

Aspects of

SYMMETRY

*Robert B. Howlett*

## Contents

### Chapter 1: Symmetry

§1a	An example of abstract symmetry	2
§1b	Structure preserving transformations	3
§1c	The symmetries of some structured sets	7
§1d	Some groups of transformations of a set with four elements	10

### Chapter 2: Introductory abstract group theory

§2a	Group axioms	15
§2b	Basic deductions from the axioms	18
§2c	Subgroups and cosets	21
§2d	On the number of elements in a set.	26
§2e	Equivalence relations	30
§2f	Cosets revisited	34
§2g	Some examples	36
§2h	The index of a subgroup	39

### Chapter 3: Homomorphisms, quotient groups and isomorphisms

§3a	Homomorphisms	40
§3b	Quotient groups	45
§3c	The Homomorphism Theorem	52

### Chapter 4: Automorphisms, inner automorphisms and conjugacy

§4a	Automorphisms	57
§4b	Inner automorphisms	60
§4c	Conjugacy	63
§4d	On the number elements in a conjugacy class	66

### Chapter 5: Reflections

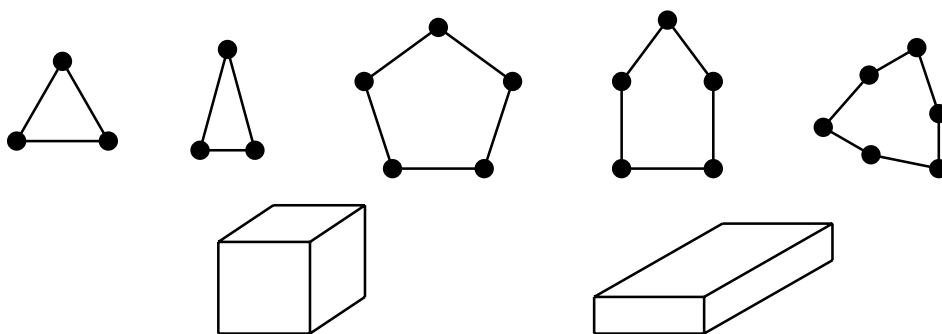
§5a	Inner product spaces	70
§5b	Dihedral groups	72
§5c	Higher dimensions	75
§5d	Some examples	81

Chapter 6: Root systems and reflection groups	
§6a Root systems	84
§6b Positive, negative and simple roots	88
§6c Diagrams	94
§6d Existence and inadmissibility proofs	105
Index of notation	110
Index	111

# 1

## Symmetry

Of the following geometrical objects, some are rather symmetrical, others less so:



Objects exhibiting a high degree of symmetry are more special, and possibly—depending on your taste—more appealing than the less symmetrical ones. They are, therefore, natural objects of mathematical interest. When studying symmetrical objects one can always exploit the symmetry to limit the amount of work one has to do. It is only necessary to measure one side of a square, since symmetry says that all other sides are the same. It is not necessary to measure any of the angles: it is a consequence of symmetry that they are all ninety degrees.

Although measuring a square is a trivial application of symmetry, it gives a tiny glimpse of the importance of symmetry in mathematics. Symmetry occurs not only in concrete, geometrical situations, but also in highly complex abstract situations, where exploitation of the symmetrical aspects of the situation can provide methods for dealing with problems that otherwise would be hopelessly intractable.

Group theory is the mathematical theory of symmetry, in which the basic tools for utilizing symmetry are developed. The purpose of these notes

## 2 Chapter One: Symmetry

is to provide an introduction to group theory, concentrating in particular on a very important class of geometrical groups: the finite Euclidean reflection groups.

### §1a An example of abstract symmetry

Before launching into a theoretical discussion of symmetry, we take a brief look at an example of the use of symmetry in an abstract situation. After this section our examples of symmetry will be almost entirely geometrical in character, but the example we consider here is drawn from number theory.

Consider the equation  $x^2 + y^2 = 1$ , and suppose we wish to find solutions of this which have the form  $x = p/q$  and  $y = r/s$ , where  $p, q, r$  and  $s$  are integers. It is not hard to see that this problem amounts to finding *Pythagorean triples*: triples  $(a, b, c)$  of integers satisfying  $a^2 + b^2 = c^2$ . The most famous of these triples,  $(3, 4, 5)$ , yields the solution  $x = 3/5, y = 4/5$ . Now, how can we bring symmetry into this?

If an equation possesses symmetry, once one solution has been found, the symmetry may present you with other solutions for free. There is an obvious symmetry to the equation  $x^2 + y^2 = 1$ , given by interchanging  $x$  and  $y$ . So of course we have the solution  $x = 4/5, y = 3/5$  also. This is not very exciting. But there are other symmetries of the equation which you probably haven't noticed yet. For example, the transformation

$$\begin{aligned}x &\mapsto x' = (3/5)x + (4/5)y \\y &\mapsto y' = (4/5)x - (3/5)y\end{aligned}$$

can be seen to be a symmetry, since if  $x^2 + y^2 = 1$  then

$$\begin{aligned}(x')^2 + (y')^2 &= ((3/5)x + (4/5)y)^2 + ((4/5)x - (3/5)y)^2 \\&= (\frac{9}{25}x^2 + \frac{24}{25}xy + \frac{16}{25}y^2) + (\frac{16}{25}x^2 - \frac{24}{25}xy + \frac{9}{25}y^2) \\&= x^2 + y^2 \\&= 1.\end{aligned}$$

So, starting with the solution  $x = 15/17, y = 8/17$ , derived from the well known Pythagorean triple  $(8, 15, 17)$ , we deduce that

$$\begin{aligned}x' &= (3/5)(15/17) + (4/5)(8/17) = 77/85 \\y' &= (4/5)(15/17) - (3/5)(8/17) = 36/85\end{aligned}$$

is another solution to our original equation. Symmetry has combined with two well-known facts to give us a somewhat less well-known one.

Before moving on, a confession is necessary: symmetry is not really needed in the study of rational solutions to the equation  $x^2 + y^2 = 1$ . A complete account of these solutions can be found in many standard texts on number theory, symmetry not being mentioned.

### §1b Structure preserving transformations

The two most basic concepts of modern mathematics are the concepts of *set* and *function*. In most formal treatments of the foundations of mathematics, *set* is a primitive—that is, undefined—concept. You can't define everything! But a set is to be thought of as a collection of things, called the *elements* of the set, and the axioms that sets are assumed to satisfy are consistent with this intuitive description of sets.

In these formal treatments, which first became fashionable at the start of the twentieth century, every mathematical object is, ultimately, a set. Thus, for example, the number zero is commonly defined as the empty set (the set with no elements), the number one as the set whose single element is the number zero, two as the set whose two elements are zero and one, and so on. Of course, in practice no one but researchers into the foundations of mathematics works directly with such definitions. The purpose of the work on foundations is to create theories which are proved to be free from internal contradictions, and which accurately model intuitive approaches to mathematics. The rest of us can then continue using the intuitive approach, happy in the knowledge that contradictions will never arise.

The concept of *function*, unlike that of set, is not usually regarded as a primitive concept. Functions are defined in terms of sets. Indeed, like everything else in mathematics, a function is, ultimately—at least in the formal treatments—a set. But the formal definition is irrelevant for our purposes; instead we should focus on the basic properties which characterize functions, relying on the fact that consistent formal theories have been created which validate this.

A function consists of two sets, the *domain* and the *codomain* of the function, and a rule that associates with each element of the domain a unique element of the codomain. Thus, for example, there is a function whose domain is the set of all people and whose codomain is the set  $\mathbb{R}$  of all real

## 4 Chapter One: Symmetry

numbers, the rule being to find the height of the person in centimetres. Observe that this rule does indeed give an element of the codomain (a number) for each element of the domain (each person).

The words *map* and *mapping* are commonly used synonymously with *function*, as also is *transformation*. However, for the purposes of this course only those functions for which the codomain is the same as the domain will be called transformations. So a transformation is a function from a set to itself. Transformations are important for us as they enable us to explicate the notion of symmetry:

**1.1 DEFINITION** A symmetry of an object is a transformation of the object which leaves it unchanged in its essential features.

This definition is not really acceptable, except as a guide to intuition, since it is too vague and imprecise. Accepting that all mathematical objects are (ultimately) sets, the notion of a transformation of an object is satisfactory, but what does it mean to say that a function from a set to itself leaves the set unchanged in its essential features? To answer this objection, we need some more concepts.

**1.2 DEFINITION** Let  $X_1, X_2, \dots, X_n$  be arbitrary sets. The *Cartesian product*  $X_1 \times X_2 \times \dots \times X_n$  is the set all ordered  $n$ -tuples  $(x_1, x_2, \dots, x_n)$  such that  $x_i \in X_i$  for each  $i$ . That is,

$$X_1 \times X_2 \times \dots \times X_n = \{ (x_1, x_2, \dots, x_n) \mid x_1 \in X_1, x_2 \in X_2, \dots, x_n \in X_n \}$$

If  $X_i = S$  for all  $i$  then  $X_1 \times X_2 \times \dots \times X_n$  is called the  *$n$ -fold Cartesian power* of the set  $S$ , sometimes written as  $S^n$ .

**1.3 DEFINITION** Let  $n$  be a nonnegative integer and  $S$  a set. An  *$n$ -ary relation* on  $S$  is a function  $\rho$  from the  $n$ -fold Cartesian power  $S^n$  to the two element set  $\{\text{true}, \text{false}\}$ . The elements  $x_1, x_2, \dots, x_n$  of  $S$  are said to satisfy the relation  $\rho$  if  $\rho(x_1, x_2, \dots, x_n) = \text{true}$ , and not satisfy  $\rho$  if  $\rho(x_1, x_2, \dots, x_n) = \text{false}$ .

Thus, for example, if  $S$  is the Euclidean plane then we can define a ternary (or 3-ary) relation  $\text{Perp}$  on  $S$  as follows: for all  $a, b, c \in S$ ,

$$\text{Perp}(a, b, c) = \begin{cases} \text{true} & \text{if } \angle abc = 90^\circ \\ \text{false} & \text{otherwise.} \end{cases}$$

Binary (or 2-ary) relations are the most familiar kind; in this case the symbol for the relation is commonly written between the two arguments. For example, the “less than” relation on the set of all real numbers:  $a < b$  is true if  $b - a$  is a positive number, false otherwise. Single argument relations—that is, unary relations—are simply properties which a single object may or may not possess. Thus there is a unary relation **Green** on the set of all visible objects: **Green**(jumper) is true if the jumper is green.

If  $R$  is an  $n$ -ary relation on  $S$ , then we will abbreviate the statement “ $R(x_1, x_2, \dots, x_n)$  is true” to “ $R(x_1, x_2, \dots, x_n)$ ”. For example, “ $a < b$ ” has the same meaning as “ $a < b$  is true”.

Virtually anything that one might want to say about sets and their elements can be reformulated in terms of relations on sets. For example, the concept of an operation on a set can be recast so that it becomes an example of a relation on the set. We could define a ternary relation **Plus** on the set  $\mathbb{R}$  by the rule

$$\text{Plus}(a,b,c) \text{ if and only if } a + b = c.$$

A vector space could be defined as a set equipped with a ternary relation **Plus** and binary relations  $(\text{Mult})_\lambda$  for each scalar  $\lambda$  (where  $(\text{Mult})_\lambda(u, v)$  if and only if, in the usual notation,  $v = \lambda u$ ), such that the appropriate axioms are satisfied.

Be warned that the definition of the term “relation” that we have employed is somewhat non-standard: most mathematical authors define a relation to be a subset of  $S^n$ , namely the subset consisting of all  $n$ -tuples  $(x_1, x_2, \dots, x_n)$  which satisfy the relation.

Another non-standard definition which is useful for our present purposes is the following:

**1.4 DEFINITION** A *structured set* is a pair  $(S, \mathcal{R})$ , consisting of a set  $S$  together with a collection  $\mathcal{R}$  of relations on  $S$ .

This definition is useful since it is easily seen that a great variety of commonplace mathematical objects are examples of structured sets. So, for example, a square is a set of four elements  $a, b, c$  and  $d$  which has (amongst other relations) a ternary relation **Perp** such that **Perp**( $a, b, c$ ), **Perp**( $c, b, a$ ), **Perp**( $b, c, d$ ), **Perp**( $d, c, b$ ), **Perp**( $c, d, a$ ), **Perp**( $a, d, c$ ), **Perp**( $b, a, d$ ), and **Perp**( $d, a, b$ ) are all true, while **Perp**( $x, y, z$ ) is false in the remaining 16



## 6 Chapter One: Symmetry

cases. As we have seen above, a vector space is an example of a structured set, since it is a set with a ternary addition relation and a collection of binary scalar multiplication relations. And in the case when the collection  $\mathcal{R}$  of relations is empty, the structured set is just a set, with no extra structure.

When talking about structured sets, the most natural kinds of transformations to consider are those that preserve the relations, in the following sense:

**1.5 DEFINITION** Let  $\rho$  be an  $n$ -ary relation on the set  $S$ . A transformation  $f: S \rightarrow S$  is said to *preserve*  $\rho$  if  $\rho(f(x_1), f(x_2), \dots, f(x_n))$  whenever  $\rho(x_1, x_2, \dots, x_n)$ .

Consider, for example, a transformation of a vector space which preserves the addition relation and all the scalar multiplication relations. If  $V$  is the vector space and  $T$  the transformation, preservation of addition says that  $\text{Plus}(T\underline{u}, T\underline{v}, T\underline{w})$  whenever  $\text{Plus}(\underline{u}, \underline{v}, \underline{w})$ . Re-expressing this in the usual notation, it says that  $T\underline{w} = T\underline{u} + T\underline{v}$  whenever  $\underline{w} = \underline{u} + \underline{v}$ . More succinctly,  $T(\underline{u} + \underline{v}) = T\underline{u} + T\underline{v}$ , for all  $\underline{u}, \underline{v} \in V$ . This is the way “preservation of addition” is usually defined in texts on vector spaces. Similarly, preservation of the binary relation  $(\text{Mult})_\lambda$  says that  $(\text{Mult})_\lambda(T\underline{u}, T\underline{v})$  whenever  $(\text{Mult})_\lambda(\underline{u}, \underline{v})$ ; in other words,  $T\underline{v} = \lambda T\underline{u}$  whenever  $\underline{v} = \lambda \underline{u}$ , or, more succinctly,  $T(\lambda \underline{u}) = \lambda T\underline{u}$  for all  $\underline{u} \in V$ . So preservation of all the relations  $(\text{Mult})_\lambda$  is preservation of scalar multiplication in the usual sense from vector space theory. Of course, a transformation which does preserve  $\text{Plus}$  and  $(\text{Mult})_\lambda$  for each  $\lambda$  is nothing other than a linear transformation on the space  $V$ .

We are now at last able to explain the intended meaning of Definition 1.1, and give a satisfactorily precise definition of symmetry. Certainly a transformation cannot be said to leave an object essentially unchanged if it is not possible to undo the effects of the transformation. In other words, the transformation must have an inverse. Recall that a function has an inverse if and only if it is bijective (one to one and onto). So a symmetry of an object must at least be a bijective transformation of the object. Now, what about the vague term “essential features” used in Definition 1.1? Of course, the features which are essential are whatever relations we happen to be interested in. We therefore arrive at the following improved definition:

**1.6 DEFINITION** A symmetry of a structured set  $(S, \mathcal{R})$  is a bijective transformation  $T: S \rightarrow S$  such that  $T$  preserves  $\rho$  for all  $\rho \in \mathcal{R}$ .

### §1c The symmetries of some structured sets

The simplest case is, of course, a set with no extra structure to worry about. So consider the structured set  $(S, \mathcal{R})$  where  $R = \emptyset$ , the empty set. If  $T: S \rightarrow S$  is any transformation it is vacuously true that  $T$  preserves all the relations in  $\mathcal{R}$ , since there are no such relations. Hence a symmetry of the structured set  $(S, \emptyset)$  is simply a bijective function from  $S$  to itself. Recall that bijective transformations of a set are commonly known as permutations. The set of all permutations of  $S$  is called the *symmetric group* on  $S$ , and in this course we will use the notation  $\text{Sym}(S)$  for the symmetric group on  $S$ . This traditional name—the symmetric group—is a little unfortunate for us, since it is not particularly compatible with our definition of symmetry. When  $\mathcal{R}$  is not empty, it will not be the case that all elements of the symmetric group  $\text{Sym}(S)$  are symmetries of the structured set  $(S, \mathcal{R})$ . As for the word “group”, we shall not give the general definition of this term until we have seen some more examples.

We digress briefly to discuss permutations, using the example of the set  $S = \{1, 2, 3, 4, 5\}$  to illustrate the terminology and notation. The permutation  $p: S \rightarrow S$  which satisfies

$$p(1) = 4, p(2) = 3, p(3) = 5, p(4) = 1, p(5) = 2$$

is commonly denoted by

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

The elements of the set are listed along the top row, the image under the permutation of each element is written underneath the element.

Recall that multiplication of permutations  $p$  and  $q$  of a set  $S$  is defined to be composition of functions. That is,  $pq$  is the permutation of  $S$  defined by the rule

$$(pq)(x) = p(q(x)) \quad \text{for all } x \in S.$$

So if  $p$  is the permutation of  $\{1, 2, 3, 4, 5\}$  given above and if

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

## 8 Chapter One: Symmetry

then a short calculation yields

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

For example,  $(pq)(3) = p(q(3)) = p(1) = 4$ .

A briefer and hence more convenient notation for permutations is the so-called “disjoint cycle” notation. In this notation the permutation  $q$  above is written as  $(1, 5, 2, 4, 3)$ , indicating that  $q(1) = 5$ ,  $q(5) = 2$ ,  $q(2) = 4$ ,  $q(4) = 3$  and  $q(3) = 1$ . Note that  $q$  could equally well be written as  $(5, 2, 4, 3, 1)$ , or  $(2, 4, 3, 1, 5)$ , or  $(4, 3, 1, 5, 2)$ , or  $(3, 1, 5, 2, 4)$ . It is only the cyclic ordering of the elements that matters. The permutation  $p$  is written as  $(1, 4)(2, 3, 5)$ . (Again there are other possibilities, such as  $(5, 2, 3)(4, 1)$ , obtained by varying the starting elements within each cycle and the order in which the cycles are written.) Multiplication of permutations is just as easy in this notation as the long notation; it can be checked that  $pq = (1, 2)(3, 4, 5)$ . It is normal in this notation to omit cycles of length 1. Thus we write  $p^2 = (2, 5, 3)$  rather than  $p^2 = (1)(4)(2, 5, 3)$ . The identity permutation, in which all cycles have length 1, will be written simply as  $i$ .

Returning now to the primary topic of this section, consider the set of all symmetries of a square  $S$  with vertices  $a, b, c, d$ . Any permutation of  $\{a, b, c, d\}$  which preserves the perpendicularity relation will be a symmetry of  $S$ . There are only 24 permutations altogether in the symmetric group  $\text{Sym}\{a, b, c, d\}$ , and it is a straightforward task to find which of them preserve Perp. They are

$$\begin{aligned} p_1 &= i & p_2 &= (a, b, c, d) & p_3 &= (a, c)(b, d) & p_4 &= (a, d, c, b) \\ p_5 &= (a, c) & p_6 &= (a, d)(b, c) & p_7 &= (b, d) & p_8 &= (a, b)(c, d). \end{aligned}$$

It is clear that if two transformations  $p$  and  $q$  both have the property of leaving an object essentially unchanged, then the composite transformation  $pq$  (obtained by applying  $q$  first and following it by  $p$ ) will also leave the object essentially unchanged. Thus, if we define  $G = \{p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8\}$ , it must be true that every product  $p_i p_j$  of permutations in the set  $G$  will also be in  $G$ . With some calculation, it can be verified that the following

multiplication table is satisfied:

	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$	$p_7$	$p_8$
$p_1$	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$	$p_7$	$p_8$
$p_2$	$p_2$	$p_3$	$p_4$	$p_1$	$p_6$	$p_7$	$p_8$	$p_5$
$p_3$	$p_3$	$p_4$	$p_1$	$p_2$	$p_7$	$p_8$	$p_5$	$p_6$
$p_4$	$p_4$	$p_1$	$p_2$	$p_3$	$p_8$	$p_5$	$p_6$	$p_7$
$p_5$	$p_5$	$p_8$	$p_7$	$p_6$	$p_1$	$p_4$	$p_3$	$p_2$
$p_6$	$p_6$	$p_5$	$p_8$	$p_7$	$p_2$	$p_1$	$p_4$	$p_3$
$p_7$	$p_7$	$p_6$	$p_5$	$p_8$	$p_3$	$p_2$	$p_1$	$p_4$
$p_8$	$p_8$	$p_7$	$p_6$	$p_5$	$p_4$	$p_3$	$p_2$	$p_1$

It is also clear that the inverse of a transformation which leaves an object essentially unchanged will also have the same property. And, indeed, it is easily verified that  $p_i^{-1} \in G$  for each  $i$ . Specifically,  $p_2^{-1} = p_4$  and  $p_4^{-1} = p_2$ , while  $p_i^{-1} = p_i$  in all the other cases.

The two properties of  $G$  which we have mentioned—that  $pq \in G$  whenever  $p, q \in G$  and  $p^{-1} \in G$  whenever  $p \in G$ —are the key properties of sets of symmetries. Any nonempty set of transformations possessing these two properties is called a *group* of transformations.

**1.7 DEFINITION** Let  $S$  be any set. A set  $G$  of transformations of  $S$  is called a *group* of transformations if  $G \neq \emptyset$  and

- (i)  $pq \in G$  for all  $p, q \in G$ , (*closure under multiplication*)
- (ii)  $p$  is bijective and  $p^{-1} \in G$  for all  $p \in G$ . (*closure under inversion*)

If  $S$  is any object whatever, then the set of all symmetries of  $S$  will be a group of transformations. Observe that the identity transformation is always a symmetry.

For our final example in this section we consider the group of all symmetries of a Euclidean vector space  $V$ . A Euclidean vector space, or real inner product space, is a vector space equipped with an inner product, which is a symmetric, bilinear and positive definite function  $V \times V \rightarrow \mathbb{R}$ . In these notes we will always use the “dot” notation for inner products: the inner product of vectors  $u, v \in V$  will be written as  $u \cdot v$ . The three defining properties of inner products are

- (i)  $u \cdot v = v \cdot u$  for all  $u, v \in V$ ,
- (ii)  $(\lambda u + \mu v) \cdot w = \lambda u \cdot w + \mu v \cdot w$  for all  $\lambda, \mu \in \mathbb{R}$  and  $u, v, w \in V$ ,

(iii)  $v \cdot v > 0$  for all nonzero  $v \in V$ .

To describe a Euclidean space  $V$  as a structured set we need the Plus and  $(\text{Mult})_\lambda$  relations introduced previously for vector spaces, and further binary relations  $(\text{Dot})_\lambda$  defined as follows:  $(\text{Dot})_\lambda(u, v)$  if  $u \cdot v = \lambda$ . A transformation  $T: V \rightarrow V$  preserves  $(\text{Dot})_\lambda$  if and only if  $Tu \cdot Tv = \lambda$  whenever  $u \cdot v = \lambda$ . So  $T$  preserves all these relations if and only if

$$(1.7.1) \quad Tu \cdot Tv = u \cdot v \quad \text{for all } u, v \in V.$$

Thus a symmetry of  $V$  is a bijective linear transformation satisfying (1.7.1). This is the usual definition of an *orthogonal transformation*: a bijective linear transformation which preserves the inner product. The set of all orthogonal transformations of a Euclidean space  $V$  is called the *orthogonal group* of  $V$ , and is denoted by  $O(V)$ .

The group  $O(V)$  has infinitely many elements; so it is not possible to write out a complete multiplication table for  $O(V)$  as we could for the eight-element group considered above. In these notes we will mainly concentrate on groups with only finitely many elements, but even so it will not usually be practicable to write out multiplication tables, as the groups will be too big. In the next chapter we will look at some of the conceptual tools which group theorists use in their attempts to make big groups understandable.

Bijjective structure-preserving transformations of algebraic systems such as fields, vector spaces and the like, are usually called *automorphisms* rather than symmetries. Thus the usual terminology is to call the orthogonal group  $O(V)$  of a Euclidean space  $V$  not the symmetry group, but the *automorphism group*, of  $V$ . In these notes we will use the terms “symmetry” and “automorphism” interchangeably, but with a tendency to prefer “symmetry” in geometric situations, “automorphism” in algebraic ones.

### §1d Some groups of transformations of a set with four elements

In the previous section we exhibited a set of eight transformations which form a group of transformations of the set  $\{a, b, c, d\}$ , and we wrote down the multiplication table for this eight-element group. In this section we will list some more examples.

Recall that for a set  $G$  of transformations of a set  $S$  to be a group, all that is necessary is for the inverse of every element of  $G$  to also be in  $G$

and the product of every pair of elements of  $G$  to also be in  $G$ . Clearly, one method of finding examples of such sets  $G$  is to start with a set containing just a few bijective transformations, randomly chosen, calculate  $x^{-1}$  and  $xy$  for all choices of  $x$  and  $y$  in this initial set, and add all of the resulting transformations to the set; then repeat this process, and keep on going until the set stops getting bigger at each stage. Maybe the process will not stop until the set contains every bijective transformation of  $S$ ; in this case the initial set of transformations is said to generate the full symmetric group  $\text{Sym}(S)$ . But sometimes we can find other examples of groups by this method.

If we start with just a single transformation  $g$ , it is not hard to see that the group of transformations that we end up with will just consist of all powers of  $g$  and all powers of  $g^{-1}$ , and the identity. These single-generator groups are known as *cyclic groups*. Some of the groups of transformations of  $S = \{1, 2, 3, 4\}$  that can be obtained like this are

$$\begin{aligned} G_1 &= \{i\} \\ G_2 &= \{i, (1, 2)\} \\ G_3 &= \{i, (1, 2)(3, 4)\} \\ G_4 &= \{i, (1, 2, 3), (1, 3, 2)\} \\ G_5 &= \{i, (1, 2, 3, 4), (1, 3)(2, 4), (1, 4, 3, 2)\}. \end{aligned}$$

The element  $(1, 2, 3, 4)$  generates the group  $G_5$ , in the sense that every element of  $G_5$  can be expressed as a power of  $(1, 2, 3, 4)$ . The element  $(1, 4, 3, 2)$  is also a generator, but  $(1, 3)(2, 4)$  is not; in fact, the group generated by  $(1, 3)(2, 4)$  is just  $\{i, (1, 3)(2, 4)\}$ , a proper subset of  $G_5$ .

1.8 DEFINITION The *order* of a group  $G$  is the number of elements of  $G$ .

The multiplication table for a cyclic group of order four looks like this:

	$i$	$x$	$x^2$	$x^3$
$i$	$i$	$x$	$x^2$	$x^3$
$x$	$x$	$x^2$	$x^3$	$i$
$x^2$	$x^2$	$x^3$	$i$	$x$
$x^3$	$x^3$	$i$	$x$	$x^2$

Observe that each row of the table is the same as the previous row moved across one place to the left, the leftmost element jumping across to become

the rightmost element. It is easily seen that the same will be true for the multiplication table of a cyclic group with any number of elements, provided the elements are initially listed in the order  $i, x, x^2, x^3, \dots$ . Note, however, that if we take the elements in a different order then the multiplication table will look a little different: it may take more than just a glance at a multiplication table to determine whether or not a group of transformations is a cyclic group.

Two groups which have the same multiplicative structure, in the sense that the same multiplication table applies to both groups, are said to be *isomorphic*. (We will give a more precise definition of this term in a later chapter.) The groups  $G_2$  and  $G_3$  above are isomorphic, since they are both cyclic groups of order two. So although  $(1, 2)$  and  $(1, 2)(3, 4)$  are rather different as transformations, nevertheless the groups they generate are isomorphic to each other. Internally, so to speak, the two groups are the same. A slightly more elaborate example of the same phenomenon occurs with transformations of the set  $\{1, 2, 3, 4, 5, 6\}$ . Here the transformations  $(1, 2)(3, 4, 5)$  and  $(1, 2, 3, 4, 5, 6)$  both generate cyclic groups of order six, although in some other ways they are not particularly alike. Isomorphism is a valuable concept, since any facts we may discover about the internal structure of a group will automatically be true also for any isomorphic group.

Returning to our discussion of groups of transformations of  $\{1, 2, 3, 4\}$ , we give some examples of groups which are not cyclic, but can be generated by two elements:

$$\begin{aligned} G_6 &= \{i, (1, 2), (3, 4), (1, 2)(3, 4)\} \\ G_7 &= \{i, (1, 2)(3, 4), (1, 3)(2, 4), (1, 4)(2, 3)\} \\ G_8 &= \{i, (1, 2), (1, 3), (2, 3), (1, 2, 3), (1, 3, 2)\}. \end{aligned}$$

The groups  $G_6$  and  $G_7$  are isomorphic: it can be checked easily that the following multiplication table is applicable to both of them:

	$i$	$a$	$b$	$c$
$i$	$i$	$a$	$b$	$c$
$a$	$a$	$i$	$c$	$b$
$b$	$b$	$c$	$i$	$a$
$c$	$c$	$b$	$a$	$i$

Groups with this multiplication table are said to be isomorphic to *Klein's four group*. It can be shown quite easily that any group of order four must either be cyclic or else isomorphic to Klein's four group. Hence group theorists

often say “there are only two groups of order four” (namely, the cyclic group and the Klein four group). This is somewhat slipshod—we probably ought to say “there are only two isomorphism types of groups of order four—but makes the point that there are only two possibilities for the internal structure of a group with four elements.

The group  $G_8$  above consists of all the permutations of  $\{1, 2, 3, 4\}$  which fix 4. It is clearly essentially the same as  $\text{Sym}\{1, 2, 3\}$ . It can be checked that the two elements  $(1, 2)$  and  $(2, 3)$  generate this group. That is, every element of  $G_8$  can be expressed in terms of these two. For example,  $(1, 3) = (1, 2)(2, 3)(1, 2)$ . There are several other two-element generating sets for  $G_8$  as well.

