 1.6.*

Proof: Our definition of addition and multiplication of two $n \times n$ matrices involves adding and multiplying entries to insert into new matrices. The results are always matrices in $\mathrm{M}_n(R)$ and so addition and multiplication are binary operations on the set of $n \times n$ matrices over R. We must verify that the Conditions A1–A4, M2 and D in the definition of a ring. [To save time, some of these will be omitted during lectures.]

Let $A = [a_{ij}]$, $B = [b_{ij}]$ and $C = [c_{ij}]$ be arbitrary $n \times n$ matrices over R.

A1: By definition,

$$A + B = [a_{ij} + b_{ij}] \qquad \text{and} \qquad B + A = [b_{ij} + a_{ij}];$$

that is, the (i, j)th entry of $A + B$ is $a_{ij} + b_{ij}$ and that of $B + A$ is $b_{ij} + a_{ij}$. However, here we are adding elements of the original ring R and we know that $a_{ij} + b_{ij} = b_{ij} + a_{ij}$ since R itself satisfies Condition A1. Hence the entries of $A + B$ and $B + A$ are the same, so $A + B = B + A$.

A2: This is similar. We know that R is a ring so satisfies Condition A2, so $a+(b+c) = (a+b)+c$ for all $a, b, c \in R$. We apply this in the following calculation to the (i,j)th entry of the matrix arising:

$$\begin{aligned}(A+B)+C &= \big([a_{ij}]+[b_{ij}]\big)+[c_{ij}]\\ &= [a_{ij}+b_{ij}]+[c_{ij}]\\ &= [(a_{ij}+b_{ij})+c_{ij}]\\ &= [a_{ij}+(b_{ij}+c_{ij})]\\ &= [a_{ij}]+[b_{ij}+c_{ij}]\\ &= [a_{ij}]+\big([b_{ij}]+[c_{ij}]\big)\\ &= A+(B+C).\end{aligned}$$

Hence Condition A2 holds in $\mathrm{M}_n(R)$.

A3: Let $\mathbf{0}$ denote the $n \times n$ matrix all of whose entries are 0 (the zero of the ring R). Since $0+a = a+0 = a$ for all $a \in R$, we now calculate

$$\begin{aligned}\mathbf{0}+A &= \mathbf{0}+[a_{ij}]\\ &= [0+a_{ij}]\\ &= [a_{ij}] = A\end{aligned}$$

and similarly $A+\mathbf{0} = A$. Hence $\mathbf{0}$ is a zero in $\mathrm{M}_n(R)$.

A4: Let us write $-A$ for the matrix whose entries are the negatives of the entries of A; that is, $-A = [-a_{ij}]$. Then

$$A+(-A) = [a_{ij}]+[-a_{ij}] = [a_{ij}+(-a_{ij})] = [0] = \mathbf{0}.$$

Similarly $(-A)+A = \mathbf{0}$. This shows Condition A4 holds in $\mathrm{M}_n(R)$.

M2: We shall denote the (i,j)th entry of the product AB by $(AB)_{ij}$. Recall this is given by the formula

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}.$$

Similar formulae will be used for the (i,j)th entry of other products of matrices. Consequently, the (i,j)th entry of the product $(AB)C$ is:

$$\begin{aligned}\big((AB)C\big)_{ij} &= \sum_{k=1}^{n}(AB)_{ik}c_{kj}\\ &= \sum_{k=1}^{n}\Big(\sum_{\ell=1}^{n}a_{i\ell}b_{\ell k}\Big)c_{kj}\\ &= \sum_{k=1}^{n}\sum_{\ell=1}^{n}(a_{i\ell}b_{\ell k})c_{kj} \qquad \text{(since Condition D holds in the ring } R\text{)}\\ &= \sum_{k=1}^{n}\sum_{\ell=1}^{n}a_{i\ell}(b_{\ell k}c_{kj}) \qquad \text{(since multiplication is associative in } R \text{ (M2))}\\ &= \sum_{\ell=1}^{n}\sum_{k=1}^{n}a_{i\ell}(b_{\ell k}c_{kj})\end{aligned}$$

$$= \sum_{\ell=1}^{n} a_{i\ell} \left(\sum_{k=1}^{n} b_{\ell k} c_{kj} \right) \quad \text{(again using the distributive law (D) in } R\text{)}$$
$$= \sum_{\ell=1}^{n} a_{i\ell} (BC)_{\ell j}$$
$$= \big(A(BC)\big)_{ij}.$$

It follows that $(AB)C = A(BC)$ since their entries are the same; that is, multiplication is associative in $\mathrm{M}_n(R)$.

D: Finally we establish that the distributive law holds in $\mathrm{M}_n(R)$. The (i, j)th entry of $A(B + C)$ is

$$\sum_{k=1}^{n} a_{ik}(b_{kj} + c_{kj}) = \sum_{k=1}^{n} (a_{ik} b_{kj} + a_{ik} c_{kj}) \qquad \text{(as the distributive law holds in } R\text{)}$$
$$= \sum_{k=1}^{n} a_{ik} b_{kj} + \sum_{k=1}^{n} a_{ik} c_{kj} \quad \text{(using A1 and A2 for addition in } R\text{)}$$
$$= (AB)_{ij} + (AC)_{ij};$$

that is, this equals the (i, j)th entry of $AB + AC$. We conclude that $A(B + C) = AB + AC$, as required.

These calculations demonstrate that the matrix ring $\mathrm{M}_n(R)$ is indeed a ring. □

Note that we have explicitly depended upon the fact that R is a ring (each axiom was used at some point, often multiple times, within the arguments) when demonstrating that the matrix ring over R is also a ring. We have established Theorem 1.7 over an arbitrary ring R, which means that we know that all of

$$\mathrm{M}_n(\mathbb{Z}), \qquad \mathrm{M}_n(\mathbb{Q}), \qquad \mathrm{M}_n(\mathbb{R}), \qquad \mathrm{M}_n(\mathbb{C})$$

are all themselves examples of rings. We could have dealt with each of these matrix rings individually, but by establishing the proposition for an arbitrary ring R means that we have got them all in one go. It also means that if R is a more complicated, much less familiar ring that we find at some point in the future, then the matrix ring $\mathrm{M}_n(R)$ is definitely also a ring without us having to repeat the work of Theorem 1.7 again.

We can make another comment about the axioms of a ring in Definition 1.1. Note that we have not assumed that multiplication is commutative in the definition. Indeed, if we consider multiplying 2×2 matrices, say over $\mathbb{Z}$, then, for example:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$
$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

that is,

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

Hence matrix multiplication is not, in general, commutative.

We give a special name to rings where the multiplication is commutative:

Definition 1.8 A *commutative ring* R is a ring that satisfies the additional condition:

M1: $ab = ba$ for all $a, b \in R$.

Example 1.9 (i) We know that multiplication in our familiar number systems is commutative. Thus, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are examples of commutative rings.

(ii) The above calculation shows that the matrix rings $\mathrm{M}_2(\mathbb{Z})$, $\mathrm{M}_2(\mathbb{Q})$, $\mathrm{M}_2(\mathbb{R})$ and $\mathrm{M}_2(\mathbb{C})$ are *not* commutative rings. The same idea shows that, for $n \geqslant 2$, none of the matrix rings $\mathrm{M}_n(\mathbb{Z})$, $\mathrm{M}_n(\mathbb{Q})$, $\mathrm{M}_n(\mathbb{R})$ and $\mathrm{M}_n(\mathbb{C})$ are commutative rings.

Polynomial rings

Definition 1.10 Let R be any ring. A *polynomial* over R is an expression of the form

$$f(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n$$

where n is a non-negative integer ($n \geqslant 0$) and $a_0, a_1, \ldots, a_n \in R$. The symbol X is called an *indeterminate*. If $f(X) \neq 0$ (that is, not all coefficients are 0) and we choose the expression for $f(X)$ so that $a_n \neq 0$, then we say that $f(X)$ has *degree* n.

We shall view two polynomials with indeterminate X as the same if they have the same coefficients, but one issue does arise. We wish to view the polynomials

$$a_0 + a_1X + a_2X^2 + \cdots + a_nX^n$$

and

$$a_0 + a_1X + a_2X^2 + \cdots + a_nX^n + 0X^{n+1} + \cdots + 0X^N$$

as being the same; that is, padding with lots of terms with 0 as coefficient should not make any difference. Consequently, suppose that

$$f(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n$$
$$g(X) = b_0 + b_1X + b_2X^2 + \cdots + b_mX^m$$

are two polynomials with indeterminate X and coefficients from the same ring R. Then we say $f(X)$ and $g(X)$ are *equal* (that is, $f(X) = g(X)$) if $a_k = b_k$ for all k *except* that additionally one polynomial has further terms all with 0 as coefficient.

This issue about padding using terms with 0 coefficients is one of the main complications when working with polynomials. Allenby's textbook [1] gets round this issue by treating polynomials as infinite sequences where from some point all subsequent entries are 0. (Thus he works with the sequence of coefficients in a polynomial, rather than the polynomial itself.) This is mathematically quite clean, but it feels we have moved some distance from the intuitive idea behind polynomials. Here we shall follow the usual route of working with polynomials as defined above as this fits with our already existing experience. One needs to keep track of "padding with 0 coefficient terms," but hopefully this will not prove to be too much of an obstacle for understanding.

To simplify notation, we shall often write

$$f(X) = \sum a_kX^k$$

to denote the polynomial $f(X) = a_0 + a_1X + \cdots + a_nX^n$. This notation always denotes a *finite* sum with the understanding that any terms with 0 as coefficient can be inserted or omitted without changing the element.

Having discussed what is meant by polynomials, let us now perform algebra with them.

Definition 1.11 Let R be a ring. The *polynomial ring* with indeterminate X over R is denoted by $R[X]$ and is the set of all polynomials in X with coefficients in R and is endowed with the following binary operations: If $f(X) = \sum a_kX^k$ and $g(X) = \sum b_kX^k$, we define

Addition: $f(X) + g(X) = \sum(a_k + b_k)X^k$;

Multiplication: $f(X)\,g(X) = \sum c_kX^k$, where

$$c_k = a_0b_k + a_1b_{k-1} + \cdots + a_kb_0 = \sum_{i=0}^{k} a_ib_{k-i}.$$

The formula for addition of polynomials is relatively straightforward: we are simply adding the coefficients of corresponding terms in each polynomial. To motivate the formula for multiplication, observe that if we multiply two bracketed expressions

$$(a_0 + a_1X + \cdots + a_mX^m)(b_0 + b_1X + \cdots + b_nX^n) \tag{1.2}$$

then the coefficient in front of X^k in the product will be

$$c_k = a_0b_k + a_1b_{k-1} + \cdots + a_kb_0.$$

Note here that we are implicitly making use of the potential to pad the polynomials $f(X)$ and $g(X)$ with terms having 0 coefficient to make the above formulae make sense. For example, if $f(X)$ has degree m and $g(X)$ has degree n, then in the product in (1.2), the coefficient c_{m+n} is actually given by

$$c_{m+n} = a_mb_n$$

since $a_i = 0$ for $i > m$ and $b_j = 0$ for $j > n$ in the product.

We have called $R[X]$ a "polynomial ring." As before, one should therefore verify that it is indeed a ring. However, polynomials will arise only rarely in this module so although the full details appear in the lecture notes, they will be omitted in the lectures.

Theorem 1.12 *Let R be a ring. Then the polynomial ring $R[X]$ is indeed a ring with respect to the addition and multiplication given in Definition 1.11.*

Proof: [This proof will be omitted during lectures.] Our definition of addition and multiplication specifies the coefficients as being built from adding and multiplying coefficients from the ingredients. Thus our addition and multiplication are binary operations on the set of polynomials over R. We must verify that the Conditions A1–A4, M2 and D in the definition of a ring.

Let $f(X) = \sum a_kX^k$, $g(X) = \sum b_kX^k$ and $h(X) = \sum c_kX^k$ be arbitrary polynomials in the indeterminate X over the ring R.

A1: We use the fact that addition is commutative in the ring R: $a + b = b + a$ for all $a, b \in R$. We apply this to the coefficients appearing in the following calculation:

$$\begin{aligned} f(X) + g(X) &= \sum a_k X^k + \sum b_k X^k \\ &= \sum (a_k + b_k) X^k \\ &= \sum (b_k + a_k) X^k \\ &= g(X) + f(X) \end{aligned}$$

This verifies that addition is commutative in $R[X]$.

A2: We use the fact that addition is associative in the ring R: $(a+b)+c = a+(b+c)$ for all $a, b, c \in R$. We apply this to the coefficients appearing in the following calculation:

$$\begin{aligned} \big(f(X) + g(X)\big) + h(X) &= \left(\sum a_k X^k + \sum b_k X^k\right) + \sum c_k X^k \\ &= \sum (a_k + b_k) X^k + \sum c_k X^k \\ &= \sum \big((a_k + b_k) + c_k\big) X^k \\ &= \sum \big(a_k + (b_k + c_k)\big) X^k \\ &= \sum a_k X^k + \sum (b_k + c_k) X^k \\ &= \sum a_k X^k + \left(\sum b_k X^k + \sum c_k X^k\right) \\ &= f(X) + \big(g(X) + h(X)\big) \end{aligned}$$

This verifies that addition is associative in $R[X]$.

A3: We shall duplicate use of notation by also writing 0 for the polynomial with a single term, namely the constant term with coefficient 0. Since we can pad with any number of terms with 0 as coefficient, this is also equal to

$$0 = 0 + 0X + 0X^2 + \cdots + 0X^n$$

for any choice of $n \geqslant 0$. Then, for $f(X)$ as above,

$$\begin{aligned} 0 + f(X) &= \sum 0X^k + \sum a_i X^k \\ &= \sum (0 + a_k) X^k \\ &= \sum a_k X^k = f(X). \end{aligned}$$

Similarly $f(X) + 0 = f(X)$. This shows that this polynomial 0 is a zero in $R[X]$.

A4: Given $f(X)$ as above, we write $-f(X)$ for the polynomial whose coefficients are $-a_k$:

$$-f(X) = \sum (-a_k) X^k.$$

Then

$$\begin{aligned} f(X) + \big(-f(X)\big) &= \sum a_k X^k + \sum (-a_k) X^k \\ &= \sum \big(a_k + (-a_k)\big) X^k \end{aligned}$$

$$= \sum 0X^k = 0$$

(the zero polynomial appearing in our verification of A3). Similarly $\big(-f(X)\big) + f(X) = 0$. This shows Condition A4 holds in $R[X]$.

M2: To verify associativity of the multiplication in $R[X]$ is a bit cumbersome. We can simplify the argument a little by observing that the formula for multiplication is

$$f(X)g(X) = \sum_m \Bigg(\sum_{\substack{(i,j) \text{ with} \\ i+j=m}} a_i b_j \Bigg) X^m;$$

that is, the X^m-coefficient is the sum of all products $a_i b_j$ where $i + j = m$. The calculation is then as follows:

$$\begin{aligned}
\big(f(X)g(X)\big)h(X) &= \left[\left(\sum_i a_i X^i \right) \left(\sum_j b_j X^j \right) \right] \left(\sum_k c_k X^k \right) \\
&= \left(\sum_m \Bigg(\sum_{\substack{(i,j) \text{ with} \\ i+j=m}} a_i b_j \Bigg) X^m \right) \left(\sum_k c_k X^k \right) \\
&= \sum_n \left(\sum_{m=0}^{n} \Bigg(\sum_{\substack{(i,j) \text{ with} \\ i+j=m}} a_i b_j \Bigg) c_{n-m} \right) X^n \\
&= \sum_n \Bigg(\sum_{\substack{(i,j,k) \text{ with} \\ i+j+k=n}} a_i b_j c_k \Bigg) X^n \\
&\qquad \text{(using associativity and distributivity of multiplication in } R\text{)} \\
&= \sum_n \left(\sum_{i=0}^{n} a_i \Bigg(\sum_{\substack{(j,k) \text{ with} \\ j+k=n-i}} b_j c_k \Bigg) \right) X^n \\
&= \left(\sum_i a_i X^i \right) \left(\sum_m \Bigg(\sum_{\substack{(j,k) \text{ with} \\ j+k=m}} b_j c_k \Bigg) X^m \right) \\
&= \left(\sum_i a_i X^i \right) \left[\left(\sum_j b_j X^j \right) \left(\sum_k c_k X^k \right) \right] \\
&= f(X)\big(g(X)h(X)\big).
\end{aligned}$$

This shows that Condition M2 holds in $R[X]$.

D: Finally we demonstrate that the distributive law holds in $R[X]$ (and this is more straightforward):

$$\begin{aligned}
f(X)\big(g(X) + h(X)\big) &= \left(\sum_i a_i X^i \right) \left[\left(\sum_j b_j X^j \right) + \left(\sum_k c_k X^k \right) \right] \\
&= \left(\sum_i a_i X^i \right) \left(\sum_j (b_j + c_j) X^j \right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_i \Big(\sum_{k=0}^{i} a_k (b_{i-k} + c_{i-k}) \Big) X^i \\
&= \sum_i \Big(\sum_{k=0}^{i} (a_k b_{i-k} + a_k c_{i-k}) \Big) X^i \\
&= \sum_i \Big(\sum_{k=0}^{i} a_k b_{i-k} \Big) X^i + \sum_i \Big(\sum_{k=0}^{i} a_k c_{i-k} \Big) X^i \\
&= \Big(\sum_i a_i X^i \Big) \Big(\sum_j b_j X^j \Big) + \Big(\sum_i a_i X^i \Big) \Big(\sum_j c_j X^j \Big) \\
&= f(X)g(X) + f(X)h(X).
\end{aligned}
$$

This shows that the polynomial ring $R[X]$ is indeed a ring, as claimed. □

As observed with matrix rings, it now follows that the sets of polynomials with integer, rational, real and complex coefficients (that is, $\mathbb{Z}[X]$, $\mathbb{Q}[X]$, $\mathbb{R}[X]$ and $\mathbb{C}[X]$) are rings under the addition and multiplication we have defined.

Fields

We have commented earlier that division in the real numbers $\mathbb{R}$ can only be permitted when we do not attempt to divide by 0. Consequently, division is *not* a binary operation on $\mathbb{R}$. There is a way to place division as a concept within the algebraic framework that we are presenting. This is usually done in the context of what is known as a field:

Definition 1.13 A *field* is a commutative ring F (that is, it has addition and multiplication defined upon it such that Conditions A1–A4, M1, M2 and D all hold) with two additional properties:

M3: there is an element $1 \in F$ with $1 \neq 0$ such that $1a = a1 = a$ for all $a \in F$;

M4: for every $a \in F$ with $a \neq 0$, there is some element $a^{-1} \in F$ such that $aa^{-1} = a^{-1}a = 1$.

Thus a field is a commutative ring that has a multiplicative identity 1 with $1 \neq 0$ and such that every non-zero element has a multiplicative inverse. The majority of students taking this module will have already met the term "field", either having taken the module *MT2501 Linear Mathematics* or be taking that in parallel with this one.

Example 1.14 (i) From our familiarity with number systems, we know that Conditions M3 and M4 hold in the rational numbers $\mathbb{Q}$, real numbers $\mathbb{R}$ and complex numbers $\mathbb{C}$. Consequently, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are our typical examples of fields.

(ii) If a is any integer, then $2a$ is an even number and so, in particular, $2a \neq 1$ for all $a \in \mathbb{Z}$. Hence there is no multiplicative inverse for 2 in $\mathbb{Z}$ and we conclude that the ring of integers $\mathbb{Z}$ is not a field.

Subrings

This is only a brief introduction to the topic of rings. They are considered in much more detail in the Honours module *MT3505 Algebra: Rings & Fields*. The last thing we mention is the concept of a *subring*:

Definition 1.15 Let R be a ring and S be a subset of R that forms a ring under the same operation as defined on R. We then say that S is a *subring* of R.

Example 1.16 (i) The ring of integers $\mathbb{Z}$ is a subring of $\mathbb{Q}$, which in turn is a subring of $\mathbb{R}$, which in turn is a subring of $\mathbb{C}$. They are subsets ($\mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$) and the operations are the same: addition and multiplication of numbers.

(ii) For a fixed n, we have the following chain of subrings of matrix rings:

$$\mathrm{M}_n(\mathbb{Z}) \subseteq \mathrm{M}_n(\mathbb{Q}) \subseteq \mathrm{M}_n(\mathbb{R}) \subseteq \mathrm{M}_n(\mathbb{C}).$$

(iii) For a fixed indeterminate X, the following polynomial rings are subrings of each other:

$$\mathbb{Z}[X] \subseteq \mathbb{Q}[X] \subseteq \mathbb{R}[X] \subseteq \mathbb{C}[X].$$

In later chapters, we shall introduce another type of algebraic structure, the *group*, and then explore in detail what we mean by a *subgroup*. For a fuller study of rings, the same thing should be done, but that is a topic for another module, namely *MT3505*.

Chapter 2

Greatest Common Divisors and the Euclidean Algorithm

The material we present in this chapter concerns certain properties of the integers. It can, however, be placed in a more general setting: There is a special type of ring called a *Euclidean domain* where the same behaviour occurs. The study of Euclidean domains is found in a later module (see *MT3505 Algebra: Rings & Fields*), but here we present sufficient information concerning division of integers for what we shall need later in this one. Some of this material may well have been encountered before (for example, in the module *MT1003 Pure & Applied Mathematics* or in your studies before arriving at St Andrews). Sufficient information will be included here to ensure that everyone taking the module has access to what is used later.

Definition 2.1 Let $a, b \in \mathbb{Z}$.

(i) We say that a *divides* b if $b = ac$ for some $c \in \mathbb{Z}$. We also then say that a is a *divisor* of b. We write $a \mid b$ to indicate that a divides b.

(ii) Suppose that at least one of a and b is non-zero. The *greatest common divisor* of a and b is the largest integer d which divides both a and b. We write $\gcd(a, b)$ to denote the greatest common divisor of a and b.

(iii) If it is the case that $\gcd(a, b) = 1$, then we say that a and b are *coprime*.

Comments:

(i) The symbol $a \mid b$ is a statement about the way that a and b are related. It is not equal to a number and the "value" it takes is either "True" or "False." In particular, note that $a \mid b$ is not the same thing as the quotient a/b. The latter is a fraction, that is, an element of the rationals $\mathbb{Q}$, while $a \mid b$ is "True" when a is a divisor of b and is "False" when it is not.

(ii) Note that $0 = a \times 0$ for all integers a. Therefore every integer divides 0. This is why we assume that at least one of a and b is non-zero when defining the greatest common divisor since there is no largest divisor of 0. However, if a is any non-zero integer, then any divisor d of a satisfies $d \leqslant |a|$. Hence there is a greatest common divisor of a and b when a and b are not both zero.

(iii) The divisors of b and $-b$ coincide, since if $b = ac$, then $-b = a(-c)$. Similarly if a divides b, then so does $-a$. Consequently,

$$\gcd(-a, b) = \gcd(a, -b) = \gcd(a, b)$$

whenever a and b are not both zero. In view of this, we shall work entirely with *positive* integers in this chapter.

(iv) For small integers, it is possible to calculate the greatest common divisor simply by listing the divisors of the two integers concerned and selecting the largest. For example,

$$\gcd(2, 5) = 1$$
$$\gcd(3, 9) = 3$$
$$\gcd(15, 20) = 5.$$

(v) In particular, 2 and 5 are coprime integers. When a and b are coprime integers, it means that the only positive divisor of a and b is 1.

The following is a key observation about the greatest common divisor of two integers. It will be used at various points during this module and, in particular, it is a key idea that will enable us to construct a particular finite field (see Theorem 4.10 below).

Theorem 2.2 (Bézout's Identity[1]) *Let a and b be integers, at least one of which is non-zero. Then the greatest common divisor of a and b is the smallest positive integer d that can be written in the form*

$$d = ua + vb$$

for some integers $u, v \in \mathbb{Z}$.

Corollary 2.3 *If a and b are coprime integers, then there exist integers $u, v \in \mathbb{Z}$ such that $ua + vb = 1$.*

Proof: Let $d = \gcd(a, b)$ (which we have observed is defined because a and b are not both zero). There do exist positive integers in the given form and so we can define d_1 to be the *smallest* positive integer such that $d_1 = ua + vb$ for some $u, v \in \mathbb{Z}$. We shall show that $d = d_1$.

On the one hand, since d divides both a and b, we can write $a = q_1 d$ and $b = q_2 d$, so

$$d_1 = (uq_1 + vq_2)d$$

is a multiple of d. Hence $d_1 \geqslant d$.

On the other hand, let us divide a by d_1 to yield a quotient and a remainder: $a = qd_1 + r$ where $0 \leqslant r < d_1$. Then

$$r = a - qd_1 = a - q(ua + vb) = (1 - qu)a + (-qv)b.$$

Hence r can be written in the same form as d_1. Since d_1 is the smallest positive integer that can be written in this form, it must be the case that $r = 0$. Therefore d_1 divides a. Applying the same argument to b, we deduce that d_1 divides b also. Finally, as d is the *greatest* common divisor of a and b, we conclude $d_1 \leqslant d$.

Hence $d = d_1$, which establishes the theorem. The corollary follows immediately since, by definition, $d = 1$ when a and b are coprime. □

[1] Named after the French mathematician *Étienne Bézout* (1730–1783).

The above proof indicates a process that can be used to calculate the greatest common divisor of two positive integers a and b. It is particularly useful when a and b are relatively large and we cannot immediately spot the divisors of the two integers, or when we wish to determine the Bézout coefficients u and v appearing in the theorem (and this is something that we shall often intend to do).

Algorithm 2.4 (Euclidean Algorithm)

Input: Two integers a and b with $a \geqslant b > 0$.

Output: The greatest common divisor $\gcd(a, b)$.

Method:

- *Step 1:* Define $a_1 = a$ and $b_1 = b$.
  Divide a_1 by b_1 to find quotient and remainder:
  $$a_1 = q_1 b_1 + r_1 \qquad \text{where } 0 \leqslant r_1 < b_1.$$
- *Step n:* Define $a_n = b_{n-1}$ and $b_n = r_{n-1}$ arising from the previous step.
  Divide a_n by b_n to find quotient and remainder:
  $$a_n = q_n b_n + r_n \qquad \text{where } 0 \leqslant r_n < b_n.$$
- Repeat until $r_k = 0$.
- The last non-zero remainder r_{k-1} is $\gcd(a, b)$.