One final group of transformations of  $\{1, 2, 3, 4\}$  that we wish to mention at this point is the *alternating group* on this set. It consists of all the even permutations. It is a well known fact that the product of two even permutations is also even, and the inverse of an even permutation is also even; so it follows that the set of all even permutations is indeed a group. In this case—dealing with a set of size four—the even permutations are the identity, the eight 3-cycles (such as  $(1, 2, 3)$ ,  $(1, 2, 4)$ , and so on), and the three permutations which are the product of two disjoint two cycles (namely,  $(1, 2)(3, 4)$ ,  $(1, 3)(2, 4)$  and  $(1, 4)(2, 3)$ ). This makes 12 even permutations altogether, which is as it should be, since we would naturally expect the numbers of even and odd permutations to be the same, and the order of  $\text{Sym}\{1, 2, 3, 4\}$ —the total number of permutations—is 24. Here is the multiplication table for  $\text{Alt}\{1, 2, 3, 4\}$ :

	$i$	$a$	$b$	$c$	$t_1$	$t_2$	$t_3$	$t_4$	$s_1$	$s_2$	$s_3$	$s_4$
$i$	$i$	$a$	$b$	$c$	$t_1$	$t_2$	$t_3$	$t_4$	$s_1$	$s_2$	$s_3$	$s_4$
$a$	$a$	$i$	$c$	$b$	$t_2$	$t_1$	$t_4$	$t_3$	$s_4$	$s_3$	$s_2$	$s_1$
$b$	$b$	$c$	$i$	$a$	$t_3$	$t_4$	$t_1$	$t_2$	$s_2$	$s_1$	$s_4$	$s_3$
$c$	$c$	$b$	$a$	$i$	$t_4$	$t_3$	$t_2$	$t_1$	$s_3$	$s_4$	$s_1$	$s_2$
$t_1$	$t_1$	$t_4$	$t_2$	$t_3$	$s_1$	$s_3$	$s_4$	$s_2$	$i$	$a$	$b$	$c$
$t_2$	$t_2$	$t_3$	$t_1$	$t_4$	$s_4$	$s_2$	$s_1$	$s_3$	$a$	$i$	$c$	$b$
$t_3$	$t_3$	$t_2$	$t_4$	$t_1$	$s_2$	$s_4$	$s_3$	$s_1$	$b$	$c$	$i$	$a$
$t_4$	$t_4$	$t_1$	$t_3$	$t_2$	$s_3$	$s_1$	$s_2$	$s_4$	$c$	$b$	$a$	$i$
$s_1$	$s_1$	$s_2$	$s_3$	$s_4$	$i$	$a$	$b$	$c$	$t_1$	$t_4$	$t_2$	$t_3$
$s_2$	$s_2$	$s_1$	$s_4$	$s_3$	$a$	$i$	$c$	$b$	$t_3$	$t_2$	$t_4$	$t_1$
$s_3$	$s_3$	$s_4$	$s_1$	$s_2$	$b$	$c$	$i$	$a$	$t_4$	$t_1$	$t_3$	$t_2$
$s_4$	$s_4$	$s_3$	$s_2$	$s_1$	$c$	$b$	$a$	$i$	$t_2$	$t_3$	$t_1$	$t_4$



**14** *Chapter One: Symmetry*

In this table  $a = (1, 2)(3, 4)$  and  $t_1 = (1, 2, 3)$ . We leave it to the reader to identify the other permutations.

Observe that for every group  $G$  of transformations of  $\{1, 2, 3, 4\}$  that we have been able to construct, the order of  $G$  is a divisor of 24, the order of the whole symmetric group on the set.

# 2

## Introductory abstract group theory

In Chapter 1 we defined the concept of a group of transformations. Since the whole purpose of group theory is to study groups of transformations, there is really no need to define any other kinds of groups. But, on the other hand, we can obtain a theory which is (superficially) more general by defining groups by means of a set of axioms. Elements of a group then do not have to be transformations, they can be any kind of thing, just so long as the axioms are satisfied. Nevertheless, we should keep firmly in our minds the fact that the way groups arise in practice, is as groups of symmetries of objects.

### §2a Group axioms

**2.1 DEFINITION** A group is a set  $G$  equipped with an operation  $(g, h) \mapsto gh$  (usually called “multiplication”), satisfying the following axioms:

- (i)  $(xy)z = x(yz)$  for all  $x, y, z \in G$ , (*Associativity*)
- (ii) there exists an element  $e \in G$  such that
  - (a)  $ex = xe = x$  for all  $x \in G$ , (*Existence of an identity element*)
  - (b) for each  $x \in G$  there exists  $y \in G$  such that  $xy = yx = e$ .  
(*Existence of inverses*)

By an *operation* on a set  $G$  we simply mean a rule which gives an element of  $G$  for every pair of elements of  $G$ . This is, in fact, the same thing as a function from  $G \times G$  to  $G$ . The only difference is notational: we would usually write  $f(x, y)$  for the result of applying a function  $f$  to a given pair of elements  $(x, y)$ , but if we were calling the function an operation we would use a notation like  $x * y$ , or  $x + y$ , or  $x \circ y$ , or simply  $xy$ , for the result of applying the operation to the pair  $(x, y)$ . In fact, for groups we usually opt for the last of these alternatives, as indeed we have done in Definition 2.1 itself.

One immediate consequence of part (ii) of Definition 2.1 is that the empty set cannot be a group. A group must always have at least one element, namely, the element  $e$  which satisfies Definition 2.1 (ii). In fact, a group need not have any other elements; it is easy to check that a set with a single element, with multiplication defined in the only way possible, does satisfy Definition 2.1, and is therefore a group.

An element  $e$  which satisfies Definition 2.1 (ii) (a) is called an *identity element*. If  $e$  is an identity element then elements  $x$  and  $y$  satisfying  $xy = yx = e$  (as in 2.1 (ii) (b)) are said to be *inverses* of each other. It is not necessary for an identity element to be called  $e$ ; indeed, for groups of transformations the identity element has to be the identity transformation, and we have already decided to denote the identity transformation by  $i$ .

A group of transformations of a set  $S$ , as defined in Definition 1.7, is easily seen to be a group in the sense of Definition 2.1. Indeed, suppose  $G$  satisfies the requirements of Definition 1.7. Observe first of all that  $G$  satisfies 2.1 (i), since it is trivial—we omit the proof—that composition of functions is associative. Since 1.7 requires that  $G$  is nonempty, we can choose an element  $p \in G$ . By part (ii) of 1.7 we know that  $p$  is a bijective transformation of  $S$  and that  $p^{-1} \in G$ . Now since  $p, p^{-1} \in G$  we can apply part (i) of 1.7 and deduce that  $pp^{-1} \in G$ . So  $i$ , the identity transformation on  $S$ , is in the set  $G$ . Since it is readily verified that  $ix = xi = x$  for all transformations  $x$  of  $S$ , we can conclude that part (ii) (a) of 2.1 is satisfied, with  $e = i$ . Finally, part (ii) (b) of 2.1 is an immediate consequence of 1.7 (ii).

Something which is notable for its absence from Definition 2.1 is the commutative law:  $xy = yx$  for all  $x$  and  $y$ . Groups do not have to satisfy this, and indeed we have already met examples of groups which do not. The elements  $(1, 2)$  and  $(2, 3)$  in the group  $\text{Sym}\{1, 2, 3\}$  satisfy

$$(1, 2)(2, 3) = (1, 2, 3) \neq (1, 3, 2) = (2, 3)(1, 2).$$

A glance at the multiplication table of the group of symmetries of a square (calculated in Chapter 1) shows that this group also fails to satisfy the commutative law. Groups which do satisfy the commutative law are called *Abelian* groups. Cyclic groups, for example, are Abelian, and so is the Klein four group.

Groups of transformations are not the only groups. At least, there are groups which, at first sight, do not appear to be groups of transformations,

although it may only take minor changes in nomenclature to make them into groups of transformations. The set of all nonzero real numbers, with the operation being ordinary multiplication of real numbers, is our first example.

Define  $\mathbb{R}^\times = \{x \in \mathbb{R} \mid x \neq 0\}$ . We must check first of all that multiplication of real numbers does give an operation on the set  $\mathbb{R}^\times$ . This is slightly less obvious than you might think. Certainly you can multiply any two elements of  $\mathbb{R}^\times$  and get an answer. But for multiplication to be an operation on  $\mathbb{R}^\times$  we need also that this answer is always in the set  $\mathbb{R}^\times$ . It is true, of course: the product of two nonzero real numbers is always a nonzero real number. It now only remains to check that the axioms (i) and (ii) of 2.1 are satisfied, and this is trivial. The associative law is well known to hold for multiplication of real numbers, the number 1 has the property required of an identity element, and for each  $x \in \mathbb{R}^\times$  the number  $y = 1/x$  has the property required for axiom (ii) (b). Note that this group,  $\mathbb{R}^\times$ , is an Abelian group, since multiplication of real numbers is commutative.

Note that everything we have just said about the set of all nonzero real numbers applies equally to the set of all nonzero elements of any field. If  $F$  is a field then  $F^\times = \{\alpha \in F \mid \alpha \neq 0\}$  is a group under the multiplication operation of  $F$ . This group is generally known as the *multiplicative group of  $F$* .

Fields, of course, have two operations, addition and multiplication. Under its addition operation, the set of all elements of a field  $F$  forms a group. This is called the *additive group of  $F$* . So the field and its additive group are the same set of elements, the difference is merely that multiplication is ignored when thinking of the field as an additive group. Checking that the addition operation of a field does satisfy the group axioms is trivial: associativity of addition is a field axiom, the zero element of the field is an additive identity (since a field axiom states that  $0 + x = x + 0 = x$  for all  $x$  in the field), for each  $x$  the element  $y = -x$  has the property that  $x + y = y + x = 0$ , as required for group axiom (ii) (b). Note that both the additive and the multiplicative groups of a field are Abelian groups.

We can generalize the preceding examples by considering matrices over a field  $F$ . If  $m$  and  $n$  are positive integers then, as usual, we define the sum of two  $m \times n$  matrices  $A$  and  $B$  by the rule  $(A+B)_{ij} = A_{ij} + B_{ij}$ . This addition operation makes the set of all  $m \times n$  matrices over  $F$  into an Abelian group, as can be easily checked. In the case  $m = n = 1$  we recover the additive group of the field.

Define  $\text{GL}_n(F)$  to be the set of all  $n \times n$  invertible matrices over the field  $F$ . Matrix multiplication, defined as usual by  $(AB)_{ij} = \sum_{k=1}^n A_{ik}B_{kj}$ , gives an operation on  $\text{GL}_n(F)$ , since the product of two invertible  $n \times n$  matrices is always an invertible  $n \times n$  matrix. Associativity of matrix multiplication is well known; so group axiom (i) holds. The  $n \times n$  identity matrix  $I$  has the property that  $XI = IX = X$  for all  $X \in \text{GL}_n(F)$ ; so group axiom (ii) (a) is satisfied. And for each invertible matrix  $X$  there is an invertible matrix  $Y$  (namely,  $Y = X^{-1}$ ) with the property that  $XY = YX = I$ ; so group axiom (ii) (b) holds. Hence  $\text{GL}_n(F)$  is a group. It is known as the *general linear group of degree  $n$*  over the field  $F$ . Note that  $\text{GL}_n(F)$  is not Abelian—since matrix multiplication is not commutative—except in the case  $n = 1$ . When  $n = 1$  we recover the multiplicative group of  $F$ . Note also that, in view of the correspondence between matrices and linear transformations, as described in linear algebra textbooks,  $\text{GL}_n(F)$  is essentially the same as the set of all invertible linear transformations on a vector space of dimension  $n$  over  $F$ ; so we really are just talking about transformations in a different guise.

## §2b Basic deductions from the axioms

Let  $G$  be a group. As a trivial notational matter, observe (by axiom (i)) that for  $x, y, z \in G$  we may use  $xyz$ , without any brackets, to unambiguously denote the element  $x(yz) = (xy)z$ . We have to be careful about the order in which the three elements are written, since  $G$  may not be Abelian, but the bracketing is unimportant. If  $w$  is also an element of  $G$  then applying axiom (i) with  $z$  replaced by  $zw$  gives  $x(y(zw)) = (xy)(zw)$ , and similarly  $(xy)(zw) = ((xy)z)w$ . So the expression  $xyzw$  is also unambiguous, as all possible bracketings yield the same result. The same applies for products with any number of factors, and so henceforth we will usually omit brackets from long products.

The most important basic property of groups is that left and right cancellation are both valid.

**2.2 PROPOSITION** *Let  $x, y, z \in G$ . Then*

- (i) *if  $xy = xz$  then  $y = z$ ,*
- (ii) *if  $yx = zx$  then  $y = z$ .*

**Proof.** Suppose that  $xy = xz$ . By 2.1 (ii) (a) there exists  $e \in G$  such that  $eg = ge = g$  for all  $g \in G$ , and by 2.1 (ii) (b) there exists  $w \in G$  such that

$xw = wx = e$ . For this element  $w$  we have that  $w(xy) = w(xz)$ , since  $xy$  and  $xz$  are the same element. So we have

$$ey = (wx)y = w(xy) = w(xz) = (wx)z = ez,$$

and we deduce that  $y = z$  since the basic property of  $e$  tells us that  $ey = y$  and  $ez = z$ .

The proof of the right cancellation property is totally analogous to that just given for left cancellation. Assuming that  $yx = zx$  we find, with  $w$  as above, that

$$y = ye = y(xw) = (yx)w = (zx)w = z(xw) = ze = z.$$

□

Note that Proposition 2.2 does not say that if  $xy = zx$  then  $y = z$ . Remember that groups do not necessarily satisfy the commutative law, and hence it is important keep the factors in products in their rightful order.

A trivial but important fact is that the identity element of a group is unique.

**2.3 PROPOSITION** *If  $e, f \in G$  and*

(i)  $ex = xe = x$  for all  $x \in G$

(ii)  $fx = xf = x$  for all  $x \in G$

then  $e = f$ .

**Proof.** By hypothesis (i) with  $f$  in place of  $x$  we have that  $ef = f$ . But by hypothesis (ii) with  $e$  in place of  $x$  we have that  $ef = e$ . □

Note that group axiom (ii) (a) guarantees the existence of an identity element, and Proposition 2.3 says that there cannot be two different identity elements. So we can, henceforth, safely talk about *the* identity element of a group. Similarly, each element of  $G$  has a unique inverse.

**2.4 PROPOSITION** *Let  $x \in G$  and let  $e$  be the identity element of  $G$ . If  $y, z \in G$  satisfy*

(i)  $xy = yx = e$ , and

(ii)  $xz = zx = e$ ,

then  $y = z$ .

**Proof.** Since hypotheses (i) and (ii) give  $xy = xz$  (both equal to  $e$ ), the conclusion  $y = z$  is immediate from Proposition 2.2. □

We know from Definition 2.1 (ii) (b) that each  $x \in G$  has an inverse (an element satisfying  $xy = yx = e$ ), and Proposition 2.4 says that  $x$  cannot have two different inverses. So we can, henceforth, safely talk about *the* inverse of  $x$ . It is customary to denote the inverse of  $x$  by  $x^{-1}$ . We should also observe the following trivial fact, the proof of which is virtually the same as that of Proposition 2.4.

2.5 PROPOSITION *If  $x, y \in G$  and either  $xy = e$  or  $yx = e$  then  $y = x^{-1}$ .*

The following is also easy to prove.

2.6 PROPOSITION *If  $x, y \in G$  then  $(xy)^{-1} = y^{-1}x^{-1}$ .*

Another similar property of groups, easy but important, concerns the solvability of simple kinds of equations.

2.7 PROPOSITION *Let  $g, h \in G$ . Then*

- (i) *the equation  $gx = h$  has a unique solution  $x \in G$ , and*
- (ii) *the equation  $yg = h$  has a unique solution  $y \in G$ .*

**Proof.** If we define  $x = g^{-1}h$  then  $gx = g(g^{-1}h) = (gg^{-1})h = eh = h$ . This establishes the existence of a solution to  $gx = h$ . Uniqueness of the solution follows easily from 2.2: if  $z$  is another solution then  $gx = gz$  (both equal to  $h$ ), and cancelling  $g$  gives  $x = z$ .

The proof of the second part is totally analogous, and is omitted.  $\square$

An  $n \times n$  array of  $n$  symbols is called a *Latin square* if each symbol occurs exactly once in each row of the array and exactly once in each column of the array. It is a consequence of Proposition 2.7 that the multiplication table of a group with  $n$  elements is a Latin square. For example, consider an arbitrary row of a group multiplication table corresponding to an arbitrary element,  $g$ . That is, consider the row which gives the values of all the products  $gx$ , as  $x$  varies over all elements of the group. The statement that an arbitrary element  $h$  occurs exactly once in this row is precisely the statement that  $gx = h$  has exactly one solution, and this is part (i) of 2.7. The reader can easily observe that all the multiplication tables that appear in Chapter 1 really are Latin squares. Unfortunately, the converse statement, that all Latin squares are group multiplication tables, is not true; the smallest counterexample occurs when  $n = 5$ . The problem of describing all Latin squares, and the problem

of describing all multiplication tables that satisfy the group axioms, are both unsolved.

To end this section we discuss some more notational matters. There is a convention, almost universally observed throughout mathematics, that the symbol “+” is only ever used for operations which are commutative. So additive groups—that is, groups for which the operation is designated by the “+” symbol—are always Abelian. Hand in hand with this is another convention: when the operation is written as addition the identity element is always written as 0, and the inverse of an element  $x$  is always written as  $-x$  rather than  $x^{-1}$ . Indeed, for additive groups the term “identity element” is not used: instead, 0 is called the *zero element*. Similarly, for additive groups it is usual to call  $-x$  the *negative* of  $x$ , and not use the word “inverse”.

## §2c Subgroups and cosets

**2.8 DEFINITION** Let  $G$  be a group and  $H$  a subset of  $G$ . We say that  $H$  is a *subgroup* of  $G$  if  $H$  is itself a group, and for all  $x, y \in H$  the product  $xy$  is the same whether calculated via  $G$ 's operation or  $H$ 's.

If  $*$  is an operation on a set  $S$  we say that a subset  $T$  of  $S$  is *closed under the operation  $*$*  if  $x * y \in T$  whenever  $x, y \in T$ . Clearly if  $T$  is closed under  $*$  then an operation  $\circ$  can be defined on  $T$  by the rule  $x \circ y = x * y$  for all  $x, y \in T$ . This does not define an operation on  $T$  if  $T$  is not closed, since the definition of the term operation requires that if  $\circ$  is to be an operation on  $T$ , then  $x \circ y$  must be in  $T$  for all  $x, y \in T$ . If  $T$  is closed and if  $\circ$  is defined in this way, we say that  $\circ$  is the operation  $T$  *inherits* from the operation  $*$  on  $G$ . It is immediate from Definition 2.8 that a subgroup  $H$  of a group  $G$  has to be closed under the operation that makes  $G$  into a group, and the operation for  $H$  must be the inherited operation.

To avoid any further use of the tedious phrase “the operation that makes  $G$  into a group” we will henceforth call the operation “multiplication”. Everything we prove will still be able to be applied to additive groups by replacing the word “multiplication” by “addition” wherever appropriate.

**2.9 PROPOSITION** Let  $H$  be a subgroup of the group  $G$ . Then the identity element of  $H$  is the same as the identity element of  $G$ . Furthermore, if  $h$  is



any element of  $H$  then the inverse of  $h$  in the group  $G$  is the same as the inverse of  $h$  in the group  $H$ .

**Proof.** Let  $e$  be the identity element of  $G$  and  $f$  the identity element of  $H$ . Then  $fh = h$  for all  $h \in H$ , and so in particular  $f^2 = f$ . But  $ex = x$  for all  $x \in G$ , and so in particular  $ef = f$  (since  $f \in H \subseteq G$ ). Hence  $ef = ff$ , and by Proposition 2.2 it follows that  $e = f$ .

Let  $h$  be an arbitrary element of  $H$ . Let  $g$  be the inverse of  $h$  in the group  $G$  and let  $k$  be the inverse of  $h$  in the group  $H$ . The definition of inverse says that  $hg = e$  and  $hk = f$ , but since we have already seen that  $e = f$  we can conclude that  $hg = hk$ . By Proposition 2.2 again,  $g = k$ .  $\square$

The second part of Proposition 2.9 shows that if  $H$  is a subgroup of  $G$  and  $h \in H$  then  $h^{-1}$  is unambiguously defined: it doesn't matter whether you are thinking of  $h$  as an element of  $H$  or as an element of  $G$ , its inverse is the same. This is just as well, for we would have had serious notational problems otherwise.

It will be convenient to say that a subset  $S$  of a group  $G$  is *closed under inversion* if  $x^{-1} \in S$  whenever  $x \in S$ . In particular, if  $H$  is a subgroup of  $G$  then  $H$  must be closed under inversion, since for  $H$  to satisfy group axiom (ii) (b) it is necessary for the inverse of each element of  $H$  to also be in  $H$ .

It turns out that the subsets of a group  $G$  which are subgroups are precisely those nonempty subsets which satisfy the two closure properties we have just discussed, closure under multiplication and closure under inversion.

**2.10 THEOREM** *Let  $G$  be a group and  $H$  a subset of  $G$ . Then  $H$  is a subgroup of  $G$  if and only if  $H$  is nonempty and closed under multiplication and inversion.*

**Proof.** We observed earlier that the empty set is not a group, and we have just shown that a subgroup is necessarily closed under multiplication and inversion; so it remains only to show that a nonempty subset of  $G$  which is closed under multiplication and inversion is a subgroup.

Suppose that  $\emptyset \neq H \subseteq G$  and that  $H$  is closed under multiplication and inversion. Then as explained earlier in this section,  $H$  inherits a multiplication from  $G$ ; so our task is simply to show that this operation on  $H$

satisfies (i) and (ii) of Definition 2.1. We know that the multiplication on  $G$  itself does satisfy these axioms.

Axiom (i) is trivially disposed of: we know that  $(xy)z = x(yz)$  for all  $x, y, z \in G$ , and so it follows that  $(xy)z = x(yz)$  for all  $x, y, z \in H$  (since  $H \subseteq G$ ).

Since  $H \neq \emptyset$  we know that there exists at least one element in  $H$ . Fix one such element, and call it  $h$ . By closure under inversion we deduce that  $h^{-1} \in H$ , and so by closure under multiplication it follows that  $hh^{-1} \in H$  also. That is,  $e \in H$ , where  $e$  is the identity element of  $G$ . Now by group axiom (ii) (a) applied to  $G$  we know that  $ex = xe = x$  for all  $x \in G$ , and so it is certainly true that  $ex = xe = x$  for all  $x \in H$  (since  $H \subseteq G$ ). Since  $e \in H$  this shows that group axiom (ii) (a) holds for  $H$ .

Group axiom (ii) (b) is all that remains: we must that for all  $x \in H$  there exists  $y \in H$  with  $xy = yx = e$ . Let  $x \in H$ , and put  $y = x^{-1}$ , the inverse of  $x$  in  $G$ . Then by closure of  $H$  under inversion we know that  $y \in H$ , and by definition of  $x^{-1}$  we know that  $xy = yx = e$ .  $\square$

Comparing Theorem 2.10 and Definition 1.7, it can be seen that a group of transformations of a set  $S$  is just the same thing as a subgroup of  $\text{Sym}(S)$  (the group of all invertible transformations of  $S$ ).

If  $H$  is a subgroup of the group  $G$  and  $x \in G$  then we define

$$xH = \{ xh \mid h \in H \},$$

and similarly

$$Hx = \{ hx \mid h \in H \}.$$

Those subsets of  $G$  of the form  $xH$  for some  $x \in G$  are called the *left cosets* of  $H$  in  $G$ , and similarly the subsets  $Hx$  are the *right cosets* of  $H$  in  $G$ . In our initial discussion of cosets, we give detailed proofs of various properties of left cosets only, although there are corresponding results for right cosets which can be proved by totally analogous arguments. In a subsequent section we will cover the same ground again using some alternative arguments, and in this subsequent section we will, for variety, use right cosets rather than left cosets.

It is important to realize that if  $x_1$  and  $x_2$  are two distinct elements of  $G$ , it is nevertheless possible that  $x_1H = x_2H$ . Our next proposition describes precisely when this happens.

2.11 PROPOSITION Let  $H$  be a subgroup of the group  $G$  and let  $x_1, x_2 \in G$ . Then the following conditions are all equivalent to each other:

- (i)  $x_1H = x_2H$ ,
- (ii)  $x_2 \in x_1H$ ,
- (iii)  $x_2 = x_1h$  for some  $h \in H$ ,
- (iv)  $x_1^{-1}x_2 \in H$ .

**Proof.** Our strategy for proving this is as follows. First of all we will prove that if (i) holds then (ii) holds also. Then we will give a separate proof that if (ii) holds then (iii) holds. Then we will give a separate proof that if (iii) holds then (iv) holds, and finally we will give yet another separate proof that if (iv) holds then (i) holds. Once all of these implications have been established it will follow that if any one of the four conditions holds then all the others do too.

First of all, then, assume that (i) holds; that is,  $x_1H = x_2H$ . By Proposition 2.9 we have that  $e \in H$  (where  $e$  is the identity element of  $G$ ), and so

$$x_2 = x_2e \in \{x_2h \mid h \in H\} = x_2H.$$

By our hypothesis that  $x_1H = x_2H$  we conclude that  $x_2 \in x_1H$ ; that is, (ii) holds. So we have established that (ii) holds whenever (i) holds; this was our first objective.

Dispense now with the assumption that (i) holds, but assume that (ii) holds. Then  $x_2 \in x_1H = \{x_1h \mid h \in H\}$ , and so  $x_2 = x_1h$  for some  $h \in H$ . That is, (iii) holds. So we have shown that (ii) implies (iii).

Now assume that (iii) holds. Then there is an element  $h \in H$  with  $x_2 = x_1h$ . Multiplying both sides this equation on the left by  $x_1^{-1}$  gives  $x_1^{-1}x_2 = x_1^{-1}x_1h = h$ , and so it follows that  $x_1^{-1}x_2 \in H$ . Hence (iv) holds.

Finally, assume that (iv) holds, so that  $x_1^{-1}x_2 \in H$ . Theorem 2.10 guarantees that  $H$  is closed under inversion, and so using Proposition 2.6 we can say that  $x_2^{-1}x_1 = (x_1^{-1}x_2)^{-1} \in H$ . Our aim is to prove that  $x_1H = x_2H$ , and we do this by first showing that  $x_2H \subseteq x_1H$ , then that  $x_1H \subseteq x_2H$ .

Let  $g$  be an arbitrary element of  $x_2H$ . Then  $g = x_2h$  for some  $h \in H$ , and it follows that

$$g = ex_2h = x_1x_1^{-1}x_2h = x_1k$$

where  $k = (x_1^{-1}x_2)h$ . But  $x_1^{-1}x_2$  and  $h$  are both elements of  $H$ , and by Theorem 2.10 we know that  $H$  is closed under multiplication. Hence  $k \in H$ ,

and so  $g = x_1k \in x_1H$ . We have now shown that every element of  $x_2H$  is also in  $x_1H$ , and therefore we have shown that  $x_2H \subseteq x_1H$ .

For the reverse inclusion, assume instead that  $g$  is an arbitrary element of  $x_1H$ . Then we have  $g = x_1h$  for some  $h \in H$ , and so

$$g = ex_1h = x_2x_2^{-1}x_1h \in x_2H$$

since  $(x_2^{-1}x_1)h \in H$  by closure of  $H$  under multiplication. Thus all elements of  $x_1H$  are in  $x_2H$ ; that is,  $x_1H \subseteq x_2H$ .

We have now shown that (iv) implies both  $x_2H \subseteq x_1H$  and  $x_1H \subseteq x_2H$ , hence  $x_1H = x_2H$ . So (iv) implies (i), and the proof is complete.  $\square$

An important corollary of Proposition 2.11 is that two unequal left cosets cannot have any elements in common.

**2.12 PROPOSITION** *Let  $H$  be a subgroup of the group  $G$  and let  $x_1, x_2 \in G$ . If  $x_1H \cap x_2H \neq \emptyset$  then  $x_1H = x_2H$ .*

**Proof.** Suppose  $x_1H \cap x_2H \neq \emptyset$ . Then there exists an element  $g \in G$  with  $g \in x_1H \cap x_2H$ . This gives  $g \in x_1H$  and  $g \in x_2H$ . Applying Proposition 2.11 with  $g$  in place of  $x_2$ , using in particular that 2.11 (ii) implies 2.11 (i), we deduce that  $x_1H = gH$ . By the similar reasoning,  $g \in x_2H$  gives  $x_2H = gH$ . So  $x_1H = x_2H$ , since both equal  $gH$ .  $\square$

The other important fact about left cosets is that each left coset of the subgroup  $H$  has the same number of elements as the subgroup  $H$  itself.

**2.13 PROPOSITION** *Let  $H$  be a subgroup of the group  $G$  and let  $x \in G$ . Then the function  $f: H \rightarrow xH$  defined by  $f(h) = xh$  for all  $h \in H$ , is bijective.*

**Proof.** Let  $h_1, h_2 \in H$  and suppose that  $f(h_1) = f(h_2)$ . Then  $xh_1 = xh_2$ , and so by 2.2 it follows that  $h_1 = h_2$ . We have shown that  $f(h_1) = f(h_2)$  implies  $h_1 = h_2$ ; that is,  $f$  is injective.

Let  $g$  be an arbitrary element of the coset  $xH$ . Then  $g = xh$  for some  $h \in H$ ; that is,  $g = f(h)$ . We have shown that every element of  $xH$  has the form  $f(h)$  for some  $h \in H$ ; that is,  $f: H \rightarrow xH$  is surjective.

We have shown that  $f$  is both injective and surjective; that is,  $f$  is bijective, as claimed.  $\square$

### §2d On the number of elements in a set.

In this section we digress completely from group theory. We shall return to group theory later.

In the Proposition 2.13 we proved the existence of a bijective function from one set to another. We had claimed that we were going to show that the two sets in question had the same number of elements. The situation is plain enough for sets which have a finite number of elements. It is clear, for example, that if two sets both have five elements then a one-to-one correspondence can be set up between the elements of the two sets. In other words, a bijective function exists from one set to the other. (Indeed,  $5! = 120$  such bijective functions exist.) On the other hand, if one set has five elements and the other six, then no bijective function exists from one set to the other. (Indeed, a function from a five element set to a six element set cannot be surjective, while a function from a six element set to a five element set cannot be injective.) For finite sets, then, there is no doubt that two sets have the same number of elements if and only if a bijective function from one to the other can be found.

For infinite sets, the situation is by no means clear. The concept of “the number of elements” of a set  $S$  is much less familiar if  $S$  in fact has infinitely many elements. But even without defining what is meant by the *number of elements* of a set  $S$ , we can still easily say what it means for two sets  $S$  and  $T$  to have *the same number of elements*. By analogy with the finite case we make the following definition.

2.14 DEFINITION Sets  $S$  and  $T$  have *equal cardinality*, or *the same number of elements*, if there exists a bijective function from  $S$  to  $T$ .

The slightly paradoxical thing about infinite sets is that it is possible for one set to be strictly contained in another, and yet have the same number of elements as the set that contains it. From one point of view, one of the sets is clearly smaller than the other; from another point of view the two sets have the same size! We will illustrate this, and some similar matters, by means of examples.

First of all, consider the set  $S = \{n \in \mathbb{Z} \mid n \geq 0\}$ , the set of all nonnegative integers, and the set  $T = \{n \in \mathbb{Z} \mid n \geq 1\}$ , the set of all positive integers. Then  $T \subset S$  and  $T \neq S$ , but the function  $S \rightarrow T$  given by  $a \mapsto a + 1$  is clearly bijective, and so  $S$  and  $T$  have the same number of elements.

Next, consider  $\mathbb{Q}$ , the set of all rational numbers. Recall that a number is rational if it can be expressed as  $n/m$ , where  $n$  and  $m$  are integers. Every real number can be expressed by means of an infinite decimal expansion, and it is well known that the number is rational if and only if, from some point onwards in the decimal expansion, the same sequence of digits repeats indefinitely. For example,  $.5000\dots$  is rational, as the 0 repeats forever, and similarly the decimal expansion for  $1/7$  consists of the same block of six digits repeating forever:  $1/7 = .142857142857142857\dots$ . Some real numbers are not rational, as for example  $\pi = 3.14159\dots$ , whose decimal expansion never repeats. But of course, given any real number it is always possible to find rational numbers which are as close as you please to the given real number. For example, the difference between  $\pi$  and the rational number  $314/100$  is less than  $2 \times 10^{-3}$ ; if you want a better approximation than that, the rational number  $31415/10000$  differs from  $\pi$  by less than  $10^{-4}$ , while  $314159/100000$  is closer still, and so on. The technical term used to describe this situation is *denseness*: the set of all rational numbers is dense in the set of all real numbers. Since obviously the set of all integers is not dense in the set of all real numbers, it is quite plain that there are many more rational numbers than there are integers.

Not so! The set  $\mathbb{Q}$  and the set  $\mathbb{Z}$  have the same number of elements. This can be proved readily by a famous argument due to Georg Cantor (1845–1918), which we present in a moment. First, though, let us state a definition.

**2.15 DEFINITION** A set  $S$  is said to be *countable*, or *enumerable*, if there exists a sequence  $s_1, s_2, s_3, \dots$  of elements of  $S$  which includes every element of  $S$  at least once.

That is, a set is countable if its elements can be listed. By crossing out repetitions one can obtain a sequence, which could be finite or could be infinite, such that each element of  $S$  occurs exactly once in the list. Of course the list will be finite if and only if  $S$  is a finite set; so we see that an infinite set  $S$  is countable if and only if there is an infinite list  $s_1, s_2, s_3, \dots$  of elements of  $S$  in which each element of  $S$  appears exactly once. Under these circumstances the function  $f$  from the set  $\mathbb{Z}^+ = \{n \in \mathbb{Z} \mid n > 0\}$  to  $S$  defined by  $f(n) = s_n$  is bijective: surjective since every element of  $S$  occurs in the list, injective since each element occurs once only. So an infinite countable set has the same number of elements as the set of all positive integers.

By means of a simple trick, Cantor was able to write down a list of all the positive rational numbers, thereby proving that the set  $\mathbb{Q}^+$  of all positive rational numbers is countable. The list is

$$1/1, 2/1, 1/2, 3/1, 2/2, 1/3, 4/1, 3/2, 2/3, 1/4, 5/1, \dots$$

and it comes from the diagonals of the rectangular array

$$\begin{array}{cccccc} 1/1 & 2/1 & 3/1 & 4/1 & 5/1 & 6/1 & \dots \\ 1/2 & 2/2 & 3/2 & 4/2 & 5/2 & \dots & \\ 1/3 & 2/3 & 3/3 & 4/3 & \dots & & \\ 1/4 & 2/4 & 3/4 & \dots & & & \\ 1/5 & 2/5 & \dots & & & & \\ 1/6 & \dots & & & & & \\ \dots & & & & & & \end{array}$$

That is, Cantor first lists those fractions  $n/m$  with  $n+m=2$ , then moves on to those with  $n+m=3$ , then  $n+m=4$ , and so on, and clearly in this way he catches every positive rational number at least once. (In fact, he catches them all infinitely often.)

So there exists a bijective function  $f: \mathbb{Z}^+ \rightarrow \mathbb{Q}^+$ . This immediately gives a bijective function  $F: \mathbb{Z} \rightarrow \mathbb{Q}$ , defined by

$$F(n) = \begin{cases} f(n) & \text{if } n > 0 \\ 0 & \text{if } n = 0 \\ -f(-n) & \text{if } n < 0, \end{cases}$$

and this justifies the claim we made above, that the set of all integers and the set of all rational numbers have the same number of elements.

At this point one might think, since  $\mathbb{Z}$  has the same number of elements as  $\mathbb{Q}$ , a set which initially looked much bigger, that all infinite sets have the same number of elements as each other. For example, we have seen that  $\mathbb{Q}$  is dense in  $\mathbb{R}$ ; surely there cannot be too much difference between the number of elements of  $\mathbb{Q}$  and the number of elements of  $\mathbb{R}$ !

We know that intuition can be misleading, and, sure enough, though the set  $\mathbb{Q}$  is countable, the set  $\mathbb{R}$  is not. Cantor proved this also, by another famous argument with a diagonal theme. Since  $\mathbb{Q}$  is a subset of  $\mathbb{R}$ , we can

conclude that the number of elements of  $\mathbb{Q}$  is definitely less than the number of elements of  $\mathbb{R}$ . (We have not actually defined what it means for a set  $S$  to have fewer elements, or lesser cardinality, than another set  $T$ . A suitable definition is as follows: the cardinality of  $S$  is less than the cardinality of  $T$  if there exists an injective function from  $S$  to  $T$  but no bijective function from  $S$  to  $T$ .)

If it were possible to list all the real numbers it would be possible, by striking out those numbers on the list which do not lie between 0 and 1, to obtain a list of all positive real numbers less than 1. Each such number  $x$  has an infinite decimal expansion,

$$x = .a_1 a_2 a_3 a_4 a_5 \dots$$

where the  $a_i$  are integers from the set  $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ . Suppose now that we have a list  $x_1, x_2, x_3, \dots$  of all the real numbers between 0 and 1. Writing out decimal expansions for them all, we have

$$\begin{aligned} x_1 &= .a_1^{(1)} a_2^{(1)} a_3^{(1)} a_4^{(1)} a_5^{(1)} a_6^{(1)} \dots \\ x_2 &= .a_1^{(2)} a_2^{(2)} a_3^{(2)} a_4^{(2)} a_5^{(2)} a_6^{(2)} \dots \\ x_3 &= .a_1^{(3)} a_2^{(3)} a_3^{(3)} a_4^{(3)} a_5^{(3)} a_6^{(3)} \dots \\ x_4 &= .a_1^{(4)} a_2^{(4)} a_3^{(4)} a_4^{(4)} a_5^{(4)} a_6^{(4)} \dots \\ x_5 &= .a_1^{(5)} a_2^{(5)} a_3^{(5)} a_4^{(5)} a_5^{(5)} a_6^{(5)} \dots \\ &\vdots \end{aligned}$$

Now let us define another infinite decimal expansion  $.a_1 a_2 a_3 \dots$ , as follows: for each  $i$ ,

$$a_i = \begin{cases} 1 & \text{if } a_i^{(i)} \neq 1 \\ 2 & \text{if } a_i^{(i)} = 1. \end{cases}$$

So this decimal expansion consists of an infinite string of 1's and 2's, presumably with a preponderance of 1's, although that is irrelevant. What is certain is that the number  $x$  defined by this decimal expansion lies between 0 and 1. Furthermore, the decimal expansions  $.a_1^{(i)} a_2^{(i)} a_3^{(i)} a_4^{(i)} \dots$  and  $.a_1 a_2 a_3 a_4 \dots$  cannot be the same, for any value of  $i$ , since by the construction  $a_i \neq a_i^{(i)}$ . In other words, we have chosen things so that the decimal expansion of  $x$  differs from the decimal expansion of  $x_1$  in the first decimal place,



differs from the decimal expansion of  $x_2$  in the second decimal place, differs from the decimal expansion of  $x_3$  in the third decimal place, and so on. So  $x \neq x_i$  for any value of  $i$ , and we have found a number between 0 and 1 which is not on the list. This is a contradiction, since the list was meant to include all the numbers between 0 and 1. So no such list can exist, and so the set of all real numbers is not countable.†

## §2e Equivalence relations

The matters discussed in the previous section have little relevance for anything else that will be discussed in these notes, since we will mostly be concerned with finite sets rather than infinite ones. The present section, however, though still digressing from group theory, is concerned with matters which will be important for us later on.

2.16 DEFINITION Let  $R$  be a binary relation on a set  $S$ . Then

- (i)  $R$  is said to be *reflexive* if  $aRa$  is true for all  $a \in S$ ,
- (ii)  $R$  is said to be *symmetric* if  $bRa$  implies  $aRb$ ,
- (iii)  $R$  is said to be *transitive* if  $aRb$  and  $bRc$  together imply  $aRc$ .

If the relation  $R$  is reflexive, symmetric and transitive, then it is called an equivalence relation.

The typical way equivalence relations arise in mathematics is as follows: if  $S$  is a set and  $\mathcal{F}$  a collection of functions defined on  $S$ , let  $\sim$  be the relation on  $S$  defined by the rule that for all  $x, y \in S$ ,

$$x \sim y \text{ if and only if } f(x) = f(y) \text{ for all } f \in \mathcal{F}.$$

In other words, we say that  $x$  is equivalent to  $y$  if all the functions take the same value at  $x$  as at  $y$ . It is trivial to check that  $\sim$  is reflexive, symmetric and transitive. Furthermore, this is all very natural and intuitive, since in everyday situations we would regard two things as equivalent if they are the same in all the aspects that interest us. In the mathematical context, “the aspects that interest us” are the values of the functions we are dealing with.

The purpose of the concept of equivalence, in mathematics as in everyday life, is to limit the number of different things we have to deal with. All

---

† We have slurred over a minor point: it is not quite true that every real number has a unique decimal expansion. For example,  $.5000\dots = .4999\dots$ . It is easily checked that this does not invalidate the argument.

equivalent things can be treated as one. Rather than make separate rules for each separate object, we can have one rule applying to a whole class of equivalent objects. Thus we are led naturally to the following definition.

**2.17 DEFINITION** Let  $\sim$  be an equivalence relation on a set  $S$ , and let  $x \in S$ . The *equivalence class* containing  $x$  is the set  $\{y \in S \mid y \sim x\}$ .

Suppose, for example, that the set  $S$  consists of a large assortment of coloured pencils, and suppose that the relation  $\sim$  is defined on  $S$  by the rule

$$p_1 \sim p_2 \text{ if } p_1 \text{ and } p_2 \text{ correspond to the same colour.}$$

It is readily seen that  $\sim$  is an equivalence relation. There will be one equivalence class for each colour that is represented in the set of pencils, the equivalence class consisting of all pencils of that colour. It is conceivable that an equivalence class might have only one member (if there is only one pencil of that colour), or that a single equivalence class includes most of the elements of  $S$ . But note that every element of  $S$  lies in some equivalence class, and no element can lie in two different equivalence classes.† We can prove these facts quite generally.

**2.18 PROPOSITION** *Let  $\sim$  be an equivalence relation on a set  $S$ . Then the equivalence classes of  $\sim$  form a partitioning of  $S$ . That is, the equivalence classes are pairwise disjoint nonempty sets, and their union is the whole of  $S$ .*

**Proof.** For each  $x \in S$  define  $C(x) = \{y \in S \mid y \sim x\}$ , the equivalence class containing  $x$ , and let  $\mathcal{Q} = \{C(x) \mid x \in S\}$ , the set of all equivalence classes of  $\sim$ . We show first that two distinct equivalence classes in  $\mathcal{Q}$  cannot overlap.

Let  $x, y \in S$  and suppose that  $C(x) \cap C(y) \neq \emptyset$ . We will show that  $C(x) = C(y)$ . (Note that this does not mean that  $x = y$ .) Choose an element  $z \in C(x) \cap C(y)$ . Then  $z \in C(x)$ , whence  $z \sim x$ , and similarly  $z \sim y$ . Since  $\sim$  is symmetric it follows that  $y \sim z$ , and combining this with  $z \sim x$  gives, by transitivity of  $\sim$ , that  $y \sim x$ . Now if  $t$  is an arbitrary element of  $C(y)$  we have that  $t \sim y$ , and with  $y \sim x$  this gives  $t \sim x$ , whence  $t \in C(x)$ . So all elements of  $C(y)$  are in  $C(x)$ ; that is,  $C(y) \subseteq C(x)$ . On the other hand, if  $t$  is an arbitrary element of  $C(x)$  then  $t \sim x$ , and since we also have  $x \sim y$  we can conclude that  $t \sim y$ , whence  $t \in C(y)$ . So  $C(x) \subseteq C(y)$ , and therefore  $C(x) = C(y)$ , as claimed.

---

† I had better assume that the pencils are monochromatic!

All that remains is to show that the equivalence classes are all nonempty, and that each element of  $S$  lies in some equivalence class. Both of these facts are clear. Firstly, if  $x \in S$  is arbitrary then  $x \in C(x)$ , since  $x \sim x$ , and so  $x$  lies in an equivalence class; secondly, if  $C \in \mathcal{Q}$  is arbitrary then by definition  $C = C(x)$  for some  $x$ , and since  $x \in C(x)$  we conclude that  $C \neq \emptyset$ .  $\square$

**2.19 DEFINITION** If  $\sim$  is an equivalence relation on the set  $S$ , then the set  $\mathcal{Q}$ , consisting of all equivalence classes of  $\sim$  on  $S$ , is called the *quotient* of  $S$  by the equivalence relation.

Initially the concept of a quotient may seem unduly abstract, dealing as it does with sets whose elements are sets. Nevertheless, it is not far removed from everyday practice. Consider numbers, for example. It is a straightforward task to decide whether or not a set has three elements, but what is the number three itself? What kind of thing is it? A plausible answer to this is to say that the number three is exactly the set of all three-element sets. Certainly it is an abstract thing that is intimately allied with the set of all three-element sets, and so to identify it with this set of sets seems quite natural, and avoids postulating the existence of an additional abstract object of uncertain nature. According to the theory we are suggesting here, the set of all natural numbers is an example of a quotient set. If  $S$  is the set of all finite sets of physical objects, an equivalence relation  $\sim$  can be defined on  $S$  by the rule that  $X \sim Y$  if the sets  $X$  and  $Y$  have the same number of elements. There is then one equivalence class for each natural number, and, according to this theory, the number and the equivalence class are one and the same thing. The set of all natural numbers is the set of all equivalence classes, and this is precisely the quotient of  $S$  by the equivalence relation  $\sim$ . It is worth noting, incidentally, that these considerations provide a method of defining infinite numbers, an issue which we sidestepped in the previous section.

We have been talking philosophy rather than mathematics. Mathematicians do not really have to answer the philosophical question “what is a number”, they just have to know what rules numbers obey. But whether or not the set of natural numbers is an example of a quotient by an equivalence relation, the fact is that when one has an equivalence relation, considering the set of equivalence classes rather than the original set itself reduces the number of objects one has to contend with, and this is the whole purpose of equivalence relations.

Our next proposition, although still extremely general and abstract, is at least mathematics rather than philosophy.

**2.20 PROPOSITION** *Let  $f: A \rightarrow B$  be an arbitrary function. Then there exists sets  $Q$  and  $I$ , and a surjective function  $\eta: A \rightarrow Q$ , a bijective function  $\psi: Q \rightarrow I$  and an injective function  $\sigma: I \rightarrow B$ , such that  $f$  is equal to the composite  $\sigma\psi\eta$ . That is, the following sequence*

$$A \xrightarrow{\eta} Q \xrightarrow{\psi} I \xrightarrow{\sigma} B$$

*provides a factorization of  $f$  as a surjective function, followed by a bijective, followed by an injective. Furthermore, the set  $Q$  is the quotient of  $A$  by an equivalence relation, and  $I$  is a subset of  $B$ .*

**Proof.** Let  $\sim$  be the equivalence relation defined on  $A$  by the rule that  $x \sim y$  if and only if  $f(x) = f(y)$ , and let  $Q$  be the quotient of  $A$  by  $\sim$ . Let  $I = \{f(x) \mid x \in A\}$ , the image of the function  $f$ . For each  $x \in A$  let  $C(x) = \{y \in A \mid f(y) = f(x)\}$ , the equivalence class of  $x$ .

Define  $\eta: A \rightarrow Q$  by the rule that  $\eta(x) = C(x)$  for all  $x \in A$ . Since by definition  $Q = \{C(x) \mid x \in A\} = \{\eta(x) \mid x \in A\}$ , which is the image of  $\eta$ , it follows that  $\eta$  is surjective.

Observe that  $I \subseteq B$ . Hence we can define  $\sigma: I \rightarrow B$  by the rule that  $\sigma(b) = b$  for all  $b \in I$ . With this definition it is immediate that if  $\sigma(b_1) = \sigma(b_2)$  then  $b_1 = b_2$ ; so  $\sigma$  is injective.

If  $C \in Q$  and  $x, y \in C$ , then  $x \sim y$  (since  $C$  is an equivalence class), and so  $f(x) = f(y)$  (by the definition of  $\sim$ ). So the function  $f$  is constant on  $C$ : all elements  $x \in C$  give the same value for  $f(x)$ . Define  $\psi(C)$  to be this constant value; that is, define  $\psi(C) = f(x)$ , where  $x \in C$  is chosen arbitrarily. Because equivalence classes are always nonempty (see 2.18) there is always an  $x \in C$  to choose; so  $\psi(C)$  is always defined, and defined unambiguously since  $f$  is constant on  $C$ . Furthermore,  $\psi(C) = f(x) \in \text{im } f = I$ . Hence  $C \mapsto \psi(C)$  defines a function  $\psi: Q \rightarrow I$ . It remains to show that  $\psi$  is bijective and that  $f = \sigma\psi\eta$ .

For all  $x \in A$  we have that  $x \in C(x)$ , and so, by the definition of  $\eta$ , we have  $\psi(C(x)) = f(x)$ . Hence

$$(\sigma\psi\eta)(x) = \sigma(\psi(\eta(x))) = \psi(\eta(x)) = \psi(C(x)) = f(x)$$

and therefore  $f = \sigma\psi\eta$ . If  $b \in I$  is arbitrary then, by the definition of  $I$ , there exists an  $x \in A$  such that  $b = f(x)$ . This gives  $b = \psi(C(x))$ , and so

$b$  is in the image of  $\psi$ . So  $\psi$  is surjective. Finally, suppose that  $C_1, C_2 \in \mathcal{Q}$  are such that  $\psi(C_1) = \psi(C_2)$ . Choose  $x_1, x_2 \in A$  such that  $C_1 = C(x_1)$  and  $C_2 = C(x_2)$ . Then

$$f(x_1) = \psi(C(x_1)) = \psi(C_1) = \psi(C_2) = \psi(C(x_2)) = f(x_2),$$

whence  $x_1 \sim x_2$ , by the definition of  $\sim$ . Now since  $x_1$  and  $x_2$  are equivalent it follows that their equivalence classes are the same (as in the proof of 2.18 above). Hence  $C_1 = C_2$ . So we have shown that  $\psi(C_1) = \psi(C_2)$  implies  $C_1 = C_2$ ; that is,  $\psi$  is injective.  $\square$

## §2f Cosets revisited

The material on equivalence relations that we have just been discussing provides us with a slightly improved way of describing cosets.

**2.21 DEFINITION** Let  $G$  be a group and  $H$  a subgroup of  $G$ . We say that elements  $g_1, g_2 \in G$  are *left congruent modulo  $H$*  if  $g_1 = hg_2$  for some  $h \in H$ . Similarly,  $g_1$  and  $g_2$  are *right congruent modulo  $H$*  if  $g_1 = g_2h$  for some  $h \in H$ .

Note that  $g_1 = hg_2$  if and only if  $h = g_1g_2^{-1}$ , and so we see that  $g_1$  and  $g_2$  are left congruent modulo  $H$  if and only if  $g_1g_2^{-1} \in H$ . The first thing to observe is that both of these congruence relations are equivalence relations on  $G$ . We present only the proofs for left congruence, since those for right congruence are totally similar.

**2.22 PROPOSITION** Let  $H$  be a subgroup of the group  $G$ , and let  $\sim$  be the relation of left congruence modulo  $H$ , so that for all  $g_1, g_2 \in G$  we have  $g_1 \sim g_2$  if and only if  $g_1g_2^{-1} \in H$ . Then  $\sim$  is an equivalence relation on  $G$ .

**Proof.** Let  $g \in G$  be arbitrary. Then  $g = eg$ , where  $e$  is the identity element of  $G$ . By Proposition 2.9 we know that  $e \in H$ , and so we have shown that  $g = hg$  for some  $h \in H$ . Hence  $g \sim g$ ; that is,  $\sim$  is reflexive.

Let  $g_1, g_2 \in G$  with  $g_1 \sim g_2$ . Then we have that  $g_1 = hg_2$  for some  $h \in H$ , and multiplying both sides of this on the left by  $h^{-1}$  gives  $g_2 = h^{-1}hg_2 = h^{-1}g_1$ . Since  $H$  is closed under inversion (by Theorem 2.10) we know that  $h^{-1} \in H$ , and hence  $g_2 \sim g_1$ . So  $\sim$  is symmetric.

Let  $g_1, g_2, g_3 \in G$  with  $g_1 \sim g_2$  and  $g_2 \sim g_3$ . Then there exist  $h, k \in H$  with  $g_1 = hg_2$  and  $g_2 = kg_3$ . Substituting the value for  $g_2$  given by the second

of these two equations into the first gives  $g_1 = hkg_3$ . Now closure of  $H$  under multiplication (Theorem 2.10) tells us that  $hk \in H$ , and so  $g_1 \sim g_3$ . Hence  $\sim$  is transitive.  $\square$

Maintaining the notation of Proposition 2.22, the fact that  $\sim$  is an equivalence relation allows us to conclude (in view of Theorem 2.18) that  $G$  is the disjoint union of the equivalence classes of  $\sim$ . But if  $x \in G$  then the equivalence class containing  $x$  is

$$\{y \in G \mid y \sim x\} = \{y \in G \mid y = hx \text{ for some } h \in H\} = \{hx \mid h \in H\},$$

which is exactly the definition of the right coset  $Hx$ . Thus the equivalence classes of  $\sim$  are exactly the right cosets of  $H$  in  $G$ . The analogues for right cosets of Propositions 2.11 and 2.12 now follow almost immediately.

**2.23 PROPOSITION** *Let  $H$  be a subgroup of the group  $G$  and let  $x_1, x_2 \in G$ . Then the following conditions are all equivalent to each other:*

- (i)  $Hx_1 = Hx_2$ ,
- (ii)  $x_2 \in Hx_1$ ,
- (iii)  $x_2 = hx_1$  for some  $h \in H$ ,
- (iv)  $x_2x_1^{-1} \in H$ .

**Proof.** Parts (iii) and (iv) are different formulations of the statement that  $x_2 \sim x_1$ , whereas (ii) says that  $x_2$  is in the equivalence class containing  $x_1$  and part (i) that the equivalence class containing  $x_1$  is the same as the equivalence class containing  $x_2$ . Thus all four are equivalent.  $\square$

**2.24 PROPOSITION** *Let  $H$  be a subgroup of the group  $G$  and let  $x_1, x_2 \in G$ . If  $x_1H \cap x_2H \neq \emptyset$  then  $x_1H = x_2H$ .*

**Proof.** This simply says that distinct equivalence classes are disjoint (which follows from 2.18).  $\square$

On the other hand, the third basic property of cosets is not a consequence of general facts about equivalence classes, but must instead be proved by the method we used previously.

**2.25 PROPOSITION** *Let  $H$  be a subgroup of the group  $G$  and let  $x \in G$ . Then  $H$  and  $Hx$  have the same number of elements.*

The proof consists of showing that  $h \mapsto hx$  is a bijective function from  $H$  to  $xH$ .

## §2g Some examples

Let  $G$  be a group and  $x \in G$ . Then  $x$  generates a cyclic subgroup of  $G$ , consisting of the identity and all the powers of  $x$  and  $x^{-1}$ . We will denote this subgroup by  $\langle x \rangle$ . For our first example, let  $G = \text{Sym}\{1, 2, 3\}$  and let  $H = \langle (1, 2) \rangle = \{i, (1, 2)\}$ , the cyclic subgroup generated by the transposition  $(1, 2)$ . Since  $(1, 2)^2 = i$  we see that  $H$  has exactly two elements:

$$H = \{i, (1, 2)\}.$$

If  $x \in \text{Sym}\{1, 2, 3\}$  then the left coset  $xH$  is the two element set  $\{x, x(1, 2)\}$ . If  $x = i$  or  $x = (1, 2)$  then  $xH$  coincides  $H$ . If we choose  $x = (2, 3)$  we find that the two elements of  $xH$  are  $(2, 3)$  and  $(1, 3, 2)$ :

$$(2, 3)H = \{(2, 3), (1, 3, 2)\}.$$

Since  $(1, 3, 2) \in (2, 3)H$  we must have  $(1, 3, 2)H = (2, 3)H$ , a fact which is also easy to check by direct calculation. The remaining two elements of  $G$  give a third coset:

$$(1, 3)H = \{(1, 3), (1, 2, 3)\}.$$

Since we knew from Proposition 2.13 that each of the coset would turn out to have the same number of elements as  $H$ , namely two, we could have predicted that the total number of cosets would have to be the number of elements of  $G$  divided by the number of elements of  $H$ .

If we had worked with right cosets rather than left cosets we would have found the same number of cosets, but the cosets themselves would be different. In fact, if  $y$  is right congruent to  $x$  modulo  $H$ , so that  $y = xh$  for some  $h \in H$ , then  $y^{-1} = h^{-1}x^{-1}$ , and since  $h^{-1} \in H$  this means that  $y^{-1}$  is left congruent to  $x^{-1}$  modulo  $H$ . Conversely, and by similar reasoning, if  $y^{-1}$  is left congruent to  $x^{-1}$  then  $y$  is right congruent to  $x$ . It follows that the right coset  $Hx^{-1}$  consists of the inverses of the elements in the left coset  $xH$ . It is a general fact—true for all groups  $G$  and subgroups  $H$ —that taking inverses changes right cosets into left cosets, and vice versa. Applying this in the particular example we have been considering, we find that

$$\begin{aligned} \{i, (1, 2)\} &= H = Hi = H(1, 2) \\ \{(1, 3), (1, 2, 3)\} &= H(1, 3) = H(1, 2, 3) \\ \{(2, 3), (1, 3, 2)\} &= H(2, 3) = H(1, 3, 2). \end{aligned}$$

For our second example, let  $G = \text{Sym}\{1, 2, 3\}$  and let  $H = \langle(1, 2, 3)\rangle$ . This time,  $H$  has three elements:

$$H = \{i, (1, 2, 3), (1, 3, 2)\}.$$

If  $x$  is any element of  $G$  which is not in  $H$  then we know that the coset  $xH$  is different from  $H$ , and hence has no elements in common with  $H$ , and also has the same number of elements as  $H$ . Since there are altogether only three elements of  $G$  which are not in  $H$ , the coset must consist exactly of these three elements. So in this case there are two left cosets of  $H$  in  $G$ :

$$\begin{aligned} \{i, (1, 2, 3), (1, 3, 2)\} &= H = iH = (1, 2, 3)H = (1, 3, 2)H \\ \{(1, 2), (1, 3), (2, 3)\} &= (1, 2)H = (1, 3)H = (2, 3)H. \end{aligned}$$

In this example, if we had chosen to work with right cosets rather than left cosets the reasoning would have been exactly the same, and we would have reached the same conclusion: there are two right cosets of  $H$  in  $G$ , one being  $H$  itself, the other consisting of all the elements of  $G$  which are not in  $H$ . So in this case, unlike the previous one, the partitioning of  $G$  into right cosets modulo  $H$  is exactly the same as the partitioning of  $G$  into left cosets modulo  $H$ . Subgroups which have this property—that every right coset is a left coset and vice versa—are known as *normal subgroups*. They play an important role in group theory, and in the next chapter we will investigate them further.

Every group has a trivial subgroup, consisting of just one element, the identity. If  $G = \text{Sym}\{1, 2, 3\}$  and  $H = \{i\}$  then we see that

$$(1, 2)H = \{(1, 2)\} = H(1, 2),$$

and a similar statement applies for all the other elements of  $G$  too. So once again left cosets are the same as right cosets: the subgroup is normal. There are six cosets altogether, each consisting of just a single element.

At the other extreme, still with  $G = \text{Sym}\{1, 2, 3\}$ , suppose that  $H = G$ . It is clear from the definition that a group is always a subgroup of itself! In this case there is only one coset, namely  $H$  itself. The subgroup is therefore normal.

For our next example, let  $G = \text{Sym}\{1, 2, 3, 4\}$  and let  $H$  be the subgroup of  $G$  generated by the permutations  $(1, 2, 3, 4)$  and  $(1, 3)$ . Then  $H$  has eight



elements. In fact,  $H$  is just the group of all symmetries of a square, as we considered in Chapter 1, except that we have here changed  $a, b, c$  and  $d$  to 1, 2, 3 and 4. Because  $G$  has twenty-four elements and  $H$  has eight we will find that there are three left cosets altogether. After writing down all the elements of  $H$ , choose some element  $x \in G$  such that  $x \notin H$  and calculate all the products  $xh$  for  $h \in H$ . This will give us a coset  $xH$  different from  $H$ . The remaining eight elements of  $G$  will constitute the third left coset. The result of these calculations is as follows: the three left cosets of  $H$  in  $G$  are

$$\begin{aligned} \{i, (1,2,3,4), (1,3)(2,4), (1,4,3,2), (1,3), (1,4)(2,3), (2,4), (1,2)(3,4)\} &= H \\ \{(1,2), (2,3,4), (1,3,2,4), (1,4,3), (1,3,2), (1,4,2,3), (1,2,4), (3,4)\} &= (1,2)H \\ \{(1,4), (1,2,3), (1,3,4,2), (2,4,3), (1,3,4), (2,3), (1,4,2), (1,2,4,3)\} &= (1,4)H. \end{aligned}$$

Taking the inverses of all the elements we find the decomposition of  $G$  into right cosets modulo  $H$ :

$$\begin{aligned} \{i, (1,4,3,2), (1,3)(2,4), (1,2,3,4), (1,3), (1,4)(2,3), (2,4), (1,2)(3,4)\} &= H \\ \{(1,2), (2,4,3), (1,4,2,3), (1,3,4), (1,2,3), (1,3,2,4), (1,4,2), (3,4)\} &= H(1,2) \\ \{(1,4), (1,3,2), (1,2,4,3), (2,3,4), (1,4,3), (2,3), (1,2,4), (1,3,4,2)\} &= H(1,4). \end{aligned}$$

Since it is not true that all left cosets are right cosets, the subgroup  $H$  is not normal in  $G$ .