Example 2.5 *Compute the greatest common divisor of* 76 *and* 92.

We shall present a more challenging example towards the end of the chapter. There we shall employ a more convenient way to present the calculation. In the following solution, we shall merely follow the steps in Algorithm, relying on the fact that the numbers involved are relatively small.

Solution: We apply the steps of the Euclidean Algorithm:

Step 1: Take $a_1 = a = 92$ and $b_1 = b = 76$. Then $92 = 1 \times 76 + 16$, so $r_1 = 16$.

Step 2: Take $a_2 = b_1 = 76$ and $b_2 = r_1 = 16$. Then $76 = 4 \times 16 + 12$, so $r_2 = 12$.

Step 3: Take $a_3 = b_2 = 16$ and $b_3 = r_2 = 12$. Then $16 = 1 \times 12 + 4$, so $r_3 = 4$.

Step 4: Take $a_4 = b_3 = 12$ and $b_4 = r_3 = 4$. Then $12 = 3 \times 4 + 0$, so $r_4 = 0$.

We have completed the application of the Euclidean Algorithm, so

$$\gcd(76, 92) = \gcd(b, a) = r_3 = 4. \qquad \square$$

Proof that the Euclidean Algorithm Works: Since at each stage $b_n = r_{n-1} < b_{n-1}$, the sequence of integers $b_n \geqslant 0$ must eventually reach 0. Hence the Euclidean Algorithm will stop after a finite number of steps.

Furthermore, if $a = qb + r$ with $a \neq 0$, then $\gcd(a, b) = \gcd(b, r)$ (see Question 2 on Problem Sheet II). Hence

$$\gcd(a, b) = \gcd(a_1, b_1) = \gcd(a_2, b_2) = \cdots = \gcd(a_k, b_k)$$

and, since $a_k = q_k b_k$ as $r_k = 0$, we deduce

$$\gcd(a, b) = \gcd(a_k, b_k) = b_k = r_{k-1},$$

as claimed. $\square$

The Extended Euclidean Algorithm

We finish our discussion of the Euclidean Algorithm by demonstrating a method to present the calculations that also keeps track of the Bézout coefficients u and v from Theorem 2.2 directly. In this process, we implement exactly the same steps as in Algorithm 2.4 but the presentation is more compact and we express each term arising in the form $ua + vb$.

The process is as follows. First create a table including the value of a and b that we are working with:

$$\begin{array}{r|l} a = \ldots & \\ \hline & \ldots = b \end{array}$$

where the two dotted entries are replaced by the numbers with which we are actually working. Having performed some stages in the calculation, we shall be in the situation where the bottom line of the table will either have an empty space in the left- or right-hand column. The entry above the empty space will express some integer in the form $ua + vb$ and the other entry in the bottom line will express some integer in the form $u'a + v'b$; that is, the table will have the form:

$$\begin{array}{r|l} \vdots & \vdots \\ \hline ua + vb = x & \vdots \\ & \hline & y = u'a + v'b \end{array}$$

(Here the symbols x, y, u, v, u' and v' will be replaced by specific integers and perhaps the left- and right-hand columns will be interchanged.) The next stage is to divide x by y to find a quotient q. Multiply everything in the filled right-hand entry by q, write this in the empty left-hand entry and then subtract to express the remainder $r = x - qy$ in the form $u''a + v''b$:

$$\begin{array}{rl|l} & \vdots & \vdots \\ \hline ua & + \; vb = \; x & \vdots \\ qu'a & + qv'b = qy & y = u'a + v'b \\ \hline (u - qu')a + (v - qv')b = & \; r & \end{array}$$

We now repeat these steps until we reach remainder 0. The Euclidean Algorithm then tells us that the last non-zero remainder is the greatest common divisor d of the original a and b and we can read off its expression as $d = ua + vb$ from the table.

We shall illustrate this process by computing the greatest common divisor of a particular pair of positive integers.

Example 2.6 In this example, we start with $a = 1232$ and $b = 546$. We then construct the following table. (This is done by using the steps described above, but it is probably most easily understood by watching it live during the lecture!)

$$\begin{array}{rr|ll} a & = 1232 & & \\ & 2b = 1092 & 546 = & b \\ \hline a - 2b = & 140 & 420 = & 3a - 6b \\ -3a + 7b = & 126 & \multicolumn{2}{l}{126 = -3a + 7b} \\ \hline 4a - 9b = & \mathbf{14} & 126 & \\ & & \hline & 0 \end{array}$$

The last non-zero remainder is 14 and hence

$$\gcd(1232, 546) = 14 = 4a - 9b.$$

(There was no need to express 126 in the form $u'a + v'b$ at the last stage of the above calculation since this would not be used in a future step.)

Chapter 3

Equivalence Relations

In this chapter, we shall introduce some terminology that can be found throughout mathematics. This will be used during this module but also in many of the Honours modules that follow.

Definition 3.1 Let A be a set. A *relation* on A is some rule, that we shall denote by $a \sim b$ here, that specifies when two elements a and b are related under the rule.

Example 3.2 We have actually met lots of examples of relations already:

(i) "Divides," denoted $|$, is a relation on the set $\mathbb{Z}$ of integers. In Definition 2.1, we defined $a \mid b$ when a divides b (that is, when $b = ac$ for some $c \in \mathbb{Z}$).

(ii) "Less than," denoted $<$, is a relation on the set $\mathbb{Z}$ (as well as on $\mathbb{Q}$ and on $\mathbb{R}$). We write $a < b$ when a is less than b.

(iii) If A is any set, "equals," denoted $=$, is a relation on A. Here $a = b$ when a and b are the same element.

(iv) If A is any set, the *universal relation* is defined by $a \sim b$ for *all* choices of a and b in A.

Note that a relation is, as we defined, just a rule that says that two elements a and b in some set. It does not take a value in the set, but rather, if anything, take the value "True" or "False." We shall usually write $a \sim b$ when the rule is True (that is, when a and b satisfy the rule) and we write $a \nsim b$ when the rule is False (that is, when a and b do not satisfy the rule). Thus, for the examples of relations that we have just listed the following hold:

(i) $2 \mid 4$, $6 \mid 120$, $3 \nmid 8$, $4 \nmid 2$.

(ii) $2 < 4$, $6 < 120$, $8 \not< 3$, $4 \not< 2$.

(iii) For equals $(=)$ on the set $\mathbb{Z}$, $2 = 2$, $4 = 4$, $2 \neq 4$, $4 \neq 2$.

(iv) For the universal relation on the set $\mathbb{Z}$, $2 \sim 4$, $6 \sim 120$, $8 \sim 3$, $4 \sim 2$. Indeed for the universal relation on any set, there are *no examples* of a and b with $a \nsim b$.

One method that could be used to specify a relation $\sim$ on a set A is to simply list all the pairs (a, b) of elements from A that satisfy $a \sim b$. This means that every relation can be defined by specifying a subset R of the set $A \times A$ of all pairs of elements:

$$R = \{\, (a, b) \mid a \sim b \,\}$$

With this definition of R, the following holds:

$$a \sim b \qquad \text{if and only if} \qquad (a, b) \in R.$$

Many textbooks formally define a relation as being any subset of $A \times A$. Via the above formula, this is equivalent to the one given in Definition 3.1.

Many different types of relation occur in mathematics. For example, one important type of relation is called a *partial order*. In this chapter, we will concentrate on what is known as an *equivalence relation*. Such a relation satisfies three specific properties:

Definition 3.3 Let $\sim$ be a relation defined on a set A.

(i) We say that $\sim$ is *reflexive* (**R**) if $a \sim a$ holds for all $a \in A$.

(ii) We say that $\sim$ is *symmetric* when the following condition holds:

S: if $a \sim b$ holds from some $a, b \in A$, then also $b \sim a$ holds.

(iii) We say that $\sim$ is *transitive* when the following condition holds:

T: if $a \sim b$ and $b \sim c$ both hold for some $a, b, c \in A$, then also $a \sim c$ holds.

(iv) We say that $\sim$ is an *equivalence relation* if it is reflexive, symmetric and transitive.

Thus an equivalence relation is one that satisfies all three of the conditions (**R**), (**S**) and (**T**) hold.

Comment: Note that both the symmetry (**S**) and transitivity (**T**) conditions are statements of the form "if ... then ...". To verify that a particular relation $\sim$ is symmetric, the argument should have roughly the following structure:

> Let a and b be elements of the set A and assume that they satisfy the condition $a \sim b$. Then use whatever mathematics you have access to and (most likely) explicit reference to the fact that you know $a \sim b$ to then deduce also that $b \sim a$.

A common error that is observed in some students' work is to present an argument that finishes with the statement along the lines of "... so $a \sim b$ and $b \sim a$." This is not correct and fails to adequately reflect the "if ... then ..." nature of the symmetry (**S**) condition.

Example 3.4 (i) Consider the relation $|$ ("divides") on the integers. If $a \in \mathbb{Z}$, then $a = 1 \times a$, which shows that $a \mid a$. Hence $|$ is reflexive (**R**).

Suppose $a, b, c \in \mathbb{Z}$ and that $a \mid b$ and $b \mid c$. The first of these tells us that $b = ax$ for some $x \in \mathbb{Z}$, while the second tells us that $c = by$ for some $y \in \mathbb{Z}$. Then

$$c = (ax)y = a(xy)$$

which tells us that $a \mid c$. Hence $|$ is transitive (**T**).

However, $|$ is not symmetric. To see this condition fails, we simply have to find a choice of $a, b \in \mathbb{Z}$ such that $a \mid b$ but for which $b \nmid a$. For example, we know that $2 \mid 4$ but $4 \nmid 2$ (since $4 = 2 \times 2$ but there is no *integer* x satisfying $2 = 4x$).

In particular, $|$ is *not* an equivalence relation. (In fact, "divides" is an example of the type of relation called a *partial order*.)

(ii) Let A be any set and consider the relation $=$ ("equals"). If a is any element of A, then $a = a$ holds (since that is what equals means!) and hence $=$ is reflexive (**R**).

If $a, b \in A$ are such that $a = b$, then a and b are the same element of A (we just haven't happened to have labelled them with the letter) and so $b = a$. Thus $=$ is symmetric (**S**).

If $a, b, c \in A$ are such that $a = b$ and $b = c$, then it must be that $a = c$ also (as all three are the same element). This shows that $=$ is transitive (**T**).

Hence "equals" is an equivalence relation. (Indeed, the concept of equivalence relation can be viewed as a generalization of the concept of "equals." One should interpret any particular equivalence relation $\sim$ as saying that something about a and b is the same when they are related under $\sim$. In the case of "equals," *everything* about a and b is the same when $a = b$.)

(iii) Let $\sim$ be the universal relation on the set A. Then $a \sim b$ for all $a, b \in A$.

Now if a is any element of A, then $a \sim a$ (as every pair of elements of A is related). Hence $\sim$ is reflexive (**R**).

Let $a, b \in A$ and suppose $a \sim b$. (This assumption actually tells us nothing beyond the fact that a and b are elements of A.) It then follows that $b \sim a$ (since every pair of elements are related under $\sim$). Hence $\sim$ is symmetric (**S**).

Let $a, b, c \in A$ and suppose $a \sim b$ and $b \sim c$. Then $a \sim c$ (because every pair of elements are related under $\sim$). Hence $\sim$ is transitive (**T**).

This shows that the universal relation $\sim$ on a set A is an equivalence relation.

(iv) Define a relation $\sim$ on the set of integers $\mathbb{Z}$ by the following rule:

$$a \sim b \qquad \text{if } b - a \text{ is an even number.} \tag{3.1}$$

Thus $3 \sim 7$ because $7 - 3 = 4$ is even. On the other hand, $4 \nsim 1$ because $1 - 4 = -3$ is odd, not even. We shall show that $\sim$ is an equivalence relation on $\mathbb{Z}$. (We shall generalize this example in the next chapter.)

(**R**): If $a \in \mathbb{Z}$, then $a - a = 0$, which is even. Hence $a \sim a$ for all $a \in \mathbb{Z}$. Thus $\sim$ is reflexive.

(**S**): Let $a, b \in \mathbb{Z}$ and suppose that $a \sim b$. This means that $b - a$ is even, say $b - a = 2x$ for some integer x. Then $a - b = -(b - a) = 2(-x)$ is also even, which shows that $b \sim a$. Hence $\sim$ is symmetric.

(**T**): Let $a, b, c \in \mathbb{Z}$ and suppose that $a \sim b$ and $b \sim c$. This means that both $b - a$ and $c - b$ are even, say $b - a = 2x$ and $c - b = 2y$ for some integers x and y. Then

$$c - a = (c - b) + (b - a) = 2y + 2x = 2(x + y)$$

is even and hence $a \sim c$. Hence $\sim$ is transitive.

Thus the relation $\sim$ defined by (3.1) is an equivalence relation.

(v) Let $A = \{\, (a, b) \mid a, b \in \mathbb{Z},\ b \neq 0 \,\}$, the set of ordered pairs of integers of which the second is always non-zero. Define a relation on A by the following rule:

$$(a, b) \sim (c, d) \qquad \text{if } ad = bc. \tag{3.2}$$

We shall show that $\sim$ is an equivalence relation on A.

(**R**): Let $(a, b) \in A$. Then $ab = ba$, so $(a, b) \sim (a, b)$ according to the rule (3.2). Hence $\sim$ is reflexive.

(**S**): Let $(a, b), (c, d) \in A$ and suppose $(a, b) \sim (c, d)$. This means that $ad = bc$. Since multiplication in the integers is commutative, it follows $cb = da$ and hence $(c, d) \sim (a, b)$. It follows that $\sim$ is symmetric.

(**T**): Let $(a, b), (c, d), (e, f) \in A$ and suppose that $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$. These mean that $ad = bc$ and $cf = de$. Now

$$afd = (ad)f = (bc)f = b(cf) = b(de) = bed.$$

However, d is non-zero, by assumption, and therefore we can divide by it and conclude $af = be$. Hence $(a, b) \sim (e, f)$. This shows that $\sim$ is transitive.

Therefore $\sim$, as defined in (3.2), is indeed an equivalence relation on A.

This set of examples gives us some indication that we might expect to see equivalence relations appear in many places in mathematics. We have met two relatively trivial examples ("equals" and the universal relation), but we have two further examples. One will be generalized further in the next chapter, while the final example (Example 3.4(v)) has a link to how one might view rational numbers. Note that under this relation $\sim$

$$(a, b) \sim (c, d) \qquad \Longleftrightarrow \qquad ad = bc \qquad \Longleftrightarrow \qquad \frac{a}{b} = \frac{c}{d}.$$

Hence this last equivalence relation is about the different ways one might express a rational number as the quotient of a pair of integers.

We shall now develop further the theory about equivalence relations so that we can use this technology in future mathematics.

Definition 3.5 Let $\sim$ be an equivalence relation on a set A. If a is an element of A, the *equivalence class* of a (with respect to $\sim$) is

$$[a] = \{\, b \in A \mid a \sim b \,\};$$

that is, $[a]$ is the subset of A consisting of all elements that are related to a under $\sim$.

Note: Note that $[a]$ denotes a particular set, even though the main part of this symbol is a lowercase letter (which is what we usually use for elements of a set).

The basic properties of equivalence classes are the following:

Theorem 3.6 *Let $\sim$ be an equivalence relation on a set A. Then*

(i) *$a \in [a]$ for every $a \in A$;*

(ii) *A is the union of all the equivalence classes of $\sim$;*

(iii) *if $a, b \in A$, then either $[a] = [b]$ or $[a] \cap [b] = \varnothing$.*

Part (iii) of the theorem says that two equivalence classes (for the given equivalence relation $\sim$) are either equal or are disjoint; that is, *distinct* equivalence classes are disjoint; i.e., have no elements in common.

PROOF: (i) Let $a \in A$. Since $\sim$ is reflexive, we know that $a \sim a$. This tells us that $a \in [a]$.

(ii) Every element of A lies in at least one equivalence class by part (i). Hence A is the union of the equivalence classes.

(iii) Let $a, b \in A$ and suppose that $[a] \cap [b] \neq \varnothing$. This means that there is at least one element, say c, belonging to both $[a]$ and $[b]$. By definition then

$$a \sim c \qquad \text{and} \qquad b \sim c.$$

Now use the fact that $\sim$ is symmetric to deduce that $c \sim b$. Then since $a \sim c$ and $c \sim b$, we deduce that $a \sim b$ by the transitivity of $\sim$. Symmetry of $\sim$ then also implies $b \sim a$.

We shall now use this observation to deduce that $[a] = [b]$. We do this by showing that every element of $[a]$ is also an element of $[y]$ and *vice versa*.

First let $x \in [a]$. This means that $a \sim x$. We then can apply transitivity of $\sim$ to the pair of observations $b \sim a$ and $a \sim x$ to deduce $b \sim x$. Hence $x \in [b]$. Thus every element of $[a]$ is also in $[b]$; that is, $[a] \subseteq [b]$.

Conversely let $x \in [b]$, so that $b \sim x$. We have already noted $a \sim b$ and so transitivity of $\sim$ implies that $a \sim x$. Hence $x \in [a]$. We have now shown that every element of $[b]$ is also in $[a]$; that is, $[b] \subseteq [a]$.

In conclusion, we have shown that if $[a] \cap [b] \neq \varnothing$, then $[a] = [b]$. This means that any pair of equivalence classes either satisfy $[a] = [b]$ or $[a] \cap [b] = \varnothing$. □

We record two sets of observations about what we have observed. This first follows immediately since part (ii) of the theorem says that the set is the union of the equivalence classes and part (iii) says that distinct equivalence classes are disjoint:

Corollary 3.7 *Let $\sim$ be an equivalence relation on a set A. Then A is the disjoint union of the equivalnce classes of $\sim$.* □

Corollary 3.8 *Let $\sim$ be an equivalence relation on a set A and let $a, b \in A$. Then*

(i) *$[a] = [b]$ if and only if $a \sim b$;*

(ii) *$[a] \cap [b] = \varnothing$ if and only if $a \nsim b$.*

PROOF: (i) If $[a] = [b]$, then $b \in [b] = [a]$ by part (i) of Theorem 3.6. Hence $a \sim b$ by definition of $[a]$.

Conversely if $a \sim b$, then b lies in $[a]$ while it also lies in $[b]$ by part (i) of Theorem 3.6. Then $b \in [a] \cap [b]$ and so $[a] = [b]$ by part (iii) of the Theorem.

(ii) This follows by use of part (iii) of Theorem 3.6. Indeed

$$a \nsim b \qquad \Longleftrightarrow \qquad [a] \neq [b] \qquad \Longleftrightarrow \qquad [a] \cap [b] = \varnothing$$

by what we have already established. □

We have another name for the situation where a set is expressed as a disjoint union of a collection of subsets:

Definition 3.9 Let A be a set and let $\mathcal{P} = \{ B_i \mid i \in I \}$ be a collection of non-empty subsets of A. (We would usually call I an *index set*. It can be any suitable choice to index the members of $\mathcal{P}$. Common choices include the positive integers $\mathbb{N}$ or all integers $\mathbb{Z}$ or some other infinite set if there are infinitely many subsets in $\mathcal{P}$, or $I = \{1, 2, \ldots, n\}$ if there are finitely many — specifically n — subsets in $\mathcal{P}$.)

If the following conditions hold:

(i) A is the union of the subsets in $\mathcal{P}$, $A = \bigcup_{i \in I} B_i$, and

(ii) any two distinct members of $\mathcal{P}$ are disjoint, $B_i \cap B_j = \varnothing$ for $i \neq j$,

then we say that $\mathcal{P}$ is a *partition* of A.

We have observed in Corollary 3.7 that if $\sim$ is an equivalence relation on a set A, then the collection of equivalence classes forms a partition on A. In fact, equivalence relations are essentially the same thing as partitions, since given any partition $\mathcal{P}$ we can define an equivalence relation whose equivalence classes are precisely the subsets in $\mathcal{P}$.

Theorem 3.10 *Let $\mathcal{P} = \{ B_i \mid i \in I \}$ be a partition of the set A. Define a relation $\sim$ on the set A by $a \sim b$ when a and b lie in the same part B_i of the partition. Then $\sim$ is an equivalence relation on A and the equivalence classes of $\sim$ are the subsets B_i in the partition $\mathcal{P}$.*

The proof of this proposition is reasonably straightforward and is left to an exercise on a problem sheet.

We finish this chapter by examining the equivalence classes for the equivalence relations appearing in Example 3.4.

Example 3.11 (i) Let A be any set and consider equals ($=$). We observed in Example 3.4(ii) that $=$ is an equivalence relation on A. Observe that

$$b \in [a] \qquad \Longleftrightarrow \qquad a = b$$

and hence $[a] = \{a\}$, the *singleton* set that has just one element, namely a itself.

(ii) Let A be any set and let $\sim$ denote the universal relation. Then $a \sim b$ for all $a, b \in A$ and therefore $b \in [a]$ for all $b \in A$. Hence there is just one equivalence class, namely,

$$[a] = A \qquad \text{for any choice of } a \in A.$$

(iii) Now consider the equivalence relation $\sim$ on $\mathbb{Z}$ from Example 3.4(iv); that is,

$$a \sim b \qquad \text{if } a - b \text{ is an even number.}$$

Note that $0 \sim b$ for all even integers b, that $1 \sim c$ for all odd integers c, and that $0 \nsim 1$. Hence there are two equivalence classes: $[0]$ consists of all even integers and $[1]$ consists of all odd integers.

(iv) Finally consider the equivalence relation $\sim$ on the set $A = \{ (a, b) \mid a, b \in \mathbb{Z},\ b \neq 0 \}$ from Example 3.4(v) given by

$$(a, b) \sim (c, d) \qquad \text{if } ad = bc.$$

Recall that $(a, b) \sim (c, d)$ if and only if $a/b = c/d$, so each equivalence class corresponds to all the pairs (a, b) such that a/b evaluates to the same rational number. The equivalence classes are therefore parametrized by rational numbers $q \in \mathbb{Q}$. If $q = m/n$ where $m, n \in \mathbb{Z}$ with $n \neq 0$, then

$$X_q = \big[(m, n)\big] = \{ (a, b) \in A \mid a/b = q \}.$$

Chapter 4

Congruences and Modular Arithmetic

In this chapter we shall be considering a specific example of an equivalence relation on the set of integers. The positive integer $m > 1$ that appears in the following definition will be fixed throughout the chapter, but in various examples we will use specific choices for this parameter.

Definition 4.1 Let m be an integer with $m > 1$. We say that two integers a and b are *congruent modulo* m if $b - a$ is divisible by m. We write $a \equiv b \pmod{m}$ when this holds.

Thus $a \equiv b \pmod{m}$ means that $b - a = mq$ for some integer $q \in \mathbb{Z}$. When $m = 2$, we note that $a \equiv b \pmod{2}$ when $b - a$ is even. Thus the relation when $m = 2$ is precisely the equivalence relation defined in Example 3.4(iv). This chapter is consequently concerned with a generalization of that example.

Example 4.2 (i) Observe

$$\begin{aligned} 4 &\equiv 31 \pmod{9}, \\ -11 &\equiv 34 \pmod{9}, \\ 20 &\not\equiv 100 \pmod{9}. \end{aligned}$$

(ii) Note $a \equiv 0 \pmod{m}$ if and only if $m \mid a$.

(iii) Two positive integers are congruent modulo 10 if and only if they have the same final digit (when expressed in base 10 as usual).

Theorem 4.3 *Congruence modulo m is an equivalence relation on $\mathbb{Z}$.*

Proof: We need to verify the three conditions for an equivalence relation from Definition 3.3. When expressed for congruence, these conditions are:

(R) $a \equiv a \pmod{m}$ for all $a \in \mathbb{Z}$;

(S) if $a \equiv b \pmod{m}$ for some $a, b \in \mathbb{Z}$, then also $b \equiv a \pmod{m}$;

(T) if $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$ for some $a, b, c \in \mathbb{Z}$, then also $a \equiv c \pmod{m}$.

We now verify that each of these conditions holds.

(R) Let $a \in \mathbb{Z}$. Then $a - a = 0 = 0 \times m$, so $m \mid (a - a)$ and hence $a \equiv a \pmod{m}$.

(S) Let $a, b \in \mathbb{Z}$ and suppose $a \equiv b \pmod{m}$. This means that $m \mid (b-a)$, so $b-a = mq$ for some $q \in \mathbb{Z}$. Then $a - b = -(b - a) = -mq = m(-q)$ and so $m \mid (a - b)$ also. Hence $b \equiv a \pmod{m}$. This shows that congruence modulo m is symmetric.

(T) Let $a, b, c \in \mathbb{Z}$ and suppose $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$. This means that m divides both $b-a$ and $c-b$. Therefore there exist integers q and r such that $b-a = mq$ and $c - b = mr$. Then

$$c - a = (c - b) + (b - a) = mr + mq = m(q + r)$$

and so $m \mid (c - a)$. Hence $a \equiv c \pmod{m}$. This shows that congruence modulo m is transitive.

In conclusion, congruence modulo m is indeed an equivalence relation on $\mathbb{Z}$. □

Since congruence modulo m is an equivalence relation on $\mathbb{Z}$, we know (by Corollary 3.7) that the integers $\mathbb{Z}$ is partitioned into the resulting equivalence classes. We shall discuss these later in the chapter. Before we do that, however, we shall make some observations about how congruences interact with arithmetic operations on $\mathbb{Z}$.

Theorem 4.4 (Congruence Arithmetic) *Let m be an integer with $m > 1$. Let a, b, c and d be integers. Then:*

(i) *If $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, then*

$$\begin{aligned} a + c &\equiv b + d \pmod{m} \\ ac &\equiv bd \pmod{m}. \end{aligned}$$

(ii) *If $a \equiv b \pmod{m}$, then*

$$\begin{aligned} a + c &\equiv b + c \pmod{m} \\ ac &\equiv bc \pmod{m}. \end{aligned}$$

Proof: (i) Suppose $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$. This means that $m \mid (b - a)$ and $m \mid (d - c)$. Hence there are integers q and r such that

$$b - a = mq \qquad \text{and} \qquad d - c = mr.$$

Now

$$\begin{aligned} (b + d) - (a + c) &= (b - a) + (d - c) \\ &= mq + mr \\ &= m(q + r) \end{aligned}$$

which shows that m divides $(b + d) - (a + c)$; that is,

$$a + c \equiv b + d \pmod{m}.$$

Also

$$\begin{aligned} bd - ac &= bd - bc + bc - ac \\ &= b(d - c) + (b - a)c \end{aligned}$$

$$\begin{aligned} &= bmr + mqc \\ &= m(br + qc) \end{aligned}$$

which shows that m divides $bd - ac$; that is,

$$ac \equiv bd \pmod{m}.$$

(ii) Suppose $a \equiv b \pmod{m}$ and let $c \in \mathbb{Z}$. Since congruence modulo m is, in particular, reflexive, certainly $c \equiv c \pmod{m}$. Hence we can take $d = c$ in part (i) to give

$$\begin{aligned} a + c &\equiv b + c \pmod{m} \\ ac &\equiv bc \pmod{m}, \end{aligned}$$

as required. □

We shall now perform a construction that is rather common in mathematics. Here we take an already existing mathematical structure and define an associated structure upon the set of equivalence classes of some equivalence relation. The latter is termed a "quotient structure" and examples of this are indeed found throughout mathematics. It will happen occur once more towards the end of this lecture course.

To be precise, what we shall do here is the following:

- Start with the ring structure on the set of integers $\mathbb{Z}$ (defined by the usual addition and multiplication, as we observed in Example 1.3).
- Let $m > 1$ be a positive integer and consider congruence modulo m (as given in Definition 4.1).
- Observe that we can define an addition and multiplication on the *set of equivalence classes*. We shall rely heavily upon Theorem 4.4 to do this.
- Observe that the conditions to be a ring are inherited.

We begin by identifying the equivalence classes for the congruence relation.

Theorem 4.5 *Let m be an integer with $m > 1$. The equivalence relation "congruent modulo m" has precisely m equivalence classes, namely*

$$[r] = \{\, km + r \mid k \in \mathbb{Z} \,\}$$

for $r = 0, 1, \ldots, m - 1$.

We shall refer these equivalence classes as the *congruence classes modulo m*.

Proof: First fix an integer r with $0 \leqslant r < m$. Observe that if a is an arbitrary integer, then

$$\begin{aligned} a \in [r] \quad &\iff \quad r \equiv a \pmod{m} \\ &\iff \quad m \mid (a - r) \\ &\iff \quad a - r = km \qquad \text{for some } k \in \mathbb{Z} \\ &\iff \quad a = km + r \qquad \text{for some } k \in \mathbb{Z}. \end{aligned}$$

Thus we have identified the elements in the equivalence class $[r]$:

$$[r] = \{\, km + r \mid k \in \mathbb{Z} \,\}.$$

We have shown that the equivalence classes have the claimed form. We still need to show that the collection $[0]$, $[1]$, $\ldots$, $[m-1]$ are all the equivalence classes and that they are distinct.

If $a \in \mathbb{Z}$, divide a by m and obtain some remainder:

$$a = qm + r \qquad \text{where } 0 \leqslant r < m.$$

Then $a \equiv r \pmod{m}$ and so $a \in [r]$. This shows that every integer in $\mathbb{Z}$ lies in one of the equivalence classes

$$[0],\ [1],\ \ldots,\ [m-1]. \tag{4.1}$$

This means that this list includes all the equivalence classes.

Finally suppose $[r] = [s]$ where $0 \leqslant r, s < m$. This means that $r \equiv s \pmod{m}$, so m divides $r - s$. However $-m < r - s < m$ by the choice of r and s, this forces $r - s = 0$; that is, $r = s$. Hence the equivalence classes listed in (4.1) are distinct.

In conclusion, there are precisely m congruence classes and they have the form specified in the statement. □

We now know that there are m equivalence classes for the relation $\equiv \pmod{m}$. We want to define arithmetic operations on the set of congruence classes. We shall achieve this by making use of the arithmetic operation on the integers and applying them to the *representatives* of the congruence classes.

Consequently, if a and b are integers, we define

$$[a] + [b] = [a + b]$$
$$[a] \cdot [b] = [ab].$$

There is an issue here: We have defined the operations in terms of "representatives" for the equivalence classes; that is, we have picked elements in each of the two equivalence classes and defined the operation in terms of these choices. However, it is possible to choose different representatives and we want to show that resulting equivalence classes is the same irrespective of the choice. Indeed, if a is any integer, Theorem 4.5 describes the congruence class $[a]$ and, in particular, tells us there are infinitely many elements in this class. As we know congruence classes are either disjoint or equal (by Theorem 3.6), then $[a] = [c]$ for all choices of c in the congruence class $[a]$. We need to verify that our definitions of addition and multiplication does not give a different answer if replace the element a by a different element in the same congruence class. The term that is used throughout mathematics is that we shall show our operations are "well-defined."

Lemma 4.6 *Let m be an integer with $m > 1$. Then there are well-defined addition and multiplication operations on the set of congruence classes modulo m given by*

$$[a] + [b] = [a + b] \qquad \textit{and} \qquad [a] \cdot [b] = [ab].$$

Proof: Let a, b, c and d be integers. Suppose that a and c are chosen from the same congruence class and that b and d are chosen from the same congruence class (possibly a different class). This means that

$$a \equiv c \pmod{m} \qquad \text{and} \qquad b \equiv d \pmod{m}.$$

Theorem 4.4(i) tells us that

$$a + b \equiv c + d \pmod{m} \qquad \text{and} \qquad ab \equiv cd \pmod{m},$$

so

$$[a+b] = [c+d] \qquad \text{and} \qquad [ab] = [cd].$$

Hence if $[a] = [c]$ and $[b] = [d]$, then

$$[a] + [b] = [c] + [d] \qquad \text{and} \qquad [a] \cdot [b] = [c] \cdot [d].$$

This shows that the addition and multiplication is *well-defined*: it does not depend upon our choice of representatives from each of the equivalence classes. □

Theorem 4.7 *Let m be an integer with $m > 1$. Then the set $R = \{[0], [1], \ldots, [m-1]\}$ of congruence classes modulo m is a commutative ring under the addition and multiplication defined by*

$$[a] + [b] = [a+b] \qquad \textit{and} \qquad [a] \cdot [b] = [ab].$$

Proof: We have shown in Lemma 4.6 that these operations do define addition and multiplication on the set R of equivalence classes. It remains to show that the conditions to be a ring hold. Let a, b and c be arbitrary integers, so $[a]$, $[b]$ and $[c]$ are arbitrary elements of R.

A1: Observe

$$[a] + [b] = [a+b] = [b+a] = [b] + [a]$$

(using the fact that A1 holds in $\mathbb{Z}$). Hence Condition A1 holds in R with our given addition.

A2: Observe

$$\begin{aligned}[a] + \big([b] + [c]\big) = [a] + [b+c] &= [a + (b+c)] \\ &= [(a+b) + c] = [a+b] + [c] = \big([a] + [b]\big) + [c],\end{aligned}$$

so Condition A2 holds in R with our given addition.

A3: Observe

$$[0] + [a] = [0+a] = [a] \qquad \text{and} \qquad [a] + [0] = [a+0] = [a].$$

This holds for all choice of a and hence $[0]$ is the zero for the addition operation on R.

A4: Given our arbitrary element $a \in \mathbb{Z}$, we can consider the equivalence class $[-a]$ of its negative. We observe

$$[a] + [-a] = [a + (-a)] = [0] \qquad \text{and} \qquad [-a] + [a] = [(-a) + a] = [0].$$

Hence Condition A4 holds in R.

M1: Observe

$$[a] \cdot [b] = [ab] = [ba] = [b] \cdot [a]$$

using the fact that (M1) holds in $\mathbb{Z}$. Hence multiplication is commutative in R.

M2: Similarly

$$[a] \cdot \big([b] \cdot [c]\big) = [a] \cdot [bc] = [a(bc)] = [(ab)c] = [ab] \cdot [c] = \big([a] \cdot [b]\big) \cdot [c],$$

so multiplication is associative in R.

D: Finally

$$\begin{aligned}[a] \cdot \big([b] + [c]\big) = [a] \cdot [b+c] &= [a(b+c)] \\ &= [ab + ac] = [ab] + [ac] = [a] \cdot [b] + [a] \cdot [c].\end{aligned}$$

This checks all conditions required for a commutative ring. □

Comment: Observe that verification of the Conditions A1–A4, M1, M2 and D for a commutative ring were actually quite straightforward. These are inherited somewhat directly from the fact the same conditions hold in the original ring $\mathbb{Z}$ of integers. The most difficult step in verifying the above theorem is actually Lemma 4.6 where we established that addition and multiplication on the set of congruence classes is well-defined. In future similar situations, establishing that operations on equivalence classes are well-defined will usually be the most important and difficult step.

Definition 4.8 We shall write $\mathbb{Z}/m\mathbb{Z}$ for the ring consisting of the congruence classes modulo m under the addition and multiplication in the above theorem. (Another common notation is $\mathbb{Z}_m$ and this is found in many texts.)

We normally simplify the notation and omit the brackets around the elements of $\mathbb{Z}/m\mathbb{Z}$. Although strictly speaking they are equivalence classes of integers, we usually denote the m elements of $\mathbb{Z}/m\mathbb{Z}$ by

$$0,\ 1,\ \ldots,\ m-1.$$

We then also use $\equiv \pmod m$ as the "equals" in this ring. For example, if $m = 6$, we would denote the addition in $\mathbb{Z}/6\mathbb{Z}$ by writing

$$3 + 5 \equiv 2 \pmod 6.$$

Similar notation can be used to denote the multiplication in such rings.

We shall now give further examples of the results of performing the addition and multiplication in these rings for the values $m = 4$ and $m = 5$. In this example, we shall use what is called a *Cayley table* or *addition/multiplication table*. In the case of the addition table, we place $a + b$ in the entry occurring in the row with label a and in the column with label b. The analogous convention is used for the multiplication table.

Example 4.9 (i) In the first example, we take $m = 4$. The addition and multiplication tables are then:

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

$\times$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

(ii) Now we do the same thing for $m = 5$. The addition and multiplication tables are then:

$+$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$\times$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Note two properties that we observe in the above examples, specifically within the multiplication tables. Property M3 holds in both $\mathbb{Z}/4\mathbb{Z}$ and $\mathbb{Z}/5\mathbb{Z}$:

$$1a = a1 = a \qquad \text{for all } a;$$

that is, 1 is a multiplicative identity. Property M4 (existence of multiplicative inverses) does not hold in $\mathbb{Z}/4\mathbb{Z}$: there does not exist any choice of a such that $2a \equiv 1 \pmod 4$.

However, since 1 occurs in all except the first row and in all except the first column of the multiplicative table of $\mathbb{Z}/5\mathbb{Z}$, we conclude that every non-zero element of $\mathbb{Z}/5\mathbb{Z}$ has a multiplicative inverse. Indeed:

$$1^{-1} = 1, \qquad 2^{-1} = 3, \qquad 3^{-1} = 2, \qquad 4^{-1} = 4$$

The crucial thing that distinguishes between the two examples is that 5 is a prime number, whereas 4 is not. More generally, we shall finish this chapter with the following observation:

Theorem 4.10 *Let p be any prime number. Then $\mathbb{Z}/p\mathbb{Z}$ is a field.*

Recall that a *prime number* is a positive integer $p > 1$ whose only positive divisors are 1 and p. When p is prime, we shall often write $\mathbb{F}_p$ for the field $\mathbb{Z}/p\mathbb{Z}$.

Proof: Let p be a prime number and $F = \mathbb{Z}/p\mathbb{Z}$, the ring of congruence classes modulo p. We know by Theorem 4.7 that F is a commutative ring under addition and multiplication performed modulo p. According to Definition 1.13, we must verify that Conditions M3 (existence of a multiplicative identity) and M4 (existence of multiplicative inverses) hold in F.

M3: If $a \in \{0, 1, 2, \ldots, p-1\}$, then $a1 = 1a = a$. Hence 1 is a multiplicative identity in F; that is, Condition M3 holds in F.

M4: Now let a be an integer that represents a non-zero element of F. Thus $a \not\equiv 0 \pmod{p}$, so p does not divide a. Consider now the greatest common divisor of the integers a and p. The only positive divisors of p are 1 and p, but p does not divide a, and hence $\gcd(a, p) = 1$; that is, a and p are coprime. Theorem 2.2 tells us there are integers u and v with $ua + vp = 1$. Hence

$$ua \equiv 1 \pmod{p}.$$

This shows that u is a multiplicative inverse for a in F. Hence Condition M4 holds in F.

In conclusion, F is indeed a field. □

As a comment, one place that fields occur is in the study of linear mathematics. You will recall from the module *MT2501* that one of the primary objects of study is the vector space and that every vector space occurs *over some field*. In *MT2501*, the typical examples of fields that were used were the real numbers $\mathbb{R}$ and the complex numbers $\mathbb{C}$. However, one could study vector spaces over the rational numbers $\mathbb{Q}$ and also over the field $\mathbb{F}_p$ of congruence classes modulo p for any prime p. Thus, we could consider vector spaces over $\mathbb{F}_2$, $\mathbb{F}_3$, $\mathbb{F}_5$, $\mathbb{F}_7$, ... The case of the field $\mathbb{F}_2$ of two elements is particularly important in discrete mathematics, coding theory, theoretical computing, etc.

We finish the chapter by discussing solution of congruence equations. The natural place to consider these is in the ring $\mathbb{Z}/m\mathbb{Z}$ of congruence classes. Indeed, if x satisfies $ax \equiv b \pmod{m}$ and $y \equiv x \pmod{m}$, then Theorem 4.4 tells us that $ay \equiv ax \equiv b \pmod{m}$. Hence solutions to the equation $ax \equiv b \pmod{m}$ consist of entire congruence classes and so we really are solving the equation in $\mathbb{Z}/m\mathbb{Z}$. We shall illustrate such solutions in the following example.

Example 4.11 *Determine all the solutions (if any) in the ring $\mathbb{Z}/49\mathbb{Z}$ of the following equations:*

(i) $4x \equiv 9 \pmod{49}$*;*

(ii) $7x \equiv 0 \pmod{49}$*;*

(iii) $7x \equiv 9 \pmod{49}$.

In this example, we are deliberately working in a ring $\mathbb{Z}/m\mathbb{Z}$ where m is not prime. In the case of a finite field $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$, the solution to part (i) will adapt (and there will be no examples arising in the form of parts (ii) and (iii)).

Solution: (i) Note here that 4 is coprime to 49: The divisors of 49 are 1, 7 and 49, so $\gcd(4, 49) = 1$. Indeed, we shall apply the Extended Euclidean Algorithm with $a = 49$ and $b = 4$:

$$\begin{array}{rr|l} a & = 49 & \\ & 12b = 48 & 4 = b \\ \hline a - 12b = & 1 & 4 \\ & & \hline & & 0 \end{array}$$

Hence the Bézout coefficients are $u = 1$ and $v = -12$:

$$1 \times 49 - 12 \times 4 = 1.$$

We reduce this equation modulo 49:

$$-12 \times 4 \equiv 1 \pmod{49};$$

that is,

$$37 \times 4 \equiv 1 \pmod{49}$$

and so 37 is a multiplicative inverse for 4 in $\mathbb{Z}/49\mathbb{Z}$. We now turn to the equation under consideration:

$$4x \equiv 9 \pmod{49}$$

Multiply both sides by 37:

$$x \equiv 37 \times 4x \equiv 37 \times 9 = 333 \equiv 39 \pmod{49}$$

(The last step is achieved by determining the remainder when one divides 333 by 49.) Hence the given solution has a unique solution in $\mathbb{Z}/49\mathbb{Z}$:

$$x \equiv 39 \pmod{49}$$

(ii) Note that $7x \equiv 0 \pmod{49}$ if and only if 49 divides $7x$; that is, $7x = 49q$ for some $q \in \mathbb{Z}$. Thus $x = 7q$ for some $q \in \mathbb{Z}$. Hence the solutions of $7x \equiv 0 \pmod{49}$ are

$$x = 0,\ 7,\ 14,\ 21,\ 28,\ 35,\ 42.$$

In particular, there are seven solutions to the equation in $\mathbb{Z}/49\mathbb{Z}$.

(iii) Suppose x is a solution of $7x \equiv 9 \pmod{49}$. Then 7 also divides $7x - 9$, so $7x \equiv 9 \pmod{7}$. Hence

$$2 \equiv 9 \equiv 7x \equiv 0 \pmod{7},$$

which is a contradiction. Hence the equation $7x \equiv 9 \pmod{49}$ has no solutions. □

Chapter 5

Groups

So far we have met two of the major types of algebraic structure: rings and fields (and fields are just rings with additional properties). We shall now turn to the main topic of the module, namely the algebraic structure called a *group*. We shall be studying this for the rest of the course. We begin with the definition:

Definition 5.1 A *group* is a set G together with a binary operation

$$(x, y) \mapsto xy$$

(usually denoted, as here, by multiplication) such that the following conditions hold:

(i) $(xy)z = x(yz)$ for all $x, y, z \in G$;

(ii) there is some element $1 \in G$, called the *identity*, such that $1x = x1 = x$ for all $x \in G$;

(iii) for every $x \in G$, there is some element $x^{-1} \in G$, called the *inverse* of x, such that $xx^{-1} = x^{-1}x = 1$.

For a general group, we shall use the notation of Definition 5.1, namely writing the binary operation as multiplication. In specific examples, we shall use notation appropriate for the example. One possible alternative choice of notation might be addition $x + y$ (see, for example, Example 5.4 below). The three conditions (i)–(iii) appearing in the definition are often called the "axioms of a group." As with rings, we shall simply say "G is a group" to mean that G is a set with a binary operation defined upon it so that the above conditions hold. Consequently, we often do not distinguish explicitly between a group and the set upon which the group structure is defined.

As with the multiplication in a ring, note that we have not assumed the binary operation in a group is commutative. Groups where the operation is commutative are very special and they are given the following name:

Definition 5.2 Let G be a group. We say that G is *abelian*[1] if

$$xy = yx \qquad \text{for all } x, y \in G.$$

An additional piece of terminology that we introduce at this point is the "order" of a group so that we can use it as we present our first examples.

Definition 5.3 The *order* of a group G is the number of elements in the group. It is denoted by $|G|$.

[1] This concept is named after the Norwegian mathematician *Niels Henrik Abel* (1802–1829).

We begin with some examples that arise immediately from what we have already done.

Example 5.4 (i) Let R be any ring. Then Conditions A1–A4 tell us that if we just use the addition on R, then the conditions for an *abelian* group hold. Thus $(R, +)$, the set R equipped with the addition $+$ as its binary operation, is an abelian group.

(ii) In particular, the set of integers $\mathbb{Z}$ is an abelian group under addition. Equally, the set of rationals $\mathbb{Q}$, the set of real numbers $\mathbb{R}$ and the set of complex numbers $\mathbb{C}$ are all groups under addition. These are examples of infinite groups.

(iii) Let m be an integer with $m > 1$. Theorem 4.7 tells us that $\mathbb{Z}/m\mathbb{Z} = \{0, 1, \ldots, m-1\}$ forms a ring when we use addition and multiplication modulo m. Consequently, this set together with addition performed modulo m is also an abelian group. It has order m: $|\mathbb{Z}/m\mathbb{Z}| = m$.

In conclusion, any example of a ring gives us an example of an abelian group by simply forgetting the multiplication. However, there are lots of examples of groups that do not arise in this way. If we find an example of a group that is not abelian, then we know it does not come from the addition of a ring. We shall construct our first example using matrices.

Example 5.5 Consider the set of 2×2 matrices over the real numbers with non-zero determinant

$$\mathrm{GL}_2(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\middle|\; a, b, c, d \in \mathbb{R},\ ad - bc \neq 0 \right\}.$$

Recall that if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is a 2×2 matrix then

$$\det A = ad - bc.$$

There are two properties of determinant that we shall use:

- $\det(AB) = (\det A)(\det B)$ for all square matrices A and B of the same size over the same field, and
- if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has $\det A \neq 0$, then

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

has the property that

$$AA^{-1} = A^{-1}A = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Now if $A, B \in \mathrm{GL}_2(\mathbb{R})$, then

$$\det(AB) = (\det A)(\det B) \neq 0,$$

so $AB \in \mathrm{GL}_2(\mathbb{R})$. Hence matrix multiplication defines a binary operation on $\mathrm{GL}_2(\mathbb{R})$.
We now verify the group axioms:

(i) We have already shown that matrix multiplication is associative: We verified Condition M2 holds in the ring $\mathrm{M}_2(\mathbb{R})$ in Theorem 1.7. In particular, if A, B and C are 2×2 matrices over $\mathbb{R}$ with non-zero determinant then

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

(ii) Consider the 2×2 identity matrix

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

Note $\det I = 1$, so $I \in \mathrm{GL}_2(\mathbb{R})$. This has the property that, for $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$,

$$IA = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = A$$

and

$$AI = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = A.$$

Hence I is an identity for matrix multiplication on $\mathrm{GL}_2(\mathbb{R})$.

(iii) Let $A \in \mathrm{GL}_2(\mathbb{R})$. Then $\det A \neq 0$, so we can construct the inverse matrix A^{-1} that has the property

$$AA^{-1} = A^{-1}A = I.$$

Note

$$(\det A)(\det A^{-1}) = \det I = 1,$$

so $\det A^{-1} = 1/\det A \neq 0$. This shows that $A^{-1} \in \mathrm{GL}_2(\mathbb{R})$ and we have shown that every $A \in \mathrm{GL}_2(\mathbb{R})$ has an inverse.

This shows that $\mathrm{GL}_2(\mathbb{R})$ is a group.
Moreover

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \in \mathrm{GL}_2(\mathbb{R})$$

(since both have determinant 1) and we have already calculated that

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \neq \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Hence this is an example of a *non-abelian* group.

Let us place this example in context by making the following definition.

Definition 5.6 Let F be a field and n be a positive integer. We define $\mathrm{GL}_n(F)$ to be the set of all $n \times n$ matrices over the field F that have non-zero determinant,

$$\mathrm{GL}_n(F) = \{\, A \in \mathrm{M}_n(F) \mid \det A \neq 0 \,\},$$

together with matrix multiplication as binary operation. We call $\mathrm{GL}_n(F)$ the *general linear group* of degree n over F.

A very similar argument to Example 5.5 shows that $\mathrm{GL}_n(F)$ is a group and that it is non-abelian if $n \geqslant 2$. Associativity of the matrix multiplication comes from Theorem 1.7. The main step remaining in verifying that $\mathrm{GL}_n(F)$ is a group is to remember that if A is an $n \times n$ matrix with $\det A \neq 0$ then A has an inverse. (This does not depend on $n = 2$: it just happens to be easier to write down the formula for the inverse when $n = 2$.)

We showed that we can construct additive groups from a ring. In the case of a field, we can also construct a group from the multiplication.

Lemma 5.7 (Multiplicative group of a field) *Let F be a field. The set of non-zero elements in F is an abelian group under the multiplication of F.*

We write $F^* = F \setminus \{0\}$ for the non-zero elements of the field F viewed as a group under multiplication. It is called the *multiplicative group* of F.

PROOF: We attempt to define a binary operation on $F^* = F \setminus \{0\}$ using the multiplication of the field:

$$(x, y) \mapsto xy.$$

Suppose $xy = 0$ for some $x, y \in F^*$. Multiply by the multiplicative inverse x^{-1} of x:

$$y = x^{-1}xy = x^{-1}0 = 0.$$

This contradicts the assumption that x and y are non-zero. Hence $xy \neq 0$ for all $x, y \in F^*$, so multiplication does define a binary operation on the set of non-zero elements of the field F.

The multiplication is associative (since F satisfies Condition M2). The multiplicative identity 1 provided by Condition M3 for a field is the identity element for this binary operation and Condition M4 guarantees that each element of F^* has an inverse under multiplication. Finally a field is, in particular, a commutative ring and hence F^* is abelian. □

In particular, $\mathbb{Q}^*$, $\mathbb{R}^*$ and $\mathbb{C}^*$ are infinite abelian groups under multiplication and, for each prime p, the set $\mathbb{F}_p^*$ of non-zero members of the finite field $\mathbb{F}_p$ is an abelian group of order $p - 1$.

Cayley tables

One way that we can specify the binary operation in a finite group is via its *multiplication table* (also called its *Cayley*[2] *table*). We used these earlier in Example 4.9.

If G is a *finite* group, then we construct a table whose rows and columns are labelled by the elements of G. The entry in row x and column y is the element xy.

Example 5.8 (Trivial Group) Consider the set $\{1\}$ containing a single element and with the only possible Cayley table:

	1
1	1

This defines a binary operation on the set and it is immediate that this is a group. Associativity is merely observed in the equation $(11)1 = 1 = 1(11)$, while 1 is an identity and its own inverse. This is the *trivial group*.

In particular, we now have a group of order 1, while $\mathbb{Z}/m\mathbb{Z}$ as a group under addition has order m for each $m > 1$. Hence we have at least one group of each possible finite order.

Example 5.9 (Klein 4-group) Consider the table

	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

(5.1)

[2]Named after the British mathematician *Arthur Cayley* (1821–1895).

In fact this defines a group. (It is not automatic that a multiplication table defines a group: one needs to check whether or not the axioms of a group hold.)

Let $G = \{e, a, b, c\}$.

Binary operation: Since only e, a, b and c occur in the table, it is the case that $xy \in G$ for all choices of x and y from G. Hence the table does define a binary operation on G.

Associativity: This is not easy to check from the table. We would need to check that

$$(xy)z = x(yz)$$

for all $4^3 = 64$ choices of x, y and z from G. For example,

$$(aa)b = eb = b$$
$$a(ab) = ac = b.$$

By a bit of ingenuity one can significantly reduce the number of checks required, but we shall leave associativity of the multiplication till later. In general, determining associativity for the multiplication given by an arbitrary Cayley table is not straightforward. Consequently such tables are not quite so useful as one might hope.

Identity: This can be read off the first row and first column of the table. Observe $ee = e$, $ea = ae = a$, $eb = be = b$ and $ec = ce = c$. So e is the identity element for the binary operation in the Table (5.1).

Inverses: We can find inverses by seeking the occurrence of the identity element e in the entries in the table. Observe $ee = e$, $aa = e$, $bb = e$ and $cc = e$. Hence $e^{-1} = e$, $a^{-1} = a$, $b^{-1} = b$ and $c^{-1} = c$ are the inverses.