As our final example, let  $G = \text{Alt}\{1, 2, 3, 4\}$ , the group of all even permutations of  $\{1, 2, 3, 4\}$ , and let  $K$  be the subgroup of  $G$  consisting of the identity and the three permutations  $(1, 2)(3, 4)$ ,  $(1, 3)(2, 4)$  and  $(1, 4)(2, 3)$ . We find that the three left cosets of  $K$  in  $G$  are as follows:

$$\begin{aligned} K &= \{i, (1, 2)(3, 4), (1, 3)(2, 4), (1, 4)(2, 3)\} \\ (1, 2, 3)K &= \{(1, 2, 3), (1, 3, 4), (2, 4, 3), (1, 4, 2)\} \\ (1, 3, 2)K &= \{(1, 3, 2), (1, 4, 3), (2, 3, 4), (1, 2, 4)\}. \end{aligned}$$

In the notation we used in Chapter 1 when discussing the multiplication table of  $\text{Alt}\{1, 2, 3, 4\}$ , the cosets are  $\{i, a, b, c\}$ ,  $\{t_1, t_2, t_3, t_4\}$  and  $\{s_1, s_2, s_3, s_4\}$ .

The right coset  $K(1, 3, 2)$  consists of the inverses of the elements of the left coset  $(1, 2, 3)K$ , but on calculating these inverses we find that they are exactly the elements of  $(1, 3, 2)K$ . So  $K(1, 3, 2) = (1, 3, 2)K$ . Similarly  $K(1, 2, 3) = (1, 2, 3)K$ . Since the left cosets are also right cosets, the subgroup  $K$  is normal in  $G$ .

## §2h The index of a subgroup

**2.26 DEFINITION** Let  $G$  be a group and  $H$  a subgroup of  $G$ . The *index* of  $H$  in  $G$ , denoted by  $[G : H]$ , is defined to be the number of left cosets of  $H$  in  $G$ .

As we observed above, taking inverses of elements changes left cosets into right cosets. So there is a one-to-one correspondence between the set of all left cosets of  $H$  in  $G$  and the set of all right cosets of  $H$  in  $G$ , and hence the index could equally well be defined as the number of right cosets of  $H$  in  $G$ .

If  $S$  is any set, we will write  $|S|$  for the number of elements in  $S$ . Recall that if  $G$  is a group then  $|G|$  is called the order of  $G$ .

Suppose now that  $G$  is a group with  $|G|$  finite, so that if  $H$  is a subgroup of  $G$  then  $|H|$  and  $[G : H]$  will also be finite. Since we know that  $|xH| = |H|$  for all  $x \in G$ , we can conclude that  $G$  is the disjoint union of  $[G : H]$  sets each of which has  $|H|$  elements. It follows that  $|G| = [G : H]|H|$ . This result is known as *Lagrange's Theorem*.

**2.27 THEOREM** Let  $G$  be a finite group and  $H$  a subgroup of  $G$ . Then

$$|G| = [G : H]|H|,$$

and, in particular, the order and index of  $H$  are both divisors of  $|G|$ .

We know that if  $x \in G$  is arbitrary then the set of all powers of  $x$  forms a subgroup  $\langle x \rangle$ , known as the cyclic subgroup generated by  $x$ . The order of  $\langle x \rangle$  is also called the order of the element  $x$ ; it is the least positive integer  $n$  such that  $x^n$  is the identity element of  $G$ . As a corollary of Theorem 2.27 we deduce the following fact about orders of elements of  $G$ .

**2.28 COROLLARY** If  $G$  is a finite group and  $x \in G$ , then the order of  $x$  is a divisor of  $|G|$ .

# 3

## Homomorphisms, quotient groups and isomorphisms

As explained in Chapter 1, the symmetries of a structured set always form a group, and this is the reason for the importance of groups in mathematics. But from the axiomatic description of group theory given in Chapter 2 we see that groups themselves are also examples of structured sets, since the multiplication operation that a group  $G$  must possess can alternatively be described as a ternary relation on  $G$ . As always when considering structured sets, the most important kind of functions to consider are those which preserve the structure.

### §3a Homomorphisms

**3.1 DEFINITION** If  $G$  and  $H$  are groups then a function  $\phi: G \rightarrow H$  is called a *homomorphism* if  $\phi(xy) = \phi(x)\phi(y)$  for all  $x, y \in G$ .

To express the multiplication operation as a relation, as suggested in the introduction above, we would define a relation **Mult** on a group by

$$\mathbf{Mult}(x, y, z) \text{ if and only if } xy = z.$$

A homomorphism  $\phi$  from  $G$  to  $H$  is then a function that “preserves multiplication”, in the sense that if  $\mathbf{Mult}(x, y, z)$  then  $\mathbf{Mult}(\phi(x), \phi(y), \phi(z))$  for all  $x, y, z \in G$ .

Before continuing with theoretical matters, we give a few examples of homomorphisms. First of all, let  $F$  be a field and  $G = \mathrm{GL}_n(F)$ , the group of all invertible  $n \times n$  matrices over  $F$ . For each  $X \in \mathrm{GL}_n(F)$  let  $\det(X)$  be the determinant of  $X$ , and note that  $\det(X) \neq 0$  since  $X$  is invertible. The function  $\det: G \rightarrow F^\times$  defined by  $X \mapsto \det(X)$  is a homomorphism, since

$$\det(XY) = \det(X) \det(Y)$$

for all  $n \times n$  matrices  $X$  and  $Y$ .

Recall that if  $z = x + iy$  is a complex number, with real part  $x$  and imaginary part  $y$ , then the modulus of  $z$  is the real number  $|z| = \sqrt{x^2 + y^2}$ . Since  $|z| = 0$  if and only if  $z = 0$ , the rule  $z \mapsto |z|$  defines a function from  $\mathbb{C}^\times$  to  $\mathbb{R}^\times$ . Furthermore, it is a well known property of complex numbers that  $|z_1 z_2| = |z_1| |z_2|$  for all  $z_1, z_2 \in \mathbb{C}$ ; so this function is a homomorphism.

If  $p \in \text{Sym}\{1, 2, \dots, n\}$  define  $\ell(p)$  to be the number of ordered pairs  $(i, j)$  of integers in the set  $\{1, 2, \dots, n\}$  such that  $i < j$  and  $p(i) > p(j)$ . For example, if

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

then it can be seen that  $\ell(p) = 14$ . Indeed, the pairs  $(i, j)$  with  $i < j$  and  $p(i) > p(j)$  are  $(1, 2)$ ,  $(1, 4)$ ,  $(1, 6)$ ,  $(1, 8)$ ,  $(2, 6)$ ,  $(3, 4)$ ,  $(3, 5)$ ,  $(3, 6)$ ,  $(3, 7)$ ,  $(3, 8)$ ,  $(4, 6)$ ,  $(5, 6)$ ,  $(5, 8)$  and  $(7, 8)$ . (These correspond to the cases where a number in the second row of  $p$  has a smaller number somewhere to its right.) The sign, or parity, of  $p$  is  $\varepsilon(p) = (-1)^{\ell(p)}$ ; if  $\varepsilon(p) = 1$  then  $p$  is an even permutation, otherwise it is an odd permutation. The most important property of parity, and the reason for its importance, is the fact that if two permutations have the same parity then their product is even, while if they have opposite parity then their product is odd. That is,

$$\varepsilon(pq) = \varepsilon(p)\varepsilon(q)$$

for all  $p, q \in \text{Sym}\{1, 2, \dots, n\}$ . It is trivial that the numbers  $\pm 1$  form a cyclic group of order 2 (the operation being ordinary multiplication of numbers), and the above equation shows that  $\varepsilon: \text{Sym}\{1, 2, \dots, n\} \rightarrow \{\pm 1\}$  is a homomorphism.

Bridge is a game played by four people, two against two in partnership. Given four people  $P_1, P_2, P_3$  and  $P_4$ , there are precisely three different ways to choose the partnerships:  $P_4$  can partner  $P_1, P_2$  or  $P_3$ , and once  $P_4$ 's partner has been chosen the partnerships are completely determined. A more formal mathematical way of expressing this fact is to say that there are three equivalence relations on the set  $\{1, 2, 3, 4\}$  having the property that there are exactly two elements in each equivalence class. We shall call these equivalence relations  $R_1, R_2$  and  $R_3$ ; they correspond to the following partitionings of  $\{1, 2, 3, 4\}$  into equivalence classes:

$$\begin{aligned} \{1, 2, 3, 4\} &= \{4, 1\} \cup \{2, 3\} \\ \{1, 2, 3, 4\} &= \{4, 2\} \cup \{1, 3\} \\ \{1, 2, 3, 4\} &= \{4, 3\} \cup \{1, 2\}. \end{aligned}$$

Thus, for the relation  $R_1$  we have  $4R_11$  and  $2R_13$ , but not  $1R_12$  or  $1R_13$ . The equivalence classes correspond to the partnerships; for the relation  $R_i$ , player  $P_i$  is the partner of player  $P_4$ .

Now if  $p$  is any permutation of  $\{1, 2, 3, 4\}$  then we see that permuting the players in accordance with  $p$  will permute the three equivalence relations in some manner. To be precise, if  $p$  is a permutation of  $\{1, 2, 3, 4\}$ , then for each equivalence relation  $R$  on  $\{1, 2, 3, 4\}$  there is another equivalence relation  $S$  on  $\{1, 2, 3, 4\}$  such that

$$iRj \text{ if and only if } p(i)Sp(j).$$

If we choose  $R = R_1$  we will find that  $S = R_l$  for some  $l$ , then choosing  $R = R_2$  will give  $S = R_m$  for some  $m \neq l$ , and finally  $R = R_3$  will give  $S = R_n$  with  $n \neq l$  and  $n \neq m$ . In this way the permutation  $p$  has given rise to a permutation  $\begin{bmatrix} 1 & 2 & 3 \\ l & m & n \end{bmatrix}$  of  $\{1, 2, 3\}$ . Calling this permutation  $\phi p$ , the following formula summarizes all that we have said: for each  $p \in \text{Sym}\{1, 2, 3, 4\}$  there is a permutation  $\phi p \in \text{Sym}\{1, 2, 3\}$  such that

$$(3.1.1) \quad iR_kj \text{ if and only if } p(i)R_{(\phi p)(k)}p(j).$$

We shall show in a moment that this function  $\phi$  from  $\text{Sym}\{1, 2, 3, 4\}$  to  $\text{Sym}\{1, 2, 3\}$  is a homomorphism. But before doing so, let us calculate  $\phi p$  for several different permutations  $p$ ; this may make the situation more understandable.

First of all, choose  $p = (1, 2)(3, 4)$ . The equivalence relation  $R_3$  has players  $P_4$  and  $P_3$  as one partnership,  $P_1$  and  $P_2$  the other, and so the permutation  $p$  says that each player swaps places with his/her partner. The partnerships are thus unchanged. So  $p$  in fact preserves the equivalence relation  $R_3$ . Next consider  $R_2$ , for which the partnerships are  $\{P_4, P_2\}$  and  $\{P_1, P_3\}$ . The permutation  $(1, 2)(3, 4)$  makes each player swap places with someone from the other partnership. The total effect of this is that the partnerships are the same as before:

$$\begin{array}{cc} \{P_4, P_2\} & \{P_1, P_3\} \\ \downarrow \downarrow & \downarrow \downarrow \\ \{P_3, P_1\} & \{P_2, P_4\} \end{array}$$

Thus  $p$  preserves  $R_2$  as well. Similarly we find that  $p$  preserves  $R_1$ :

$$\begin{array}{cc} \{P_4, P_1\} & \{P_2, P_3\} \\ \downarrow \downarrow & \downarrow \downarrow \\ \{P_3, P_2\} & \{P_1, P_4\} \end{array}$$

We have thus shown that if  $p = (1, 2)(3, 4)$  then  $\phi p = i$ , the identity permutation of  $\{1, 2, 3\}$ .

Next consider  $p = (1, 2)$ . For the relation  $R_3$ , where  $P_4$  and  $P_3$  are partners, the permutation  $p$  tells  $P_4$  and  $P_3$  to stay where they are, while the other partners,  $P_1$  and  $P_2$ , swap places. This does not affect the partnerships. So  $p$  preserves  $R_3$ . But for  $R_2$  the players  $P_4$  and  $P_2$  are partners, and  $p$  keeps  $P_4$  in the same place but makes  $P_2$  swap with  $P_1$ . So  $p$  changes  $R_2$  into  $R_1$  (where  $P_4$  partners  $P_1$ ). Similarly,  $p$  transforms  $R_1$  into  $R_2$ , and we conclude that  $p = (1, 2)$  gives  $\phi p = (1, 2)$ .

More generally, let  $p \in \text{Sym}\{1, 2, 3, 4\}$  be any permutation such that  $p(4) = 4$ . Of course, there are exactly six of these, namely  $i$ ,  $(1, 2)$ ,  $(1, 3)$ ,  $(2, 3)$ ,  $(1, 2, 3)$  and  $(1, 3, 2)$ . The equivalence relation  $R_i$  has  $P_4$  partnering  $P_i$ ; apply  $p$  and we have  $P_{p(4)}$  partnering  $P_{p(i)}$ . But since  $p(4) = 4$ , we have  $P_4$  partnering  $P_{p(i)}$ , and the relation corresponding to this is  $R_{p(i)}$ . In other words, for these permutations  $p$  which leave 4 fixed, the equivalence relations  $R_1$ ,  $R_2$  and  $R_3$  are permuted in exactly the same way as  $P_1$ ,  $P_2$  and  $P_3$  are permuted. So we have  $\phi(1, 3) = (1, 3)$ ,  $\phi(1, 3, 2) = (1, 3, 2)$  and so on.

Finally, let us calculate  $\phi p$  when  $p = (1, 4, 3, 2)$ . Consider  $R_3$  first: here  $P_1$  and  $P_2$  are partners. Apply  $p$  and we find that  $P_{p(1)}$  and  $P_{p(2)}$  are partners; that is,  $P_4$  and  $P_1$  are partners. So  $R_3$  has been transformed to  $R_1$ . For  $R_1$ , where  $P_1$  and  $P_4$  are partners, applying  $p$  makes  $P_4$  and  $P_3$  partners, showing that  $R_1$  is transformed to  $R_3$ . We see that  $R_2$  must be fixed, and so  $\phi(1, 4, 3, 2) = (1, 3)$ .

Now for the promised proof that  $\phi$  is a homomorphism. Let  $p$  and  $q$  be two permutations in  $\text{Sym}\{1, 2, 3, 4\}$ , and apply (3.1.1) with  $k$  replaced by  $(\phi q)(k)$  and  $i, j$  replaced by  $p^{-1}(i), p^{-1}(j)$ . This gives

$$p^{-1}(i)R_{(\phi q)(k)}p^{-1}(j) \text{ if and only if } iR_{((\phi p)(\phi q))(k)}j.$$

Now apply (3.1.1) again with  $i, j$  replaced by  $(q^{-1}p^{-1})(i), (q^{-1}p^{-1})(j)$ , and  $p$  replaced by  $q$ . This gives

$$(q^{-1}p^{-1})(i)R_k(q^{-1}p^{-1})(j) \text{ if and only if } p^{-1}(i)R_{(\phi q)(k)}p^{-1}(j).$$

Finally, apply (3.1.1) again, this time with  $p$  replaced by  $pq$  and  $i, j$  replaced by  $(q^{-1}p^{-1})(i), (q^{-1}p^{-1})(j)$ . We find that

$$(q^{-1}p^{-1})(i)R_k(q^{-1}p^{-1})(j) \text{ if and only if } iR_{(\phi(pq))(k)}j.$$

Combining these three equivalences shows that for all  $i, j \in \{1, 2, 3, 4\}$ ,

$$iR_{(\phi(pq))(k)}j \text{ if and only if } iR_{((\phi p)(\phi q))(k)}j,$$

and hence the relations  $R_{(\phi(pq))(k)}$  and  $R_{((\phi p)(\phi q))(k)}$  are the same. Thus

$$(\phi(pq))(k) = ((\phi p)(\phi q))(k)$$

for all  $k \in \{1, 2, 3\}$ , and hence the permutations  $\phi(pq)$  and  $(\phi p)(\phi q)$  are equal. Thus we have shown that  $\phi$  preserves multiplication, as required.

The preceding example is fairly typical of the way homomorphisms arise. Given a group  $G$  of symmetries of some object  $X$ , if one can find another associated object  $Y$ —perhaps a part of  $X$ —then it may turn out that every symmetry of  $X$  in the group  $G$  gives rise to a symmetry of  $Y$ . In such cases there will be a homomorphism from  $G$  to the group of symmetries of  $Y$ .

For example, consider the set  $G$  of all permutations of  $\{1, 2, 3, 4, 5\}$  that preserve the subset  $\{1, 2, 3\}$ . That is,

$$G = \{p \in \text{Sym}\{1, 2, 3, 4, 5\} \mid p(i) \in \{1, 2, 3\} \text{ for all } i \in \{1, 2, 3\}\}.$$

Then  $G$  is a group, and for each element  $p \in G$  we can define  $\psi p$  to be that permutation of  $\{1, 2, 3\}$  such that  $(\psi p)(i) = p(i)$  for all  $i \in \{1, 2, 3\}$ . Thus, if  $p = (1, 3)(4, 5)$  then  $p$  preserves  $\{1, 2, 3\}$ , and so  $p \in G$ , and we see that  $\psi p$ , the permutation of  $\{1, 2, 3\}$  to which  $p$  gives rise, is just  $(1, 3)$ . This mapping  $\psi$ , given by  $p \mapsto \psi p$ , is a homomorphism.

To close this section we return briefly to general theoretical matters. We know that a bijective function from one set to another is just a one-to-one correspondence between the elements of the two sets. If the sets are groups and the function a homomorphism, this one-to-one correspondence also preserves the group structure. So, for example, if  $G$  is a finite group, with elements  $g_1, g_2, \dots, g_n$ , then  $H$  also has  $n$  elements, namely  $\phi(g_1), \phi(g_2), \dots, \phi(g_n)$ , and if the product in  $G$  of  $g_i$  and  $g_j$  is  $g_k$ , then in  $H$  we have

$$\phi(g_i)\phi(g_j) = \phi(g_i g_j) = \phi(g_k).$$

So the multiplication table of  $G$  becomes the multiplication table of  $H$  if, for each  $i$ , the element  $g_i$  is replaced by  $\phi(g_i)$  wherever it occurs. As we mentioned in Chapter 1, groups  $G$  and  $H$  which are related in this way are said to be isomorphic to each other.

**3.2 DEFINITION** A bijective homomorphism is called an *isomorphism*, and groups  $G$  and  $H$  are said to be *isomorphic* if there is an isomorphism from  $G$  to  $H$ . We will write  $G \cong H$  to indicate that  $G$  and  $H$  are isomorphic. The relation  $\cong$ , defined on the class of all groups, is also called *isomorphism*.

It would clearly be undesirable to use the terminology we have just introduced were  $\cong$  not an equivalence relation, or, at the very least, symmetric. We would not be prepared to say “ $G$  and  $H$  are isomorphic”, if we were not also prepared to say “ $H$  and  $G$  are isomorphic”. But fortunately it is true that isomorphism of groups is an equivalence relation. We leave the proof of this as an exercise.

Finally, we prove two simple general properties of homomorphisms.

**3.3 PROPOSITION** Let  $G, H$  be groups and  $\phi: G \rightarrow H$  a homomorphism.

- (i) If  $e_G$  is the identity element of  $G$  and  $e_H$  the identity element of  $H$ , then  $\phi(e_G) = e_H$ .
- (ii) If  $x$  is any element of  $G$ , then  $\phi(x^{-1}) = \phi(x)^{-1}$ .

**Proof.** Since  $\phi(e_G) \in H$ , we have

$$\phi(e_G)e_H = \phi(e_G) = \phi(e_G^2) = \phi(e_G)^2,$$

and cancellation yields  $e_H = \phi(e_G)$ , as desired. Now for all  $x \in G$ ,

$$\phi(x^{-1})\phi(x) = \phi(x^{-1}x) = \phi(e_G) = e_H,$$

and by Proposition 2.5 it follows that  $\phi(x^{-1}) = \phi(x)^{-1}$ . □

### §3b Quotient groups

It will often be convenient for us to deal with the product of two subsets of a group, as defined in the following natural sense.

**3.4 DEFINITION** If  $G$  is a group and  $S, T \subseteq G$ , then we define

$$ST = \{ st \mid s \in S \text{ and } t \in T \}.$$

Observe that this multiplication of subsets of  $G$  satisfies the associative law.



3.5 PROPOSITION If  $G$  is a group and  $S, T, U \subseteq G$ , then  $(ST)U = S(TU)$ .

**Proof.** By the definition of the product of two subsets of  $G$ ,

$$\begin{aligned} (ST)U &= \{xu \mid x \in ST \text{ and } u \in U\} \\ &= \{(st)u \mid s \in S, t \in T \text{ and } u \in U\} \\ &= \{s(tu) \mid s \in S, t \in T \text{ and } u \in U\} \\ &= \{sy \mid s \in S \text{ and } y \in TU\} \\ &= S(TU), \end{aligned}$$

as required. □

Of course, Proposition 3.5 allows us to omit the bracketing from products of three or more subsets of  $G$ .

If  $x \in G$  and  $T \subseteq G$  then we define  $xT = \{x\}T = \{xt \mid t \in T\}$ . Similarly,  $Tx$  is defined to equal  $T\{x\}$ . Note that this is consistent with, and generalizes, the notation for cosets which we introduced previously: now  $xT$  and  $Tx$  are defined for all  $T \subseteq G$ , whereas previously they were only defined if  $T$  were a subgroup.

In the previous chapter we mentioned normal subgroups, describing them as subgroups with the property that every right coset of the subgroup is also a left coset of the subgroup, and vice versa. If  $K$  is a normal subgroup of  $G$  and  $x \in G$ , then the left coset of  $K$  containing  $x$  is  $xK$ , and the right coset of  $K$  containing  $x$  is  $Kx$ . But the left coset containing  $x$  must be a right coset, and so it must be the right coset containing  $x$ . It follows that  $xK = Kx$  for all  $x \in G$ ; this would be a suitable alternative definition of normality. However, let us choose the following third alternative as our official definition of normality in these notes.

3.6 DEFINITION Let  $G$  be a group and  $K$  a subgroup of  $G$ . We say that  $K$  is *normal* in  $G$  if  $x^{-1}kx \in K$  for all  $k \in K$  and  $x \in G$ .

It is not totally obvious that this actually is equivalent to the others; so let us prove it.

3.7 PROPOSITION Let  $G$  be a group and  $K$  a subgroup of  $G$ . The following are equivalent:

- (i)  $x^{-1}kx \in K$  for all  $k \in K$  and  $x \in G$ ,
- (ii)  $x^{-1}Kx \subseteq K$  for all  $x \in G$ ,
- (iii)  $x^{-1}Kx = K$  for all  $x \in G$ ,
- (iv)  $xK = Kx$  for all  $x \in G$ ,
- (v) every left coset of  $K$  in  $G$  is also a right coset of  $K$  in  $G$ .

**Proof.** Let  $x \in G$ . Note that, in accordance with our definitions concerning products of subsets,  $x^{-1}Kx = \{x^{-1}kx \mid k \in K\}$ . Thus (i) says that every element of  $x^{-1}Kx$  is in  $K$ ; that is,  $x^{-1}Kx \subseteq K$ . Hence (i) and (ii) are equivalent.

The equivalence of (iv) and (v) was proved in the preamble above. If (iv) holds then using 3.5 we see that for all  $x \in G$ ,

$$x^{-1}Kx = x^{-1}(Kx) = x^{-1}(xK) = (x^{-1}x)K = eK = K,$$

whence (iii) holds. Conversely, if (iii) holds then

$$Kx = (xx^{-1})Kx = x(x^{-1}Kx) = xK \quad (\text{for all } x \in G),$$

and so (iv) holds. Hence the equivalence of (iii) and (iv) is established, and all that remains is to prove that (ii) and (iii) are equivalent; furthermore, it is trivial that (iii) implies (ii), and so the task reduces to proving that (ii) implies (iii).

Assume that (ii) holds, and let  $x \in G$ . By (ii) we have that  $x^{-1}Kx \subseteq K$ . But the same must also be true with  $x$  replaced by  $x^{-1}$ , since the formula in (ii) is assumed to hold for all  $x \in G$ . So we also have that  $xKx^{-1} \subseteq K$ , and it follows that  $x^{-1}(xKx^{-1})x \subseteq x^{-1}Kx$ . But since

$$x^{-1}(xKx^{-1})x = (x^{-1}x)K(x^{-1}x) = eKe = K$$

we have that  $K \subseteq x^{-1}Kx$ . Combining this with the previously established fact that  $x^{-1}Kx \subseteq K$ , we can conclude that  $x^{-1}Kx = K$ , and therefore that (iii) holds, as required.  $\square$

Let  $K$  be a normal subgroup of the group  $G$ . Since the left and right cosets of  $K$  in  $G$  coincide, it follows that the equivalence relations on  $G$  of left congruence modulo  $K$  and right congruence modulo  $K$  must coincide also. Indeed, if  $x, y \in G$  are left congruent modulo  $K$ , then  $x = ky$  for some  $k \in K$ ; but this gives  $x = y(y^{-1}ky)$ , and since normality of  $K$  yields

that  $y^{-1}ky \in K$ , we can conclude that  $x, y$  are right congruent modulo  $K$  as well. In this situation we say simply that  $x$  and  $y$  are *congruent* modulo  $K$ .

The crucial property of congruence modulo a normal subgroup, which makes it far more valuable than left or right congruence modulo a non-normal subgroup, is that it interacts with multiplication in the best way imaginable.

**3.8 PROPOSITION** *Let  $K$  be a normal subgroup of the group  $G$ , and let  $\equiv$  be the relation on  $G$  of congruence modulo  $K$ . If  $x, x', y, y' \in G$  are such that  $x' \equiv x$  and  $y' \equiv y$ , then  $x'y' \equiv xy$ .*

**Proof.** Given  $x' \equiv x$  and  $y' \equiv y$ , there exist  $h, k \in K$  such that  $x' = xh$  and  $y' = yk$ , and hence

$$x'y' = (xh)(yk) = xy(y^{-1}hy)k.$$

But since  $K$  is normal we know that  $y^{-1}hy \in K$ , and since  $k \in K$  it follows from closure of  $K$  under multiplication that  $(y^{-1}hy)k \in K$ . Thus the equation  $x'y' = xy(y^{-1}hy)k$  says that  $x'y' \equiv xy$ , as required.  $\square$

As an almost immediate consequence of 3.8 we see that if  $C_1, C_2 \subseteq G$  are equivalence classes for the equivalence relation  $\equiv$ , then the product  $C_1C_2$  is also an equivalence class. For, let  $g$  be an arbitrary element of  $C_1C_2$ . Then we have  $g = xy$  for some  $x \in C_1$  and  $y \in C_2$ . If  $g'$  is also an arbitrary element of  $C_1C_2$  then, similarly,  $g' = x'y'$ , where  $x' \in C_1$  and  $y' \in C_2$ . Since  $x$  and  $x'$  are in the same equivalence class we have that  $x' \equiv x$ , and similarly  $y' \equiv y$ , and now by 3.8 we conclude that  $g' \equiv g$ . So every element of  $C_1C_2$  is congruent to  $g$ . Conversely, suppose that  $g''$  is an arbitrary element of  $G$  that is congruent to  $g$ . Then  $g'' = gk$  for some  $k \in K$ , and this gives  $g'' = x(yk) \in C_1C_2$ , since  $x$  is in  $C_1$ , and  $yk$ , being congruent to  $y$ , is in  $C_2$ . So an element of  $G$  is congruent to  $g$  if and only if it is in the set  $C_1C_2$ ; thus  $C_1C_2$  is an equivalence class, as claimed.

**3.9 PROPOSITION** *Let  $K$  be a normal subgroup of the group  $G$ . Then  $(xK)(yK) = xyK$  for all  $x, y \in G$ .*

**Proof.** As above, let  $\equiv$  be the relation of congruence modulo  $K$ . From our discussion of congruence modulo a subgroup in Chapter 2, we know that  $xK$  is the equivalence class containing  $x$ , and  $yK$  the equivalence class containing  $y$ . By the discussion above it follows that  $(xK)(yK)$  is an equivalence class; but since  $x \in xK$  and  $y \in yK$  it follows that  $xy \in (xK)(yK)$ , whence  $(xK)(yK)$  is the equivalence class that contains  $xy$ . So  $(xK)(yK) = xyK$ , as claimed.  $\square$

Continuing with the assumption that  $K$  is a normal subgroup of  $G$  and  $\equiv$  the relation of congruence modulo  $K$ , let  $\mathcal{Q}$  be the quotient of  $G$  by  $\equiv$ . That is, by Definition 2.19,  $\mathcal{Q}$  is the set of all equivalence classes of  $G$  under  $\equiv$ . We have seen that the product of two equivalence classes is an equivalence class; so subset multiplication defines an operation on  $\mathcal{Q}$ . It turns out that this operation makes  $\mathcal{Q}$  into a group; however, before proving this in general, let us look at an example.

Let  $G = \text{Alt}\{1, 2, 3, 4\}$  and  $K$  the subgroup consisting of the identity permutation  $i$  and the three permutations  $a = (1, 2)(3, 4)$ ,  $b = (1, 3)(2, 4)$  and  $c = (1, 4)(2, 3)$ . We investigated the cosets of  $K$  in  $G$  at the end of Chapter 2, and concluded that  $K$  is normal in  $G$ , and that the cosets of  $K$  in  $G$  are, in the notation used in the multiplication table in Chapter 1,

$$\begin{aligned} iK &= \{i, a, b, c\} \\ (1, 2, 3)K &= \{t_1, t_2, t_3, t_4\} \\ (1, 3, 2)K &= \{s_1, s_2, s_3, s_4\}. \end{aligned}$$

Since when writing out the multiplication table we grouped the elements of  $G$  according to their cosets modulo  $K$ , it is easy to observe from a glance at the table that the assertion of Proposition 3.8 is indeed satisfied. The product of two elements of the coset  $\{t_1, t_2, t_3, t_4\}$  is an element of the coset  $\{s_1, s_2, s_3, s_4\}$ , the product of an element in  $\{s_1, s_2, s_3, s_4\}$  with one in  $\{t_1, t_2, t_3, t_4\}$  lies in  $\{i, a, b, c\}$ , and so on. If we write the cosets as  $I = \{i, a, b, c\}$ ,  $T = \{t_1, t_2, t_3, t_4\}$  and  $S = \{s_1, s_2, s_3, s_4\}$  then we see that the multiplication table for the cosets is as follows.