This shows (except for the deferral of verifying associativity) that the Table (5.1) endows $G = \{e, a, b, c\}$ with the structure of a group. Note also that $xy = yx$ for all $x, y \in G$ since the (x, y)- and (y, x)-entries are the same in the Table (5.1). Therefore the group constructed is *abelian*. This is visibly an example of a group of order 4.

Definition 5.10 The group given in Example 5.9 is called the *Klein 4-group*. In this set of lecture notes, it will be denoted by V_4 (though many sources denote it by K_4).

So $V_4 = \{e, a, b, c\}$ with multiplication as given in Table (5.1). The choice of notation is split in the literature: many sources use K_4 and many use V_4, so the lecturer is choosing his preference. To finish for now our work with this group, we shall illustrate one reasonably easy way to verify associativity:

Lemma 5.11 *The binary operation in the Klein 4-group V_4 is associative.*

The method of proof here is to observe that the multiplication is essentially the same as a binary operation that we already know is associative.

Proof: Consider the following 2×2 matrices over $\mathbb{R}$:

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad C = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

We shall construct the multiplication table for these four elements:

	I	A	B	C
I	I	A	B	C
A	A	I	C	B
B	B	C	I	A
C	C	B	A	I

This has exactly the same shape as the multiplication table for the Klein 4-group (Table (5.1)). We know that matrix multiplication is associative, so

$$(XY)Z = X(YZ)$$

for all $X, Y \in \{I, A, B, C\}$. Hence the multiplication given in the table just given is associative and so the same must hold for the Table (5.1), which has the same pattern of entries. Hence we conclude

$$(xy)z = x(yz) \qquad \text{for all } x, y, z \in V_4;$$

that is, multiplication is associative in V_4. □

Non-examples

So far we have encountered several examples of groups and seen how we show a particular set with given binary operation is indeed a group. One might start to wonder whether basically everything is a group (in which case it would not be that interesting a concept), so here we give some examples that are not groups.

Example 5.12 Consider the set of integers $\mathbb{Z}$ and consider subtraction as a binary operation on $\mathbb{Z}$:

$$(a, b) \mapsto a - b.$$

This is a binary operation on $\mathbb{Z}$ because $a - b \in \mathbb{Z}$ for all integers a and b. However, subtraction is not associative:

$$(4 - 2) - 1 = 2 - 1 = 1$$
$$4 - (2 - 1) = 4 - 1 = 3$$

So $(a - b) - c = a - (b - c)$ is *not* true for all $a, b, c \in \mathbb{Z}$. Hence the set of integers is *not* a group under the binary operation of subtraction.

Example 5.13 Let $E = \{\, 2a \mid a \in \mathbb{Z} \,\}$ be the set of even integers and consider multiplication as a binary operation on E. Note that the product of two even integers is still even, so multiplication is a binary operation on E. In this case, associativity is ok: multiplication of integers is already known to be associative.

Suppose $e \in E$ is an identity for multiplication, so $ex = xe = x$ for all $x \in E$. In particular,

$$2e = 2.$$

Therefore $e = 1$, which is a contradiction because $1 \notin E$. Hence E is not a group under multiplication because there is no identity element in E.

Example 5.14 Consider the set of real numbers $\mathbb{R}$ and consider multiplication as a binary operation on $\mathbb{R}$:

$$(x, y) \mapsto xy$$

We already know that multiplication is an associative binary operation on $\mathbb{R}$ and that 1 is an identity for this operation:

$$1x = x1 = x \qquad \text{for all } x \in \mathbb{R}.$$

In this case, the problem is with inverses: the real number 0 has no multiplicative inverse. Indeed, $0y = y0 = 0$ for all $y \in \mathbb{R}$, so there is no choice of y such that these products equal 1. Hence $\mathbb{R}$ is not a group under multiplication.

One might raise an objection to the previous example. Perhaps there is another choice of identity element for which we can find inverses for all elements. This is, however, not the case. If there is an identity for a binary operation, then it is unique, so there cannot be such a choice of alternative identity element.

Basic properties of groups

To address what we had just discussed, let us start by observing that identities are unique.

Lemma 5.15 *Let $*$ be a binary operation on a set A. If there is an identity for $*$, then it is unique.*

Proof: Suppose that e and f are both identities for $*$. Then they satisfy

$$ea = ae = a \qquad \text{and} \qquad fa = af = a$$

for all $a \in A$. In particular, substituting $a = f$ in the first gives

$$ef = f$$

and subsituting $a = e$ in the second gives

$$ef = e.$$

Hence $e = f$. This shows that the identity, if it exists, is unique. □

This now completes the justification of Example 5.14. Having shown 1 is an identity for the multiplication in $\mathbb{R}$, we know it is *the only* choice of identity and there is therefore no multiplicative inverse for 0 in the real numbers $\mathbb{R}$.

The following summarizes the basic properties of the various elements appearing in the definition of a group. They will be used throughout the course.

Theorem 5.16 *Let G be a group. Then*

(i) *the identity element 1 of G is unique;*

(ii) *the inverse of each element $x \in G$ is unique;*

(iii) $(x^{-1})^{-1} = x$ *for all $x \in G$;*

(iv) $(xy)^{-1} = y^{-1}x^{-1}$ *for all* $x, y \in G$. *More generally, if* $x_1, x_2, \ldots, x_n \in G$, *then*

$$(x_1 x_2 \ldots x_n)^{-1} = x_n^{-1} \ldots x_2^{-1} x_1^{-1}.$$

PROOF: (i) This was established (in greater generality) in Lemma 5.15.

(ii) Let $x \in G$ and suppose that $a, b \in G$ are elements have the property required to be an inverse for x:

$$xa = ax = 1 \qquad \text{and} \qquad xb = bx = 1.$$

We calculate the product axb using two choices of bracketing:

$$(ax)b = 1b = b$$

and

$$a(xb) = a1 = a.$$

Associativity of the binary operation tells us these products are equal. Hence $a = b$. This shows that x has a unique inverse that henceforth we denote by x^{-1}.

(iii) The equations $x^{-1}x = xx^{-1} = 1$ tells us that x has the property of being an inverse for x^{-1}. (It is what we multiply x^{-1} by in order to produce the identity 1.) Part (ii) tells us that x must be the unique inverse of x^{-1}:

$$(x^{-1})^{-1} = x.$$

(iv) Observe

$$(xy)(y^{-1}x^{-1}) = x(yy^{-1})x^{-1} = x1x^{-1} = xx^{-1} = 1$$

and

$$(y^{-1}x^{-1})(xy) = y^{-1}(x^{-1}x)y = y^{-1}1y = y^{-1}y = 1.$$

Hence $y^{-1}x^{-1}$ is an element that has the property of being an inverse for xy. Part (ii) tells us that it is the unique inverse of xy:

$$(xy)^{-1} = y^{-1}x^{-1}. \tag{5.2}$$

We now prove the formula

$$(x_1 x_2 \ldots x_n)^{-1} = x_n^{-1} \ldots x_2^{-1} x_1^{-1}$$

by induction on n. It is trivially true when $n = 1$: It then merely says $x_1^{-1} = x_1^{-1}$. Suppose that the formula holds for $n - 1$:

$$(x_1 x_2 \ldots x_{n-1})^{-1} = x_{n-1}^{-1} \ldots x_2^{-1} x_1^{-1}.$$

Take $x = x_1 x_2 \ldots x_{n-1}$ and $y = x_n$ in Equation (5.2). We then determine

$$\begin{aligned}
(x_1 x_2 \ldots x_n)^{-1} &= (xy)^{-1} \\
&= y^{-1}x^{-1} && \text{by Equation (5.2)} \\
&= x_n^{-1}(x_1 x_2 \ldots x_{n-1})^{-1} \\
&= x_n^{-1} x_{n-1}^{-1} \ldots x_2^{-1} x_1^{-1} && \text{by the inductive hypothesis.}
\end{aligned}$$

This establishes the claimed formula by induction. □

There are two main observations that we can make using the above basic properties, the first relating to solutions of equations in a group and the second relating to the shape of group multiplication tables.

Lemma 5.17 (Solution of equations) *Let G be a group.*

(i) *Let $a, b, x \in G$. Then*

Right cancellativity: *if $ax = bx$, then $a = b$;*

Left cancellativity: *if $xa = xb$, then $a = b$.*

(ii) *Let $a, b \in G$. The equations*

$$ax = b \qquad \text{and} \qquad ya = b$$

have unique solutions in G, namely $x = a^{-1}b$ and $y = ba^{-1}$, respectively.

PROOF: (i) We verify only right cancellativity in lectures. The proofs are similar, but are both presented in these notes.

Right cancellativity: Suppose $ax = bx$. Multiply on the right by x^{-1}:

$$(ax)x^{-1} = (bx)x^{-1}$$

Hence, by associativity,

$$a(xx^{-1}) = b(xx^{-1});$$

that is, by the definition of inverses,

$$a1 = b1;$$

so, by the definition of the identity, we conclude

$$a = b.$$

Left cancellativity: This is similar. Suppose $xa = xb$. Multiply on the left by x^{-1} to give $x^{-1}(xa) = x^{-1}(xb)$. Hence, by associativity, $(x^{-1}x)a = (x^{-1}x)b$; that is, $1a = 1b$, which implies $a = b$.

(ii) Similarly we shall only establish the solution of the second of the equations in lectures, but both are presented here in these notes.

For the second equation, note that $(ba^{-1})a = b(a^{-1}a) = b1 = b$, so $y = ba^{-1}$ is a solution of the equation $ya = b$. Furthermore, if y is any solution of this equation then $ya = b$ and multiplying on the right by a^{-1} gives

$$(ya)a^{-1} = ba^{-1}$$

so, by associativity,

$$y(aa^{-1}) = ba^{-1};$$

that is,

$$y1 = ba^{-1}$$

and so

$$y = ba^{-1}.$$

Hence $y = ba^{-1}$ is the unique solution to $ya = b$.

For the first equation, observe $a(a^{-1}b) = (aa^{-1})b = 1b = b$, so $x = a^{-1}b$ is a solution of $ax = b$. Conversely, if x is any solution then $ax = b$, so $a^{-1}(ax) = a^{-1}b$; that is, $(a^{-1}a)x = a^{-1}b$. We deduce $1x = a^{-1}b$ and hence $x = a^{-1}b$. This shows that $x = a^{-1}b$ is the unique solution of the equation $ax = b$. □

The last part of this lemma illustrates how we manipulate formulae involving elements of a group. If we have a formula such as

$$ya = b$$

then multiplying on the right by a^{-1} produces

$$y = ba^{-1}$$

and has the effect of moving a from the left-hand side of the equation to the right-hand side.

Corollary 5.18 (Form of Cayley tables) *Let G be a group. Then every element of G occurs precisely once in each row and precisely once in each column of the Cayley table of G.*

Proof: In lectures, we shall just show that each element occurs precisely once in each row. The arguments are similar and that for columns is given in these notes.

Fix an element $g \in G$. Then g occurs in row a of the Cayley table (that is, the row labelled with the element a) when there exists some $x \in G$ satisfying

$$ax = g.$$

Lemma 5.17(ii) tells us this equation has a unique solution, namely $x = a^{-1}g$. Hence g occurs precisely once in row a namely in the $(a, a^{-1}g)$-entry of the table.

Similarly, g occurs in column b of the Cayley table when there exists x satisfying $xb = g$. Lemma 5.17(ii) says this equation has a unique solution, namely $x = gb^{-1}$. We conclude that g occurs precisely once in column b namely in the (gb^{-1}, b)-entry. □

Powers of elements

The axioms of a group enable us to define what we mean by powers of its elements:

Definition 5.19 Let G be a group, x be an element of G and n be any integer. We define the power x^n as follows:

(i) x^n is the product $xx \dots x$ of n copies of the element x if n is a positive integer;

(ii) $x^0 = 1$, the identity element of G;

(iii) x^n is the product $x^{-1}x^{-1} \dots x^{-1}$ of $|n|$ copies of the inverse of x if n is negative.

We are using the fact that the binary operation in the group is associative when making this definition. This means that the product $xx\ldots x$ of n copies of x is uniquely defined and it does not depend upon choice of bracketing. With the above definition, the standard properties of powers all hold in a group:

Theorem 5.20 (Power Laws) *Let G be a group, x be an element of G and m and n be integers. Then*

$$x^m x^n = x^{m+n} \qquad \text{and} \qquad (x^m)^n = x^{mn}.$$

We omit the proof. When m and n are positive integers, the formulae are actually quite easy to establish. For example,

$$x^m x^n = \underbrace{xx\ldots x}_{m \text{ times}}\underbrace{xx\ldots x}_{n \text{ times}} = x^{m+n}$$

since this is a product of $m+n$ copies of x. To prove Theorem 5.20 in general needs a detailed case analysis of when m, n and $m+n$ are positive or negative. This is relatively complicated, though each step within it is straightforward. It is for this reason that we choose to omit the proof.

Use of additive notation

In some situations it is appropriate to use addition ($+$) to denote the binary operation on a group. For example, in the case of the integers under addition, or more generally the additive group of any ring, it would be confusing to use another notation to denote the operation and so we will use addition at that point. To maintain consistency, when addition is used for the binary operation on a group we shall also use the following:

- the identity element will be written as 0,
- the inverse of an element x will be written as $-x$,
- instead of using powers, we shall write nx for $x + x + \cdots + x$ (n times) when $n > 0$ and a similar adjustment using multiples instead of powers when $n \leqslant 0$.

The power laws then become

$$mx + nx = (m+n)x \qquad \text{and} \qquad m(nx) = (mn)x$$

for $m, n \in \mathbb{Z}$ and x an element in an additively written group.

Finally, it should be noted that addition is usually *only* used for certain *abelian groups*, namely those for which the binary operation is normally written as $+$.

Chapter 6

Permutations and Symmetric Groups

In this chapter, we shall present one of the most important examples of a group. It is much closer to being a typical example than those met in the previous chapter (for a start, this type of group is generally not abelian) and it also occurs throughout the study of groups. The elements of a symmetric group are special types of functions. Consequently we begin by discussing some properties of functions.

Functions

Notation: In this module we shall follow a common practice in algebra of writing functions *on the right*. If $f\colon X \to Y$ is a function from a set X to a set Y, instead of writing $f(x)$ for the image of $x \in X$ under f (as is usual elsewhere in mathematics) we shall write

$$xf$$

for the image of x under f.

Definition 6.1 Let $f\colon X \to Y$ and $g\colon Y \to Z$ be functions from X to Y and from Y to Z (respectively). The *composite* of f and g is the function $fg\colon X \to Z$ defined by

$$x\,fg = (xf)g \qquad \text{for each } x \in X.$$

Thus fg is the function that results from first applying f and then applying g. The definition of the composite indicates what we gain by our convention of writing functions on the right: fg means first apply f and then apply g, and the multiple composite

$$f_1 f_2 \dots f_n$$

means first apply the function f_1, then apply f_2, then f_3, etc. This contrasts with the convention of writing functions on the left where

$$f_1 \circ f_2 \circ \dots \circ f_n$$

would mean first apply f_n, then f_{n-1}, etc. Consequently, the advantage of choosing to write function on the right is that we can read the composite of a collection of functions from left to right rather than in reverse. This will make various calculations more straightforward in what follows.

To describe the elements of symmetric groups, we need the following terminology. This is used throughout mathematics and students are recommended to become very familiar with these terms.

Definition 6.2 Let $f\colon X \to Y$ be a function.

(i) We say that f is *injective* (or *one-one*) if, for any $x, y \in X$,

$$xf = yf \qquad \text{implies} \qquad x = y.$$

(ii) We say that f is *surjective* (or *onto*) if for every $y \in Y$ there exists some $x \in X$ with $xf = y$.

(iii) We say that f is *bijective* (or *invertible*) if it is both injective and surjective.

Thus, $f\colon X \to Y$ being injective means that distinct elements of X are mapped to distinct elements of Y; that is, if $x_1, x_2 \in X$ then

$$x_1 \neq x_2 \qquad \text{implies} \qquad x_1 f \neq x_2 f.$$

In the case of f being surjective this states that every element of Y is the image of some element of X under f. Putting these together, we see that if f is bijective then this means that every element of Y is the image of some element of X and, moreover, this element of X is unique.

We shall, in a moment, explain why "invertible" is an alternative name for bijective functions, but first we illustrate these concepts with a few examples.

Example 6.3 (i) Recall that $\mathbb{Z}/2\mathbb{Z}$ is the ring of congruence classes modulo 2, so consists of two elements 0 and 1. Define a function $f\colon \mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$ by

$$xf = \begin{cases} 0 & \text{if } x \text{ is even,} \\ 1 & \text{if } x \text{ is odd.} \end{cases}$$

This function f is surjective because there are elements mapping to both elements of $\mathbb{Z}/2\mathbb{Z}$; indeed, $0f = 0$ and $1f = 1$. It is not injective because, for example, $2f = 0f = 0$.

(ii) Define a function $g\colon \mathbb{Z} \to \mathbb{Z}$ by

$$xg = 2x \qquad \text{for each } x \in \mathbb{Z}.$$

We shall first observe that g is injective. Indeed, let $x, y \in \mathbb{Z}$ and suppose $xg = yg$. This means that $2x = 2y$ which forces $x = y$ (by dividing by 2).

This function g is not surjective: there does not exist any $x \in \mathbb{Z}$ such that $xg = 1$.

(iii) The function $h\colon \mathbb{Z} \to \mathbb{Z}$ defined by $xh = -x$ for each $x \in \mathbb{Z}$ is bijective.

Indeed, if $xh = yh$, then $-x = -y$ and negating gives $x = y$, so h is injective. If $y \in \mathbb{Z}$ is arbitrary, then $(-y)h = -(-y) = y$, so h is surjective.

(iv) Finally, note that whether or not a function is injective or surjective depends on the choice of domain as well as the formula that defines it. For example, consider the function $f_1\colon \mathbb{Z} \to \mathbb{Z}$ given by

$$xf_1 = x + 1.$$

This is surjective, for if $y \in \mathbb{Z}$ is arbitrary, then $(y-1)f_1 = (y-1) + 1 = y$.

On the other hand, if $f_2\colon \mathbb{N} \to \mathbb{N}$ (where $\mathbb{N} = \{1, 2, \dots\}$) is defined by the same formula

$$xf_2 = x + 1,$$

then this function is *not* surjective. There does not exist any $x \in \mathbb{N}$ with $xf_2 = 1$.

Theorem 6.4 *Let $f\colon X \to Y$ be a function. Then f is bijective if and only if there is a function $f^{-1}\colon Y \to X$ such that*

$$ff^{-1} = \mathrm{id}_X \qquad \text{and} \qquad f^{-1}f = \mathrm{id}_Y$$

(the identity maps on X and Y, respectively).

The function f^{-1} appearing in this theorem is called the *inverse* of f. The theorem justifies the use of the term "invertible": a function f is invertible if and only if it has an inverse.

PROOF: Suppose that f is bijective. If $y \in Y$, there exists some element $x = x_y$ in X such that $xf = y$. Moreover, this x is unique because f is injective. Define $f^{-1}\colon Y \to X$ by $yf^{-1} = x$; that is, f^{-1} maps each $y \in Y$ to the unique choice of $x \in X$ with $xf = y$. Then

$$yf^{-1}f = xf = y,$$

so $f^{-1}f = \mathrm{id}_Y$ (that is, this composition is given by $y \mapsto y$).

If $x \in X$, then set $y = xf$. Then the unique element of X that maps to this y is x, so $yf^{-1} = x$ according to the above definition of f^{-1}. Hence

$$xff^{-1} = yf^{-1} = x.$$

Hence $ff^{-1} = \mathrm{id}_X$. We have verified that this function f^{-1} satisfies the required equations.

Conversely, suppose such a function f^{-1} exists. Let $x_1, x_2 \in X$ and suppose $x_1f = x_2f$. Apply f^{-1} to both sides of this equation:

$$x_1ff^{-1} = x_2ff^{-1};$$

that is,

$$x_1 = x_2$$

since ff^{-1} is the identity map on X. Hence f is injective. Now let $y \in Y$. Define $x = yf^{-1} \in X$. Then

$$xf = yf^{-1}f = y$$

(since $f^{-1}f$ is the identity map on Y). Hence f is surjective. This shows that if f has an inverse, then it is bijective. □

Symmetric groups

Now that the word "inverse" has arisen, we can introduce the groups that we are interested in here.

Definition 6.5 Let X be a set.

(i) A *permutation* of X is any bijective function $\sigma\colon X \to X$.

(ii) The *symmetric group* on X, denoted $\mathrm{Sym}(X)$, is the set of all permutations of X with composition as the binary operation.

So the elements of the symmetric group $\mathrm{Sym}(X)$ are the invertible functions from X to itself and if σ and τ are elements of $\mathrm{Sym}(X)$, their "product" is the composite $\sigma\tau$ defined as in Definition 6.1. We should start by verifying that we do indeed have a group:

Theorem 6.6 *Let X be a set. Then the symmetric group $\mathrm{Sym}(X)$ is a group.*

PROOF: The first thing we need to do is verify that composition does define a binary operation; that is, show that if σ and τ are permutations of X, then so is the composite.

Let $\sigma, \tau \in \mathrm{Sym}(X)$. The composite $\sigma\tau$ is a function $X \to X$. Let $x, y \in X$ and suppose that $x\sigma\tau = y\sigma\tau$; that is,

$$(x\sigma)\tau = (y\sigma)\tau.$$

Hence we deduce

$$x\sigma = y\sigma,$$

since τ is injective, and then

$$x = y$$

since σ is injective. This shows that $\sigma\tau$ is injective.

Now let $y \in X$ be arbitrary. Since τ is surjective, there exists $z \in X$ with $z\tau = y$. Then as σ is surjective, there exists $x \in X$ with $x\sigma = z$. Then

$$x\sigma\tau = z\tau = y.$$

This shows that $\sigma\tau$ is surjective. In conclusion, if σ and τ are bijective, then so is $\sigma\tau$. Hence composition does define a binary operation on $\mathrm{Sym}(X)$.

We now need to verify the axioms of a group.

Associativity: Let ρ, σ and τ be permutations of X. We shall show that $(\rho\sigma)\tau = \rho(\sigma\tau)$ by showing that they have the same effect on any $x \in X$. Indeed

$$x\big((\rho\sigma)\tau\big) = \big(x(\rho\sigma)\big)\tau = \big((x\rho)\sigma\big)\tau$$

and

$$x\big(\rho(\sigma\tau)\big) = (x\rho)(\sigma\tau) = \big((x\rho)\sigma\big)\tau;$$

that is, both choices of bracketing produce functions that mean first apply ρ, then σ, and then τ. Hence

$$(\rho\sigma)\tau = \rho(\sigma\tau) \qquad \text{for all } \rho, \sigma, \tau \in \mathrm{Sym}(X).$$

This shows that composition is an associative binary operation on $\mathrm{Sym}(X)$.

Identity: Let us write ε for the identity map $X \to X$; that is,

$$x\varepsilon = x \qquad \text{for all } x \in X.$$

Then ε is a bijective map (easy checks!). Indeed, $\varepsilon\varepsilon = \varepsilon$ $(= \mathrm{id}_X)$, so even Theorem 6.4 says ε is bijective with inverse equal to itself. This is the identity element in $\mathrm{Sym}(X)$: if $\sigma \in \mathrm{Sym}(X)$ then, for any $x \in X$,

$$x(\varepsilon\sigma) = (x\varepsilon)\sigma = x\sigma$$

and

$$x(\sigma\varepsilon) = (x\sigma)\varepsilon = x\sigma.$$

So $\varepsilon\sigma = \sigma\varepsilon = \sigma$ for all $\sigma \in \mathrm{Sym}(X)$, as required.

Inverses: We use Theorem 6.4. If $\sigma \in \mathrm{Sym}(X)$, then there is an inverse $\sigma^{-1}\colon X \to X$ with the property that

$$\sigma\sigma^{-1} = \sigma^{-1}\sigma = \varepsilon.$$

This formula also tells us that $\sigma^{-1} \in \mathrm{Sym}(X)$ (that is, is bijective) since it is invertible with σ as its inverse. (Alternatively, it is not too hard to check that if σ is bijective with inverse σ^{-1}, then σ^{-1} is also bijective.)

In conclusion, $\mathrm{Sym}(X)$ is a group with respect to the binary operation of composition. □

If one looks carefully at the proof of the above theorem, one will see that the following observations hold:

Proposition 6.7 *Let $f\colon W \to X$, $g\colon X \to Y$ and $h\colon Y \to Z$ be functions. Then*

(i) $(fg)h = f(gh)$;

(ii) *if f and g are injective, then the composite fg is injective;*

(iii) *if f and g are surjective, then the composite fg is surjective.* □

Here we are simply noting that (i) verifying that composition is associative in Theorem 6.6 did not depend upon the functions being bijective, (ii) verifying that the composite of two injective functions is injective does not depend upon surjectivity, and (iii) similarly for surjectivity.

Symmetric groups of finite degree

The set X appearing in Theorem 6.6 can be anything that suits. For example, we could consider $\mathrm{Sym}(\mathbb{N})$ or $\mathrm{Sym}(\mathbb{R})$, both of which are rather complicated groups. In this module, we shall concentrate our effort on the case when X is finite. This leads us to make the following definition.

Definition 6.8 Let n be a positive integer and take $X = \{1, 2, \ldots, n\}$ (a set with n elements). We shall write S_n for the symmetric group $\mathrm{Sym}(X)$ and call it the *symmetric group of degree n*.

The integer n will be fixed and we shall write $X = \{1, 2, \ldots, n\}$ throughout the rest of the chapter. If $\sigma \in S_n$, that is, σ is a permutation of X (a bijection $X \to X$), we sometimes use two-row notation to denote σ. In this notation, we write

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ 1\sigma & 2\sigma & 3\sigma & \ldots & n\sigma \end{pmatrix}$$

where we write the numbers 1, 2, …, n in the first row and then, in the second row, we write the image $k\sigma$ of k under the function σ beneath the value k. Thus

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix}$$

denotes the permutation of $\{1, 2, 3, 4\}$ which maps 1 to 2, 2 to 4, 3 to 3 and finally 4 to 1.

Theorem 6.9 *Let n be a positive integer. The symmetric group S_n of degree n is a group of order $n!$.*

PROOF: The fact that S_n is a group follows immediately from Theorem 6.6. (It is just a special case of that theorem when we take $X = \{1, 2, \ldots, n\}$.) Consider a function $X \to X$ written in two-row notation:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ 1\sigma & 2\sigma & 3\sigma & \ldots & n\sigma \end{pmatrix}$$

Now this σ is a bijection precisely when the values 1σ, 2σ, $\ldots$, $n\sigma$ are distinct and include all values from X. Provide we do list all the n elements from X in the second row, we do define a permutation of X. We could choose 1σ to be any element of X (and so there are n choices). Then having made this first choice, we can choose 2σ to be any element of X *except* the value 1σ (and so there are $n-1$ choices). Similarly there are then $n-2$ choices for 3σ, and so on. In conclusion, the number of different permutations in S_n is

$$n(n-1)(n-2)\cdots 2 \cdot 1;$$

that is,

$$|S_n| = n!.$$ □

We can calculate the composite of two permutations σ and τ using the two-row notation. For example, to perform the following calculation of the composite of two permutations

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 2 & 4 & 1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 2 & 4 & 1 \end{pmatrix} = ?$$

all we need to do is calculate the effect of applying the first permutation to an element x from X and then apply the second. (Recall that $\sigma\tau$ means first apply σ then apply τ *because* we are writing functions on the *right*.) In the above case, the first permutation maps $1 \mapsto 5$ and then the second maps $5 \mapsto 1$. The composite has the effect of doing the first and then the second, so results in $1 \mapsto 1$. Thus, the entry below 1 in the composite is also 1. We can calculate the effect of the composite on the other points in X as follows:

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

Hence

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

Note that this composite is a permutation of $X = \{1, 2, 3, 4, 5\}$: the entries in the second row are the elements of X, each occurring precisely once. This is totally unsurprising: In the proof of Theorem 6.6 we showed that the composite of two permutations is again another permutation and the above calculation is just a special case of that general fact.

Let us also calculate the composite of the above two permutations but in the reverse order:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 2 & 4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 2 & 4 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 2 & 1 & 5 \end{pmatrix}$$

(since the effect of the product is $1 \mapsto 3 \mapsto 4$, $2 \mapsto 5 \mapsto 3$, $3 \mapsto 2 \mapsto 2$, $4 \mapsto 4 \mapsto 1$ and $5 \mapsto 1 \mapsto 5$). Note that this is a different permutation to the one calculated above. Our calculations therefore show that for these two permutations σ and τ:

$$\sigma\tau \neq \tau\sigma$$

These are two permutations from the symmetric group S_5 of degree 5 and we have observed that they *do not commute*. Hence S_5 is a non-abelian group. In fact, this observation can be made more generally:

Theorem 6.10 *Let n be a positive integer. Then S_n is abelian if and only if $n = 1$ or 2.*

PROOF: The verification that S_n is abelian when $n = 1$ or 2 is left as an exercise on the problem sheet.

Suppose that $n \geqslant 3$. Consider the following two permutations from S_n:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & \dots & n \\ 2 & 1 & 3 & 4 & \dots & n \end{pmatrix} \qquad \text{and} \qquad \tau = \begin{pmatrix} 1 & 2 & 3 & 4 & \dots & n \\ 1 & 3 & 2 & 4 & \dots & n \end{pmatrix}$$

Note in particular that $k\sigma = k\tau = k$ for $k \geqslant 4$ and consequently $k\sigma\tau = k\tau\sigma = k$ for such k. We calculate the effect of the products on the points 1, 2 and 3 directly:

$$\sigma\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & \dots & n \\ 3 & 1 & 2 & 4 & \dots & n \end{pmatrix}$$
$$\tau\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & \dots & n \\ 2 & 3 & 1 & 4 & \dots & n \end{pmatrix}$$

Hence $\sigma\tau \neq \tau\sigma$ which is enough to show that S_n is non-abelian. □

Cycle notation

Although the two-row notation is relatively straightforward to understand, it is rather inefficient and it is difficult to obtain insight into the actual behaviour of permutations from it. If we were to list all $4! = 24$ elements of S_4 in two-row notation, then they would look pretty much the same and we would get little understanding. It would be nice to have a more useful way to denote permutations that gives us a bit more information. For example, the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 3 & 1 \end{pmatrix}$$

fixes 2 and 3 (that is, maps each back to themself) and swaps 1 and 4. We would like a more efficient way to describe this element that makes this more transparent. The solution is to use what is called *cycle notation*.

Definition 6.11 (i) Let $i_1, i_2, \dots, i_r$ be r distinct elements from $X = \{1, 2, \dots, n\}$ (so $1 \leqslant r \leqslant n$). The *$r$-cycle* $(i_1\ i_2\ \dots\ i_r)$ is the permutation in S_n which maps

$$i_1 \mapsto i_2, \qquad i_2 \mapsto i_3, \qquad \dots, \qquad i_{r-1} \mapsto i_r, \qquad i_r \mapsto i_1$$

and which fixes all other points in X. We also say that an r-cycle is a cycle of *length* r.

(ii) Two cycles $(i_1\ i_2\ \dots\ i_r)$ and $(j_1\ j_2\ \dots\ j_s)$ from S_n are called *disjoint* if the sets of points involved are disjoint:

$$\{i_1, i_2, \dots, i_r\} \cap \{j_1, j_2, \dots, j_s\} = \varnothing$$

(iii) The term *transposition* is used for a cycle of length 2; that is, a permutation of the form $(i\ j)$ for some distinct pair of points $i, j \in X$.

This last term, the special name used for 2-cycle, refers to the behaviour that a transposition $(i\ j)$ "transposes" the values i and j while fixing all other points in the set $X = \{1, 2, \dots, n\}$.

Example 6.12 (i) In the symmetric group S_5 of degree 5:

$$(2\ 4\ 5) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 3 & 5 & 2 \end{pmatrix}$$

(ii) Consider the following element of the symmetric group S_8 of degree 8:

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

We can see that

$$1\sigma = 3, \qquad 3\sigma = 6, \qquad 6\sigma = 2, \qquad 2\sigma = 1$$

while

$$5\sigma = 7, \qquad 7\sigma = 5$$

and

$$4\sigma = 4, \qquad 8\sigma = 8.$$

This means that σ has the effect of moving the elements of $X = \{1, 2, \ldots, 8\}$ in the same way as the product of four cycles

$$(1\ 3\ 6\ 2)\,(4)\,(5\ 7)\,(8);$$

that is,

$$\sigma = (1\ 3\ 6\ 2)\,(4)\,(5\ 7)\,(8).$$

This expresses σ as a product of four *disjoint* cycles. To justify this equation — that "equals" is the appropriate notation — we observe that the product of the cycles appearing on the right-hand side has the same effect on each point of X as σ does. Hence these functions are the same.

(iii) If $k \in X = \{1, 2, \ldots, n\}$, then the 1-cycle (k) has the effect of mapping k to itself and mapping all other elements of X also to themselves; that is, $(k) = \varepsilon = \mathrm{id}_X$ (the identity element from S_n). This means that all 1-cycles equal the identity.

As a consequence, the calculation in (ii) can be simplified:

$$\sigma = (1\ 3\ 6\ 2)\,(5\ 7)$$

expressing $\sigma \in S_8$ as a product of two disjoint cycles.

Another notation that many authors use is to write ( ) for the identity element. This fits well with the cycle notation described above.

Lemma 6.13 *Disjoint cycles commute; that is, if σ and τ are disjoint cycles in S_n then $\sigma\tau = \tau\sigma$.*

PROOF: Write $\sigma = (i_1\ i_2\ \ldots\ i_r)$ and $\tau = (j_1\ j_2\ \ldots\ j_s)$. Assume that σ and τ are disjoint:

$$\{i_1, i_2, \ldots, i_r\} \cap \{j_1, j_2, \ldots, j_s\} = \varnothing$$

Let $x \in X = \{1, 2, \ldots, n\}$. There are three possibilities (since the cycles are disjoint) and we consider them in turn:

$x = i_k$ for some k: Then $x\sigma = i_{k+1}$ (or i_1 if $k = r$), so both x and $x\sigma$ are fixed by τ. Hence $x\sigma\tau = x\sigma$ and $x\tau\sigma = x\sigma$.

$x = j_k$ for some k: Then $x\tau$ is also in $\{j_1, j_2, \ldots, j_s\}$, so x and $x\tau$ are fixed by σ. Hence $x\sigma\tau = x\tau = x\tau\sigma$.

x is not in either cycle: Then x is fixed by both σ and τ, so $x\sigma\tau = x\tau = x$ and $x\tau\sigma = x\sigma = x$.

In conclusion, $x\sigma\tau = x\tau\sigma$ for all $x \in X$. This shows $\sigma\tau = \tau\sigma$, as required. □

Theorem 6.14 *Let n be a positive integer. Every permutation from S_n can be written as a product of disjoint cycles.*

PROOF: Let $\sigma \in S_n$. Our first task is to construct the cycles that comprise σ. Define a relation $\sim$ on the set $X = \{1, 2, \ldots, n\}$ by the rule that

$$x \sim y \qquad \text{when } y = x\sigma^k \text{ for some } k \in \mathbb{Z}.$$

Thus, x is related to y when applying one of the permutations from the list

$$\ldots,\ \sigma^{-2},\ \sigma^{-1},\ \varepsilon,\ \sigma,\ \sigma^2,\ \sigma^3,\ \ldots$$

moves x to y. (Recall the convention that σ^0 is the identity map $\varepsilon = (\,)$.)

Claim: $\sim$ is an equivalence relation on X.

We verify the three conditions required of an equivalence relation.

(R) Let $x \in X$. Then $x\sigma^0 = x\varepsilon = x$ by definition of the identity map, so $x \sim x$. Hence $\sim$ is reflexive.

(S) Suppose $x, y \in X$ satisfy $x \sim y$. Therefore there exists some $k \in \mathbb{Z}$ such that $y = x\sigma^k$. Apply σ^{-k}, which is the inverse of σ^k: This yields $x = y\sigma^{-k}$ and therefore $y \sim x$. Hence $\sim$ is symmetric.

(T) Suppose $x, y, z \in X$ satisfy $x \sim y$ and $y \sim z$. Therefore there exist $j, k \in \mathbb{Z}$ such that $y = x\sigma^j$ and $z = y\sigma^k$. Then $x\sigma^{j+k} = (x\sigma^j)\sigma^k = y\sigma^k = z$, which shows that $x \sim z$. Hence $\sim$ is transitive.

Since $\sim$ is an equivalence relation, it now follows that X is the disjoint union of its equivalence classes. Suppose

$$X = C_1 \cup C_2 \cup \cdots \cup C_m$$

is the expression of X as this disjoint union. Consider one of these equivalence classes C_i (where $1 \leqslant i \leqslant m$) and pick some element $x_i \in C_i$. Consider the list of images of x_i under non-negative powers of σ:

$$x_i,\ x_i\sigma,\ x_i\sigma^2,\ x_i\sigma^3,\ \ldots \tag{6.1}$$

As our set X is finite, there must exist a smallest value r such that $x_i\sigma^r$ is a repeat of one of the earlier values in the list. Suppose that

$$x_i\sigma^r = x_i\sigma^s$$

where $0 \leqslant s < r$. Applying σ^{-s} yields $x_i\sigma^{r-s} = x_i$. The fact that r is chosen to be the *smallest* value indexing a repeat in the list (6.1) forces $r - s = r$ and hence $s = 0$. Thus

$$x_i\sigma^r = x_i. \tag{6.2}$$

Claim: $C_i = \{x_i, x_i\sigma, x_i\sigma^2, \ldots, x_i\sigma^{r-1}\}$.

Let $y \in C_i$. Then $y \sim x_i$ and so there exists $k \in \mathbb{Z}$ such that $y = x\sigma^k$. If we divide k by r, we obtain a quotient and a remainder: $k = qr + t$ where $0 \leqslant t < r$. Then

$$y = x_i\sigma^k = x_i(\sigma^r)^q\sigma^t = x_i\sigma^t$$

by repeated use of the fact that $x_i\sigma^r = x_i$ (which also implies $x_i(\sigma^r)^{-1} = x_i$ by applying the inverse). Hence $y \in \{x_i, x_i\sigma, x_i\sigma^2, \ldots, x_i\sigma^{r-1}\}$.

Conversely, each element of the form $x_i\sigma^j$ is related to x_i under $\sim$ by definition of the relation, so certainly belongs to C_i.

Now write r_i instead of r for the value appearing in the above argument. This means that for each i, this is the value such that

$$C_i = \{x_i, x_i\sigma, x_i\sigma^2, \ldots, x_i\sigma^{r_i-1}\}.$$

Now define

$$\tau = \prod_{i=1}^{m}(x_i \;\; x_i\sigma \;\; x_i\sigma^2 \;\; \ldots \;\; x_i\sigma^{r_i-1}),$$

the product in order of the cycles obtained by taking the elements in the equivalence class C_i together in each cycle for $i = 1, 2, \ldots, m$. Note that the ith cycle contains the elements from C_i and hence the permutation τ is a product of disjoint cycles because $C_i \cap C_j = \varnothing$ when $i \neq j$.

Claim: $\sigma = \tau$.

We prove these permutations are equal by showing that they have the same effect on an element $x \in X$. Now since X is the union of the equivalence classes of $\sim$, we know $x \in C_i$ for some i and hence $x = x_i\sigma^t$ where $0 \leqslant t < r_i$. Now if $t < r_i - 1$, then

$$x\sigma = (x_i\sigma^t)\sigma = x_i\sigma^{t+1},$$

while if $t = r_i - 1$, then

$$x\sigma = (x_i\sigma^{r_i-1})\sigma = x_i\sigma^{r_i} = x_i$$

(by Equation (6.2)).

In the case of the permutation τ, the cycles indexed by j, for $j \neq i$, fix all the points in C_i (since these cycles are disjoint). Hence the first $i - 1$ cycles comprising τ fix the element x, the ith cycle then moves x in the same way as σ does (as just calculated), and the remaining cycles comprising τ then fix this image. Thus

$$x\tau = x\sigma.$$

This shows $\tau = \sigma$. In conclusion, σ is indeed a product of disjoint cycles. □

Uniqueness of the decomposition: There are several obvious rearrangements that we can perform to the decomposition into cycles that Theorem 6.14 provides. Firstly since disjoint cycles commute (Lemma 6.13), we can rearrange the individual cycles within the product. Secondly, if

$$\tau = (i_1 \; i_2 \; i_3 \; \ldots \; i_r)$$

is one of the cycles that appears, we could start the cycle at any of its entries:

$$\tau = (i_k\ i_{k+1}\ \ldots\ i_r\ i_1\ \ldots\ i_{k-1})$$

for all choices of k. (This rearrangement still produces a cycle that maps each i_m to the next one.) Finally, any cycle of length 1 equals the identity and so it can be omitted from the product.

In fact, the decomposition of a permutation σ into disjoint cycles is unique up to the above possible changes. The method of proof is close to our proof of Theorem 6.14. One should observe that the equivalence classes C_i appearing in the proof are determined by the decomposition of σ into the disjoint cycles. The changes listed above arise by (i) changing the numbering of the equivalence classes C_i, (ii) changing the choice of element x_i from C_i, and (iii) not including a cycle when $|C_i| = 1$. As the C_i are specified by the permutation σ, it will follow (with a bit more work) that the decomposition into cycles is essentially unique.

Example 6.15 In this example, we shall express two permutations in S_8 as products of cycles and then in the third step show how multiplication (that is, composition) of these two permutations can be calculated in terms of the cycle decomposition.

(i)

$$\begin{aligned}\sigma &= \begin{pmatrix}1&2&3&4&5&6&7&8\\1&5&6&8&3&2&4&7\end{pmatrix}\\ &= (1)\,(2\ 5\ 3\ 6)\,(4\ 8\ 7)\\ &= (2\ 5\ 3\ 6)\,(4\ 8\ 7)\end{aligned}$$

(ii)

$$\begin{aligned}\tau &= \begin{pmatrix}1&2&3&4&5&6&7&8\\5&6&1&8&4&3&7&2\end{pmatrix}\\ &= (1\ 5\ 4\ 8\ 2\ 6\ 3)\,(7)\\ &= (1\ 5\ 4\ 8\ 2\ 6\ 3)\end{aligned}$$

(iii) We now calculate the product of σ and τ using the cycle notation by simply following the image of each element of $X = \{1, 2, \ldots, 8\}$ in the product:

$$\begin{aligned}\sigma\tau &= (2\ 5\ 3\ 6)\,(4\ 8\ 7)\,(1\ 5\ 4\ 8\ 2\ 6\ 3)\\ &= (1\ 5)\,(2\ 4)\,(3)\,(4)\,(6)\,(7\ 8)\\ &= (1\ 5)\,(2\ 4)\,(7\ 8)\end{aligned}$$

(iv) Observe

$$(1\ 2)\,(1\ 3)\ldots(1\ n) = (1\ 2\ \ldots\ n).$$

This is established by checking where each $x \in \{1, 2, \ldots, n\}$ is mapped to under the product on the left-hand side. Indeed, the same argument then shows

$$(i_1\ i_2)\,(i_1\ i_3)\ldots(i_1\ i_r) = (i_1\ i_2\ \ldots\ i_r)$$

for any distinct $i_1, i_2, \ldots, i_r \in X$. This shows that every cycle can be expressed as a product of transpositions.

Theorem 6.16 *Let n be a positive integer. Every permutation from S_n can be written as a product of some transpositions.*

PROOF: Let $\sigma \in S_n$. By Theorem 6.14, $\sigma = \sigma_1\sigma_2 \dots \sigma_m$ as a product of (disjoint) cycles. Now use the calculation in Example 6.15(iv) to write each σ_i as a product of transposition. Putting this together shows that σ is a product of transpositions. □

There are quite a few variations on this theme. We have already observed that every permutation can be expressed as a product of cycles (Theorem 6.14) and also as a product of transpositions (Theorem 6.16). The following are examples of these types of result:

Theorem 6.17 *Let n be a positive integer. Every permutation from S_n can be written as a product involving the transpositions* $(1\ 2)$, $(2\ 3)$, $\dots$, $(n{-}1\ n)$.

PROOF: In view of Theorem 6.16, it is sufficient to show that every transposition can expressed as a product involving only transpositions from the set

$$A = \{(1\ 2),\ (2\ 3),\ \dots, (n{-}1\ n)\}.$$

Accordingly, we shall establish the following claim:

Claim: The transposition $(i\ \ i{+}k)$ can be expressed as a product of transpositions in A.

We prove this claim by induction on k. When $k = 1$, there is nothing to do since $(i\ \ i{+}1) \in A$. Assume the result holds for some value k. Observe that, since the points i, $i + k$ and $i + k + 1$ are distinct, by direct calculation

$$(i{+}k\ \ i{+}k{+}1)\,(i\ \ i{+}k)\,(i{+}k\ \ i{+}k{+}1) = (i\ \ i{+}k{+}1).$$

Substituting the product involving transpositions from A that equals $(i\ \ i{+}k)$ into the left-hand side now yields the required formula for the next term $(i\ \ i{+}k{+}1)$. Hence the claim holds by induction.

In conclusion, every transposition $(i\ j)$ can be expressed as a product involving only those in A and this therefore also extends to every permutation in S_n using Theorem 6.16. □

Theorem 6.18 *Let n be a positive integer. Every permutation from S_n can be written as a product involving only the transposition* $(1\ 2)$ *and the n-cycle* $(1\ 2\ \dots\ n)$.

PROOF: Observe that

$$(1\ 2\ \dots\ n)^{-1}\,(i\ \ i{+}1)\,(1\ 2\ \dots\ n) = (i{+}1\ \ i{+}2) \tag{6.3}$$

by direct calculation. First apply the left-hand side of (6.3) to $i + 1$:

$$\begin{aligned}(i+1)\,(1\ 2\ \dots\ n)^{-1}\,(i\ \ i{+}1)\,(1\ 2\ \dots\ n) &= i\,(i\ \ i{+}1)\,(1\ 2\ \dots\ n)\\ &= (i+1)\,(1\ 2\ \dots\ n) = i+2\end{aligned}$$

and similarly the product on the left-hand side of (6.3) moves $i{+}2$ to $i{+}1$. If $k \neq i{+}1, i{+}2$, then a similar argument shows the product fixes k. Hence the formula (6.3) holds.

Repeated application of this calculation shows that every transposition of the form $(i\ \ i{+}1)$ can be expressed as a product involving only $(1\ 2)$ and $(1\ 2\ \dots\ n)$. Now using Theorem 6.17 shows that we can build every permutation using these two transpositions. □

Chapter 7

Isometries

In Theorem 6.6, we showed that the set of all permutations of a set X forms a group under composition. This, together with Proposition 6.7(i) (which says that composition is always an associative operation), starts to indicate why groups are important objects in mathematics. If, instead of just taking a set, X is any mathematical structure (say, a polyhedron, or a vector space, or a ring, or a geometrical space, etc.) then the set of bijective maps from X to itself that *preserve the structure* usually forms a group with composition as the binary operation. There is still work to be done here to verify this assertion, but associativity will follow from Proposition 6.7(i). The group concerned will consist of some of the permutations of the set X (since we are dealing with certain bijections $X \to X$) and so occurs as a subset of the group $\mathrm{Sym}(X)$. Indeed, this actually provides an example of what is called a *subgroup*, which is something that we will discuss later in Chapter 8.

Consequently, groups arise throughout mathematics as the "symmetries" of other mathematical objects and the theory we develop in this module (and those that follow) is of significance beyond algebra. In this chapter of the notes, we give examples of how groups arise in certain geometrical settings.

Isometries of the plane and of subsets of the plane

Consider distance between points in the real plane $\mathbb{R}^2$. If $\boldsymbol{v}_1 = (x_1, y_1)$ and $\boldsymbol{v}_2 = (x_2, y_2)$, then the distance between the two points is given by

$$|\boldsymbol{v}_1 - \boldsymbol{v}_2| = \sqrt{(x_2 - x_1)^2 - (y_2 - y_1)^2}.$$

In this chapter, we consider bijective functions $\mathbb{R}^2 \to \mathbb{R}^2$ that preserve this distance. Accordingly, we make the following definitions:

Definition 7.1 (i) An *isometry* of the real plane $\mathbb{R}^2$ is a bijective function $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that

$$|\boldsymbol{v}_1 f - \boldsymbol{v}_2 f| = |\boldsymbol{v}_1 - \boldsymbol{v}_2| \qquad \text{for all } \boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^2.$$

(ii) The *isometry group* $\mathrm{Isom}(\mathbb{R}^2)$ is the collection of all isometries of $\mathbb{R}^2$ with composition as binary operation.

Thus an isometry of $\mathbb{R}^2$ is an invertible map f with the property that the distance between two points is always unchanged when we apply f. They are often also called *symmetries*, but we choose not to use that term so as not to confuse with the terminology used to refer to the symmetric group.

Example 7.2 The following are examples of isometries of the plane $\mathbb{R}^2$:

(i) a *translation* $f_{\boldsymbol{a}}$ given, for fixed vector $\boldsymbol{a}$, by $\boldsymbol{v} \mapsto \boldsymbol{v} + \boldsymbol{a}$;

(ii) a *rotation* about a given point by a given angle;

(iii) a *reflection* in a given line.

Theorem 7.3 (i) *The set of isometries* $\mathrm{Isom}(\mathbb{R}^2)$ *of the plane is a group under composition.*

(ii) *Every isometry of* $\mathbb{R}^2$ *is one of: a translation, a rotation, a reflection, or the product of a translation and a reflection.*

The term *glide-reflection* is sometimes used for the composite of a translation and a reflection.

We shall not present the proof of Theorem 7.3 in this course. To complete the proof of (i), the following steps should be followed:

- Show that the composite of two isometries is again an isometry. Hence composition is a binary operation on $\mathrm{Isom}(\mathbb{R}^2)$.
- Use Proposition 6.7(i) to conclude that composition is associative as a binary operation on $\mathrm{Isom}(\mathbb{R}^2)$.
- Observe that the identity map is certainly an isometry on $\mathbb{R}^2$.
- Show that if f is an isometry, then its inverse (which exists since by assumption f is a bijection) is also an isometry.

These steps are not very difficult. The required arguments make explicit use of the definition of an isometry (Definition 7.1). Part (ii) of the theorem, however, requires more effort.

We can see that $\mathrm{Isom}(\mathbb{R}^2)$ is a rather large group. There are infinitely many choices of translations, or rotations, or reflections. We can produce smaller groups that are easier to describe by focusing on isometries that preserve a figure in the plane.

Definition 7.4 Let X be a figure in the plane. The set of isometries of $\mathbb{R}^2$ that map X to itself is called the *group of isometries* of X and we shall denote it similarly by $\mathrm{Isom}(X)$:

$$\mathrm{Isom}(X) = \{\, f \in \mathrm{Isom}(\mathbb{R}^2) \mid Xf = X \,\}$$

Similarly to Theorem 7.3, one can check that the group of isometries of the figure X is also a group. This is, however, most easily verified using the concept of a "subgroup," which is introduced in the next chapter. Accordingly, we leave the verification that the group of isometries of X is a group to Problem Sheet VIII.

We shall illustrate various groups that arise in this form.

Example 7.5 Let T be an equilateral triangle (as shown in Figure 7.1). There are six isometries of T:

Reflections: σ_a, σ_b and σ_c in the lines a, b and c, respectively;

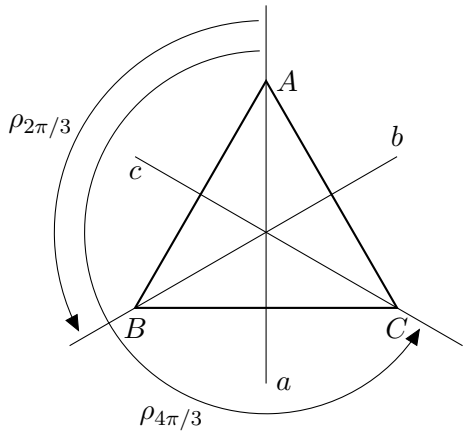

Figure 7.1: Isometries of an equilateral triangle

Rotations: $\rho_{2\pi/3}$ and $\rho_{4\pi/3}$ about the centre anticlockwise through angles of $2\pi/3$ and $4\pi/3$ radians (that is, 120° and 240°), respectively;

Identity: the identity map 1.