	$I$	$T$	$S$
$I$	$I$	$T$	$S$
$T$	$T$	$S$	$I$
$S$	$S$	$I$	$T$

Since this is exactly the same as the multiplication table for a cyclic group of order three, we conclude that in this case at least,  $\mathcal{Q}$ , the set of all equivalence classes for the relation congruence modulo  $K$ , is a group. Now let us prove it in general.

**3.10 PROPOSITION** *Let  $G$  be a group and let  $K$  be a normal subgroup of  $G$ . Let  $\mathcal{Q} = \{xK \mid x \in G\}$ , the set of all cosets of  $K$  in  $G$ , which equals the set of all equivalence classes of  $G$  under congruence modulo  $K$ . Then  $\mathcal{Q}$  forms*

a group, with multiplication in  $\mathcal{Q}$  satisfying the rule that  $(xK)(yK) = xyK$  for all  $x, y \in G$ . The identity element of this group is the coset  $K = eK$  (where  $e$  is the identity element of  $G$ ), and for all  $x \in G$  the inverse in  $\mathcal{Q}$  of the coset  $xK$  is the coset  $x^{-1}K$ .

**Proof.** We have seen that subset multiplication yields an operation on  $\mathcal{Q}$ , and in 3.9 above we have seen that it satisfies  $(xK)(yK) = xyK$  for all  $x, y \in G$ . Furthermore, we know from Proposition 3.5 that it is associative. So all that remains is to check the group axioms (ii) (a) and (b) pertaining to the identity element and inverses (see Definition 2.1).

Let  $C$  be an arbitrary element of  $\mathcal{Q}$ , so that  $C = xK$  for some  $x \in G$ . By 3.9 we have that

$$KC = (eK)(xK) = exK = xK = C,$$

and similarly

$$CK = (xK)(eK) = xeK = xK = C.$$

So the element  $K \in \mathcal{Q}$  has the property that  $CK = KC = C$  for all  $C \in \mathcal{Q}$ ; that is,  $K$  is an identity element for  $\mathcal{Q}$ .

Again, let  $C = xK$ , an arbitrary element of  $\mathcal{Q}$ . By 3.9

$$C(x^{-1}K) = (xK)(x^{-1}K) = xx^{-1}K = eK = K,$$

and similarly

$$(x^{-1}K)C = (x^{-1}K)(xK) = x^{-1}xK = eK = K,$$

showing that  $x^{-1}K$  has the property required of an inverse of  $C$ . So the group axioms are satisfied, and all parts of the proposition are proved.  $\square$

**3.11 DEFINITION** The group  $\mathcal{Q}$  in Proposition 3.10 is called the *quotient group*, or *factor group*,  $G/K$ .

Surely, one might think, this same construction will work for any subgroup  $K$  of a group  $G$ ; the restriction to normal subgroups must be unnecessary. For, with  $\mathcal{Q} = \{xK \mid x \in G\}$  and with multiplication defined by  $(xK)(yK) = xyK$ , all the group axioms are satisfied. The associative law is trivial, since

$$(xKyK)zK = xyKzK = (xy)zK = x(yz)K = xKyzK = xK(yKzK),$$

the coset  $K = eK$  is an identity element, since by our definition of multiplication

$$(xK)(eK) = xeK = xK = exK = (eK)(xK),$$

and  $x^{-1}K$  is the inverse of  $xK$  since

$$(x^{-1}K)(xK) = (x^{-1}x)K = eK = (xx^{-1})K = (xK)(x^{-1}K).$$

The error in this reasoning comes right at the start, where we blithely asserted that multiplication could be defined on  $\mathcal{Q}$  in such a way that the formula  $(xK)(yK) = xyK$  holds. For the subsequent reasoning to work, we need this formula for all  $x, y \in G$ . However, if  $K$  is not normal in  $G$  this formula is inconsistent with itself. If  $x, x' \in G$  are right congruent modulo  $K$  then  $x$  and  $x'$  correspond to the same element  $C \in \mathcal{Q}$ , the coset  $C = xK = x'K$ . If we multiply this element of  $\mathcal{Q}$  by another,  $yK$  say, then on the one hand the product  $C(yK)$  must be  $(xK)(yK) = xyK$ , on the other hand it must also be  $(x'K)(yK) = x'yK$ . But the product  $C(yK)$  cannot be defined in two different ways: the reasoning will be invalid unless  $x'yK = xyK$ , for all  $y \in G$ , whenever  $x'K = xK$ . Rearranging this using 2.11, the required condition becomes that  $(xy)^{-1}(x'y)$ , which equals  $y^{-1}(x^{-1}x')y$ , must be in  $K$  whenever  $x^{-1}x' \in H$ . Thus putting  $x' = xk$  we see that our multiplication rule is unambiguously defined only if  $y^{-1}ky \in K$  whenever  $k \in K$ ; that is,  $K$  must be normal in  $G$ .

An example should make this clearer. Let  $G = \text{Sym}\{1, 2, 3\}$  and let  $K = \{i, (1, 2)\}$ , the cyclic subgroup generated by  $(1, 2)$ . The left cosets of  $K$  in  $G$  all have two elements, and we can check readily that  $\{(1, 3), (1, 2, 3)\}$  and  $\{(2, 3), (1, 3, 2)\}$  are both left cosets. Let us attempt to evaluate the product of these by means of the so-called formula  $(xK)(yK) = xyK$ . We have that

$$\begin{aligned} \{(1, 3), (1, 2, 3)\} &= (1, 3)K \\ \{(2, 3), (1, 3, 2)\} &= (2, 3)K \end{aligned}$$

and therefore

$$\begin{aligned} \{(1, 3), (1, 2, 3)\}\{(2, 3), (1, 3, 2)\} &= (1, 3)(2, 3)K \\ &= (1, 3, 2)K \\ &= \{(2, 3), (1, 3, 2)\}. \end{aligned}$$

But equally we have

$$\begin{aligned} \{(1, 3), (1, 2, 3)\} &= (1, 2, 3)K \\ \{(2, 3), (1, 3, 2)\} &= (1, 3, 2)K \end{aligned}$$

and therefore

$$\begin{aligned} \{(1, 3), (1, 2, 3)\}\{(2, 3), (1, 3, 2)\} &= (1, 2, 3)(1, 3, 2)K \\ &= iH \\ &= \{i, (1, 2)\}. \end{aligned}$$

Clearly, the attempted definition of coset multiplication is unsatisfactory in this case.

Return now to the case when  $K$  is normal in  $G$ , so that the quotient  $G/K$  has a well-defined multiplication, and is a group. Given that the construction works in this case, the following observation is trivial.

**3.12 PROPOSITION** *Let  $K$  be a normal subgroup of the group  $G$ . Then the function  $\eta: G \rightarrow G/K$ , defined by the rule that  $\eta(x) = xK$  for all  $x \in G$ , is a surjective homomorphism.*

**Proof.** By 3.9,  $\eta(x)\eta(y) = (xK)(yK) = xyK = \eta(xy)$ , for all  $x, y \in G$ . Hence  $\eta$  is a homomorphism. It is surjective since by definition of  $G/K$ , every element of  $G/K$  has the form  $xK = \eta(x)$  for some  $x \in G$ .  $\square$

The homomorphism  $\eta$  appearing in Proposition 3.12 is called the *natural* or *canonical* homomorphism from  $G$  to  $G/K$ .

### §3c The Homomorphism Theorem

Our main objective in this section is to analyse homomorphisms. The next definition provides a key concept for this task.

**3.13 DEFINITION** Let  $G, H$  be groups and  $\phi: G \rightarrow H$  a homomorphism. The *kernel* of  $\phi$  is the set  $\ker \phi = \{s \in S \mid \phi(s) = e_H\}$ , where  $e_H$  is the identity element of  $H$ .

There is an intimate connection between normal subgroups and homomorphisms, the first aspect of which is the following proposition.

**3.14 PROPOSITION** *Let  $G, H$  be groups and  $\phi: G \rightarrow H$  a homomorphism. Then the kernel of  $\phi$  is a normal subgroup of  $G$ .*

**Proof.** We first use Proposition 2.10 to prove that  $\ker \phi$  is a subgroup of  $G$ . By Proposition 3.3 we know that  $e_G \in \ker \phi$ , and so  $\ker \phi \neq \emptyset$ . If  $x, y \in \ker \phi$  then  $\phi(x) = \phi(y) = e_H$ , and it follows that

$$\phi(xy) = \phi(x)\phi(y) = e_H^2 = e_H,$$

whence  $e_H \in \ker \phi$ . Thus  $\ker \phi$  is closed under multiplication. And if  $x \in \ker \phi$  then since  $\phi(x) = e_H$  we deduce from 3.3 that

$$\phi(x^{-1}) = \phi(x)^{-1} = e_H^{-1} = e_H,$$

whence  $\ker \phi$  is closed under inversion. We have now verified that the hypotheses of 2.10 are satisfied, and can conclude that  $\ker \phi$  is a subgroup.

Let  $k \in \ker \phi$  and  $x \in G$ . Then  $\phi(k) = e_H$ , and since  $\phi(x^{-1}) = \phi(x)^{-1}$  (by 3.3), we have

$$\phi(x^{-1}kx) = \phi(x^{-1})\phi(k)\phi(x) = \phi(x)^{-1}e_H\phi(x) = \phi(x)^{-1}\phi(x) = e_H,$$

whence  $x^{-1}kx \in \ker \phi$ . Since this holds for all  $k \in \ker \phi$  and all  $x \in G$ , we have shown that the subgroup  $\ker \phi$  is normal in  $G$ .  $\square$

A homomorphism from  $G$  to  $H$  also gives rise to a subgroup of  $H$ , as the next proposition shows. In this case, however, the subgroup need not be normal.

**3.15 PROPOSITION** *Let  $G, H$  be groups and  $\phi: G \rightarrow H$  a homomorphism. Then the image of  $\phi$ ,*

$$\text{im } \phi = \{ \phi(g) \mid g \in G \}$$

*is a subgroup of  $H$ .*

**Proof.** In view of 2.10, our task is to show that  $\text{im } \phi$  is nonempty and closed under multiplication and inversion.

Since  $G \neq \emptyset$ , since  $G$  must at least have an identity element, it follows that  $\text{im } \phi \neq \emptyset$ . Specifically, we know that  $\phi(e_G) \in \text{im } \phi$ . (In fact, by Proposition 3.3, we know that  $\phi(e_G) = e_H$ , although technically this fact is not needed here.)



Let  $h, k \in \text{im } \phi$ . Then, by definition of  $\text{im } \phi$ , there exist  $x, y \in G$  with  $h = \phi(x)$  and  $k = \phi(y)$ . Now

$$hk = \phi(x)\phi(y) = \phi(xy) \in \text{im } \phi$$

and we have shown that the product of two arbitrary elements of  $\text{im } \phi$  always lies in  $\text{im } \phi$ . That is,  $\text{im } \phi$  is closed under multiplication.

Let  $h \in \text{im } \phi$ . Then  $h = \phi(x)$  for some  $x \in G$ , and by 3.3

$$h^{-1} = \phi(x)^{-1} = \phi(x^{-1}) \in \text{im } \phi,$$

and so we see that  $\text{im } \phi$  is also closed under inversion, as required.  $\square$

Obviously, by the definition of surjectivity, a homomorphism  $\phi$  from a group  $G$  to a group  $H$  is surjective if and only if the subgroup  $\text{im } \phi$  is equal to the whole group  $H$ . Parallel to this, but a little less trivial, we have the following criterion for injectivity of  $\phi$ .

**3.16 PROPOSITION** *Let  $G$  and  $H$  be groups and  $\phi: G \rightarrow H$  a homomorphism. Then  $\phi$  is injective if and only if the identity element of  $G$  is the only element of  $\ker \phi$ .*

**Proof.** Assume first that  $\phi$  is injective. We know from 3.3 that  $\phi(e_G) = e_H$ ; so  $e_G$  is an element of  $\ker \phi$ . To prove that it is the only element, let  $x \in \ker \phi$  be arbitrary, and observe that we then have  $\phi(x) = e_H = \phi(e_G)$ . Injectivity of  $\phi$  then gives immediately that  $x = e_G$ , as required.

Conversely, assume that  $\ker \phi = \{e_G\}$ ; we will prove that  $\phi$  is injective. Accordingly, assume that  $x, y \in G$  with  $\phi(x) = \phi(y)$ . By 3.3 we deduce that

$$\phi(x^{-1}y) = \phi(x^{-1})\phi(y) = \phi(x)^{-1}\phi(y) = \phi(x)^{-1}\phi(x) = e_H,$$

whence  $x^{-1}y \in \ker \phi$ . But this means that  $x^{-1}y = e_G$ , and hence  $x = y$ . So we have shown that  $\phi(x) = \phi(y)$  implies  $x = y$ ; that is,  $\phi$  is injective.  $\square$

We now come to the Homomorphism Theorem, or the First Isomorphism Theorem, which is the main theorem of introductory group theory.

3.17 THEOREM Let  $G$  and  $H$  be groups, and  $\phi: G \rightarrow H$  a homomorphism. Let  $K = \ker \phi$ , the kernel of  $\phi$ , and  $I = \text{im } \phi$ , the image of  $\phi$ . Then

$$G/K \cong I.$$

Indeed, there is an isomorphism  $\psi: G/K \rightarrow I$  such that  $\psi(xK) = \phi(x)$  for all  $x \in G$ .

**Proof.** Let  $C$  be an arbitrary element of  $G/K$ , and recall that  $C$  is an equivalence class for the relation of congruence modulo  $K$ . If  $x, y$  are arbitrary elements of  $C$  then  $x = yk$  for some  $k \in K$ , and so

$$\phi(x) = \phi(yk) = \phi(y)\phi(k) = \phi(y)e_H = \phi(y),$$

since  $K = \ker \phi$ . Therefore we can, without ambiguity, define  $\psi(C)$  so that  $\psi(C) = \phi(x)$  for all  $x \in C$ . Since  $\phi(x) \in \text{im } \phi = I$ , this rule defines a function  $\psi$  from  $G/K$  to  $I$ . Furthermore, if  $x \in G$  is arbitrary then  $x \in xK$ , and so  $\psi(xK) = \phi(x)$ . We have thus shown that there exists a function  $\psi: G/K \rightarrow I$  such that  $\psi(xK) = \phi(x)$  for all  $x \in G$ .

True, we could have started this proof with the apparently simply prescription “Define  $\psi: G/K \rightarrow I$  by  $\psi(xK) = \phi(x)$  for all  $x$ ”. However, this runs the risk of falling into the trap that appeared when we tried to define quotient groups for non-normal subgroups: it is possible to have  $xK = x'K$  without having  $x = x'$ , and so we would be simultaneously be defining  $\psi(xK) = \psi(x'K)$  as  $\phi(x)$  and as  $\phi(x')$ . To show that  $\psi$  is well-defined, it is necessary to show that  $\phi(x) = \phi(x')$  whenever  $xK = x'K$ ; this is in fact what we did in the first paragraph of the proof.

Given that  $\psi$  is well-defined, the rest of the proof becomes easy. First of all, if  $C_1, C_2 \in G/K$  are arbitrary, then we have  $C_1 = xK$  and  $C_2 = yK$  for some  $x, y \in G$ , and now

$$\begin{aligned} \psi(C_1C_2) &= \psi(xK yK) = \psi(xyK) = \phi(xy) \\ &= \phi(x)\phi(y) = \psi(xK)\psi(yK) = \psi(C_1)\psi(C_2). \end{aligned}$$

Hence  $\psi$  is a homomorphism.

Let  $h \in I$ . Since  $I = \text{im } \phi$  we must have  $h = \phi(x)$  for some  $x \in G$ , and hence  $\psi(xK) = \phi(x) = h$ . So we have shown that every element of  $I$ , the codomain of  $\psi$ , has the form  $\psi(C)$  for some  $C \in G/K$ . That is,  $\psi$  is surjective.

Finally, let  $C_1, C_2 \in G/K$  with  $\psi(C_1) = \psi(C_2)$ . We have  $C_1 = xK$  and  $C_2 = yK$  for some  $x, y \in G$ , and now

$$\phi(x) = \psi(xK) = \psi(C_1) = \psi(C_2) = \psi(yK) = \phi(y),$$

from which it follows that

$$\phi(x^{-1}y) = \phi(x)^{-1}\phi(y) = e_H.$$

Thus  $x^{-1}y \in \ker \phi = K$ , and, by 2.11,  $xK = yK$ . That is,  $C_1 = C_2$ . We have thus shown that  $\psi(C_1) = \psi(C_2)$  implies  $C_1 = C_2$ , and therefore that  $\psi$  is injective.

Since  $\psi$  is an injective and surjective homomorphism from  $G/K$  to  $I$ , it follows that  $G/K \cong I$ , as claimed.  $\square$

Although we did not make use of Proposition 2.20 in our proof of Theorem 3.17, it is nevertheless true that Theorem 3.17 is little more than Proposition 2.20 applied to a function which happens to be a homomorphism of groups. According to Proposition 2.20, a homomorphism  $\phi: G \rightarrow H$  can be factorized as a surjection  $\eta: G \rightarrow Q$ , followed by a bijection  $\psi: Q \rightarrow I$ , followed by an injection  $\sigma: I \rightarrow H$ . Furthermore, examining the proof of 2.20, we find that  $I$  is the image of  $\phi$  and  $Q$  the quotient of  $G$  by the equivalence relation  $\equiv$  defined by

$$x \equiv y \text{ if and only if } \phi(x) = \phi(y).$$

We showed in the proof of 3.17 that  $\phi(x) = \phi(y)$  if and only if  $x$  is congruent to  $y$  modulo  $K = \ker \phi$ ; so the equivalence relation derived from Proposition 2.20 is exactly the same as the equivalence relation used in the definition of  $G/K$ . So the  $Q$  from 2.20 is, in this context, exactly the quotient group  $G/K$ . Furthermore, looking once more at the proof of 2.20, we find that the mapping  $\psi: Q \rightarrow I$  satisfies  $\psi(C) = \phi(x)$  whenever  $x \in C$ . The mapping  $\psi$  in Theorem 3.17 is defined in exactly the same way. It can be checked easily that the proofs of the bijectivity of  $\psi$  given in the proof of 2.20 and in the proof of 3.17 are essentially the same.

It is worth noting also that the mapping  $\eta: G \rightarrow Q$ , defined in the proof of 2.20, is precisely the natural homomorphism  $G \rightarrow G/K$  as defined in Proposition 3.12. The mapping  $\sigma: I \rightarrow H$  is defined by  $\sigma(h) = h$  for all  $h \in I$ ; so clearly  $\sigma$  is a homomorphism also. So the content of Theorem 3.17, given Proposition 2.20, is really that the three factors  $\sigma$ ,  $\psi$  and  $\eta$  are all homomorphisms, if the original function  $\phi: G \rightarrow H$  is a homomorphism.

# 4

## Automorphisms, inner automorphisms and conjugacy

According to the general conventions we have adopted, if we regard a group  $G$  as a structured set then a symmetry, or automorphism, of  $G$  is a bijective transformation of  $G$  that preserves the group structure. We start this chapter by looking at a few examples of automorphisms of groups.

### §4a Automorphisms

In view of the above remarks and the definition of isomorphism given in Chapter 3, we see that the following definition is appropriate.

**4.1 DEFINITION** An *automorphism* of a group  $G$  is an isomorphism from  $G$  to itself.

Recalling that an isomorphism is a bijective homomorphism, and that a bijective function from  $G$  to itself can be called a permutation of  $G$ , the definition can be rephrased as follows: an automorphism is a permutation of  $G$  which is also a homomorphism.

The set of all symmetries of anything, no matter what, is always a group; so it is certainly true that the set of all automorphisms of a group is a group.

**4.2 DEFINITION** If  $G$  is a group then  $\text{Aut}(G)$ , the automorphism group of  $G$ , is the set  $\{\phi \mid \phi: G \rightarrow G \text{ is an automorphism}\}$ , with multiplication of elements of  $\text{Aut}(G)$  defined to be composition of functions.

The multiplication operation on  $\text{Aut}(G)$  is thus inherited from the multiplication operation on  $\text{Sym}(G)$ , the group of all permutations of the set  $G$ . If a separate proof were required that  $\text{Aut}(G)$  is indeed a group, the simplest way to do it would be to use 2.10 to show that  $\text{Aut}(G)$  is a subgroup of

$\text{Sym}(G)$ . The main task then is to show that the composite of two homomorphisms is a homomorphism, and the inverse of an isomorphism is also an isomorphism. We leave this as an exercise, and instead look at a few examples.

Let  $G = \{I, T, S\}$  be a cyclic group of order three. The multiplication table of  $G$  is as follows:

	$I$	$T$	$S$
$I$	$I$	$T$	$S$
$T$	$T$	$S$	$I$
$S$	$S$	$I$	$T$

If we interchange  $S$  and  $T$  in this table we obtain the following:

	$I$	$S$	$T$
$I$	$I$	$S$	$T$
$S$	$S$	$T$	$I$
$T$	$T$	$I$	$S$

However, this is still a multiplication table for  $G$ : the product of two elements of  $G$  will be the same whichever table you choose. For example, both tables say that  $T^2 = S$ , and both tables say that  $ST = I$ . The same would not have worked had we chosen to interchange  $T$  and  $I$  rather than  $T$  and  $S$ . The permutation

$$\phi = \begin{bmatrix} I & T & S \\ I & S & T \end{bmatrix}$$

preserves the multiplicative structure of the group  $G$ , whereas the permutation

$$\rho = \begin{bmatrix} I & T & S \\ T & I & S \end{bmatrix}$$

does not. That is,  $\phi$  is an automorphism of  $G$  and  $\rho$  is not. It is easy to show that in fact the only other automorphism of  $G$  is the identity transformation.

We proved in Proposition 3.3 that if  $\phi: G \rightarrow H$  is any homomorphism then  $\phi(e_G) = e_H$ . It follows that any automorphism of a group must take the identity element to itself. If  $G$  is the Klein 4-group then it turns out that every permutation of  $G$  which fixes the identity element is an automorphism of  $G$ . This can be seen as follows. Write  $G = \{e, a, b, c\}$ , where  $e$  is the identity element, and multiplication is given by the table which we wrote down in Chapter 1. The important thing to note here is that three simple rules

describe the multiplication completely. First,  $x^2 = e$  for all  $x \in G$ ; second,  $xe = ex = x$  for all  $x \in G$ ; third, if  $x$  and  $y$  are distinct nonidentity elements of  $G$  then  $xy$  is the third nonidentity element of  $G$ . Now let  $\phi: G \rightarrow G$  be any permutation which fixes  $e$ , and let  $x, y \in G$ . We will show, using a case-by-case analysis, that  $\phi(x)\phi(y) = \phi(xy)$ . If  $x = y$  then, using the first of the three rules,

$$\phi(x)\phi(y) = \phi(x)^2 = e = \phi(e) = \phi(x^2) = \phi(xy).$$

So suppose that  $x \neq y$ . If  $y = e$  then, by the second rule, we have

$$\phi(x)\phi(y) = \phi(x)\phi(e) = \phi(x)e = \phi(x) = \phi(xe) = \phi(xy).$$

A similar argument applies if  $x = e$ . Finally, if  $x$  and  $y$  are both non-identity elements then the third rule tells us that  $xy = z$ , where  $z$  is the third non-identity element. Furthermore, since  $\phi$  is bijective and  $\phi(e) = e$ , the elements  $\phi(x)$ ,  $\phi(y)$  and  $\phi(z)$  must be distinct nonidentity elements of  $G$ ; so the third rule also tells us that  $\phi(x)\phi(y) = \phi(z)$ . So  $\phi(x)\phi(y) = \phi(xy)$  in this case too, and so our claim is established. Since  $\phi(x)\phi(y) = \phi(xy)$  for all  $x, y \in G$ , it follows that  $\phi$  is a homomorphism, and hence an automorphism of  $G$ .

From our previous discussion of the cyclic group of order three, it can be seen that the cyclic group of order three also has the property that all permutations of it which fix the identity element are automorphisms, although in this case there are only two such permutations, as opposed to six in the case of the Klein group. It trivially holds also for the cyclic groups of orders 1 and 2, when the only permutation which fixes the identity element is the identity permutation. However, there are no other groups with this property.

Let  $G$  be the cyclic group of finite order  $n$ . Expressed in terms of some fixed generating element  $x$ , the distinct elements of  $G$  are  $e = x^0$  (the identity element),  $x, x^2, \dots, x^{n-2}$  and  $x^{n-1}$ . Subsequent powers of  $x$  give the same elements again:  $x^n = x^0$ ,  $x^{n+1} = x$ , and so on. In fact, if  $r$  and  $s$  are any integers, then  $x^r = x^s$  if and only if  $r - s$  is a multiple of  $n$ .

For each integer  $k$  there is a function  $\phi = \phi_k: G \rightarrow G$  such that  $\phi(x^r) = x^{kr}$  for all integers  $r$ . To prove this we must show that  $x^{kr} = x^{ks}$  whenever  $x^r = x^s$ ; otherwise  $\phi$  will not be uniquely defined. But if  $x^r = x^s$  then, as we remarked above,  $r - s$  must be a multiple of  $n$ . But if  $r - s$  is a multiple of  $n$  then  $kr - ks$ , which equals  $k(r - s)$ , must be a multiple of  $n$  also, and so  $x^{kr} = x^{ks}$ , as required.

Now if  $r$  and  $s$  are arbitrary integers we have that

$$\phi(x^r)\phi(x^s) = x^{kr}x^{ks} = x^{kr+ks} = x^{k(r+s)} = \phi(x^{r+s}) = \phi(x^r x^s).$$

Since all elements of  $G$  are powers of  $x$  we conclude that  $\phi(gh) = \phi(g)\phi(h)$  for all  $g, h \in G$ . So  $\phi$  is a homomorphism from  $G$  to  $G$ .<sup>†</sup> This does not mean that  $\phi$  is an automorphism: as yet we do not know whether or not  $\phi$  is bijective. To determine whether  $\phi$  is injective, we investigate the kernel of  $\phi$ .

Suppose that  $x^r \in \ker \phi$ . Then  $x^{kr} = \phi(x^r) = e$ , and so  $kr$  must be a multiple of  $n$ . Now let  $n = dm$  and  $k = dh$ , where  $d$  is the greatest common divisor of  $n$  and  $k$ . Note that  $m$  and  $h$  cannot have any factors greater than 1, since if they did we could find a common factor of  $n$  and  $k$  greater than  $d$ . Now since  $kr$  is a multiple of  $n$  it follows, after dividing through by  $d$ , that  $hr$  is a multiple of  $m$ . Since  $m$  and  $h$  have no nontrivial common factors it follows that  $r$  is a multiple of  $m$ . The converse is also true: if  $r$  is a multiple of  $m$  then  $x^r \in \ker \phi$ . So

$$\ker \phi = \{x^r \mid r \text{ is a multiple of } m = n/\gcd(n, k)\}.$$

We see therefore that if the gcd of  $n$  and  $k$  is greater than 1 then  $\ker \phi$  contains at least one element  $x^r$  such that  $r$  is not a multiple of  $n$ . So  $\ker \phi \neq \{e\}$  in this case, and hence (by Proposition 3.16)  $\phi$  is not injective. On the other hand if the gcd of  $n$  and  $k$  is 1 then

$$\ker \phi = \{x^r \mid r \text{ is a multiple of } n\} = \{e\},$$

and  $\phi$  is injective. It is easily seen that an injective transformation of a finite set is necessarily surjective also, and so we conclude that  $\phi$  is an automorphism of  $G$  if and only if the greatest common divisor of  $n$  and  $k$  is 1.

#### §4b Inner automorphisms

The groups whose automorphisms we discussed in §4a were all Abelian. This section, however, is primarily concerned with groups which are not Abelian, because the construction we present yields only the identity automorphism in the Abelian case.

---

<sup>†</sup> Homomorphisms from a group to itself are sometimes called *endomorphisms*.

**4.3 PROPOSITION** Let  $G$  be a group and  $g \in G$ , and define  $\alpha_g: G \rightarrow G$  by the rule  $\alpha_g(h) = ghg^{-1}$  for all  $h \in G$ . Then  $\alpha_g$  is an automorphism of  $G$

**Proof.** If  $h, k$  are arbitrary elements of  $G$  then

$$\alpha_g(h)\alpha_g(k) = (ghg^{-1})(gkg^{-1}) = ghekg^{-1} = g(hk)g^{-1} = \alpha_g(hk),$$

and so  $\alpha_g$  is a homomorphism.

Suppose that  $h, k \in G$  with  $\alpha_g(h) = \alpha_g(k)$ . Then  $ghg^{-1} = gkg^{-1}$ , and applying 2.2 (i) and 2.2 (ii) we conclude that  $h = k$ . Since  $\alpha_g(h) = \alpha_g(k)$  implies that  $h = k$ , it follows that  $\alpha_g$  is injective.

Let  $h \in G$  be arbitrary. Since

$$\alpha_g(g^{-1}hg) = g(g^{-1}hg)g^{-1} = (gg^{-1})h(gg^{-1}) = ehe = h$$

it follows that  $h \in \text{im } \alpha_g$ . Since  $h$  was arbitrary, this shows that  $\alpha_g$  is surjective. □

**4.4 DEFINITION** If  $G$  is a group, then an automorphism  $\xi$  of  $G$  is called an *inner automorphism* if there exists  $g \in G$  such that  $\xi(h) = ghg^{-1}$  for all  $h \in G$ .

The reason for the name “inner” is clear enough: inner automorphisms of  $G$  are produced, in some sense, by elements from within  $G$  itself.

**4.5 PROPOSITION** Let  $G$  be a group, and for each  $g \in G$  let  $\alpha_g \in \text{Aut}(G)$  be as defined in 4.3. The function  $\alpha: G \rightarrow \text{Aut}(G)$ , defined by  $\alpha(g) = \alpha_g$  for all  $g \in G$ , is a homomorphism.

**Proof.** We must show that  $\alpha(g)\alpha(h) = \alpha(gh)$ , for all  $g, h \in G$ . That is, we must show that  $\alpha_g\alpha_h = \alpha_{gh}$ , for all  $g, h \in G$ . Now since  $\alpha_g\alpha_h$  and  $\alpha_{gh}$  are functions from  $G$  to  $G$ , to say that they are equal is to say that they have the same effect on all elements of  $G$ .