It is a straightforward set of calculations to determine the Cayley table for the isometry group of T:

	1	$\rho_{2\pi/3}$	$\rho_{4\pi/3}$	σ_a	σ_b	σ_c
1	1	$\rho_{2\pi/3}$	$\rho_{4\pi/3}$	σ_a	σ_b	σ_c
$\rho_{2\pi/3}$	$\rho_{2\pi/3}$	$\rho_{4\pi/3}$	1	σ_b	σ_c	σ_a
$\rho_{4\pi/3}$	$\rho_{4\pi/3}$	1	$\rho_{2\pi/3}$	σ_c	σ_a	σ_b
σ_a	σ_a	σ_c	σ_b	1	$\rho_{4\pi/3}$	$\rho_{2\pi/3}$
σ_b	σ_b	σ_a	σ_c	$\rho_{2\pi/3}$	1	$\rho_{4\pi/3}$
σ_c	σ_c	σ_b	σ_a	$\rho_{4\pi/3}$	$\rho_{2\pi/3}$	1

This calculation can be performed by keeping track of the images of the vertices A, B and C when we apply one of these isometries followed by another. It can be abbreviated by (i) recalling the defining property of the identity, (ii) observing that each reflection squares to the identity, and (iii) using the fact that each element of the group occurs once in each row and once in each column.

Example 7.6 An isoceles triangle has only two isometries: the identity transformation 1 and the reflection σ in the perpendicular bisector of the base of the triangle (see Figure 7.2). The Cayley table for group of isometries of the isosceles triangle is easy to compute:

	1	σ
1	1	σ
σ	σ	1

Example 7.7 A scalene triangle (one where all the sides have different lengths) has only one isometry: the identity function 1. The isometry group of a scalene triangle has order 1 and is trivial.

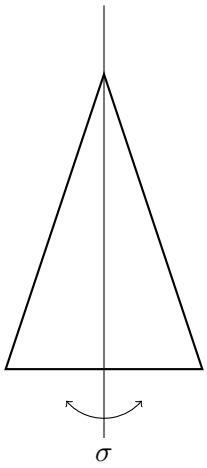

Figure 7.2: Isometries of an isoceles triangle

Example 7.8 There are infinitely many isometries of a circle: all the rotations about the centre of the circle (through any angle) and all the reflections about lines that pass through the centre of the circle.

Dihedral groups

We will now discuss an important example of the isometry group of a figure in the plane.

Definition 7.9 (Dihedral group) Let n be an integer satisfying $n \geqslant 3$. The *dihedral group* D_{2n} of order $2n$ is the isometry group of a regular polygon with n sides.

In particular, when $n = 3$, this definition defines the dihedral group D_6 to be the isometry group of an equilateral triangle T. Consequently, this group is the one we considered in Example 7.5 and which we observed is indeed a group of order 6.

Notational comment: The choice of notation for dihedral groups is not consistent in the literature. Some authors write D_n for the isometry group of a regular polygon with n sides (and so these authors are viewing the subscript as indicating the value of the parameter). Other authors use D_{2n}, as these lecture notes do, and so use the subscript to refer to the order of the group constructed.

The definition refers to the dihedral group D_{2n} (consisting of the isometries of the regular polygon with n sides) as having order $2n$. We should verify that this is indeed the case.

Theorem 7.10 *Let n be an integer with $n \geqslant 3$. The regular polygon with n sides has precisely $2n$ isometries:*

(i) ***Rotations:*** *the identity and $n - 1$ non-identity rotations about the centre;*

(ii) ***Reflections:*** *n reflections in axes through the centre.*

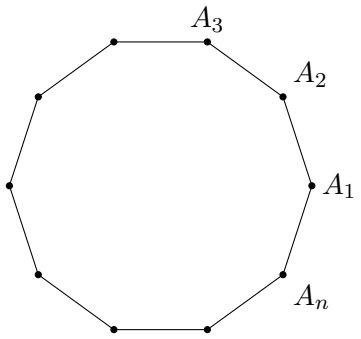

Figure 7.3: Labelling the vertices of a regular polygon

Furthermore, if ρ denotes an anticlockwise rotation through an angle $2\pi/n$ and σ denotes any of the reflections, then the elements of D_{2n} are

$$\rho^k,\ \rho^k\sigma \qquad \text{for } k = 0, 1, \ldots, n-1.$$

Proof: From the diagram of the regular polygon with n sides, one can see that there are n rotations and n reflections. This shows that $|D_{2n}| \geqslant 2n$. The question is whether there are any other isometries. Label the vertices of the regular polygon X with n sides as A_1, A_2, …, A_n (see Figure 7.3). Let $\phi \in D_{2n}$ be any isometry of X. Now ϕ must map A_1 to one of the vertices, say to A_i, and there are n possibilities for this choice. The vertices A_2 and A_n are initially adjacent to A_1 and hence ϕ must map these two vertices to the two vertices adjacent to A_i. There are at most 2 choices for which way round the images lie. The remaining vertices now have their images uniquely specified: for example, observe that if we follow a path from $A_1\phi$, through $A_2\phi$ and then to the next vertex, then this last vertex must be $A_3\phi$. Consequently there are at most $2n$ possibilities for ϕ; that is, $|D_{2n}| \leqslant 2n$.

We conclude $|D_{2n}| = 2n$ and, moreover, there are precisely n rotations and n reflections.

Furthermore, if ρ is an anticlockwise rotation through an angle of $2\pi/n$, then ρ^k is an anticlockwise rotation through an angle of $2k\pi/n$. Hence

$$1, \rho, \rho^2, \ldots, \rho^{n-1}$$

are the n distinct rotations of X. Note that they are not reflections since a reflection will reverse the cyclical order of the vertex numbering. We now claim that, for a fixed reflection σ, the elements

$$\sigma, \rho\sigma, \rho^2\sigma, \ldots, \rho^{n-1}\sigma \tag{7.1}$$

are the n reflections of X. Since $D_{2n} = \mathrm{Isom}(X)$ is a group, they are certainly elements of D_{2n}.

If $\rho^i\sigma = \rho^j\sigma$ with $0 \leqslant i, j \leqslant n-1$, then upon multiplying by σ^{-1}, we would conclude $\rho^i = \rho^j$ which forces $i = j$ (as the above rotations are distinct). If some $\rho^i\sigma$ were a rotation then $\rho^i\sigma = \rho^j$ for some j and hence $\sigma = \rho^{j-i}$, which is a contradiction as σ is a reflection. Hence the list (7.1) consists of distinct reflections and so is precisely the set of reflections.

We have now shown

$$D_{2n} = \left\{ \rho^i,\ \rho^i\sigma \;\middle|\; i = 0, 1, \ldots, n-1 \right\},$$

as required. □

Note that $\rho^n = 1$, since if one successively rotates the n-sided regular polygon n times through an angle of $2\pi/n$ then in total the angle is 2π and one is back where one started. Similarly, applying a reflection twice returns to the start so

$$\sigma^2 = 1$$

and, more generally,

$$(\rho^k \sigma)^2 = 1 \qquad \text{for all } k.$$

Note that the first of these tells us that the inverse of ρ is given by

$$\rho^{-1} = \rho^{n-1}.$$

We can also verify, by direct calculation, that the following equation holds

$$\sigma\rho = \rho^{n-1}\sigma = \rho^{-1}\sigma.$$

Using this formula, one can compute products involving the elements appearing in Theorem 7.10 directly. For example,

$$(\rho^2\sigma)(\rho\sigma) = \rho^2(\sigma\rho)\sigma = \rho^2\,\rho^{-1}\sigma\sigma = \rho.$$

Chapter 8

Subgroups

The previous chapters have introduced the concept of a group and given a variety of examples of these objects. The remainder of the module will be concerned with developing the study of groups. We shall be discussing what is often termed the "structure" of a group. The first example of such structure, discussed in this chapter, is the "subgroup." This is where one group occurs within another group and is defined formally as follows:

Definition 8.1 Let G be a group. A *subgroup* of G is a subset H of G which is itself a group under the binary operation of G. We write $H \leqslant G$ to denote that H is a subgroup of G.

Let us expand upon this definition to understand what it is actually saying. Suppose that H is a subgroup of the group G. What this means is that we have two group G and H with the properties that (i) all the elements of H are also elements of the group G, and (ii) when we calculate the product xy of two elements x and y of H we get the same answer when we calculate it within the group H and within the group G. The last point is important: It is not enough to have two groups with one inside the other; it must also be he case that the multiplications in the groups to be the same.

Example 8.2 As groups under addition, the group of integers $\mathbb{Z}$ is a subgroup of the group of rational numbers $\mathbb{Q}$, which in turn is a subgroup of the group of real numbers $\mathbb{R}$, which in turn is a subgroup of the group of complex numbers $\mathbb{C}$. Indeed,

$$\mathbb{Z} \leqslant \mathbb{Q} \leqslant \mathbb{R} \leqslant \mathbb{C}$$

and any one of these is a subgroup of any that contains it. (Note here that the binary operation is the same: It is addition.)

Example 8.3 Since the rational numbers $\mathbb{Q}$ form a field, the non-zero rational numbers form a group $\mathbb{Q}^*$ under multiplication. Certainly $\mathbb{Q}^*$ is a subset of $\mathbb{Q}$ and the latter is a group under addition. However, $\mathbb{Q}^*$ is not a subgroup of $\mathbb{Q}$, since the binary operations are different in these two groups.

We are guaranteed some subgroups always exist in a group as the following shows:

Lemma 8.4 *Let G be a group with identity element 1. Then $\{1\}$ and G are subgroups of G.*

PROOF: Both $\{1\}$ and G are certainly subsets of G. It then immediately follows that G is a subgroup of itself (for it is a group and has the same multiplication as itself). The multiplication in $\{1\}$ is inherited from G:

$$1 \times 1 = 1$$

We need to observe that $\{1\}$ is indeed a group under this operation. Associativity follows trivially, 1 is the identity for this multiplication on $\{1\}$ and $1^{-1} = 1$. Hence $\{1\}$ is a subgroup of itself. □

The proof just given is actually more complicated than it needs to be. In fact, there is a straightforward method to check whether a non-empty subset H of a group G is a subgroup. Indeed, this will be the method that we use to check for subgroups in the future.

Theorem 8.5 *The following are equivalent for a subset H of a group G:*

(i) *H is a subgroup of G;*

(ii) *H is a non-empty subset of G such that $xy \in H$ and $x^{-1} \in H$ whenever $x, y \in H$.*

This result is often phrased in the following way:

> *A subgroup of a group is a non-empty subset that is closed under multiplication and inverses.*

PROOF: (i) ⇒ (ii): Suppose that H is a subgroup of G. Then H is a group, so it contains an identity element for its operation and certainly then H is non-empty. Furthermore, if x and y are elements of H, then the multiplication of x and y as elements of H coincides with the value obtained when we multiply them as elements of G; that is, it equals the product xy (as calculated in G) and so this product belongs to H. This shows

$$xy \in H \qquad \text{for all } x, y \in H.$$

Now there is some element $e \in H$ that behaves as an identity element for the group H. In particular, $e^2 = e$ (which is the same whether we calculate in H or in G) and multiplying by the inverse of e as an element of G gives

$$e = e^2 e^{-1} = e e^{-1} = 1.$$

Thus the identity element of H coincides with that of G. In particular,

$$\textit{the identity element } 1 \textit{ of } G \textit{ belongs to the subgroup } H.$$

Now if $x \in H$, then it has some inverse $h \in H$ in the group H. Then $xh = hx = 1$ (since we have just shown the identity element of H is the identity of G). Thus this inverse of x as an element of H is also its inverse as an element of G. Uniqueness of inverses (Theorem 5.16(ii)) tells us that $h = x^{-1}$ (the inverse of x as calculated in G). Hence

$$x^{-1} \in H \qquad \text{for all } x \in H.$$

This completes this part of the proof.

(ii) ⇒ (i): Conversely suppose H is a non-empty subset of G such that $xy, x^{-1} \in H$ for all $x, y \in H$. (These are the product and inverse calculated in G. Indeed at this stage, we

have not yet shown H is a group.) Then, in particular, the multiplication of G induces a binary operation on H:

$$\begin{aligned} H \times H &\to H \\ (x, y) &\mapsto xy \end{aligned}$$

(The hypothesis is used to ensure this operation does take values in H.) As the original binary operation on G is associative, this is inherited by the multiplication on H:

$$(xy)z = x(yz) \qquad \text{for all } x, y, z \in H,$$

since this actually holds for all elements of G.

By assumption, H is non-empty, so it contains at least one element a. Then $a^{-1} \in H$ by hypothesis and then $aa^{-1} \in H$; that is, $1 \in H$. This element satisfies $1x = x1 = x$ for all $x \in H$ (as this holds for all elements of G) and we conclude that 1 is also the identity element for H.

Finally, if $x \in H$, then by hypothesis the element x^{-1} also belongs to H and satisfies $xx^{-1} = x^{-1}x = 1$. Thus H contains an inverse for each of its elements.

In conclusion, if (ii) holds, then H has the structure of a group under multiplication inherited from that from G; that is, H is a subgroup of G. □

Before we discuss how to use this characterization of what it means to be a subgroup, we shall record some observations made in the above proof:

Lemma 8.6 *Let G be a group and H be a subgroup of G. Then*

(i) *H contains the identity element 1 of G, and*

(ii) *the inverse of an element $x \in H$ is the same whether we consider it as belonging to the group H or the group G.* □

As a consequence, when we have a subgroup of a group, we can simply refer to "the identity element" and to "inverses" without having to specify which of the two groups we are working within since they are the same.

Example 8.7 Let us first illustrate how Theorem 8.5 can be used to establish what we observed in Example 8.2 and Lemma 8.4.

(i) Consider the additive group of real numbers $\mathbb{R}$ and the subset of integers $\mathbb{Z}$. Certainly $\mathbb{Z}$ is a non-empty subset and if $x, y \in \mathbb{Z}$ then the sum $x + y$ is an integer and the negative $-x$ is an integer. Hence the conditions of Theorem 8.5 tells us that $\mathbb{Z}$ is a subgroup of $\mathbb{R}$. Similar arguments apply to the other subgroups listed in Example 8.2.

(ii) If G is any group, then

- $1 \times 1 = 1$ and $1^{-1} = 1$, so $\{1\}$ is a subgroup of G by the conditions of Theorem 8.5.
- Certainly $xy, x^{-1} \in G$ for all $x, y \in G$, so G is a subgroup of itself.

Example 8.8 Recall that the collection of invertible $n \times n$ matrices over a field F forms a group, which is called the *general linear group* $\mathrm{GL}_n(F)$. Consider the following subset

$$\mathrm{SL}_n(F) = \{\, A \in \mathrm{GL}_n(F) \mid \det A = 1 \,\},$$

the set of all $n \times n$ matrices over F with determinant 1. Observe that the identity matrix I has determinant 1, so belongs to $\mathrm{SL}_n(F)$. In particular, $\mathrm{SL}_n(F)$ is a non-empty subset of $\mathrm{GL}_n(F)$. Let $A, B \in \mathrm{SL}_n(F)$. Then

$$\det(AB) = (\det A)(\det B) = 1 \times 1 = 1$$

so $AB \in \mathrm{SL}_n(F)$. Also since $AA^{-1} = I$, upon taking determinant we conclude

$$\det A^{-1} = \frac{\det I}{\det A} = \frac{1}{1} = 1,$$

so $A^{-1} \in \mathrm{SL}_n(F)$. In conclusion, $\mathrm{SL}_n(F)$ satisfies the conditions of Theorem 8.5, so it is a subgroup of $\mathrm{GL}_n(F)$. We call $\mathrm{SL}_n(F)$ the *special linear group* of degree n over F.

Example 8.9 Consider the following subset of the symmetric group S_4 of degree 4:

$$H = \{\, (\,),\ (1\ 2)(3\ 4),\ (1\ 3)(2\ 4),\ (1\ 4)(2\ 3) \,\}$$

(where we express the elements in terms of the cycle decomposition). We construct the multiplication table of the elements of H:

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

The table shows that H is closed under products and inverses: Indeed, it shows $x^{-1} = x \in H$ for all $x \in H$. Hence H is a subgroup of S_4.

Cyclic subgroups

We shall now consider a particular type of subgroup that we can find in a group.

Definition 8.10 Let G be a group and x be an element of G. The *cyclic subgroup* of G generated by x is the set

$$\langle x \rangle = \{\, x^m \mid m \in \mathbb{Z} \,\};$$

that is, $\langle x \rangle$ consists of all powers of x.

The group G is called *cyclic* when $G = \langle x \rangle$ for some x in G.

We justify the terminology:

Theorem 8.11 *Let G be a group and $x \in G$. The cyclic subgroup $\langle x \rangle$ generated by x is a subgroup of G.*

Proof: Certainly $x = x^1$ is an element of $\langle x \rangle$, so this is a non-empty subset of G. Let x^m, x^n be two elements of $\langle x \rangle$. Then, by the power laws (Theorem 5.20),

$$x^m x^n = x^{m+n} \in \langle x \rangle \qquad \text{and} \qquad (x^m)^{-1} = x^{-m} \in \langle x \rangle.$$

Hence $\langle x \rangle$ is a subgroup of G. □

To aid us in describing the elements of cyclic subgroups, we shall introduce some further terminology:

Definition 8.12 Let G be a group and x be an element of G. The *order* of x is the smallest positive integer n such that $x^n = 1$ if such n exists. If no such n exists, we say that x has *infinite order*. We shall write $o(x)$ for the order of x.

Comment: Be aware that we are using the word "order" for two things. The order of a group (or, indeed, a subgroup of a group) is the number of elements in that group. The order of an element is different as specified in the above definition. We shall explain shortly what the link is between these two apparent different uses of the same word.

Example 8.13 (i) Recall the dihedral group D_{2n} of order $2n$. It contains n rotations, which are the powers of the rotation ρ defined in Theorem 7.10, and n reflections. If σ is any reflection, then $\sigma^2 = 1$. Hence $o(\sigma) = 2$. (Note $\sigma^1 \neq 1$ since σ is not the identity.)

In the case of the basic rotation ρ, its powers

$$1, \rho, \rho^2, \ldots, \rho^{n-1}$$

are the distinct rotations in D_{2n} while $\rho^n = 1$. Consequently $o(\rho) = n$.

(ii) Let $\sigma = (1\ 2)\,(3\ 4\ 5) \in S_5$. We calculate the powers of σ:

$$\begin{aligned} \sigma^2 &= (3\ 5\ 4) & \qquad \sigma^5 &= (1\ 2)\,(3\ 5\ 4) \\ \sigma^3 &= (1\ 2) & \qquad \sigma^6 &= (\,) \\ \sigma^4 &= (3\ 4\ 5) \end{aligned}$$

Hence $o(\sigma) = 6$.

The following result gives more information about the order of an element and tells us what it tells us.

Theorem 8.14 *Let G be a group and x be an element of G.*

(i) *Suppose $o(x) < \infty$. Then, for any integer k, $x^k = 1$ if and only if the order of x divides k.*

(ii) *The number of elements in the cyclic subgroup $\langle x \rangle$ generated by x equals the order $o(x)$ of the element x.*

We now have the link between "order" as applied to a subgroup and "order" as applied to an element. The order of the subgroup generated by the element x is the same as the order of the element x.

Proof: (i) Suppose $o(x)$ divides k, so $k = q \cdot o(x)$ for some $q \in \mathbb{Z}$. Then

$$x^k = x^{q \cdot o(x)} = (x^{o(x)})^q = 1^q = 1.$$

Conversely, suppose $x^k = 1$. Divide k by $o(x)$ to produce a quotient and remainder:

$$k = q \cdot o(x) + r \qquad \text{where } 0 \leqslant r < o(x).$$

Then

$$1 = x^k = x^{q\,o(x)+r} = (x^{o(x)})^q x^r = 1^q x^r = x^r.$$

Note $o(x)$ is, by definition, the smallest positive integer such that $x^{o(x)} = 1$. Hence, since $r < o(x)$, it must be the case that $r = 0$. Thus $k = q \cdot o(x)$; that is, $o(x)$ divides k.

(ii) Consider the set of powers of x:

$$\langle x \rangle = \{\, x^m \mid m \in \mathbb{Z} \,\}$$

If the set of powers has repeats, say $x^m = x^n$ for some $m, n \in \mathbb{Z}$ with $m < n$, then upon multiplying by x^{-m} we deduce

$$x^{n-m} = x^n x^{-m} = x^m x^{-m} = x^0 = 1.$$

This shows that there exists some power of x that equals the identity and hence x has finite order. Consequently, *if x has infinite order*, then the powers of x are distinct and

$$|\langle x \rangle| = \infty = o(x).$$

Suppose now that $o(x) < \infty$. Let m be any integer and divide it by $o(x)$ to give a quotient and remainder:

$$m = q \cdot o(x) + r \qquad \text{where } 0 \leqslant r < o(x).$$

Then

$$x^m = x^{q\,o(x)+r} = (x^{o(x)})^q x^r = 1^q x^r = x^r$$

so every power of x equals one of the following elements

$$1 = x^0,\ x,\ x^2,\ \ldots,\ x^{o(x)-1}.$$

Moreover these elements are distinct, since if $x^i = x^j$ with $0 \leqslant i < j < o(x)$, then we would conclude $x^{j-i} = 1$. However this exponent satisfies $0 < j - i < o(x)$, contradicting the definition of the order $o(x)$. Thus

$$\langle x \rangle = \{1, x, x^2, \ldots, x^{o(x)-1}\}$$

contains precisely $o(x)$ many elements. □

We shall now improve on the observation made in Example 8.13(ii) by establishing the order of an arbitrary permutation in the symmetric group S_n of degree n. This result also indicates what we have gained by learning how to write a permutation as a product of disjoint cycles.

Theorem 8.15 *Let n be a positive integer.*

(i) *If σ is a cycle in S_n of length r, then $o(\sigma) = r$.*

(ii) *Let σ be an arbitrary permutation in S_n and suppose that $\sigma = \sigma_1\sigma_2 \ldots \sigma_k$ as a product of disjoint cycles where the cycle σ_i has length r_i. Then*

$$o(\sigma) = \operatorname{lcm}(r_1, r_2, \ldots, r_k),$$

the lowest common multiple *of the lengths r_i.*

The *lowest common multiple* of $r_1, r_2, \ldots, r_k$ is the smallest positive integer m that is divisible by each one of $r_1, r_2, \ldots, r_k$.

Proof: (i) Suppose that $\sigma = (i_1\ i_2\ \ldots\ i_r)$. Then

$$i_1\sigma^r = i_2\sigma^{r-1} = i_3\sigma^{r-2} = \cdots = i_r\sigma = i_1$$

and similarly $i_k\sigma^r = i_k$ for $k = 2, 3, \ldots, r$. By definition, σ fixes all other points of $X = \{1, 2, \ldots, n\}$ and we conclude that

$$x\sigma^r = x \qquad \text{for all } x \in X.$$

Thus $\sigma^r = (\,)$. On the other hand, if $1 \leqslant k < r$ then

$$i_1\sigma^k = i_{k+1} \neq i_1,$$

so $\sigma^k \neq (\,)$. We conclude $o(\sigma) = r$, as claimed.

(ii) Let $M = \operatorname{lcm}(r_1, r_2, \ldots, r_k)$ where r_i is the length of the cycle σ_i appearing in σ. Then, in particular, $r_i = o(\sigma_i)$ divides M and so

$$\sigma_i^M = (\,)$$

using part (i) and Theorem 8.14(i). Now the cycles σ_i are disjoint and so they commute by Lemma 6.13. Hence we may rearrange the terms in a product so

$$\begin{aligned}\sigma^M &= (\sigma_1\sigma_2\ldots\sigma_k)^M \\ &= \sigma_1^M\sigma_2^M\ldots\sigma_k^M \\ &= (\,).\end{aligned}$$

Hence $o(\sigma) \leqslant M$.

On the other hand, let $N = o(\sigma)$. Then, again using the fact that disjoint cycles commute

$$(\,) = \sigma^N = (\sigma_1\sigma_2\ldots\sigma_k)^N = \sigma_1^N\sigma_2^N\ldots\sigma_k^N.$$

Now if x is one of the points appearing in the cycle σ_j then x is fixed by all the other cycles involved (as they are disjoint). Therefore

$$x\sigma_j^N = x\sigma_1^N\sigma_2^N\ldots\sigma_k^N = x.$$

As σ_j^N fixes all points not occurring in the cycle σ_j, we now conclude

$$\sigma_j^N = (\,).$$

Hence the order of σ_j divides N by Theorem 8.14(i); that is, r_j divides N. This will be true for all choices of j and consequently N is a common multiple of the r_j. Therefore

$$M = \operatorname{lcm}(r_1, r_2, \ldots, r_k) \leqslant N = o(\sigma).$$

These two inequalities now established then yield the claimed result. □

This theorem would immediately produce the observation made in Example 8.13(ii). The permutation $\sigma = (1\ 2)\,(3\ 4\ 5)$ in S_5 has order 6 as it has two cycles of lengths 2 and 3, respectively, and the lowest common multiple of these lengths is 6.

Example 8.16 (i) Consider the following permutation of S_{10}:

$$\begin{aligned}\sigma &= \begin{pmatrix}1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10\\ 4 & 5 & 3 & 8 & 10 & 9 & 1 & 7 & 6 & 2\end{pmatrix} \\ &= (1\ 4\ 8\ 7)\,(2\ 5\ 10)\,(6\ 9).\end{aligned}$$

Then σ is a product of three cycles of lengths 4, 3 and 2. Hence

$$o(\sigma) = \operatorname{lcm}(4, 3, 2) = 12.$$

(ii) Consider the symmetric group S_3 of degree 3. We know from Theorem 6.9 that this is group of order $3! = 6$. We can list the elements of S_3 in cycle notation:

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

Thus S_3 contains the identity, three 2-cycles and two 3-cycles. Hence there are one element of order 1, three elements of order 2 and two elements of order 3 in S_3. We can therefore construct a number of cyclic subgroups of S_3:

$$\begin{aligned} \langle ( ) \rangle &= \{ ( ) \} \qquad \text{(the trivial subgroup)} \\ \langle (1\ 2) \rangle &= \{ ( ), (1\ 2) \} \\ \langle (1\ 3) \rangle &= \{ ( ), (1\ 3) \} \\ \langle (2\ 3) \rangle &= \{ ( ), (2\ 3) \} \\ \langle (1\ 2\ 3) \rangle &= \langle (1\ 3\ 2) \rangle = \{ ( ), (1\ 2\ 3), (1\ 3\ 2) \} \end{aligned}$$

The 3-cycles (1 2 3) and (1 3 2) are the square (and indeed also the inverse) of each other and consequently they generate the same cyclic subgroup.

Alternating groups

We shall now construct a very important example of subgroup that occurs within the finite symmetric group S_n. Recall from Theorem 6.16 that every permutation in S_n can be written as a product of transpositions. There is no claim that the transpositions occurring in this theorem are disjoint. Indeed, they are not: we needed overlapping transpositions to construct an r-cycle as in Example 6.15(iv). Equally, the expression for a permutation in terms of transpositions is not unique. For example,

$$(2\ 3) = (1\ 2)\,(2\ 3)\,(1\ 3).$$

There is, however, an aspect of uniqueness in the decomposition. We shall observe that the parity of the number of transpositions involved (that is, whether it is odd or even) is uniquely determined.

Theorem 8.17 *Let σ be a permutation in the symmetric group S_n of degree n. The number of transpositions occurring in a product that equals σ is either always odd or always even.*

Proof: To establish this result, we shall make use of some techniques relating to matrices. For $i = 1, 2, \ldots, n$, let $\boldsymbol{e}_i$ be the row vector with all entries 0 except for the ith entry which equals 1:

$$\boldsymbol{e}_1 = (1, 0, 0, \ldots, 0), \quad \boldsymbol{e}_2 = (0, 1, 0, \ldots, 0), \quad \ldots, \quad \boldsymbol{e}_n = (0, 0, \ldots, 0, 1).$$

Thus, in the language of *MT2501 Linear Mathematics*, $\mathscr{B} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ is a basis for the space $\mathbb{R}^n$ of row vectors of length n.

Now if $\sigma \in S_n$, define A_σ to be the $n \times n$ matrix with a 1 in entry $(i, i\sigma)$ for $i = 1, 2, \ldots, n$ and all other entries equal 0. (Thus each row of A_σ has a single non-zero entry and each column has a single non-zero entry.) The choice of A_σ ensures that

$$\boldsymbol{e}_i A_\sigma = \boldsymbol{e}_{i\sigma}$$

for $i = 1, 2, \ldots, n$. Observe that if $\sigma, \tau \in S_n$, then

$$\boldsymbol{e}_i A_\sigma A_\tau = \boldsymbol{e}_{i\sigma} A_\tau = \boldsymbol{e}_{(i\sigma)\tau} = \boldsymbol{e}_{i(\sigma\tau)} = \boldsymbol{e}_i A_{\sigma\tau}$$

(using the definition that the product $\sigma\tau$ of the two permutations is their composite). Hence

$$A_{\sigma\tau} = A_\sigma A_\tau \tag{8.1}$$

for all permutations $\sigma, \tau \in S_n$. (Equation (8.1) could also be established by careful examination of the entries when we multiply the two matrices A_σ and A_τ, but the calculation is a little easier if one computes the effect on the basis vectors.)

Now suppose $\sigma = \tau_1 \tau_2 \ldots \tau_m$ expresses σ as a product of m transpositions. Then, by repeated use of Equation (8.1),

$$A_\sigma = A_{\tau_1} A_{\tau_2} \ldots A_{\tau_m}.$$

Let $\tau = (k\ \ell)$ be a transposition. Then the matrix A_τ has the form

$$A_\tau = \begin{pmatrix} 1 & & & & & & & & & \\ & \ddots & & & & & & & & \\ & & 1 & & & & & & & \\ & & & 0 & & & & 1 & & \\ & & & & 1 & & & & & \\ & & & & & \ddots & & & & \\ & & & & & & 1 & & & \\ & & & 1 & & & & 0 & & \\ & & & & & & & & 1 & \\ & & & & & & & & & \ddots \\ & & & & & & & & & & 1 \end{pmatrix}$$

(where all entries not shown equal 0); that is, A_τ is obtained from the identity matrix by swapping row k and row ℓ. Hence, by the rules of determinants,

$$\det A_\tau = -\det I = -1$$

for every transposition τ. Therefore, using the rule $\det(AB) = (\det A)(\det B)$ for determinants[1], it follows that

$$\det A_\sigma = (\det A_{\tau_1})(\det A_{\tau_2}) \ldots (\det A_{\tau_m}) = (-1)^m.$$

The matrix A_σ is uniquely determined by the permutation σ and hence its determinant cannot depend upon how we write it as a product of transpositions. Therefore the number m of transpositions involved must be either always odd (when $\det A_\sigma = -1$) or always even (when $\det A_\sigma = 1$). This establishes the theorem. □

[1] If you consult a textbook on linear algebra for the proof of this rule, you might find that the book uses the formula $(*)$ $\det A = \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma) a_{1,1\sigma} a_{2,2\sigma} \ldots a_{n,n\sigma}$ for the determinant of a matrix $A = [a_{ij}]$, where $\operatorname{sgn} \sigma = \pm 1$ depending upon whether σ is an even or an odd permutation. Accordingly, the proof of Theorem 8.17 given might appear circular since it uses determinant to show that this "sign" function is defined. However, with some care one can establish the determinant rule by induction on the size of the matrix without using the formula $(*)$ and, provided that one does this, then we are permitted to use this rule to establish the above theorem.

Relying upon the above theorem to tell us the concept makes sense, we can now make the following definition:

Definition 8.18 Let σ be a permutation in the symmetric group S_n of degree n.

(i) We say that σ is an *even permutation* if it can be written as a product of an even number of transpositions.

(ii) We say that σ is an *odd permutation* if it can be written as a product of an odd number of transpositions.

Indeed, Theorem 8.17 tells us that a permutation is either odd or even and cannot be both.

Lemma 8.19 *Let σ be an r-cycle in the symmetric group S_n. Then*

(i) *if r is odd, then σ is an even permutation;*

(ii) *if r is even, then σ is an odd permutation.*

This lemma might look contradictory, but it is important to remember that being an odd or even permutation refers to how many transpositions are involved in the product, *not* the length of cycles!

Proof: This follows immediately from the formula

$$(i_1\ i_2\ \ldots\ i_r) = (i_1\ i_2)\,(i_1\ i_3)\ldots(i_1\ i_r)$$

from Example 6.15(iv) that expresses an r-cycle as a product of $r-1$ transpositions. □

Definition 8.20 Let n be a positive integer. The *alternating group* A_n of degree n is the set of even permutations viewed as a group under composition.

Theorem 8.21 *Let n be a positive integer.*

(i) *The alternating group A_n is a subgroup of the symmetric group S_n.*

(ii) *If $n \geqslant 2$, then $|A_n| = \frac{1}{2}|S_n| = \frac{1}{2}n!$.*

As a consequence, A_n is indeed a group under the binary operation inherited from S_n.

Proof: (i) We take the convention that the identity is a product of zero transpositions, so $(\,) \in A_n$. (This is only necessary when $n = 1$. For $n \geqslant 2$, the identity is also the product of two transpositions: $(\,) = (1\ 2)\,(1\ 2)$.) In particular, A_n is non-empty.

Now let $\sigma, \tau \in A_n$. Then σ is a product of an even number, say k, of transpositions and τ is a product of an even number, say ℓ, of transpositions. Then $\sigma\tau$ can be expressed as a product of $k + \ell$ transpositions and this is still even. Hence $\sigma\tau \in A_n$.

If $\sigma = \sigma_1\sigma_2\ldots\sigma_k$ as a product of k transpositions, then

$$\begin{aligned}\sigma^{-1} &= (\sigma_1\sigma_2\ldots\sigma_k)^{-1}\\ &= \sigma_k^{-1}\sigma_{k-1}^{-1}\ldots\sigma_1^{-1}\\ &= \sigma_k\sigma_{k-1}\ldots\sigma_1\end{aligned}$$

is also a product k transpositions (with use of Theorem 5.16(iv)). Hence $\sigma^{-1} \in A_n$.

An application of the subgroup test (Theorem 8.5) shows that the alternating group is a subgroup of S_n.

(ii) Let O_n be the set of odd permutations in S_n. Then $S_n = A_n \cup O_n$ expresses the symmetric group as a union of two disjoint sets: the even permutations and the odd permutations. If $\sigma \in A_n$ is any even permutation, then $\sigma\,(1\ 2)$ is an odd permutation since can be expressed as a product of one more transposition than those involved when expressing σ. Hence we can define a function $\phi\colon A_n \to O_n$ by

$$\phi\colon \sigma \mapsto \sigma\,(1\ 2).$$

Claim: ϕ is a bijection.

Use of right cancellativity (Lemma 5.17(i)) shows that ϕ is injective: if $\sigma\,(1\ 2) = \tau\,(1\ 2)$, then necessarily $\sigma = \tau$. If τ is any odd permutation, then the product $\tau\,(1\ 2)$ is even and

$$\big(\tau\,(1\ 2)\big)\phi = \tau\,(1\ 2)\,(1\ 2) = \tau.$$

Hence ϕ is surjective. Putting these together, we conclude ϕ is bijective.

It follows that A_n and O_n contain the same number of elements:

$$|O_n| = |A_n|$$

The fact that S_n is the disjoint union of A_n and O_n then gives us $|S_n| = 2|A_n|$ and hence

$$|A_n| = \tfrac{1}{2}|S_n| = \tfrac{1}{2}n!$$

as claimed. □

Chapter 9

Lagrange's Theorem

In the previous chapter, we introduced the concept of a subgroup of a group. We gave various examples of subgroups. One might wonder about what level of restriction there is on which subsets can be subgroups. The purpose of this chapter is to give a greater understanding about the structure of a group and how the presence of a subgroup actually gives quite a rigid structure. In particular, the main result of the chapter will tell us that a subgroup of a finite group G must have order dividing the order of G. This means that there is indeed restriction over which subgroups can exist.

Cosets

We will be working with a fixed group G with a fixed subgroup H throughout this chapter. We begin with the following definition:

Definition 9.1 Let G be a group and H be a subgroup of G. If $x \in G$, the *right coset* of H with *representative* x is the following subset of G:

$$Hx = \{\, hx \mid h \in H \,\};$$

that is, Hx consists of all the products hx as h ranges over all the elements of H.

Similarly, the *left coset* of H with *representative* x is the subset

$$xH = \{\, xh \mid h \in H \,\}.$$

We shall work almost entirely with right cosets in this chapter. Very similar results hold for left cosets, but we shall not state all the relevant facts here.

Example 9.2 (i) Recall from Example 8.16(ii) that

$$H = \langle (1\ 2) \rangle = \{(\,), (1\ 2)\}$$

is a subgroup of S_3. (This is the cyclic subgroup generated by an element of order 2, so it contains two elements.) We shall calculate the right cosets $H\sigma$ of H for all possible choices of representative σ from S_3:

$$\begin{aligned} H(\,) &= \{(\,), (1\ 2)\} & H(1\ 2) &= \{(1\ 2), (\,)\} \\ H(1\ 3) &= \{(1\ 3), (1\ 2\ 3)\} & H(2\ 3) &= \{(2\ 3), (1\ 3\ 2)\} \\ H(1\ 2\ 3) &= \{(1\ 2\ 3), (1\ 3)\} & H(1\ 3\ 2) &= \{(1\ 3\ 2), (2\ 3)\} \end{aligned}$$

Observe that the right cosets we have calculated have the property that any two of them are either the same or are disjoint. This will turn out to be a general property of cosets.

(ii) Take $H = \{1\}$ to be the trivial subgroup of a group G. If $x \in G$, then $Hx = \{x\}$ is a single element of G.

(iii) Take $H = G$ to be the same as the original group G. Then $Hx = G$ for any choice of x. This follows from Lemma 5.17 that tells us that every element g of G does occur as a product $yx = g$ for some $y \in G$.

We shall now establish the basic properties of right cosets.

Theorem 9.3 *Let G be a group and H be a subgroup of G.*

(i) *If $x, y \in G$, then $Hx = Hy$ if and only if $xy^{-1} \in H$.*

(ii) *Two right cosets of H in G are either equal or disjoint.*

(iii) *The group G is the disjoint union of the right cosets of H.*

(iv) *Every right coset of H in G contains the same number of elements as the subgroup H: if $x \in G$, then $|Hx| = |H|$.*

Proof: (i) Suppose first that $Hx = Hy$. Now $x = 1x \in Hx$ (because the identity element 1 belongs to H) and therefore $x \in Hy$. Therefore $x = hy$ for some $h \in H$. Then

$$xy^{-1} = h \in H.$$

Conversely, suppose $xy^{-1} \in H$. We shall show that every element of the right coset Hx also belongs to Hy and *vice versa*. Let $h \in H$ so that hx is a typical element of the coset Hx. Then

$$hx = h(xy^{-1})y \in Hy$$

because $h(xy^{-1}) \in H$ since we know $h, xy^{-1} \in H$ and the subgroup is closed under multiplication. Thus every element of Hx also belongs to Hy, so $Hx \subseteq Hy$.

Similarly, if $h \in H$, then

$$hy = h(yx^{-1})x = h(xy^{-1})^{-1}x \in Hx$$

since H is closed under products and inverses so $h(xy^{-1})^{-1} \in H$. This shows every element of Hy also belongs to Hx, so $Hy \subseteq Hx$.

In conclusion, if $xy^{-1} \in H$, then $Hx = Hy$. This completes the proof of (i).

(ii) Consider two cosets Hx and Hy of H in G. These cosets are disjoint when $Hx \cap Hy = \varnothing$. Let us suppose that $Hx \cap Hy \neq \varnothing$, so there exists some element $g \in Hx \cap Hy$. Since $g \in Hx$, we know that $g = hx$ for some $h \in H$. Equally, $g = ky$ for some $k \in H$. Then

$$hx = ky$$

so

$$xy^{-1} = h^{-1}k \in H.$$

Now part (i) tells us that $Hx = Hy$. This shows that if two cosets are not disjoint then they are equal and so we have established (ii).

(iii) If x is any element of G, then x belongs to one of the right cosets of H: Indeed, $x = 1x \in Hx$. Hence G is the union of right cosets of H and part (ii) tells us this is a disjoint union.

(iv) Define a mapping $\phi\colon H \to Hx$ by

$$\phi\colon h \mapsto hx.$$

Right cancellativity (Lemma 5.17(i)) shows that ϕ is injective: if $hx = kx$ for $h, k \in H$, then $h = k$. The map ϕ is surjective since by definition the coset Hx consists of precisely the elements that occur as images of ϕ. Hence ϕ is a bijection and therefore

$$|Hx| = |H|.$$

This completes the proof of the theorem. □

We shall use the following terminology relating to cosets:

Definition 9.4 Let G be a group and H be a subgroup of G. The *index* of H in G is the number of right cosets of H occurring in G. We shall denote it by $|G : H|$.

Comment: Some sources use square brackets, $[G : H]$, when denoting the index of H in G. It is also the case that the index of H in G is also equal to the number of left cosets that H has in G. (This is verified in Question 9 on Problem Sheet IX.) Consequently, Definition 9.4 does not actually depend upon the choice of using right cosets rather than left cosets.

Let us now interpret Theorem 9.3. It tells us that the group G is a disjoint union of the right cosets of the subgroup H. There are $|G : H|$ many of these cosets and each coset contains $|H|$ many elements. Therefore the number of elements in G equals the product of the number of cosets by this common size of cosets. This establishes the following theorem:

Theorem 9.5 (Lagrange's Theorem[1]) *Let G be a group and H be a subgroup of G. Then*

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

In particular, if G is a finite group, then the order of H divides the order of G. □

Observe that the theorem also shows that we can calculate the number of cosets of a subgroup H of a finite group G by the formula

$$|G : H| = |G|/|H|.$$

Example 9.6 (i) In Example 9.2, we observed that the subgroup $H = \langle(1\ 2)\rangle = \{(\,), (1\ 2)\}$ has precisely three different cosets in S_3:

$$H(\,) = \{(\,), (1\ 2)\}, \quad H(1\ 3) = \{(1\ 3), (1\ 2\ 3)\}, \quad H(2\ 3) = \{(2\ 3), (1\ 3\ 2)\}$$

This coincides with what Lagrange's Theorem tells us: since H has order 2, it has index 3 in the group S_3.

[1] This theorem is named after the French mathematician *Joseph-Louis Lagrange* (1736–1813). His work predates the introduction of the concept of group into mathematics. This theorem is viewed as being inspired by the work that he did concerning permutations.

(ii) Recall the proof of Theorem 8.21(ii). When $n \geqslant 2$, we observed that the set O_n of odd permutations in S_n is the image of the alternating group A_n under the map $\phi\colon \sigma \mapsto \sigma\,(1\ 2)$. Hence

$$O_n = A_n(1\ 2)$$

is the right coset of A_n with representative $(1\ 2)$ and the union

$$S_n = A_n \cup O_n = A_n \cup A_n(1\ 2)$$

is the decomposition of the symmetric group S_n into the two right cosets of A_n. In particular, $|S_n : A_n| = 2$.

(iii) Lagrange's Theorem is often used to conclude that certain subgroups cannot exist within a group. For example, the symmetric group S_3 has order 6, so it has no subgroup of order 4 and the symmetric group S_4 has order 24, so it has no subgroup of order 5.

Warning: Note that Lagrange's Theorem only indicates that certain subgroups cannot exist. The converse of Lagrange's Theorem is not true: If m is divisor of the order of a finite group G, there is no guarantee that G possesses a subgroup of order m. Indeed, one can show that the alternating group A_4 of degree 4 (which has order 12) has no subgroup of order 6.

We shall finish this chapter by noting some further consequences of Lagrange's Theorem.

Corollary 9.7 *Let G be a finite group and x be an element of G. Then $o(x)$ is finite and divides the order of G.*

Proof: Use Theorem 8.14(ii) to conclude that $o(x)$ equals the order of the subgroup $\langle x\rangle$ of G. In particular, this must be finite and, by Lagrange's Theorem, divides $|G|$. □

Corollary 9.8 *Let G be a finite group of order p, where p is a prime number. Then G is cyclic; that is, $G = \langle x\rangle$ for some $x \in G$.*

Proof: Choose any $x \in G$ that is not the identity. Then $\langle x\rangle$ is a non-trivial subgroup of G and, by Lagrange's Theorem, its order divides $|G| = p$. The only positive divisors of p are 1 and p, so we conclude

$$|\langle x\rangle| = p = |G|.$$

Therefore $G = \langle x\rangle$. □

Note, indeed, that this proof shows more than is claimed in the corollary. It shows, in fact, that $G = \langle x\rangle$ for *all* choices of $x \in G$ except for the identity element.

Chapter 10

Homomorphisms, Normal Subgroups and Quotients Groups

There are two main themes that arise in algebra. The first is the study of algebraic structure occurring with other structures and subgroups are an example of such behaviour. The second theme is to consider functions between algebraic structures that interact well with the binary operations present. We turn to this theme in this chapter.

Homomorphisms

Definition 10.1 Let G and H be groups and we assume that the binary operations are written as multiplication in both these groups. A *homomorphism* $\phi\colon G \to H$ is a function such that

$$(xy)\phi = (x\phi)(y\phi) \qquad \text{for all } x, y \in G.$$

Comment: People often say that a homomorphism "preserves the structure" of the groups. By this, one means that the above equation holds. The term comes from Greek: *homos* meaning "same" and *morphe* meaning "shape." The condition appearing in the definition means that if we take two elements x and y of G then we get the same answer when we (i) multiply them in G and then apply the function ϕ, and (ii) apply the function ϕ to both elements and then multiply the resulting elements of H.

We now give some examples of homomorphisms. We shall return to these examples throughout this chapter.

Example 10.2 (i) Recall the general linear group of degree 2 over $\mathbb{R}$ consists of all invertible 2×2 matrices with real coefficients. Define $\psi\colon \mathrm{GL}_2(\mathbb{R}) \to \mathbb{R}^*$ by

$$A\psi = \det A.$$

Note that a matrix $A \in \mathrm{GL}_2(\mathbb{R})$ has non-zero determinant, so the determinant is some element of the multiplicative group of $\mathbb{R}$. This function is a homomorphism due to the standard multiplicative property of determinants: If $A, B \in \mathrm{GL}_2(\mathbb{R})$, then

$$(AB)\psi = \det(AB) = (\det A)(\det B) = (A\psi)(B\psi).$$

(ii) The set of real numbers $\{\pm 1\}$ is a subgroup of the multiplicative group of real numbers. We can see this by constructing the multiplication table so as to observe that

this set is closed under products and inverses:

	$+1$	-1
$+1$	$+1$	-1
-1	-1	$+1$

There is a homomorphism from the symmetric group S_n of degree n to this group by defining

$$\sigma\Phi = \begin{cases} +1 & \text{if } \sigma \text{ is an even function,} \\ -1 & \text{if } \sigma \text{ is an odd function.} \end{cases}$$

To check that this Φ is a homomorphism, one needs to consider the four possible cases of whether the two permutations involved are odd or even. For example, if σ is an odd permutation and τ is an even permutation, then the product $\sigma\tau$ is odd (as the number of transpositions involved is the sum of the number in σ and the number in τ). Hence in this case

$$(\sigma\tau)\Phi = -1 = (-1)(+1) = (\sigma\Phi)(\tau\Phi).$$

Checking all four cases verifies that $(\sigma\tau)\Phi = (\sigma\Phi)(\tau\Phi)$ for all $\sigma, \tau \in S_n$, as required to show $\Phi\colon S_n \to \{\pm 1\}$ is a homomorphism.

(iii) Let $m \geqslant 2$ be an integer. We shall consider the additive group $(\mathbb{Z}, +)$ of the ring $\mathbb{Z}$ and the additive group $(\mathbb{Z}/m\mathbb{Z}, +)$ of the ring $\mathbb{Z}/m\mathbb{Z}$ of congruence classes modulo m. Define $\phi\colon \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ by

$$\phi\colon x \mapsto [x]$$

where $[x]$ denotes the congruence class of x modulo m. If $x, y \in \mathbb{Z}$, then

$$(x + y)\phi = [x + y] = [x] + [y] = x\phi + y\phi,$$

using the definition of the arithmetic in the ring of congruence classes. This shows that ϕ is a homomorphism between these two additive groups. Note also in this example that we adjust the formula required for a homomorphism because the binary operations in the two groups are written as addition, not multiplication.

(iv) Let G and H be any groups and suppose 1_H denotes the identity element of H. Define $\zeta\colon G \to H$ by

$$x\zeta = 1_H \qquad \text{for all } x \in G.$$

Thus ζ is a constant function that always takes the value 1_H. If $x, y \in G$, then

$$(xy)\zeta = 1_H = 1_H \cdot 1_H = (x\zeta)(y\zeta).$$

Hence ζ is a homomorphism.

(v) If G is any group, then the identity function $\varepsilon\colon G \to G$ is a homomorphism. Indeed, if $x, y \in G$, then

$$(xy)\varepsilon = xy = (x\varepsilon)(y\varepsilon).$$

The next step is to begin to understand what properties homomorphisms have. The basic properties are summarized in the following result.

Theorem 10.3 *Let G and H be groups and $\phi\colon G \to H$ be a homomorphism. Suppose that the identity element of G is 1_G and the identity element of H is 1_H. Then*

(i) $1_G\phi = 1_H$;

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

In part (ii), x^{-1} denotes the inverse of the element x in the group G while $(x\phi)^{-1}$ denotes the inverse of the element $x\phi$ in the group H.

PROOF: (i) Since 1_G is the identity element of G, it satisfies $1_G^2 = 1_G$. Hence upon applying the homomorphism ϕ we obtain

$$1_G\phi = (1_G^2)\phi = (1_G\phi)^2.$$

Multiplying now by the inverse of $1_G\phi$ in the group H (that is, using cancellativity), we deduce $1_G\phi = 1_H$.

(ii) Observe that

$$x\phi \cdot (x^{-1})\phi = (xx^{-1})\phi = 1_G\phi = 1_H$$

with use of part (i). Similarly

$$(x^{-1})\phi \cdot x\phi = (x^{-1}x)\phi = 1_G\phi = 1_H.$$

We conclude that $(x^{-1})\phi$ is the element that we need to multiply $x\phi$ by in order to produce the identity; that is, by uniqueness of inverses,

$$(x^{-1})\phi = (x\phi)^{-1}.$$ □

Isomorphisms

Now that we know about homomorphisms, even though have only basic information about them, we can at least discuss what it means for a pair of groups to be "essentially the same." The relevant term is the following:

Definition 10.4 Let G and H be group. An *isomorphism* between G and H is a bijective homomorphism $\phi\colon G \to H$. If there is an isomorphism between G and H, then we say that G and H are *isomorphic* and write $G \cong H$.

Recall that a bijection $\phi\colon G \to H$ is a function such that each element of H corresponds to a unique element of G under ϕ; that is, if $h \in H$ there is a unique choice of $g \in G$ with $g\phi = h$. In the case of an isomorphism $\phi\colon G \to H$ we are saying that not only do the elements of the two groups correspond, but also if $g_1, g_2 \in G$ correspond to $h_1, h_2 \in H$ (that is, $h_1 = g_1\phi$ and $h_2 = g_2\phi$), then

$$h_1h_2 = (g_1\phi)(g_2\phi) = (g_1g_2)\phi;$$

that is, the product of g_1 and g_2 corresponds to the product of h_1 and h_2. This tells us that an isomorphism between two groups is essentially just a relabelling of the elements but otherwise the multiplication is the same. Consequently, the two groups should be viewed as being the same.