So, let  $g, h$  and  $x$  be arbitrary elements of  $G$ . Since multiplication in  $\text{Aut}(G)$  is composition of functions,  $(\alpha_g\alpha_h)(x)$  is, by definition,  $\alpha_g(\alpha_h(x))$ . Therefore

$$(\alpha_g\alpha_h)(x) = \alpha_g(hxh^{-1}) = ghxh^{-1}g^{-1} = (gh)x(gh)^{-1} = \alpha_{gh}(x),$$

and since this is valid for all  $x \in G$  we conclude that the functions  $\alpha_g\alpha_h$  and  $\alpha_{gh}$  are equal, as required. □



In Chapter 3 we expended some effort proving the Homomorphism Theorem, which is applicable whenever one has a homomorphism from one group to another. In Proposition 4.5 we showed that a certain function is a homomorphism; we would be foolish not to apply the Homomorphism Theorem and see if it tells us anything interesting.

The Propositions 3.14 and 3.15 should really be regarded as part of the Homomorphism Theorem, since they have the same hypotheses as 3.17 and are needed for its statement. To apply these results to the homomorphism  $\alpha$  defined in 4.5, the first task is to calculate the kernel and the image of  $\alpha$ .

The kernel of  $\alpha$  is the set of all elements  $g \in G$  such that the function  $\alpha_g: G \rightarrow G$  is the identity function. That is,  $\ker \alpha$  is the set of all  $g \in G$  such that  $\alpha_g(h) = h$  for all  $h \in G$ .

4.6 DEFINITION The *centre* of a group  $G$  is the set

$$Z(G) = \{ g \in G \mid gh = hg \text{ for all } h \in G \}.$$

That is,  $Z(G)$  consists of those elements which commute with all elements of  $G$ .

Multiplying the equation  $gh = hg$  on the right by  $g^{-1}$  converts it into  $ghg^{-1} = h$ . Recalling that  $\alpha_g(h)$  was defined as  $ghg^{-1}$ , we conclude that the centre of  $G$  is the set of all  $g \in G$  such that  $\alpha_g(h) = h$  for all  $h \in G$ .

4.7 PROPOSITION The kernel of the homomorphism  $\alpha$  of Proposition 4.5 is  $Z(G)$ , the centre of  $G$ .

By Proposition 3.14 it follows that the centre is a normal subgroup.

4.8 PROPOSITION The centre of a group  $G$  is always a normal subgroup of  $G$ .

Note that Proposition 4.8 is also very easy to prove directly, using 2.10 and the definition of normality.

The image of the homomorphism  $\alpha$  is  $\text{Inn}(G) = \{ \alpha_g \mid g \in G \}$ , the set of all inner automorphisms of  $G$ .

4.9 PROPOSITION The set  $\text{Inn}(G)$  is a subgroup of  $\text{Aut}(G)$ .

This follows directly from 3.15, or, alternatively, can be proved readily using 2.10. We call  $\text{Inn}(G)$  the *inner automorphism group* of  $G$ .

4.10 PROPOSITION If  $G$  is any group, then the central quotient group,  $G/Z(G)$ , is isomorphic to the inner automorphism group of  $G$ .

**Proof.** This is immediate from 3.17. □

### §4c Conjugacy

4.11 DEFINITION Let  $G$  be a group and  $x, y \in G$ . Then  $x$  and  $y$  are said to be *conjugate* in  $G$  if there exists  $g \in G$  such that  $gxg^{-1} = y$ . In this situation we will sometimes say that  $g$  *transforms*  $x$  into  $y$ .

An alternative formulation of this concept is as follows:  $x, y \in G$  are conjugate if there exists  $\xi \in \text{Inn}(G)$  such that  $\xi(x) = y$ .

It is a straightforward exercise to show that conjugacy is an equivalence relation, which therefore partitions  $G$  into mutually disjoint equivalence classes. These are called the *conjugacy classes* of  $G$ .<sup>†</sup>

Observe that if  $G$  is an Abelian group then  $gxg^{-1} = y$  becomes  $x = y$ ; so, in an Abelian group,  $x$  cannot be conjugate to anything but itself. So the conjugacy classes of an Abelian group are just the single element subsets of the group. Accordingly, the first interesting example to look at is the smallest group which is not Abelian, namely, the symmetric group  $\text{Sym}\{1, 2, 3\}$ .

Note that  $(2, 3)(1, 2)(2, 3)^{-1} = (1, 3)$ , and  $(1, 2)(1, 3)(1, 2)^{-1} = (2, 3)$ . It follows that the elements  $(1, 2)$ ,  $(1, 3)$  and  $(2, 3)$  of  $G = \text{Sym}\{1, 2, 3\}$  are all conjugate to each other. Let  $p \in G$  be arbitrary, and let  $p(1) = a$  and  $p(2) = b$ . Then

$$(p(1, 2)p^{-1})(a) = (p(1, 2))(p^{-1}(a)) = (p(1, 2))(1) = p(2) = b,$$

and similarly it can be checked that  $p(1, 2)p^{-1}$  takes  $b$  to  $a$ . Since  $a$  and  $b$  are interchanged by  $p(1, 2)p^{-1}$ , the remaining element of  $\{1, 2, 3\}$  must be fixed.

---

<sup>†</sup> In fact, in the literature, the conjugacy classes of a group are usually just called the *classes* of the group.

Hence  $p(1, 2)p^{-1} = (a, b)$ . It follows that  $(1, 2)$ ,  $(1, 3)$  and  $(2, 3)$  are the only elements of this conjugacy class. By similar reasoning it can be verified that  $(1, 2, 3)$  and  $(1, 3, 2)$  (which equals  $(1, 2)(1, 2, 3)(1, 2)$ ) are the only elements of  $G$  conjugate to  $(1, 2, 3)$ . The remaining element of  $G$  is  $i$ , the identity, which must form a class by itself. Indeed, it is obvious that  $i$  cannot be conjugate to anything but itself, since  $pip^{-1} = i$  for all  $p$ . So the partitioning of  $G$  into conjugacy classes is as follows:

$$G = \{i\} \cup \{(1, 2), (1, 3), (2, 3)\} \cup \{(1, 2, 3), (1, 3, 2)\}.$$

Next we investigate conjugacy in  $\text{Sym}\{1, 2, \dots, n\}$ , for arbitrary  $n$ . The key observation is that if  $q$  is an element of  $\text{Sym}\{1, 2, \dots, n\}$  that takes  $a$  to  $b$ , and  $p \in \text{Sym}\{1, 2, \dots, n\}$  is arbitrary, then  $pqp^{-1}$  takes  $p(a)$  to  $p(b)$ . For, given  $q(a) = b$ , we find that

$$(pqp^{-1})(p(a)) = (pq)(p^{-1}(p(a))) = (pq)(a) = p(q(a)) = p(b).$$

Thus, for example, if  $q$  is the cycle  $(1, 4, 2, 3, 5)$  then  $pqp^{-1}$  is the cycle  $(p(1), p(4), p(2), p(3), p(5))$ , for since  $q$  takes 1 to 4 it follows that  $pqp^{-1}$  takes  $p(1)$  to  $p(4)$ , and since  $q$  takes 4 to 2 it follows that  $pqp^{-1}$  takes  $p(4)$  to  $p(2)$ , and so on. By choosing  $p$  appropriately, we can arrange for  $(p(1), p(4), p(2), p(3), p(5))$  to be any given 5-cycle. For example, if we want  $pqp^{-1}$  to be  $(5, 4, 3, 2, 1)$ , then  $p = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 3 & 2 & 4 & 1 \end{bmatrix}$  will do. Similarly, if  $q = (2, 6)(3, 4)$  then  $pqp^{-1} = (p(2), p(6))(p(3), p(4))$ , and by suitable choice of  $p$  we can make  $pqp^{-1}$  equal any product of two disjoint transpositions. In general, two elements of  $\text{Sym}\{1, 2, \dots, n\}$  are conjugate if and only if they are of the same cycle type. That is, when written as products of disjoint cycles, they have the same number of cycles of length 1, the same number of length 2, the same number of length 3, and so on. For example, in  $\text{Sym}\{1, 2, \dots, 24\}$  a permutation might have three cycles of length 1, one of length two, four of length three and one of length seven. Such a permutation is

$$q = (3, 7)(8, 21, 17)(11, 12, 13)(16, 6, 22)(20, 24, 23)(4, 9, 19, 10, 5, 18, 15),$$

and another is

$$r = (11, 12)(1, 2, 3)(7, 15, 14)(13, 16, 17)(19, 22, 20)(5, 6, 21, 24, 23, 10, 9).$$

Because  $q$  and  $r$  have the same cycle type, it follows that they must be conjugate in  $\text{Sym}\{1, 2, \dots, 24\}$ . Here is a permutation  $p$  such that  $pqp^{-1} = r$ :

$$\begin{bmatrix} 3 & 7 & 8 & 21 & 17 & 11 & 12 & 13 & 16 & 6 & 22 & 20 & 24 & 23 & 4 & 9 & 19 & 10 & 5 & 18 & 15 & 1 & 2 & 14 \\ 11 & 12 & 1 & 2 & 3 & 7 & 15 & 14 & 13 & 16 & 17 & 19 & 22 & 20 & 5 & 6 & 21 & 24 & 23 & 10 & 9 & 4 & 8 & 18 \end{bmatrix}$$

Arranging the numbers in the top row into increasing order (which is more usual) this becomes:

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 & 21 & 22 & 23 & 24 \\ 4 & 8 & 11 & 5 & 23 & 16 & 12 & 1 & 6 & 24 & 7 & 15 & 14 & 18 & 9 & 13 & 3 & 10 & 21 & 19 & 2 & 17 & 20 & 22 \end{bmatrix}$$

In  $\text{Sym}\{1, 2, 3, 4, 5\}$ , which has 120 elements, there are exactly seven possible cycle types, and so seven classes, as follows.

- The identity element is a one element conjugacy class.
- The elements conjugate to  $(1, 2)$  form ten element class.
- The elements conjugate to  $(1, 2, 3)$  form a twenty element class.
- The elements conjugate to  $(1, 2, 3, 4)$  form a thirty element class.
- The elements conjugate to  $(1, 2, 3, 4, 5)$  form a twenty-four element class.
- The elements conjugate to  $(1, 2)(3, 4)$  form a fifteen element class.
- The elements conjugate to  $(1, 2)(3, 4, 5)$  form a twenty element class.

Add these up and check that it comes to 120. Also, verify the numbers themselves!

For groups of symmetries of geometrical objects, elements which are conjugate invariably turn out to be geometrically similar kinds of transformations. For example, let  $G$  be the group of all symmetries of a square with vertices  $a, b, c$  and  $d$  (see Chapter 1). The clockwise rotation through  $90^\circ$  and the anticlockwise rotation through  $90^\circ$ , corresponding to the permutations  $(a, b, c, d)$  and  $(a, d, c, b)$ , are conjugate in  $G$ ; indeed, the element  $(a, c) \in G$  transforms one to the other. The element  $(a, b)(c, d)$  is a reflection in a line parallel to, and midway between, a pair of opposite sides. The element  $(a, d)(b, c)$  is also such a reflection. These elements are conjugate in  $G$ ; the element  $(a, b, c, d)$  transforms one to the other. The elements  $(a, c)$  and  $(b, d)$  are the reflections in diagonals of the square. They are conjugate in  $G$ , and  $(a, b, c, d)$  transforms one to the other. The other two elements of  $G$ , the identity and the central inversion  $(a, c)(b, d)$ , both form one element classes. So the class decomposition of  $G$  is

$$G = \{i\} \cup \{(a, c)(b, d)\} \cup \{(a, b)(c, d), (a, d)(b, c)\} \\ \cup \{(a, c), (b, d)\} \cup \{(a, b, c, d), (a, d, c, b)\}.$$

Let  $G$  be the group of all invertible  $3 \times 3$  matrices over the complex field. That is,  $G = \text{GL}_3(\mathbb{C})$ . It is proved in textbooks on linear algebra that if a matrix  $A \in G$  has three distinct eigenvalues,  $\alpha$ ,  $\beta$  and  $\gamma$  say, then  $A$  is similar (to use the linear algebra term) to the matrix

$$D = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{bmatrix}.$$

That is, there exists an invertible matrix  $T$  such that  $T^{-1}AT = D$ . In group theory terminology, the elements  $A, D \in G$  are conjugate. If the matrix  $A$  has a repeated eigenvalue the story is not quite so simple, but nevertheless it is fully understood: every matrix is similar to a *Jordan normal form* matrix. This classical fact from linear algebra describes the conjugacy classes in the group  $\text{GL}_3(\mathbb{C})$  (and, indeed, the theory applies to  $\text{GL}_n(\mathbb{C})$  for all  $n$ ).

#### §4d On the number elements in a conjugacy class

Let  $g$  be a fixed element of the group  $G$ . Since the elements of  $G$  which are conjugate to  $g$  are exactly the elements of the form  $tgt^{-1}$ , there is a function  $f$  from  $G$  to the conjugacy class containing  $g$ , given by  $t \mapsto tgt^{-1}$ . In keeping with the idea used in Proposition 2.20, let us define an equivalence relation on  $G$  by the rule that  $s \equiv t$  if and only if  $f(s) = f(t)$ . The different conjugates of  $g$  in  $G$  are then in one-to-one correspondence with the equivalence classes of  $G$  under this equivalence relation.

4.12 PROPOSITION *Let  $G$  be a group and  $g \in G$ . The set of all elements of  $G$  which commute with  $g$ ,*

$$C_G(g) = \{ t \in G \mid gt = tg \},$$

*is a subgroup of  $G$ .*

**Proof.** Since  $ge = eg$  (where  $e$  is the identity element of  $G$ ) we see that  $e \in C_G(g)$ , and so  $C_G(g)$  is nonempty.

Let  $s, t \in C_G(g)$ . Then  $gt = tg$  and  $gs = sg$ , and so

$$g(st) = (gs)t = (sg)t = s(gt) = s(tg) = (st)g,$$

and it follows that  $st \in C_G(g)$ . Hence  $C_G(g)$  is closed under multiplication.

Let  $t \in C_G(g)$ . Then  $gt = tg$ , and multiplying this equation on the left by  $t^{-1}$ , and on the right by  $t^{-1}$  as well, gives  $t^{-1}(gt)t^{-1} = t^{-1}(tg)t^{-1}$ , which simplifies to  $gt^{-1} = t^{-1}g$ . So  $t^{-1} \in C_G(g)$ . Hence  $C_G(g)$  is also closed under inversion, and by Proposition 2.10 it follows that  $C_G(g)$  is a subgroup of  $G$ .  $\square$

**4.13 DEFINITION** The subgroup  $C_G(g)$  defined in Proposition 4.12 is called the *centralizer* of  $g$  in  $G$ .

Since the equation  $gt = tg$  is equivalent to  $tgt^{-1} = g$ , the centralizer of  $g$  can also be described as the set of elements of  $G$  that transform  $g$  to itself. Note also that  $t$  is in the centralizer of  $g$  if and only if  $g$  is in the centralizer of  $t$ .

The relevance of the centralizer in the present context is clarified by the following result.

**4.14 PROPOSITION** Let  $G$  be a group and  $g \in G$ . If  $s, t$  are arbitrary elements of  $G$  then  $sgs^{-1} = tgt^{-1}$  if and only if  $sC_G(g) = tC_G(g)$ . That is,  $s$  and  $t$  transform  $g$  to the same element if and only if they lie in the same left coset of the centralizer of  $g$ .

**Proof.** The equation  $sgs^{-1} = tgt^{-1}$  is equivalent to  $t^{-1}sg = gt^{-1}s$ , which says that  $t^{-1}s \in C_G(g)$ . By 2.11, this is equivalent to  $sC_G(g) = tC_G(g)$ , as required.  $\square$

At the beginning of this section we defined an equivalence relation on  $G$  by the rule that  $s \equiv t$  if and only if  $sgs^{-1} = tgt^{-1}$ ; Proposition 4.14 shows that this equivalence relation is exactly right congruence modulo the centralizer of  $g$ , so that the equivalence classes are just the left cosets of the centralizer. Because the conjugates of  $g$  are in one-to-one correspondence with these equivalence classes, it follows that the number of conjugates of  $g$  in  $G$  equals the index in  $G$  of the centralizer. The next proposition summarizes what we have proved.

**4.15 PROPOSITION** Let  $G$  be a group and  $g \in G$ . Then the number of conjugates of  $g$  in  $G$  is equal to  $[G : C_G(g)]$ . Moreover, there is a bijective

correspondence between the conjugates of  $g$  and the left cosets of  $C_G(g)$  such that  $tC_G(g) \leftrightarrow tgt^{-1}$  for all  $t \in G$ .

An example will clarify the idea a little. Let  $G = \text{Sym}\{1, 2, 3, 4\}$  and let  $g = (1, 3)(2, 4)$ . For each  $t \in G$  we have  $tgt^{-1} = (t(1), t(3))(t(2), t(4))$ . Let us first calculate  $C_G(g)$ , which is the set of those  $t$  such that  $tgt^{-1} = g$ . There are in fact eight such elements  $t$ ; they are as follows:

- (i)  $t(1) = 1, t(3) = 3, t(2) = 2, t(4) = 4$  (the identity element);
- (ii)  $t(1) = 1, t(3) = 3, t(2) = 4, t(4) = 2$ ;
- (iii)  $t(1) = 3, t(3) = 1, t(2) = 2, t(4) = 4$ ;
- (iv)  $t(1) = 3, t(3) = 1, t(2) = 4, t(4) = 2$ ;
- (v)  $t(1) = 2, t(3) = 4, t(2) = 1, t(4) = 3$ ;
- (vi)  $t(1) = 2, t(3) = 4, t(2) = 3, t(4) = 1$ ;
- (vii)  $t(1) = 4, t(3) = 2, t(2) = 1, t(4) = 3$ ;
- (viii)  $t(1) = 4, t(3) = 2, t(2) = 3, t(4) = 1$ .

Writing them out in cycle notation, they are

$$i, (2, 4), (1, 3), (1, 3)(2, 4), (1, 2)(3, 4), (1, 2, 3, 4), (1, 4, 3, 2), (1, 4)(2, 3).$$

Multiplying all these elements on the left by  $(1, 2)$  we find that the eight elements of the left coset  $(1, 2)C_G(g)$  are

$$(1, 2), (1, 2, 4), (1, 3, 2), (1, 3, 2, 4), (3, 4), (2, 3, 4), (1, 4, 3), (1, 4, 2, 3).$$

For these eight values of  $t$  we have that  $tgt^{-1} = (1, 2)g(1, 2)^{-1} = (2, 3)(1, 4)$ . In a similar fashion we find that the eight elements of the left coset  $(1, 4)C_G(g)$  are

$$(1, 4), (1, 4, 2), (1, 3, 4), (1, 3, 4, 2), (1, 2, 4, 3), (1, 2, 3), (2, 4, 3), (2, 3).$$

These eight values of  $t$  all give  $tgt^{-1} = (1, 4)g(1, 4)^{-1} = (1, 2)(3, 4)$ . The three elements which have the same cycle type as  $(1, 3)(2, 4)$  have been matched with the left cosets of  $C_G(g)$ , in the way specified in Proposition 4.15.

Let  $G$  be a finite group and let  $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_s$  be all the conjugacy classes of  $G$ . Since conjugacy is an equivalence relation on  $G$  we know that the classes  $\mathcal{C}_i$  form a partitioning of  $G$ : each element of  $G$  lies in exactly one

class. In particular, therefore, the total number of elements of  $G$  equals the sum of the sizes of the classes:

$$|G| = \sum_{i=1}^s |\mathcal{C}_i|.$$

For each class  $\mathcal{C}_i$ , choose a representative element  $g_i \in \mathcal{C}_i$ . By Proposition 4.15,  $|\mathcal{C}_i| = [G : C_G(g_i)]$ , and so the equation above can be rewritten as

$$|G| = \sum_{i=1}^s [G : C_G(g_i)].$$

This is called the *class equation* of the group  $G$ .

It is useful to collect together the terms  $[G : C_G(g_i)]$  in the class equation which are equal to 1. Now  $[G : C_G(g_i)] = 1$  if and only if the subgroup  $C_G(g_i)$  is the whole of  $G$  (so that  $C_G(g_i)$  itself is the one and only coset of  $C_G(g_i)$  in  $G$ ), and this means that every  $t \in G$  commutes with  $g_i$ . But  $g_i t = t g_i$  for all  $t \in G$  if and only if  $g_i \in Z(G)$ . In other words, the single element conjugacy classes of  $G$  correspond to the elements of the centre of  $G$ . The class equation now becomes

$$|G| = |Z(G)| + \sum_{i=r}^s [G : C_G(g_i)],$$

where  $g_r, g_{r+1}, \dots, g_s$  are representatives of the the conjugacy classes of elements of  $G$  that lie outside the centre of  $G$ .

**4.16 PROPOSITION** *Let  $G$  be a group such that  $|G| = p^n$ , where  $p$  is a prime number and  $n > 0$ . Then the centre of  $G$  contains at least one non-identity element.*

**Proof.** Since  $|G| = p^n$  and  $p$  is prime, every divisor of  $p^n$  is also a power of  $p$ . We know that the index of any subgroup of  $G$  is necessarily a divisor of  $|G|$ , and so each of the terms  $[G : C_G(g_i)]$  appearing in the class equation is a power of  $p$ . Now we have

$$|Z(G)| = |G| - \sum_{i=r}^s [G : C_G(g_i)],$$

and all the terms on the right hand side are powers of  $p$  greater than 1. So all the terms on the right hand side of the equation are divisible by  $p$ , and it follows that  $|Z(G)|$  is divisible by  $p$  also. In particular,  $|Z(G)| \neq 1$ , whence  $Z(G)$  has more elements than just the identity.  $\square$



# 5

## Reflections

At this point we will abruptly abandon abstract group theory, although we have not even scratched the surface of that enormous subject. Instead, we turn our attention to  $n$ -dimensional Euclidean geometry.

There is an extremely naive viewpoint according to which, since we live in three-dimensional space, higher dimensional geometry has no real application, and is just some totally useless nonsense invented by mathematicians to amuse themselves as they sit in their ivory tower. The truth is that the mathematical analysis of very simple, concrete, physical problems immediately and inevitably involves higher dimensional spaces. To describe the position of a particle in three dimensional space requires three coordinates, but if you need to worry about the positions of two particles simultaneously you need six. Immediately you are working in six dimensional space. True, what mathematicians call six dimensional geometry, the person in the street would not recognise as geometry. It becomes hard to draw pictures to illustrate the arguments that are used, which therefore appear more like algebra than geometry. Indeed, there is no clear distinction between algebra and geometry. But, as we shall see, concepts from two and three dimensional geometry do have natural generalizations in higher dimensions, and it is natural to use the term “geometry” to apply to reasoning involving such concepts.

### §5a Inner product spaces

Let  $\mathbb{R}^n$  be the set of all  $n$ -component column vectors:

$$\mathbb{R}^n = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \mid x_i \in \mathbb{R} \text{ for all } i \right\}.$$

If  $\underline{x}, \underline{y} \in \mathbb{R}^n$  we define the *dot product* of  $\underline{x}$  and  $\underline{y}$  by the formula

$$\underline{x} \cdot \underline{y} = \sum_{i=1}^n x_i y_i$$

where  $x_i, y_i$  are the  $i$ th coordinates of  $\underline{x}, \underline{y}$  respectively. Note that  $\underline{x} \cdot \underline{x} \geq 0$  for all  $\underline{x}$ , with equality if and only if  $\underline{x} = \underline{0}$ . We define

$$\|\underline{x}\| = \sqrt{\underline{x} \cdot \underline{x}}.$$

When  $n = 2$  or  $3$  we can use Cartesian coordinates to identify  $\mathbb{R}^n$  with the space of position vectors of points, in two or three dimensional Euclidean space, relative to fixed origin  $O$ . It turns out that in these cases—provided we use a rectangular coordinate system—the formula

$$\text{distance} = \sqrt{(\underline{x} - \underline{y}) \cdot (\underline{x} - \underline{y})} = \|\underline{x} - \underline{y}\|$$

gives the distance between the points  $P$  and  $Q$  whose position vectors are  $\overrightarrow{OP} = \underline{x}$  and  $\overrightarrow{OQ} = \underline{y}$ . So it is natural to use this as the definition of the distance between  $\underline{x}$  and  $\underline{y}$  whatever the value of  $n$ . Similarly, if  $\theta = \angle POQ$  then

$$\cos \theta = \frac{\underline{x} \cdot \underline{y}}{\|\underline{x}\| \|\underline{y}\|},$$

and so we use this same formula to define the angle between  $\underline{x}$  and  $\underline{y}$  in general.

With the dot product as defined above,  $\mathbb{R}^n$  becomes an inner product space (as defined in Chapter 1). Furthermore, if  $V$  is any inner product space over  $\mathbb{R}$ , and if the dimension of  $V$  is  $n$ , then a basis  $v_1, v_2, \dots, v_n$  of  $V$  can be found which is *orthonormal*, in the sense that

$$v_i \cdot v_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j, \end{cases}$$

and then there is an inner product preserving vector space isomorphism between  $V$  and  $\mathbb{R}^n$  given by

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n \leftrightarrow \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix}.$$

So, effectively,  $\mathbb{R}^n$  is the only  $n$ -dimensional inner product space over  $\mathbb{R}$ . The corollary of this fact which will be important for us later is the following proposition.

**5.1 PROPOSITION** If  $v_1, v_2, \dots, v_n$  are elements of a real inner product space  $V$ , then there exist  $x_1, x_2, \dots, x_n$  in  $\mathbb{R}^n$  with  $x_i \cdot x_j = v_i \cdot v_j$  for all  $i, j$ . In particular, if  $\theta_{ij}$  is the angle between  $x_i$  and  $x_j$ , then  $\cos \theta_{ij} = \frac{v_i \cdot v_j}{\|v_i\| \|v_j\|}$ .

We will also apply the concepts of distance and angle to arbitrary real inner product spaces, defining them by the same formulas as for  $\mathbb{R}^n$ .

It will sometimes be necessary for us to deal with spaces which are almost inner product spaces, but not quite.

**5.2 DEFINITION** Let  $V$  be a vector space over  $\mathbb{R}$ . A *bilinear form* on  $V$  is a function  $f: V \times V \rightarrow \mathbb{R}$  such that

- (i)  $f(\lambda v + \mu u, w) = \lambda f(v, w) + \mu f(u, w)$  for all  $u, v, w \in V$  and  $\lambda, \mu \in \mathbb{R}$ ,
  - (ii)  $f(w, \lambda v + \mu u) = \lambda f(w, v) + \mu f(w, u)$  for all  $u, v, w \in V$  and  $\lambda, \mu \in \mathbb{R}$ .
- The bilinear form  $f$  is said to be *symmetric* if in addition
- (iii)  $f(u, v) = f(v, u)$  for all  $u, v \in V$ .

Observe that a symmetric bilinear form is an inner product if it satisfies the extra condition that  $f(v, v) > 0$  for all nonzero  $v \in V$ .

**5.3 PROPOSITION** Let  $v_1, v_2, \dots, v_n$  be a basis of a vector space  $V$  over the field  $\mathbb{R}$ , and let  $A = (a_{ij})$  be an arbitrary  $n \times n$  matrix over  $\mathbb{R}$ . Then there exists a bilinear form  $f$  on  $V$  such that  $f(v_i, v_j) = a_{ij}$  for all  $i, j$ . The form  $f$  is symmetric if  $A$  is a symmetric matrix.

**Proof.** Every element of  $V$  can be written uniquely as a linear combination of the basis vectors  $v_1, v_2, \dots, v_n$ . If  $v = \sum_{i=1}^n \lambda_i v_i$  and  $u = \sum_{i=1}^n \mu_i v_i$ , define

$$f(v, u) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i a_{ij} \mu_j.$$

It is straightforward to check that  $f$  has the required properties. □

### §5b Dihedral groups

We look first at two dimensional space, and examine the group of symmetries of a regular  $n$ -sided polygon. Number its vertices  $1, 2, \dots, n$ , where  $1$  is adjacent to  $2$ ,  $2$  adjacent to  $3$ , and so on, and  $n$  adjacent to  $1$ . Let  $O$  be the centre of the polygon; the perpendicular bisectors of all the sides, and the bisectors of all the angles, all pass through  $O$ . These lines are all axes

of symmetry of the polygon. Note that  $n$  bisectors of sides plus  $n$  bisectors of angles only makes  $n$  axes of symmetry altogether, since each is counted twice. There is a difference between the case of even  $n$ , when the bisector of an angle coincides with the bisector of the opposite angle, and similarly for sides, and odd  $n$ , where the bisector of an angle is also the bisector of the opposite side. The cases  $n = 5$  and  $n = 6$  are illustrated.



An “axis of symmetry” is a line in which the object can be reflected without being changed. Given a line  $\ell$  in the plane, the reflection in  $\ell$  is the transformation of the plane which takes each point  $P$  to its mirror image, which is that point  $P'$  such that  $\ell$  is the perpendicular bisector of the line segment  $PP'$ . Expressed as a permutation of the vertices, the reflection in the line which is the perpendicular bisector of the side joining vertices 1 and  $n$  is

$$\begin{aligned} r &= (1, n)(2, n-1)(3, n-2) \cdots \\ &= \begin{cases} (1, 2k)(2, 2k-1) \cdots (k, k+1) & \text{if } n = 2k \text{ is even,} \\ (1, 2k+1)(2, 2k) \cdots (k, k+2) & \text{if } n = 2k+1 \text{ is odd.} \end{cases} \end{aligned}$$

Similarly, the reflection in the bisector of the angle at vertex  $n$  is

$$\begin{aligned} s &= (1, n-1)(2, n-2)(3, n-3) \cdots \\ &= \begin{cases} (1, 2k-1)(2, 2k-2) \cdots (k-1, k+1) & \text{if } n = 2k \text{ is even,} \\ (1, 2k)(2, 2k-1) \cdots (k, k+1) & \text{if } n = 2k+1 \text{ is odd.} \end{cases} \end{aligned}$$

The product of these reflections is

$$rs = (1, 2, 3, \dots, n)$$

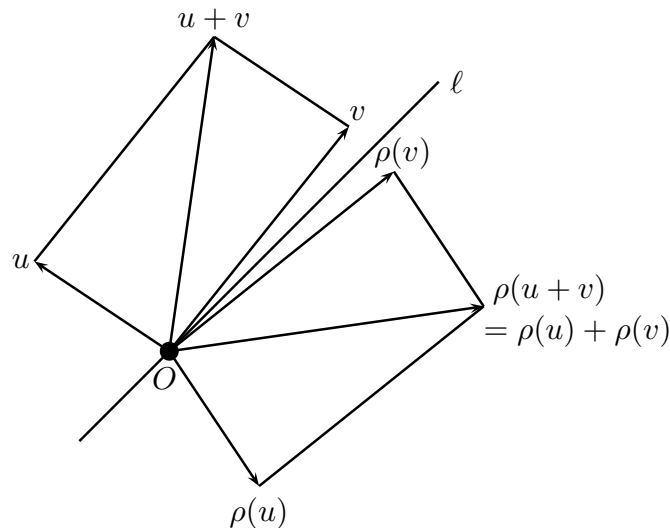
which is a rotation about  $O$  through an angle of  $\pi/n$ . The powers of  $rs$  give  $n$  rotational symmetries of the polygon, which, when combined with the  $n$  reflection symmetries corresponding to the  $n$  axes of symmetry mentioned above, give  $2n$  symmetries altogether. These  $2n$  symmetries constitute the full symmetry group of the regular  $n$ -sided polygon; the group is called the *dihedral group* of order  $2n$ .