Example 10.5 We have actually already seen the concept isomorphism appearing in these lecture notes, but have not had this terminology to describe it. For example, if $G = \{1, g_1, g_2, g_3\}$ is a group of order 4 such that $x^2 = 1$ for all $x \in G$, then the Cayley table has the form

	1	g_1	g_2	g_3
1	1	g_1	g_2	g_3
g_1	g_1	1		
g_2	g_2		1	
g_3	g_3			1

There is a unique way to complete this to the Cayley table of a group (since we must have each element appear exactly once in each row and each column):

	1	g_1	g_2	g_3
1	1	g_1	g_2	g_3
g_1	g_1	1	g_3	g_2
g_2	g_2	g_3	1	g_1
g_3	g_3	g_2	g_1	1

This is, of course, the same pattern as that of the Klein 4-group. Consequently, G is isomorphic to the Klein 4-group V_4: The isomorphism $\phi\colon G \to V_4$ is given by

$$1\phi = e, \qquad g_1\phi = a, \qquad g_2\phi = b, \qquad g_3\phi = c$$

(where $V_4 = \{e, a, b, c\}$ as in Example 5.9).

Kernels and images

We shall now describe the main tools that we use to describe the behaviour of homomorphisms.

Definition 10.6 Let $\phi\colon G \to H$ be a homomorphism from a group G to a group H.

(i) The *kernel* of ϕ is
$$\ker\phi = \{\, x \in G \mid x\phi = 1_H \,\}.$$

(ii) The *image* of ϕ is
$$\operatorname{im}\phi = \{\, x\phi \mid x \in G \,\}.$$

So, if $\phi\colon G \to H$ is a homomorphism, the kernel is the set of elements of G that are mapped to the identity element of H, while the image is the set of elements of H that arise as the image of some element of G. Consequently, $\ker\phi$ is some subset of G, while $\operatorname{im}\phi$ is some subset of H. The main properties of these subsets are as follows, specifically that they are in fact subgroups.

Theorem 10.7 *Let $\phi\colon G \to H$ be a homomorphism from a group G to a group H. Then*

(i) *the kernel of ϕ is a subgroup of G;*

(ii) *the image of ϕ is a subgroup of H;*

(iii) *$g^{-1}xg \in \ker\phi$ for all $x \in \ker\phi$ and $g \in G$.*

Part (iii) of this theorem tells us that, not only is the kernel closed under products and inverses, it is also closed under the forming of products $g^{-1}xg$ where $x \in \ker\phi$ but g can be *any* element of G. We shall discuss this property later in this chapter.

PROOF: (i) Since $1_G\phi = 1_H$, as observed in Theorem 10.3(i), the identity element 1_G of G is in the kernel of ϕ. In particular, $\ker\phi$ is non-empty. Let $x, y \in \ker\phi$. Then $x\phi = y\phi = 1_H$. Now

$$(xy)\phi = (x\phi)(y\phi) = 1_H \cdot 1_H = 1_H$$

and

$$(x^{-1})\phi = (x\phi)^{-1} = 1_H^{-1} = 1_H$$

with use of Theorem 10.3(ii). This shows $xy, x^{-1} \in \ker\phi$ for all $x, y \in \ker\phi$. The subgroup test (Theorem 8.5) tells us that the kernel, $\ker\phi$, is indeed a subgroup of G.

(ii) Since G is non-empty, the set $\operatorname{im}\phi$ of images of the elements of G under ϕ is certainly also non-empty. Let $g, h \in \operatorname{im}\phi$. Then $g = x\phi$ and $h = y\phi$ for some $x, y \in G$. Now

$$gh = (x\phi)(y\phi) = (xy)\phi \in \operatorname{im}\phi$$

and

$$g^{-1} = (x\phi)^{-1} = (x^{-1})\phi \in \operatorname{im}\phi,$$

again with use of Theorem 10.3(ii). Therefore, by Theorem 8.5, the image $\operatorname{im}\phi$ is a subgroup of H.

(iii) Let $x \in \ker\phi$ and $g \in G$. Observe

$$(g^{-1}xg)\phi = (g\phi)^{-1}(x\phi)(g\phi) = (g\phi)^{-1} \cdot 1_H \cdot (g\phi) = (g\phi)^{-1}(g\phi) = 1_H.$$

(We used the definition of the term homomorphism and, again, Theorem 10.3(ii) in this calculation.) We conclude that $g^{-1}xg \in \ker\phi$. □

Example 10.8 Let us return to the examples in Example 10.2 to see how these subgroups occur.

(i) We defined $\psi\colon \mathrm{GL}_2(\mathbb{R}) \to \mathbb{R}^*$ by $A\psi = \det A$ (the determinant of the matrix A). Every non-zero real number does occur as the determinant of an invertible matrix, indeed

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

and therefore this ψ is surjective: $\operatorname{im}\psi = \mathbb{R}^*$. The kernel is the set of matrices of determinant 1; that is, $\ker\psi = \mathrm{SL}_2(\mathbb{R})$, the special linear group (see Example 8.8).

(ii) We define $\Phi\colon S_n \to \{\pm 1\}$ by mapping an even permutation to $+1$ and an odd permutation to -1. If $n \geqslant 2$, then there are both even permutations (e.g., the identity) and odd permutations (e.g., (1 2)) in S_n. Hence $\operatorname{im}\Phi = \{\pm 1\}$; that is, Φ is surjective. The kernel of ϕ consists of all even permutations; that is, $\ker\Phi = A_n$, the alternating group of degree n.

(iii) In our third example, we defined the homomorphism of additive groups $\phi\colon \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ by $x\phi = [x]$, the congruence class of x modulo m. Since the codomain consists of all the congruence classes, this ϕ is surjective: $\operatorname{im}\phi = \mathbb{Z}/m\mathbb{Z}$. The kernel consists of all $x \in \mathbb{Z}$ such that $[x] = [0]$; that is, $x \equiv 0 \pmod m$. Hence $\ker\phi$ consists of all integers that are divisible by m.

(iv) The trivial homomorphism $\zeta\colon G \to H$ between two groups is given by $x\zeta = 1_H$ for all $x \in G$. Hence $\ker\zeta = G$ and $\operatorname{im}\zeta = \{1_H\}$ (the trivial subgroup of H).

(v) The identity map $\varepsilon\colon G \to G$, given by $x \mapsto x$ for all $x \in G$, is a bijective homomorphism. That is, it is an example of an isomorphism from the group G to itself. The image $\operatorname{im}\varepsilon = G$, while $\ker\varepsilon = \{1_G\}$.

This set of examples does illustrate how the kernel and image have significance to a homomorphism. We summarize now their general relevance to the question of whether a homomorphism is injective or surjective. Part (ii) of the following theorem is just the definition of what it means for a function to be surjective. Part (i) requires a little more effort to verify.

Theorem 10.9 *Let $\phi\colon G \to H$ be a homomorphism between two groups G and H. Then*

(i) *ϕ is injective if and only if $\ker\phi = \{1_G\}$;*

(ii) *ϕ is surjective if and only if $\operatorname{im}\phi = H$.*

PROOF: (i) Suppose first that ϕ is injective. Let $x \in \ker\phi$. Then $x\phi = 1_H = 1_G\phi$. Since ϕ is injective, we conclude that $x = 1_G$. Consequently, the only element in the kernel is the identity 1_G (and we know this element is definitely in the set). Hence $\ker\phi = \{1_G\}$.

Conversely suppose $\ker\phi = \{1_G\}$. Let $x, y \in G$ and suppose $x\phi = y\phi$. Then

$$(xy^{-1})\phi = (x\phi)(y\phi)^{-1} = (x\phi)(x\phi)^{-1} = 1_H$$

so $xy^{-1} \in \ker\phi$. By assumption, $xy^{-1} = 1_G$ and therefore $x = y$ (upon multiplying on the right by y). Hence ϕ is injective.

(ii) As note above, part (ii) follows immediately from the definitions. □

The use of part (i) of the theorem means that it is easier to check that a homomorphism $\phi\colon G \to H$ is injective compared with an arbitrary function. We do not have to check that different elements are always mapped to different values by ϕ. We just need to check that only the element of G that is mapped to the identity element of H (that is, the only element in the kernel of ϕ) is the identity element of G.

Normal subgroups

We shall now investigate the property appearing in part (iii) of Theorem 10.7. We make the following definition:

Definition 10.10 Let G be a group and N be a subgroup of G. We say that N is a *normal subgroup* of G if it satisfies the additional condition: $g^{-1}xg \in N$ for all $x \in N$ and $g \in G$. We write $N \trianglelefteq G$ to indicate that N is a normal subgroup of the group G.

Comments:

(i) The element $g^{-1}xg$ is called the *conjugate* of x by g.

(ii) To verify that a particular subset N of a group G is indeed a normal subgroup we need to first verify that N is a subgroup (so apply the subgroup test, Theorem 8.5) and then verify that N is closed under conjugation by arbitrary elements of G (so check the condition in the above definition).

(iii) If $\phi\colon G \to H$ is any homomorphism, then Theorem 10.7 tells us that the kernel of ϕ is a normal subgroup of G.

This last comment means that we have an immediate collection of examples of normal subgroups by use of Example 10.8:

Example 10.11 (i) The special linear group $\mathrm{SL}_2(\mathbb{R})$ is a normal subgroup of the general linear group $\mathrm{GL}_2(\mathbb{R})$ since it is the kernel of the determinant map. (Indeed, the same argument shows that $\mathrm{SL}_n(F)$ is a normal subgroup of $\mathrm{GL}_n(F)$ for every positive integer n and every field F.)

(ii) The alternating group A_n is a normal subgroup of the symmetric group S_n since it is the kernel of the "sign map" Φ that sends an even permutation to $+1$ and an odd permutation to -1.

(iii) If G is any group, then G is a normal subgroup of itself (as it is the kernel of the map $\zeta\colon x \mapsto 1$) and $\{1\}$ is a normal subgroup of G (as it is the kernel of the identity map).

An alternative method to verify that a subgroup is normal is to directly check that it is closed under conjugation. The above examples could all also have been established by that method.

Theorem 10.12 *Let G be a group and N be a subgroup of G. The following conditions on N are equivalent:*

(i) *N is a normal subgroup of G;*

(ii) *the set $g^{-1}Ng = \left\{ g^{-1}xg \;\middle|\; x \in N \right\}$ equals N for all $g \in G$;*

(iii) *every left coset of N in G is also a right coset of N in G and* vice versa*;*

(iv) *$Ng = gN$ for all $g \in G$.*

Proof: (i) $\Rightarrow$ (ii): Suppose that $N \trianglelefteq G$; that is, N satisfies the condition in Definition 10.10. By assumption, $g^{-1}xg \in N$ for all $x \in N$ and all $g \in G$. This shows that $g^{-1}Ng \subseteq N$ for all $g \in G$.

In particular, if $g \in G$ is fixed, then $g^{-1}Ng \subseteq N$ and $gNg^{-1} = (g^{-1})^{-1}Ng^{-1} \subseteq N$. Let $x \in N$. Then $gxg^{-1} = y \in N$, so $x = g^{-1}yg \in g^{-1}Ng$. This establishes the reverse inclusion $N \subseteq g^{-1}Ng$.

In conclusion, $g^{-1}Ng = N$ and this holds for all $g \in G$.

(ii) $\Rightarrow$ (iii): Suppose $g^{-1}Ng = N$ for all $g \in N$. Fix $g \in G$. If $x \in N$, then $g^{-1}xg = y$ for some $y \in N$ and hence $xg = gy \in gN$. Since x is an arbitrary element of N, we conclude $Ng \subseteq gN$.

Similarly, if $x \in N = g^{-1}Ng$, then $x = g^{-1}zg$ for some $z \in N$ and hence $gx = zg \in Ng$. Since x is arbitrary, we conclude $gN \subseteq Ng$.

In conclusion, $Ng = gN$ for all $g \in G$. In particular, every left coset equals a right coset (indeed the right coset with the same representative) and every right coset equals a left coset.

(iii) $\Rightarrow$ (iv): Let $g \in G$. The left coset gN equals some right coset of N, say $gN = Nh$ for some $h \in G$. Then $g = g1 \in gN = Nh$ and $g = 1g \in Ng$, so $Ng \cap Nh \neq \varnothing$. Since

cosets are either disjoint or equal (Theorem 9.3(ii)) we conclude that $Ng = Nh$. Hence $gN = Nh = Ng$, as claimed.

(iv) $\Rightarrow$ (i): Let $x \in N$ and $g \in G$. By assumption, $Ng = gN$, so $xg \in Ng = gN$ and we deduce $xg = gy$ for some $y \in N$. Then $g^{-1}xg = y \in N$. This shows $g^{-1}xg \in N$ for all $x \in N$ and $g \in G$; that is, $N \trianglelefteq G$. □

Example 10.13 We shall also give some examples relating to normal subgroups that do not, on the face of it, directly refer to kernels of homomorphisms. Recall from Question 8 on Problem Sheet VI that conjugation in the symmetric group is given, for conjugates of cycles, by:

$$\rho^{-1}\,(i_1\ i_2\ \ldots\ i_r)\,\rho = (i_1\rho\ i_2\rho\ \ldots\ i_r\rho)$$

and that the conjugate of a permutation in S_n is another permutation with the same disjoint cycle structure.

(i) Let $H = \langle(1\ 2)\rangle = \{(\,),\ (1\ 2)\}$ and consider whether or not this is a normal subgroup of the symmetric group S_3 of degree 3. We calculate

$$(1\ 3)^{-1}\,(1\ 2)\,(1\ 3) = (2\ 3) \notin H.$$

Hence H is not closed under conjugation, so H is *not* a normal subgroup of S_3.

(ii) Let $H = \{(\,),\ (1\ 2)(3\ 4),\ (1\ 3)(2\ 4),\ (1\ 4)(2\ 3)\}$. In Example 8.9, we showed that this H is a subgroup of the symmetric group S_4 of degree 4. Indeed, we showed that the Cayley table is essentially the same as that of the Klein 4-group, so H is a subgroup of S_4 that is *isomorphic* to the Klein 4-group V_4. Note that H consists of the identity element and all the permutations that can be written as a product of two disjoint transpositions.

Now if $\sigma \in H$ and $\rho \in S_4$, then $\rho^{-1}\sigma\rho$ has the same disjoint cycle structure as σ. Hence either $\rho^{-1}\sigma\rho = (\,) \in H$ or $\rho^{-1}\sigma\rho$ is a product of two disjoint transpositions and so is in H. Therefore $H \trianglelefteq S_4$.

Quotient groups

We shall now describe a construction that is very useful in the study of groups. If G is a group and N is a normal subgroup, then we can construct a new group, the quotient group, whose elements are the cosets of N. The basic idea is that this will usually be a smaller group (indeed, Lagrange's Theorem tells us that the number of elements is $|G|/|N|$) and so working with this group should in some sense be easier. We will be able to do little more than begin to work with this construction, but with further exploration one would be able to understand how to obtain information about the original group G from this quotient group.

Definition 10.14 Let G be a group and let N be a normal subgroup of G. Let

$$G/N = \{\, Nx \mid x \in G \,\},$$

the set of all right cosets of N in G. Define a multiplication on G/N by

$$Nx \cdot Ny = Nxy$$

for $x, y \in G$ (i.e., we multiply the representatives of the cosets).

We call G/N with this multiplication the *quotient group* of G by N.

Note: We only use the notation G/N when $N \trianglelefteqslant G$. If a subgroup H of a group G is *not* normal then we cannot use the above formula to define a group structure on the set of cosets of H.

Theorem 10.15 *Let G be a group and N be a normal subgroup of G. Then G/N is a group with respect to the multiplication defined above.*

PROOF: The most challenging step in the proof is the first one, verifying that we do have a binary operation on G/N, whereas checking the axioms of a group turns out to be straightforward. The issue is very similar to that in Lemma 4.6. In the case of a quotient group, the multiplication given in Definition 10.14 depends upon the representatives x and y and we know that cosets can have different representatives. We need to ensure our multiplication is well-defined; that is, that it depends only upon the cosets concerned and not on the *choice* of representatives for the cosets.

Consider two cosets of N and suppose that we have written them with different representatives: $Nx = Nx'$ and $Ny = Ny'$. This means that

$$x(x')^{-1} \in N \qquad \text{and} \qquad y(y')^{-1} \in N,$$

by Theorem 9.3(i). Let's give names to these elements of N, say $a = x(x')^{-1}$ and $b = y(y')^{-1}$, so $x = ax'$ and $y = by'$ where $a, b \in N$. To help in the following calculation we shall write g^h for the conjugate $h^{-1}gh$ of an element g by another element h. Then

$$xy = ax'by' = ax'b(x')^{-1}x'y' = ab^{(x')^{-1}}x'y'.$$

Since $N \trianglelefteqslant G$, it follows that the conjugate $b^{(x')^{-1}} \in N$ and so $ab^{(x')^{-1}} \in N$. Therefore

$$xy(x'y')^{-1} = ab^{(x')^{-1}} \in N$$

and, using Theorem 9.3(i) again,

$$Nxy = Nx'y'.$$

This tells us that we get the same answer for the product $Nx \cdot Ny$ whether we calculate it using the representatives x and y or using the representatives x' and y'. Hence we have a well-defined multiplication on G/N.

The remaining steps of the proof are more straightforward. We check the axioms of a group:

Associativity: Let $Nx, Ny, Nz \in G/N$. Then

$$(Nx \cdot Ny) \cdot Nz = Nxy \cdot Nz = N(xy)z$$

and

$$Nx \cdot (Ny \cdot Nz) = Nx \cdot Nyz = Nx(yz).$$

But the multiplication in G is associative, so $(xy)z = x(yz)$ and hence

$$(Nx \cdot Ny) \cdot Nz = Nx \cdot (Ny \cdot Nz),$$

as required.

Identity: $Nx \cdot N1 = Nx1 = Nx$ and $N1 \cdot Nx = N1x = Nx$ for all $x \in G$, so $N1$ is the identity element of G/N.

Inverses: $Nx \cdot Nx^{-1} = Nxx^{-1} = N1$ and $Nx^{-1} \cdot Nx = Nx^{-1}x = N1$, so Nx^{-1} is the inverse of Nx in G/N.

Hence G/N is a group. □

Example 10.16 The quaternion group Q_8 of order 8 contains eight elements

$$Q_8 = \{\pm 1, \pm i, \pm j, \pm k\}$$

and has multiplication given by

$$i^2 = j^2 = k^2 = -1, \qquad ij = -ji = k, \quad jk = -kj = i, \quad ki = -ik = j.$$

One should maintain the order of the elements in the above products, but one can multiply throughout by -1 to determine products such as $k(-i) = -ki = -j$. This can be shown to be a group. Indeed the easiest method is to observe that this group is isomorphic to the group Q of matrices in Question 12 on Problem Sheet VIII.

Let $N = \langle -1 \rangle = \{\pm 1\}$, the cyclic subgroup generated by the element -1. This is certainly a subgroup. Furthermore, the definition of the multiplication is that $(-1)i = -i = i(-1)$ and similarly for other products involving -1, so -1 commutes with all elements of Q_8. Therefore

$$g^{-1}(-1)g = g^{-1}g(-1) = -1 \in N$$

for all $g \in Q_8$. On the other hand, $g^{-1}1g = 1$ for all $g \in Q_8$. We conclude that $N \trianglelefteq Q_8$.

We may therefore construct the quotient group Q_8/N. Let us calculate some products in this quotient group: The coset $Ni = \{\pm i\} = N(-i)$ by definition and

$$(Ni)^2 = Ni \cdot Ni = Ni^2 = N(-1) = N1$$

since $-1 \in N$ and therefore these last two cosets are equal by Theorem 9.3(i). Exactly the same argument shows that $(Nj)^2 = (Nk)^2 = N1$ and we now conclude that every element of Q_8/N satisfies $(Ng)^2 = N1$. Example 10.5 tells us that

$$Q_8/N \cong V_4.$$

The First Isomorphism Theorem

The final thing we do in the lecture course is bring the concepts of this chapter together. Let $\phi\colon G \to H$ be a homomorphism. We know from Theorem 10.7 that the kernel of ϕ is a normal subgroup of G. According to Theorem 10.15, we can construct the quotient group $G/\ker\phi$. It is a group, but the question that should arise is what does this group have to do with the original homomorphism? The following theorem gives the answer:

Theorem 10.17 (First Isomorphism Theorem) *Let G and H be groups and $\phi\colon G \to H$ be a homomorphism. Then*

$$G/\ker\phi \cong \operatorname{im}\phi.$$

Proof: We need to construct an isomorphism (a bijective homomorphism) between $G/\ker\phi$ and $\operatorname{im}\phi$. To simply notation, we shall write $K = \ker\phi$ and define the function $\theta\colon G/K \to \operatorname{im}\phi$ by

$$(Kx)\theta = x\phi.$$

The definition of θ appears to depend on the choice of representative x for the coset Kx, so we must first check that θ is well-defined.

Suppose $Kx = Ky$. Then $xy^{-1} \in K = \ker\phi$, so

$$1 = (xy^{-1})\phi = (x\phi)(y\phi)^{-1}$$

and multiplying by $y\phi$ gives $x\phi = y\phi$. Hence θ is well-defined: we get the same value for the image of $Kx = Ky$ under θ whether we use the representative x or y.

Next

$$(Kx \cdot Ky)\theta = (Kxy)\theta = (xy)\phi = (x\phi)(y\phi) = (Kx)\theta \cdot (Ky)\theta$$

and we deduce θ is a homomorphism.

If $g \in \operatorname{im}\phi$, then $g = x\phi$ for some $x \in G$. Then

$$(Kx)\theta = x\phi = g$$

and we see that θ is surjective.

Finally if $Kx \in \ker\theta$, then $1 = (Kx)\theta = x\phi$, so $x \in \ker\phi = K$ and therefore $Kx = K1$, the identity element in G/K, by use of Theorem 9.3(i). This shows that $\ker\theta = \{K1\} = \{1_{G/K}\}$ and Theorem 10.9(i) now tells us θ is injective.

Hence θ is a bijective homomorphism; that is, an isomorphism. Hence

$$G/\ker\theta = G/K \cong \operatorname{im}\phi. \qquad \square$$

Example 10.18 (i) Recall Example 10.8(i): there is a homomorphism $\psi\colon \mathrm{GL}_2(\mathbb{R}) \to \mathbb{R}^*$ which has kernel $\ker\psi = \mathrm{SL}_2(\mathbb{R})$ (the special linear group) and image $\operatorname{im}\psi = \mathbb{R}^*$ (that is, ψ is surjective). The First Isomorphism Theorem tells us

$$\mathrm{GL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{R}) \cong \mathbb{R}^*.$$

(ii) Also recall from Example 10.8(ii) that the homomorphism $\Phi\colon \mathrm{SL}_2(\mathbb{R}) \to \{\pm 1\}$ has kernel $\ker\Phi = A_n$ (the alternating group) and, provided $n \geqslant 2$, is surjective. Hence, for $n \geqslant 2$,

$$S_n/A_n \cong \{\pm 1\} = \langle -1\rangle,$$

which is a cyclic group of order 2. (This is not so surprising: Since $|S_n : A_n| = 2$, the quotient S_n/A_n is of order 2 and so is cyclic by Corollary 9.8.)

Having shown that the groups above are isomorphic, it then follows that their properties are the same. As an example, since $\mathbb{R}^*$ is an abelian group (multiplication in the field $\mathbb{R}$ is commutative), the same is true for the quotient group $\mathrm{GL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{R})$. The general linear group $\mathrm{GL}_2(\mathbb{R})$ is non-abelian (as observed earlier) but we have found a normal subgroup for which the associated quotient group is abelian.

The following example arises in many situations. For example, if G is any finite group and $x \in G$, then x has finite order by the Corollary 9.7 of Lagrange's Theorem. Consequently $C = \langle x\rangle$ is a cyclic group of order $n = o(x)$ (and this n divides the order of the finite group G that we started with). We would then be able to apply the method presented in the following example.

Example 10.19 Let $C = \langle x\rangle$ be a cyclic group of order n (for some positive integer n). Consider the group $\mathbb{Z}$ of integers under addition and define a map $\phi\colon \mathbb{Z} \to C$ by

$$\phi\colon k \mapsto x^k \qquad \text{for } k \in \mathbb{Z}.$$

The standard power laws (see Theorem 5.20) show that

$$(k+\ell)\phi = x^{k+\ell} = x^k x^\ell = (k\phi)(\ell\phi)$$

for all $k, \ell \in \mathbb{Z}$. Hence ϕ is a homomorphism from the group $(\mathbb{Z}, +)$ to the cyclic group C. (Note that the group operation in $\mathbb{Z}$ is addition $+$ and so on the left-hand side of the above formula we use the addition operation, while on the right-hand side we use the group multiplication in C.)

Since $C = \langle x \rangle$ is cyclic, the element of C are the powers of the generator x and hence $\operatorname{im}\phi = \{ x^k \mid k \in \mathbb{Z} \} = C$. We determine the kernel of ϕ by using the observation in Theorem 8.14(i) concerning the order of the element x:

$$\begin{array}{lll} k \in \ker\phi & \text{if and only if} & k\phi = 1 \\ & \text{if and only if} & x^k = 1 \\ & \text{if and only if} & o(x) = n \text{ divides } k \end{array}$$

by use of Theorem 8.14(i). Hence $\ker\phi$ consists of all the multiples of n:

$$\ker\phi = \{ nq \mid q \in \mathbb{Z} \} = n\mathbb{Z}.$$

An application of the First Isomorphism Theorem tells us that

$$C = \langle x \rangle \cong \mathbb{Z}/n\mathbb{Z}$$

where the group on the right-hand side is the quotient group of $\mathbb{Z}$ by the (normal) subgroup $n\mathbb{Z}$. (It is safe to bracket "normal" here since $(\mathbb{Z}, +)$ is an abelian group, so *every* subgroup is normal.)

Corollary 10.20 *Any two cyclic groups of the same order are isomorphic.*

Proof: Let $C = \langle x \rangle$ and $D = \langle y \rangle$ be cyclic groups of order n. The above example shows

$$C \cong \mathbb{Z}/n\mathbb{Z} \cong D$$

and hence C and D are isomorphic. □

Bibliography

[1] R. B. J. T. Allenby, *Rings, Fields and Groups: An Introduction to Abstract Algebra, Second Edition*, Butterworth-Heinemann, 1991.

Versions

Significant updates to the notes will be listed below. Updates that are merely correcting typographic errors and similar will be indicated by appending to a letter to the version number on the front page and will not be listed below.

Version 1.0: Lecture notes from the academic year 2021–22 converted in format but otherwise unchanged.

Version 1.1: Streamlined material on the Euclidean Algorithm. Changed proof concerning the sign of a permutation. Added example and corollary concerning homomorphisms to cyclic groups.