Note that  $(rs)^n = i$ . That is,  $rsr \cdots rs = i$ , where there are  $2n$  factors altogether on the left hand side. Multiplying this equation on the right by the alternating product  $srs \cdots$  to  $n$  factors gives us the relation  $rsr \cdots = srs \cdots$ , where there are  $n$  factors on each side. (For example, if  $n = 3$  the relation  $(rs)^3 = i$  gives  $srs = (rsr srs)(srs) = rsr$ .) If  $1 \leq k < n$  then the two alternating products of length  $k$ , namely  $rsr \cdots$  and  $srs \cdots$ , correspond to two distinct elements of the dihedral group. There are  $n - 1$  possible values for  $k$ , giving  $2(n - 1)$  elements, and together with the identity and the one element of length  $n$  this accounts for all  $2n$  elements of the group. Thus, expressed in terms of the reflections  $r$  and  $s$ , the elements of the dihedral group of order  $2n$  are

$$i, r, s, rs, sr, rsr, srs, rsts, srst, \dots \dots, rsr \cdots = srs \cdots,$$

where the expressions are arranged in order of nondecreasing length, the longest expression having length  $n$ .

If  $\ell$  is a line through the origin  $O$ , and if  $\rho$  is the reflection in  $\ell$ , then it is easily seen that  $\rho$  is a linear transformation of the vector space of position vectors relative to  $O$ . Since addition in this space is governed by the parallelogram law, the following diagram illustrates the fact that  $\rho$  preserves addition.



A similar kind of diagram can be drawn to illustrate preservation of scalar multiplication.

Since the reflection symmetries of the regular polygon discussed above all correspond to lines through the centre of the polygon, if we choose the centre as our origin of coordinates then the reflections in question are all linear transformations. Since the product of two linear transformations is again linear, all the elements of the dihedral group correspond to linear transformations of the plane. Our aim in the remainder of these notes will be to classify all finite groups of linear transformations of Euclidean space which are generated by reflections. The dihedral groups are the simplest examples, and they are—in several ways—of fundamental importance in the study of the more complicated examples.

### §5c Higher dimensions

If  $\rho$  is a reflection of  $n$ -dimensional space then, as in the 2-dimensional case described above, if  $P$  is any point then  $\rho(\overrightarrow{OP}) = \overrightarrow{OP'}$ , where the mirror image point  $P'$  has the property that the mirror perpendicularly bisects the line segment  $PP'$ . We have assumed that the origin  $O$  lies on the surface of the mirror, in order to ensure that  $\rho$  is a linear transformation. In the 2-dimensional case the mirror is a line, in the 3-dimensional case it is a plane, and in  $n$ -dimensions it is an  $(n - 1)$ -dimensional subspace. This means that there is a unique direction perpendicular to the surface of the mirror.† An  $(n - 1)$ -dimensional subspace of  $n$ -dimensional space is usually called a *hyperplane*.

Let  $H$  be a hyperplane and  $\underline{a}$  be a nonzero vector perpendicular to  $H$ . Points on  $H$  correspond to vectors  $\underline{u}$  which are perpendicular to  $\underline{a}$ . Expressing this in terms of the dot product,  $\underline{u} \in H$  if and only if  $\underline{u} \cdot \underline{a} = 0$ . Now if  $P$  is the point with position vector  $\overrightarrow{OP} = \underline{a}$ , and if  $P'$  is the mirror image of  $P$ , then  $\overrightarrow{OP'} = -\overrightarrow{OP}$ . That is,  $\rho(\underline{a}) = -\underline{a}$ . If  $\underline{u}$  is the position vector of a point on the surface of the mirror, then  $\rho(\underline{u}) = \underline{u}$ . We can now easily obtain a general formula for  $\rho(\underline{v})$  for all  $\underline{v} \in \mathbb{R}^n$ .

**5.4 PROPOSITION** *Let  $\underline{a}$  be a nonzero vector in  $\mathbb{R}^n$  and let  $H$  be the hyperplane of points perpendicular to  $\underline{a}$ . If  $\rho$  is the reflection in  $H$ , then*

$$\rho(\underline{v}) = \underline{v} - \frac{2(\underline{v} \cdot \underline{a})}{\underline{a} \cdot \underline{a}} \underline{a}$$

for all  $\underline{v} \in \mathbb{R}^n$ .

---

† Or, rather, a unique pair of mutually opposite directions.

**Proof.** Let  $v \in \mathbb{R}^n$ , and let  $\lambda = (v \cdot a)/(a \cdot a)$ . Bilinearity of the dot product gives

$$\begin{aligned}(v - \lambda a) \cdot a &= v \cdot a - \lambda(a \cdot a) \\ &= v \cdot a - \frac{v \cdot a}{a \cdot a} a \cdot a \\ &= v \cdot a - v \cdot a \\ &= 0,\end{aligned}$$

and therefore  $v - \lambda a \in H$ . Hence  $\rho(v - \lambda a) = v - \lambda a$ , and it follows that

$$\begin{aligned}\rho(v) &= \rho(v - \lambda a) + \rho(\lambda a) \\ &= v - \lambda a + \lambda \rho(a) \\ &= v - \lambda a + \lambda(-a) \\ &= v - 2\lambda a\end{aligned}$$

as claimed. □

Proposition 5.4 was stated in terms of  $\mathbb{R}^n$ , but it is natural to extend the terminology to all real inner product spaces.

**5.5 DEFINITION** If  $V$  is an arbitrary real inner product space and if  $a \in V$  is nonzero, then the transformation  $\rho_a: V \rightarrow V$  defined by

$$\rho_a(v) = v - 2 \frac{v \cdot a}{a \cdot a} a$$

is called the *reflection in the hyperplane orthogonal to  $a$* .

The proof of the following easy fact is left as an exercise for the reader.

**5.6 PROPOSITION** If  $a, b$  are nonzero elements of the inner product space  $V$ , then  $\rho_a = \rho_b$  if and only if  $a$  is a scalar multiple of  $b$ . In particular,  $\rho_{-a} = \rho_a$ .

Recall that an orthogonal transformation of an inner product space is a linear transformation which preserves the inner product. Geometrically, this means that distances and angles are preserved, since distance and angle are defined in terms of the inner product. It is straightforward to show that reflections are orthogonal transformations.

**5.7 PROPOSITION** Let  $V$  be a real inner product space and  $0 \neq a \in V$ . Then  $\rho_a(u) \cdot \rho_a(v) = u \cdot v$ , for all  $u, v \in V$ .

**Proof.** Given  $u, v \in V$ , put  $\lambda = (v \cdot a)/(a \cdot a)$  and  $\mu = (u \cdot a)/(a \cdot a)$ . Then

$$\begin{aligned} \rho_a(u) \cdot \rho_a(v) &= (u - 2\mu a) \cdot (v - 2\lambda a) \\ &= u \cdot (v - 2\lambda a) - 2\mu(a \cdot (v - 2\lambda a)) \\ &= u \cdot v - 2\lambda(u \cdot a) - 2\mu(a \cdot v) + 4\mu\lambda(a \cdot a) \\ &= u \cdot v - 2\lambda(\mu(a \cdot a)) - 2\mu(\lambda(a \cdot a) + 4\lambda\mu(a \cdot a)) \\ &= u \cdot v. \end{aligned}$$

□

It is also geometrically clear that reflections do, in fact, preserve distances and angles.

**5.8 PROPOSITION** Let  $\sigma$  be an orthogonal transformation of the inner product space  $V$ , and let  $0 \neq a \in V$ . Then  $\sigma\rho_a\sigma^{-1} = \rho_{\sigma(a)}$ .

**Proof.** Let  $v \in V$  be arbitrary, and let  $x = \sigma^{-1}(v)$ . Then we have

$$\begin{aligned} (\sigma\rho_a\sigma^{-1})(v) &= \sigma(\rho_a(\sigma^{-1}(v))) \\ &= \sigma(\rho_a(x)) \\ &= \sigma\left(x - 2\frac{x \cdot a}{a \cdot a}a\right) \quad (\text{by definition of } \rho_a) \\ &= \sigma(x) - 2\frac{x \cdot a}{a \cdot a}\sigma(a) \quad (\text{as } \sigma \text{ is linear}) \\ &= \sigma(x) - 2\frac{\sigma(x) \cdot \sigma(a)}{\sigma(a) \cdot \sigma(a)}\sigma(a) \quad (\text{as } \sigma \text{ preserves the dot product}) \\ &= v - 2\frac{v \cdot \sigma(a)}{\sigma(a) \cdot \sigma(a)}\sigma(a) \quad (\text{as } x = \sigma^{-1}(v) \text{ gives } v = \sigma(x)) \\ &= \rho_{\sigma(a)}(v) \quad (\text{by definition of } \rho_{\sigma(a)}) \end{aligned}$$

Since  $\sigma\rho_a\sigma^{-1}$  and  $\rho_{\sigma(a)}$  are transformations of  $V$  which have the same effect on all elements of  $V$ , it follows that they are equal. □



Our next task is to investigate the product of two reflections. In the case of the Euclidean plane, geometrical methods can be used to show that the product of two reflections is a rotation. Alternatively, we can use linear algebra to derive the same result, as follows. Let

$$\underline{a} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} r \cos \theta \\ r \sin \theta \end{bmatrix},$$

the position vector of an arbitrary point in the plane. Assume that  $\underline{a} \neq \underline{0}$ , and let  $\rho$  be the reflection in the hyperplane orthogonal to  $\underline{a}$ . (Of course, in this situation, a hyperplane is simply a line.) Using the formula in 5.5 we find that

$$\begin{aligned} \rho \begin{bmatrix} 1 \\ 0 \end{bmatrix} &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 2 \frac{\begin{pmatrix} 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} r \cos \theta \\ r \sin \theta \end{pmatrix}}{\begin{pmatrix} r \cos \theta \\ r \sin \theta \end{pmatrix} \cdot \begin{pmatrix} r \cos \theta \\ r \sin \theta \end{pmatrix}} \begin{bmatrix} r \cos \theta \\ r \sin \theta \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} - (2r \cos \theta / r^2) \begin{bmatrix} r \cos \theta \\ r \sin \theta \end{bmatrix} \\ &= \begin{bmatrix} 1 - 2 \cos^2 \theta \\ -2 \cos \theta \sin \theta \end{bmatrix} \\ &= \begin{bmatrix} -\cos 2\theta \\ -\sin 2\theta \end{bmatrix} \\ &= -\cos 2\theta \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \sin 2\theta \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{aligned}$$

and similarly

$$\begin{aligned} \rho \begin{bmatrix} 0 \\ 1 \end{bmatrix} &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} - 2 \frac{\begin{pmatrix} 0 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} r \cos \theta \\ r \sin \theta \end{pmatrix}}{\begin{pmatrix} r \cos \theta \\ r \sin \theta \end{pmatrix} \cdot \begin{pmatrix} r \cos \theta \\ r \sin \theta \end{pmatrix}} \begin{bmatrix} r \cos \theta \\ r \sin \theta \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} - (2r \sin \theta / r^2) \begin{bmatrix} r \cos \theta \\ r \sin \theta \end{bmatrix} \\ &= \begin{bmatrix} -2 \sin \theta \cos \theta \\ 1 - 2 \sin^2 \theta \end{bmatrix} \\ &= \begin{bmatrix} -\sin 2\theta \\ \cos 2\theta \end{bmatrix} \\ &= -\sin 2\theta \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \cos 2\theta \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \end{aligned}$$

Thus the matrix of the linear transformation  $\rho$  relative to the standard basis  $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}$  of  $\mathbb{R}^2$  is

$$\begin{bmatrix} -\cos 2\theta & -\sin 2\theta \\ -\sin 2\theta & \cos 2\theta \end{bmatrix}$$

(which is the same as  $\begin{pmatrix} \cos \psi & \sin \psi \\ \sin \psi & -\cos \psi \end{pmatrix}$ , where  $\psi = (\pi/2) - 2\theta$ ). Now if  $\rho'$  is the reflection corresponding to another vector  $a'$ , then  $\rho'$  will have matrix

$$\begin{bmatrix} -\cos 2\theta' & -\sin 2\theta' \\ -\sin 2\theta' & \cos 2\theta' \end{bmatrix},$$

where  $\theta'$  is the anticlockwise angle from the  $x$ -axis to  $a'$ , and the product  $\rho\rho'$  has matrix

$$\begin{aligned} & \begin{bmatrix} -\cos 2\theta & -\sin 2\theta \\ -\sin 2\theta & \cos 2\theta \end{bmatrix} \begin{bmatrix} -\cos 2\theta' & -\sin 2\theta' \\ -\sin 2\theta' & \cos 2\theta' \end{bmatrix} \\ &= \begin{bmatrix} \cos 2\theta \cos 2\theta' + \sin 2\theta \sin 2\theta' & \cos 2\theta \sin 2\theta' - \sin 2\theta \cos 2\theta' \\ \sin 2\theta \cos 2\theta' - \cos 2\theta \sin 2\theta' & \sin 2\theta \sin 2\theta' + \cos 2\theta \cos 2\theta' \end{bmatrix}, \\ &= \begin{bmatrix} \cos 2(\theta - \theta') & -\sin 2(\theta - \theta') \\ \sin 2(\theta - \theta') & \cos 2(\theta - \theta') \end{bmatrix} \end{aligned}$$

which is well known to be the matrix of an anticlockwise rotation about the origin through an angle of  $2(\theta - \theta')$ . Note that the angle of rotation is twice the angle between  $a$  and  $a'$ .

We move now to the general situation, and consider nonzero linearly independent vectors  $a$  and  $b$  in an  $n$ -dimensional inner product space  $V$ . The vectors perpendicular to both  $a$  and  $b$  form an  $(n - 2)$ -dimensional subspace,  $K$ , which is the intersection of  $H_a$  and  $H_b$ , the reflecting hyperplanes corresponding to  $\rho_a$  and  $\rho_b$ . If  $v \in K$  then  $v$  is fixed by both reflections  $\rho_a$  and  $\rho_b$ , and hence also by the product  $\rho_a\rho_b$ . On the other hand, if  $u$  is a vector which is a linear combination of  $a$  and  $b$ , then  $\rho_a(u)$  and  $\rho_b(u)$  are also linear combinations of  $a$  and  $b$ . This says that the 2-dimensional subspace  $P$  spanned by  $a$  and  $b$  is both  $\rho_a$ -invariant and  $\rho_b$ -invariant. In fact,  $P$  is a plane, and  $\rho_a$  and  $\rho_b$  act on  $P$  as reflections in the lines  $H_a \cap P$  and  $H_b \cap P$  respectively. Hence  $\rho_a\rho_b$  acts on  $P$  as a rotation through twice the angle between  $a$  and  $b$ .

Note that the subspaces  $K$  and  $P$  are complementary to each other, meaning that every element of  $V$  is uniquely expressible in the form  $x + y$  with  $x \in K$  and  $y \in P$ , and so it follows that the effect of  $\rho_a\rho_b$  on the whole of  $V$  is determined by what it does on the subspaces  $K$  and  $P$ . Specifically,

$$(\rho_a\rho_b)(x + y) = (\rho_a\rho_b)(x) + (\rho_a\rho_b)(y) = x + \rho(y)$$

where  $\rho$  is a rotation of  $P$ . The  $K$ -component is fixed, the  $P$ -component is rotated.

**5.9 PROPOSITION** Suppose that  $a$  and  $b$  are nonzero elements of a real inner product space  $V$ , and suppose that there is a finite group  $G$  of transformations of  $V$  which contains both the reflection  $\rho_a$  and the reflection  $\rho_b$ . Then the angle between  $a$  and  $b$  is  $(p/q)\pi$  for some integers  $p$  and  $q$ .

**Proof.** Let  $P$  be the subspace spanned by  $a$  and  $b$ , and  $\theta$  the angle between  $a$  and  $b$ . If  $P$  is 1-dimensional then  $\theta$  is either 0 or  $\pi$ , satisfying the Proposition. Otherwise  $P$  is 2-dimensional, and  $\rho_a\rho_b$  acts on  $P$  as a rotation through  $2\theta$ . It follows that  $(\rho_a\rho_b)^q$  acts on  $P$  as a rotation through  $2q\theta$ . But since  $\rho_a\rho_b \in G$ , which is finite, it follows from 2.28 that  $(\rho_a\rho_b)^q = i$  (the identity transformation) for some integer  $q$  which is a divisor of  $|G|$ . So for this  $q$  the angle  $2q\theta$  must be an integral multiple of  $2\pi$ . Thus  $2q\theta = 2p\pi$  for some integer  $p$ , and the desired conclusion follows.  $\square$

**5.10 DEFINITION** We will say that a set  $S$  of vectors in an inner product space  $V$  is  $\pi$ -commensurable if all elements of  $S$  are nonzero, and for every pair  $a, b$  of elements of  $S$  there exist integers  $p, q$  such that the angle between  $a$  and  $b$  is  $(p/q)\pi$ .

There are some trivial examples of large  $\pi$ -commensurable sets of vectors in  $n$ -dimensional space. For example, it is possible to have  $n$  mutually perpendicular vectors: the elements of the standard basis have this property. In 2-dimensional space it is easy to find a set of  $2k$  vectors such that the angle between any two of them has the form  $(r/k)\pi$  for some  $r \in \mathbb{Z}$ : just take the position vectors of the vertices and the midpoints of the edges of a regular  $k$ -sided polygon with centre at the origin. We will call  $\pi$ -commensurable sets of this type *polygonal*. Now if  $V$  and  $V'$  are two inner product spaces then there exists another inner product space  $U$ , called the *orthogonal direct sum* of  $V$  and  $V'$ , which contains  $V$  and  $V'$  as subspaces, in such a way that every nonzero vector of  $V$  is perpendicular to every nonzero vector of  $V'$ . Indeed, each element of  $U$  is uniquely expressible as a sum  $v + v'$  with  $v \in V$  and  $v' \in V'$ , and the inner product of two such elements  $v_1 + v'_1$  and  $v_2 + v'_2$  is given by

$$(v_1 + v'_1) \cdot (v_2 + v'_2) = v_1 \cdot v_2 + v'_1 \cdot v'_2.$$

Now if  $S$  is a  $\pi$ -commensurable set of vectors in  $V$ , and  $S'$  a  $\pi$ -commensurable set of vectors in  $V'$ , then the subset  $S \cup S'$  of  $U$  will be  $\pi$ -commensurable too, since anything in  $S$  will make an angle of  $\pi/2$  with anything in  $S'$ . This enables us to construct  $\pi$ -commensurable sets consisting of pairwise

orthogonal subsets each of which is polygonal. But it is a nontrivial task to find examples of  $\pi$ -commensurable sets of vectors which are not of this kind.

### §5d Some examples

Let  $G$  be the group of all symmetries of a regular tetrahedron  $T$ . A tetrahedron can be described as a pyramid on a triangular base; it has four vertices and four triangular faces. Regularity means that the triangular faces are equilateral, and all congruent to each other. Every permutation of the vertices in fact corresponds to a symmetry of  $T$ , and so it follows that  $|G| = 24$ .

For every pair of vertices of  $T$  there is an edge joining the two; thus there are six edges altogether. For each edge  $\alpha$  there is a unique opposite edge  $\alpha'$ , joining the two vertices which are not endpoints of  $\alpha$ . The directions of  $\alpha$  and  $\alpha'$  are perpendicular to each other, and  $\alpha$  lies in the plane which perpendicularly bisects  $\alpha'$ . The reflection in this plane is a symmetry of the tetrahedron; it interchanges the two vertices which are the endpoints of  $\alpha'$ , and fixes the two which are the endpoints of  $\alpha$ . Thus we see that there are exactly six reflections in the group  $G$ ; each edge of  $T$  has exactly one plane of symmetry through it.

To visualize the situation adequately, it helps to embed  $T$  in a cube. The vertices of the cube can be coloured red and blue, so that each edge has one red endpoint and one blue endpoint. The four red vertices of the cube are the vertices of  $T$ , the four blue ones vertices of a second regular tetrahedron  $T'$ , which shares the same symmetry group  $G$  as  $T$ . The diagonals of the faces of the cube are the edges of the tetrahedra. The cube has twelve edges, which split into six pairs of opposite edges, and for each pair of opposite edges there is a plane that includes both the edges, and which passes through  $O$ , the centre of the cube, as well as including one edge of  $T$  and one edge of  $T'$ . These are the six planes corresponding to the reflections in  $G$ . If  $P$  is one of these planes, then the line through  $O$  perpendicular to  $P$  bisects another two edges of the cube. So we see that if  $a$  is the position vector of the midpoint of an edge of the cube, then the plane orthogonal to  $a$  is one of the six planes of symmetry we have described, and so  $\rho_a$  is an element of  $G$ . There are twelve possible choices for  $a$ , corresponding to the twelve edges of the cube (but only six reflections since  $\rho_a = \rho_{-a}$ ). Proposition 5.9 guarantees that this set of vectors is  $\pi$ -commensurable. In fact, if  $X$  and  $Y$  are midpoints of edges of the cube then the angle between  $\overrightarrow{OX}$  and  $\overrightarrow{OY}$  is  $0$ ,  $\pi$  or  $\pi/2$  if the two edges are parallel, and  $\pi/3$  or  $2\pi/3$  otherwise.

For our next example, consider the full group of symmetries of a cube. As well as the six reflections we described above, the cube has another three reflections. There are three pairs of opposite faces, and the plane which is parallel to a pair of opposite faces and midway between the two is a plane of symmetry. If  $a = O\bar{X}$ , where  $X$  is the central point of a face of the cube, then the plane orthogonal to  $a$  is one of these planes of symmetry. Six possible values of  $a$  correspond to only three reflections, since  $\rho_a = \rho - a$ . If  $X$  is the midpoint of a face and  $Y$  the midpoint of an edge then it is easily checked that the angle between  $\overrightarrow{OX}$  and  $\overrightarrow{OY}$  is either  $\pi/4$ ,  $\pi/2$  or  $3\pi/4$ .

For our final 3-dimensional example, consider a regular dodecahedron, and let  $a$ ,  $b$  and  $c$  to be the position vectors of  $A$ ,  $B$  and  $C$ , midpoints of three suitably chosen edges. We can arrange that

$$\begin{aligned} \text{angle}(a, b) &= \frac{4}{5}\pi, & \rho_a\rho_b & \text{ has order } 5, \\ \text{angle}(b, c) &= \frac{2}{3}\pi, & \rho_b\rho_c & \text{ has order } 3, \\ \text{angle}(a, c) &= \frac{1}{2}\pi, & \rho_a\rho_c & \text{ has order } 2. \end{aligned}$$

Each of the twelve pentagonal faces has five lines of symmetry bisecting it, and each such line determines a plane through  $O$ , the reflection in which is a symmetry of the dodecahedron. Each of these planes bisect four faces, and so we obtain a set  $\mathcal{H}$  of  $5 \times 12/4 = 15$  planes corresponding to fifteen reflection symmetries. Each plane in  $\mathcal{H}$  passes through a uniquely determined pair of opposite edges—note that there are thirty edges altogether—and perpendicularly bisects another pair of opposite edges. Furthermore, the line through  $O$  normal to the plane bisects a third pair of opposite edges. In particular the reflection in the plane orthogonal to the position vector of the midpoint of an edge is indeed a symmetry of the dodecahedron. The thirty position vectors of midpoints of edges of the dodecahedron form a  $\pi$ -commensurable set.

To find suitable points  $A$ ,  $B$  and  $C$ , proceed as follows. Let  $P$  be the centre of one of the faces, let  $Q$  be the midpoint of one of the edges of this face and let  $R$  be one of the vertices on this edge. The plane containing the triangle  $QRO$  is in the set  $\mathcal{H}$ , and so its normal through  $O$  passes through points  $A$  and  $A'$  which are midpoints of opposite edges. Choose  $A$  to be the one which is on the same side of the plane  $QRO$  as the point  $P$ . Similarly, choose  $B$  on the same side of  $PRO$  as  $Q$  with  $BO$  normal to  $PRO$ , and  $C$  on the same side of  $PQO$  as  $R$  with  $CO$  normal to  $PQO$ . It can be shown that the reflections  $\rho_a$ ,  $\rho_b$  and  $\rho_c$  thus determined—that is, the reflections in

the planes  $QRO$ ,  $PRO$  and  $PQO$ —generate the full group of symmetries of the dodecahedron.

(This group is in fact isomorphic to  $\text{Alt}(5) \times C_2$ . As we have seen above, each plane in  $\mathcal{H}$  determines three pairs of opposite edges; these three pairs are mutually perpendicular to each other, and the fifteen pairs of opposite edges split into five such sets of three. The symmetry group permutes these five sets and also contains the transformation  $-I$ , which fixes each of the sets by taking each edge to its opposite.)

# 6

## Root systems and reflection groups

If  $G$  is a finite group of transformations of  $n$ -dimensional Euclidean space, and if  $S$  is a set vectors such that  $\rho_a \in G$  for all  $a \in S$ , then (as we proved in Proposition 5.9) the set  $S$  is  $\pi$ -commensurable, meaning that for every pair  $a, b$  of elements of  $S$ , the angle between  $a$  and  $b$  is a rational multiple of  $\pi$ . Not all configurations of angles are feasible, however. For example, it is impossible to find three points  $P, Q, R$  in  $\mathbb{R}^3$  such that  $\angle POQ = \angle POR = \angle QOR = 3\pi/4$ . Try to do it!

In this chapter we will continue the investigation of sets of vectors in an  $n$ -dimensional inner product space  $V$  for which the corresponding reflections lie in a finite group, and determine precisely which configurations are possible. This will provide a classification of all finite groups of transformations of  $V$  which are generated by reflections.

### §6a Root systems

6.1 DEFINITION Let  $V$  be a real inner product space. A set  $S \subseteq V$  is called a *root system* in  $V$  if

- (i)  $S$  is finite,
- (ii) all elements of  $S$  are nonzero,
- (iii)  $S$  spans  $V$ , and
- (iv)  $\rho_a(b) \in S$  whenever  $a, b \in S$ .

Of course, conditions (i) and (iv) of this definition are the crucial ones. Condition (ii) has to be there, since  $\rho_0$  is undefined, but condition (iii) could reasonably be omitted. If a subset  $S$  of  $V$  satisfies (i), (ii) and (iv), but not (iii), then  $S$  is not a root system in  $V$ , but it is a root system in the subspace of  $V$  that it spans.

If  $S$  is a root system then the vectors  $a \in S$  are called *roots*.

Let  $V$  be an inner product space. Given a set  $S$  of nonzero elements of  $V$ , we can form the group  $G$  of transformations of  $V$  that is generated by  $\mathcal{R} = \{\rho_a \mid a \in S\}$ . This group consists of the identity transformation, all the reflections in  $\mathcal{R}$ , all transformations which are products of two elements of  $\mathcal{R}$ , all that are products of three elements of  $\mathcal{R}$ , and so on. That is,

$$G = \{\rho_{a_1}\rho_{a_2}\cdots\rho_{a_k} \mid k \in \mathbb{Z} \text{ is nonnegative, and } a_i \in S \text{ for each } i\}.$$

Note that we do not need to mention inverses when describing  $G$ , since each  $\rho_a$  is its own inverse. Note also that the group  $G$  is necessarily a subgroup of  $O(V)$ , the group of all orthogonal transformations of  $V$ , since 5.7 shows that  $\rho_a \in O(V)$  for all  $a$ .

It is very likely that  $G$  will be an infinite group. For example, by Proposition 5.9, if  $\{a \in V \mid \rho_a \in \mathcal{R}\}$  is not  $\pi$ -commensurable, then the group will certainly be infinite. (If the angle between  $a$  and  $b$  is not a rational multiple of  $\pi$  then  $\rho_a$  and  $\rho_b$  by themselves generate an infinite group, since  $\rho_a\rho_b$  has infinite order.) However, the next proposition shows that if  $S$  is a root system then the group is finite.

**6.2 PROPOSITION** *Let  $W$  be a set of bijective linear transformations of a vector space  $V$ , and suppose that  $S$  is a finite set of vectors which spans  $V$ . Suppose that  $S$  is preserved all elements of  $W$ , in the sense that  $g(v) \in S$  whenever  $g \in W$  and  $v \in S$ . Then  $W$  is a finite set.*

**Proof.** For each  $g \in W$  define  $\phi_g: S \rightarrow S$  by

$$\phi_g(v) = g(v)$$

for all  $v \in S$ . In other words,  $\phi_g$  is simply the restriction to  $S$  of the transformation  $g$  of  $V$ . Note that it is because of our hypothesis that  $g(v) \in S$  for all  $v \in S$  that this definition yields a function from  $S$  to  $S$ ; it is at this point of the proof that the hypothesis is used.

Let  $g \in W$ . Since  $g$  is injective it follows that  $\phi_g$  is injective. Since an injective transformation of a finite set is necessarily surjective, we conclude that  $\phi_g \in \text{Sym}(S)$ , the group of all permutations of  $S$ . So we can define a function  $\phi: W \rightarrow \text{Sym}(S)$  by  $\phi(g) = \phi_g$  for all  $g \in W$ . We will show that  $\phi$  is injective.



Suppose that  $g, h \in W$  with  $\phi(g) = \phi(h)$ . Since one of our hypotheses is that  $S$  is finite, we may write  $S = \{v_1, v_2, \dots, v_n\}$ . Since  $\phi_g = \phi_h$  we have

$$g(v_i) = \phi_g(v_i) = \phi_h(v_i) = h(v_i)$$

for all  $i$  from 1 to  $n$ . Now since  $g$  and  $h$  are linear transformations it follows that

$$g\left(\sum_{i=1}^n \lambda_i v_i\right) = \sum_{i=1}^n \lambda_i g(v_i) = \sum_{i=1}^n \lambda_i h(v_i) = h\left(\sum_{i=1}^n \lambda_i v_i\right)$$

for all choices of the scalars  $\lambda_i$ . But one of our hypotheses is that  $S$  spans  $V$ , and so for every  $v \in V$  there exist scalars  $\lambda_i$  with  $v = \sum_{i=1}^n \lambda_i v_i$ . Hence we can conclude that  $g(v) = h(v)$  for all  $v \in V$ , and since by definition  $g$  and  $h$  are transformations of  $V$ , it follows that  $g = h$ . So we have proved that  $\phi(g) = \phi(h)$  implies  $g = h$ ; that is,  $\phi$  is injective.

Since there is an injective function from  $W$  to  $\text{Sym}(S)$  it follows that the number of elements of  $W$  is less than or equal to the number of elements of  $S$ . Since the number of permutations of a set of size  $n$  is  $n!$ , a finite number, we can conclude that  $\text{Sym}(S)$  is finite, and hence  $W$  is finite, as required.  $\square$

**6.3 COROLLARY** *Let  $S$  be a root system in the inner product space  $V$ , and  $G$  the subgroup of  $O(V)$  generated by  $\{\rho_a \mid a \in S\}$ . Then  $G$  is finite.*

**Proof.** Since  $S$  is finite and spans  $V$ , all we need in order to apply Proposition 6.2 is that  $g(b) \in S$  for all  $b \in S$  and  $g \in G$ . But if  $g \in G$  then  $g = \rho_{a_1} \rho_{a_2} \cdots \rho_{a_k}$  for some  $a_i \in S$ , and so

$$g(b) = \rho_{a_1} \rho_{a_2} \cdots \rho_{a_k}(b) = \rho_{a_1}(\rho_{a_2}(\cdots(\rho_{a_k}(b))\cdots))$$

for all  $b \in S$ . An obvious induction using Part (iv) of the definition of a root system now shows that  $g(b) \in S$ .  $\square$

Corollary 6.3 shows that for every root system there is a corresponding finite subgroup of  $O(V)$  generated by reflections. In the other direction, the next proposition shows that every finite subgroup of  $O(V)$  gives rise to a root system.

6.4 PROPOSITION Let  $V$  be an inner product space and  $G$  a finite subgroup of  $O(V)$ . Then the set

$$S = \{a \in V \mid a \cdot a = 1 \text{ and } \rho_a \in G\}$$

is a root system in the subspace of  $V$  that it spans.

**Proof.** Since  $G$  is a finite group it contains only finitely many reflections. If  $a, b \in S$  are such that  $\rho_a = \rho_b$  then, by Proposition 5.6, there exists a scalar  $\lambda$  such that  $a = \lambda b$ . But  $a \cdot a = 1 = b \cdot b$  (since  $a, b \in S$ ), and so

$$1 = a \cdot a = (\lambda b) \cdot \lambda b = \lambda^2(b \cdot b) = \lambda^2.$$

Hence  $a = \pm b$ , and so there are at most two elements of  $S$  for each reflection in  $G$ . Hence  $S$  is finite.

We have shown that  $S$  satisfies condition (i) of Definition 6.1, and so by the remarks following 6.1 it remains to show that conditions (ii) and (iv) are also satisfied. Now (ii) is trivial, since if  $a \cdot a = 1$  then  $a$  is certainly nonzero. So it remains to show that if  $a, b \in S$  then  $\rho_a(b) \in S$ .

Let  $a, b \in S$ , and let  $c = \rho_a(b)$ . By Proposition 5.7 we know that  $\rho_a$  is an orthogonal transformation; so by Proposition 5.8,

$$\rho_c = \rho_{\rho_a(b)} = \rho_a \rho_b \rho_a^{-1},$$

which is in  $G$  since  $\rho_a$  and  $\rho_b$  are in  $G$ . Furthermore, the fact that  $\rho_a$  is orthogonal also tells us that

$$c \cdot c = \rho_a(b) \cdot \rho_a(b) = b \cdot b = 1.$$

Hence  $c \in S$ , as required. □

We will say that a root system  $S$  is *normalized* if  $a \cdot a = 1$  for all  $a \in S$ . It is easily checked that if  $S$  is any root system, then the set

$$\left\{ \frac{1}{\|a\|} a \mid a \in S \right\}$$

is a normalized root system. Of course, the reflections corresponding to the vectors in this normalized system are exactly the same as those corresponding to the vectors in  $S$  itself. So as far as groups generated by reflections are concerned, there is nothing lost by restricting attention to normalized root systems; hence we shall usually do so.

## §6b Positive, negative and simple roots

Let  $S$  be a root system in the inner product space  $V$ . Observe that if  $a \in S$  then, by (iv) of Definition 6.1,  $\rho_a(a) \in S$ . But  $\rho_a(a) = a - \frac{2(a \cdot a)}{(a \cdot a)}a = -a$ . So  $-a \in S$  whenever  $a \in S$ . We now propose to divide  $S$  into two halves, called  $S^+$  and  $S^-$ , in such a way that, for all  $a \in S$ , if  $a \in S^+$  then  $-a \in S^-$ , and vice versa. We do this in a rather arbitrary fashion, simply choosing some hyperplane, and declaring the elements of  $S$  on one side of it to constitute  $S^+$ , those on the other side to constitute  $S^-$ . We only need to be sure that none of the elements of  $S$  actually lie on the hyperplane, but this is a very easy requirement to meet since there are only finitely many vectors in  $S$  and an uncountably infinite supply of hyperplanes available.

**6.5 PROPOSITION** *Let  $S$  be a root system in the inner product space  $V$ . There exists a vector  $v_0 \in V$  such that  $a \cdot v_0 \neq 0$  for all  $a \in S$ .*

**Proof.** Since  $S$  is finite, for each  $v \in V$  the set  $Q(v) = \{a \in S \mid a \cdot v = 0\}$  has at most a finite number of elements. Choose  $v_0 \in V$  so that the number of elements in  $Q(v_0)$  is minimized. We prove that in fact  $Q(v_0)$  is empty.

Suppose that  $Q(v_0) \neq \emptyset$ , and let  $a \in Q(v_0)$ . Since  $S$  is a finite set, the set of numbers

$$A = \{ (b \cdot v_0) / (b \cdot a) \mid b \in S \text{ and } b \cdot a \neq 0 \}$$

is also finite. Let  $\lambda$  be any real number which is nonzero and not in the set  $A$ , and define  $v_1 = v_0 - \lambda a$ . Then

$$a \cdot v_1 = a \cdot (v_0 - \lambda a) = a \cdot v_0 - \lambda(a \cdot a) = -\lambda(a \cdot a) \neq 0$$

since  $a \neq 0$ . Hence  $a \notin Q(v_1)$ . Furthermore, if  $b \in Q(v_1)$  then

$$0 = b \cdot v_1 = b \cdot (v_0 - \lambda a) = b \cdot v_0 - \lambda(b \cdot a),$$

and if  $b \cdot a$  were nonzero this would give  $\lambda = \frac{b \cdot v_0}{b \cdot a}$ , contrary to the definition of  $\lambda$ . So  $b \cdot a = 0$ , and hence  $b \cdot v_0 = b \cdot v_1 = 0$ . So  $b \in Q(v_0)$ . Thus we have shown that  $Q(v_1)$  is a subset of  $Q(v_0)$ , and not equal to  $Q(v_0)$  since  $a \notin Q(v_1)$  whereas  $a \in Q(v_0)$ . So  $Q(v_1)$  has fewer elements than  $Q(v_0)$ , contradicting the definition of  $v_0$ . This contradiction shows that  $Q(v_0)$  must be empty.  $\square$

We now choose arbitrarily a  $v_0 \in V$  such that  $a \cdot v_0 \neq 0$  for all  $a$  in our root system  $S$ , and we keep  $v_0$  fixed for the rest of the discussion. We can now define  $S^+$  and  $S^-$ .

**6.6 DEFINITION** Let  $S$  be a root system, and fix  $v_0 \in V$  such that  $a \cdot v_0 \neq 0$  for all  $a \in S$ . Define

$$S^+ = \{ a \in S \mid a \cdot v_0 > 0 \},$$

the set of *positive roots*, and

$$S^- = \{ a \in S \mid a \cdot v_0 < 0 \},$$

the set of *negative roots*.

The following technical definition will be of great use in our investigations of root systems.

**6.7 DEFINITION** If  $B = \{b_1, b_2, \dots, b_n\}$  is a finite set of vectors in the vector space  $V$ , then a vector  $v \in V$  is said to be a *positive linear combination* of  $B$  if  $v \neq 0$  and

$$v = \lambda_1 b_1 + \lambda_2 b_2 + \cdots + \lambda_n b_n$$

for some scalars  $\lambda_i$  such that  $\lambda_i \geq 0$  for all  $i$ . The set of all  $v \in V$  which are positive linear combinations of  $B$  will be denoted by  $\text{plc}(B)$ .

**6.8 DEFINITION** Let  $S$  be a root system in  $V$ , with sets  $S^+$  and  $S^-$  of positive and negative roots relative to some fixed  $v_0 \in V$ , as in Definition 6.6. A subset  $B \subseteq S^+$  is called a *base* of  $S$  if

- (i)  $B$  is a basis of  $V$ , and
- (ii)  $S^+ \subseteq \text{plc}(B)$ .

The next result is the key to our analysis of root systems.

**6.9 THEOREM** Let  $S$  be a root system in  $V$  and  $S^+$  a set of positive roots in  $S$ . Then  $S$  has a base  $B \subseteq S^+$ .

**Proof.** Let  $\mathcal{A} = \{ B \subseteq S^+ \mid S^+ \subseteq \text{plc}(B) \}$ . That is,  $\mathcal{A}$  is the set of all subsets  $B$  of  $S^+$  with the following property, which we shall call “Property (P)”:

(P)  $B \subseteq S^+$  and every  $a \in S^+$  is a positive linear combination of  $B$ .

Observe that  $S^+$  itself has Property (P). For, let  $S^+ = \{a_1, a_2, \dots, a_k\}$ ; now for all  $a \in S^+$  we can achieve  $a = \lambda_1 a_1 + \lambda_2 a_2 + \dots + \lambda_k a_k$  with  $\lambda_i \geq 0$  by putting  $\lambda_i = 1$  if  $a_i = a$  and  $\lambda_i = 0$  otherwise. Hence  $\mathcal{A} \neq \emptyset$ . Now choose  $B \in \mathcal{A}$  with  $|B|$  as small as possible. So  $B$  has Property (P), but no proper subset of  $B$  has Property (P). We will prove that  $B$  is a basis of  $V$ .

*Step 1.* Let  $B = \{b_1, b_2, \dots, b_n\}$ . For all  $i$ , the vector  $b_i$  is not a positive linear combination of  $B \setminus \{b_i\} = \{b_1, b_2, \dots, b_{i-1}, b_{i+1}, \dots, b_n\}$ .

**Proof.** Suppose that

$$(6.9.1) \quad b_i = \mu_1 b_1 + \dots + \mu_{i-1} b_{i-1} + \mu_{i+1} b_{i+1} + \dots + \mu_n b_n$$

with all the coefficients  $\mu_j \geq 0$ . We will prove that  $B \setminus \{b_i\}$  has Property (P), contradicting minimality of  $B$ .

Let  $a \in S^+$ . Since  $B$  has (P),

$$a = \lambda_1 b_1 + \lambda_2 b_2 + \dots + \lambda_n b_n \quad \text{for some } \lambda_i \geq 0.$$

Substituting into this the value of  $b_i$  from Equation (6.9.1) gives

$$a = \lambda_1 b_1 + \dots + \lambda_i (\mu_1 b_1 + \dots + \mu_{i-1} b_{i-1} + \mu_{i+1} b_{i+1} + \dots + \mu_n b_n) + \dots + \lambda_n b_n$$

which is clearly a positive linear combination of  $B \setminus \{b_i\}$  since all the  $\lambda_i$  and  $\mu_j$  are nonnegative. So we have shown that  $S^+ \subseteq \text{plc}(B \setminus \{b_i\})$ , as required  $\square$

*Step 2.* Let  $B = \{b_1, b_2, \dots, b_n\}$ . If  $b_i \neq b_j$ , then  $b_i \cdot b_j \leq 0$ .

**Proof.** Suppose that  $b_i \cdot b_j > 0$ . Let  $a = \rho_{b_i}(b_j)$ , which is an element of the root system  $S$  since  $b_i, b_j \in S$ . Now

$$a = b_j - \alpha b_i$$

where  $\alpha = 2(b_i \cdot b_j)/(b_j \cdot b_j) > 0$ .

If  $a \in S^+$  then  $a = \lambda_1 b_1 + \lambda_2 b_2 + \dots + \lambda_n b_n$  for some coefficients  $\lambda_i \geq 0$ . So

$$b_j - \alpha b_i = \lambda_1 b_1 + \lambda_2 b_2 + \dots + \lambda_n b_n,$$

and rearranging this gives

$$(6.9.2) \quad (1 - \lambda_j) b_j = \alpha b_i + \sum_{l \neq j} \lambda_l b_l.$$

Taking the dot product with  $v$  we find that

$$\begin{aligned} (1 - \lambda_j)(b_j \cdot v_0) &= \alpha(b_i \cdot v_0) + \sum_{l \neq j} \lambda_l(b_l \cdot v_0) \\ &\geq \alpha(b_i \cdot v_0) > 0. \end{aligned}$$

Since  $b_j \cdot v_0 > 0$ , it follows that  $1 - \lambda_j > 0$ . Now by (6.9.2)

$$b_j = \frac{\alpha}{1 - \lambda_j} b_i + \sum_{l \neq j} \frac{\lambda_l}{1 - \lambda_j} b_l \in \text{plc}(B \setminus \{b_j\})$$

contradicting Step 1. Therefore  $a \notin S^+$ .

Since  $a \notin S^+$ , it follows that  $a \in S^-$ , and so  $-a \in S^+$ . So there exist coefficients  $\lambda_i \geq 0$  with  $-a = \lambda_1 b_1 + \lambda_2 b_2 + \cdots + \lambda_n b_n$ , and this gives

$$\alpha b_i - b_j = \lambda_1 b_1 + \lambda_2 b_2 + \cdots + \lambda_n b_n,$$

or, rearranging,

$$(6.9.3) \quad (\alpha - \lambda_i) b_i = b_j + \sum_{l \neq i} \lambda_l b_l.$$

Taking the dot product with  $v_0$  we deduce that

$$\begin{aligned} (\alpha - \lambda_i)(b_i \cdot v_0) &= b_j \cdot v_0 + \sum_{l \neq j} \lambda_l b_l \cdot v_0 \\ &\geq b_j \cdot v_0 > 0, \end{aligned}$$

and therefore  $\alpha - \lambda_i > 0$  (since  $b_i \cdot v_0 > 0$ ). Now dividing through by  $\alpha - \lambda_i$  in (6.9.3) we see that  $b_i \in \text{plc}(B \setminus \{b_i\})$ , contradicting Step 1.  $\square$

*Step 3.*  $B$  is a basis of  $V$ .

**Proof.** Let  $v \in V$ . Now  $S$  spans  $V$ , since it is a root system in  $V$ , and so there exist scalar coefficients  $\lambda_a$  such that  $v = \sum_{a \in S} \lambda_a a$ . But if  $a \in S$  then  $\varepsilon a \in S^+$  for some sign  $\varepsilon = \pm 1$ , and since  $B$  has Property (P) it follows that  $\varepsilon a \in \text{plc}(B)$ . So there exist scalars  $\mu_{ab}$  (which are nonnegative if  $a \in S^+$  and nonpositive if  $a \in S^-$ ) such that  $a = \sum_{b \in B} \mu_{ab} b$ , and thus

$$v = \sum_{b \in B} \left( \sum_{a \in S} \lambda_a \mu_{ab} \right) b.$$

So every element of  $V$  is a linear combination of the elements of  $B$ , and we have shown that  $B$  spans  $V$ .

It remains to show that the elements of  $B$  are linearly independent. Suppose, to the contrary, that  $\lambda_1 b_1 + \lambda_2 b_2 + \cdots + \lambda_n b_n = 0$  for some scalars  $\lambda_i$ , at least one of which is nonzero. Now let  $H$  be the set of indices  $i \in \{1, 2, \dots, n\}$  such that  $\lambda_i \geq 0$ , and  $J$  the set of those such that  $\lambda_i < 0$  (the rest). So  $\{1, 2, \dots, n\}$  is the disjoint union of  $H$  and  $J$ , whence

$$0 = \sum_{i=1}^n \lambda_i b_i = \sum_{i \in H} \lambda_i b_i + \sum_{i \in J} \lambda_i b_i = \sum_{i \in H} |\lambda_i| b_i - \sum_{i \in J} |\lambda_i| b_i,$$

and it follows that we can define

$$v = \sum_{i \in H} |\lambda_i| b_i = \sum_{i \in J} |\lambda_i| b_i.$$

Now  $v \cdot v_0 = \sum_{i \in H} |\lambda_i| (b_i \cdot v_0) = \sum_{i \in J} |\lambda_i| (b_i \cdot v_0)$ ; furthermore,  $b_i \cdot v_0 > 0$  for all  $i$  (since  $b_i \in S^+$ ), whence each term  $|\lambda_i| (b_i \cdot v_0)$  is nonnegative, and strictly positive if  $\lambda_i \neq 0$ . Since we have assumed that at least one  $\lambda_i$  is nonzero, we conclude that at least one term is strictly positive and the others nonnegative, and hence  $v \cdot v_0 > 0$ . In particular,  $v \neq 0$ . Therefore,

$$0 < v \cdot v = \left( \sum_{i \in H} |\lambda_i| b_i \right) \cdot \left( \sum_{j \in J} |\lambda_j| b_j \right) = \sum_{i \in H} \sum_{j \in J} |\lambda_i| |\lambda_j| (b_i \cdot b_j) \leq 0$$

since  $b_i \cdot b_j \leq 0$  whenever  $i \in H$  and  $j \in J$ , by Step 2.

This contradiction proves that the vectors in  $B$  are linearly independent, and therefore form a basis of  $V$ .  $\square$

Since  $B$  is a basis and has Property (P), it is a base for the root system, and Theorem 6.9 is proved.  $\square$

It can be shown that the base  $B$  is in fact uniquely determined by the set  $S^+$  of positive roots. The positive roots which lie in this uniquely determined base are called the *simple* or *fundamental* roots of the positive system. Note that, since  $B$  is a basis for  $V$ , an arbitrary root  $a$  can be expressed as a linear combination of simple roots (since  $B$  spans  $V$ ), and furthermore the expression is unique (since  $B$  is linearly independent). If  $a$  is positive then all the scalar coefficients in this unique expression will be nonnegative (since

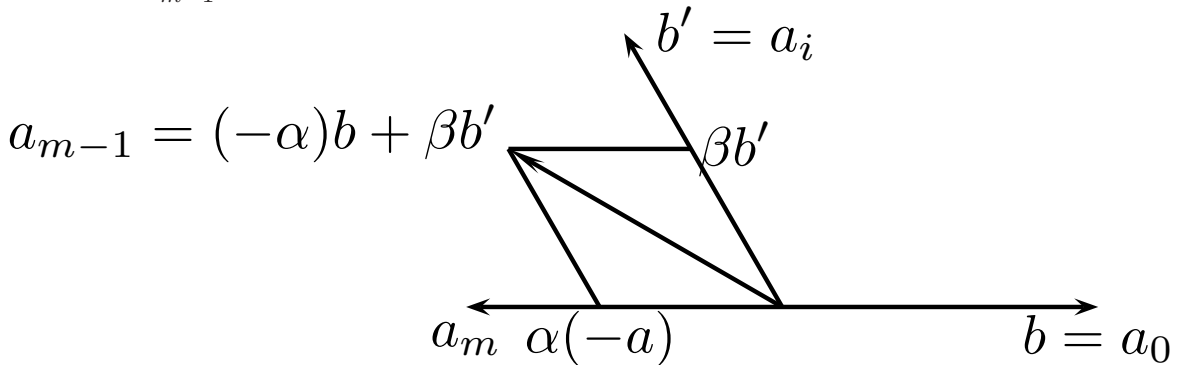
$B$  has Property (P)), while if  $a$  is negative then all the coefficients will be nonpositive (since  $-a$  is a positive root in this case).

We know from Proposition 5.9 that the angle between any two roots in a root system  $S$  is necessarily a rational multiple of  $\pi$ , and this applies in particular to any two simple roots. We also know, in fact, from Step 2 of the proof of Theorem 6.9, that the angle between any two simple roots must be obtuse or a right angle (as the cosine of the angle is nonpositive). Our next result gives even more accurate information.

**6.10 PROPOSITION** *Let  $b, b' \in B$ , where  $B$  is a base for the root system  $S$ . Then there exists a positive integer  $m \geq 2$  such that the angle between  $b$  and  $b'$  is  $\frac{m-1}{m}\pi$ .*

**Proof.** The vectors  $b$  and  $b'$  span a 2-dimensional subspace of  $V$  which can be identified with the Euclidean plane, and the transformation  $\rho_b\rho_{b'}$  acts on this plane as a rotation through  $2\theta$ , where  $\theta$  is the angle between  $b$  and  $b'$ . Let  $\sigma$  be this rotation, and let the order of  $\sigma$  be  $m$ . Applying powers of  $\sigma$  to  $b$  and  $b'$  yield  $2m$  vectors, all of which are in the root system and are linear combinations of  $b$  and  $b'$ . Clearly these roots are equally spaced (in a rotational sense) around the origin, each separated from its two nearest neighbours by an angle of  $\pi/m$ . That is to say, the unit vectors in the directions of these roots form a polygonal  $\pi$ -commensurable set, to use the terminology of Chapter 5.

Let  $b = a_0, a_1, \dots, a_{2m-1}$  be the  $2m$  roots we have been discussing, listed in the order they are encountered when circumnavigating the origin, from  $b$  back to  $b$  again. Choose the direction of circumnavigation so that  $b'$  is encountered before getting halfway round. Our aim is to show that  $b' = a_{m-1}$ .



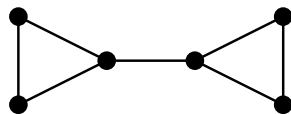


Suppose, to the contrary, that  $b' = a_i$  for some  $i$  such that  $1 \leq i \leq m-2$ . (Note that  $b' \neq a_m$ , and  $b' \neq a_0$ , since  $b$  and  $b'$  are linearly independent.) In this situation it is geometrically clear that when  $a_{m-1}$  is expressed as  $\lambda b + \mu b'$  then  $\lambda$  is negative and  $\mu$  positive. This is because all the vectors which are rotationally between  $a_m = -b$  and  $a_i = b'$  must be positive linear combinations of  $-b$  and  $b'$ , as illustrated in the diagram above (in which  $\alpha$  and  $\beta$  are positive scalars). However, we know that in the unique expression for the root  $a_{m-1}$  as a linear combination of simple roots, the coefficients must either be all nonnegative or all nonpositive. You cannot have mixed coefficients (some positive, others negative). This contradicts the fact that  $a_{m-1} = \lambda b + \mu b'$  with  $\lambda$  negative and  $\mu$  positive. So our assumption that  $b' \neq a_{m-1}$  cannot be sustained:  $b'$  has to be  $a_{m-1}$ , whence the angle between  $b$  and  $b'$  is  $\frac{m-1}{m}\pi$ , as required.  $\square$

### §6c Diagrams

A *graph* is a set  $\text{Vert}(\Gamma)$ , whose elements are called the *vertices* of  $\Gamma$ , with a binary relation  $\alpha$ , which in our cases will always be assumed to be symmetric and irreflexive. That is,  $\alpha(a, b)$  is true if and only if  $\alpha(b, a)$  is true, and  $\alpha(a, a)$  is always false. The pairs  $\{a, b\}$  of vertices such that  $\alpha(a, b)$  is true are called the *edges* of  $\Gamma$ .

Graphs are most conveniently represented by drawings consisting of dots joined by lines. There should be one dot for each vertex, and a line joining the dots corresponding to the vertices  $a$  and  $b$  if and only if  $\{a, b\}$  is an edge. Here is a graph with six vertices and seven edges.



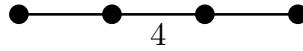
If  $B$  is a base for the root system  $S$ , then, by Proposition 6.10, for all  $a, b \in B$  there exists an integer  $m_{ab} \geq 2$  such that the angle between  $a$  and  $b$  is  $\frac{m_{ab}-1}{m_{ab}}\pi$ . We now associate to  $S$  a graph  $\Gamma$ , called the *Coxeter diagram* of the root system, as follows.

- (i)  $\text{Vert}(\Gamma)$  is in one-to-one correspondence with  $B$ .
- (ii) If  $a, b \in B$  are not perpendicular, then the vertices corresponding to  $a$  and  $b$  are joined by an edge labelled with the integer  $m_{ab}$ .

Note that  $a, b \in B$  are perpendicular if and only if  $m_{ab} = 2$ . The vertices corresponding to  $a$  and  $b$  are not joined by an edge if  $m_{ab} = 2$ . It is also customary to omit the label on the edge if  $m_{ab} = 3$ . So, for example, if  $B = \{a, b, c, d\}$ , with

$$\begin{aligned} m_{ac} = m_{ad} = m_{bd} &= 2, \\ m_{ab} = m_{cd} &= 3, \\ m_{bc} &= 4, \end{aligned}$$

then the associated Coxeter diagram is



where, from left to right, the vertices correspond to  $a, b, c$  and  $d$ . On the other hand, no root system can give rise to the diagram



since, as we commented at the start of this chapter, it is impossible to arrange three vectors in Euclidean space so that the angle between any two of them is  $(3/4)\pi$ .

It is our aim in this section to determine exactly which graphs can occur as Coxeter diagrams of root systems, for this will essentially classify finite Euclidean reflection groups.

Suppose that  $\Gamma$  is a graph which resembles a Coxeter diagram, in that its edges are labelled with integers greater than 2 (unlabelled edges being understood to have the label 3). Define a function  $m = m_\Gamma$ , to be called the *labelling function*, as follows: the domain of  $m$  is  $\text{Vert}(\Gamma) \times \text{Vert}(\Gamma)$ , and for all  $x, y \in \text{Vert}(\Gamma)$ ,

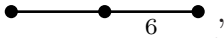
$$m(x, y) = \begin{cases} 1 & \text{if } x = y \\ 2 & \text{if } x \neq y \text{ and } \{x, y\} \text{ is not an edge,} \\ l & \text{if } \{x, y\} \text{ is an edge labelled } l. \end{cases}$$

Suppose that in fact  $\Gamma$  has  $n$  vertices; indeed, with no loss of generality, let us assume that  $\text{Vert}(\Gamma) = \{1, 2, \dots, n\}$ , and (for brevity) write  $m_{ij}$  for  $m_\Gamma(i, j)$ .

Let  $V_\Gamma$  be a vector space with basis  $v_1, v_2, \dots, v_n$ , and let  $f_\Gamma$  be the bilinear form defined on  $V_\Gamma$  and having the property that  $f_\Gamma(v_i, v_j) = \cos\left(\frac{m_{ij}-1}{m_{ij}}\pi\right)$  for all  $i, j$ . Note that the existence and uniqueness of  $f_\Gamma$  are guaranteed by 5.3. Note also that  $f_\Gamma(v_i, v_i) = 1$  for all  $i$ . We will call  $V_\Gamma$  the *space associated with  $\Gamma$* , and  $f_\Gamma$  the *form associated with  $\Gamma$* . We will also call  $v_1, v_2, \dots, v_n$  the *canonical basis* of  $V_\Gamma$ .

If this bilinear form is positive definite, it means that one can find linearly independent vectors  $b_i$  in Euclidean space such that the angle between  $b_i$  and  $b_j$  is  $\frac{m_{ij}-1}{m_{ij}}\pi$  (see Proposition 5.1). If the form is not positive definite then it is impossible to find such vectors in Euclidean space.

6.11 DEFINITION We say that a diagram  $\Gamma$  is *admissible* if the form  $f_\Gamma$  associated with  $\Gamma$  is positive definite, *inadmissible* otherwise.

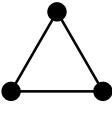
Suppose, for example, that  $\Gamma$  is the diagram , with the vertices numbered from left to right, so that  $m_{12} = 3$ ,  $m_{13} = 2$  and  $m_{23} = 6$ . The vector space  $V_\Gamma$  then has basis  $v_1, v_2, v_3$ , and the matrix of the form  $f_\Gamma$  relative to this basis—that is, the matrix whose  $(i, j)$ -entry is  $f_\Gamma(v_i, v_j)$ —is

$$\begin{bmatrix} 1 & -1/2 & 0 \\ -1/2 & 1 & -\sqrt{3}/2 \\ 0 & -\sqrt{3}/2 & 1 \end{bmatrix}.$$

It turns out that the form  $f_\Gamma$  is not positive definite, since it is possible to find a nonzero vector  $v \in V_\Gamma$  with  $f_\Gamma(v, v) = 0$ . Indeed, let  $v = v_1 + 2v_2 + \sqrt{3}v_3$ . By the definition of  $V_\Gamma$ , the vectors  $v_1, v_2, v_3$  form a basis, and are therefore linearly independent. So  $v \neq 0$ . Moreover,

$$\begin{aligned} f_\Gamma(v, v) &= [1 \quad 2 \quad \sqrt{3}] \begin{bmatrix} 1 & -1/2 & 0 \\ -1/2 & 1 & -\sqrt{3}/2 \\ 0 & -\sqrt{3}/2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ \sqrt{3} \end{bmatrix} \\ &= [0 \quad 0 \quad 0] \begin{bmatrix} 1 \\ 2 \\ \sqrt{3} \end{bmatrix} = 0. \end{aligned}$$

Hence this diagram is inadmissible.

In a similar way we can see that the diagram  is also inadmissible. In this case the matrix of the form is

$$\begin{bmatrix} 1 & -1/2 & -1/2 \\ -1/2 & 1 & -1/2 \\ -1/2 & -1/2 & 1 \end{bmatrix}$$

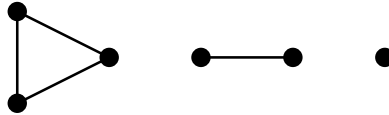
and it can be seen that

$$[1 \ 1 \ 1] \begin{bmatrix} 1 & -1/2 & -1/2 \\ -1/2 & 1 & -1/2 \\ -1/2 & -1/2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = [0 \ 0 \ 0] \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 0.$$

On the other hand, the diagram  $\bullet \text{---} \bullet \text{---} \bullet$  is admissible. To prove this one has simply to find three linearly independent vectors  $a, b, c$  in Euclidean space such that the angle between  $a$  and  $b$  is  $2\pi/3$ , the angle between  $a$  and  $c$  is  $\pi/2$ , and the angle between  $b$  and  $c$  is  $3\pi/4$ . Three suitable vectors are

$$\begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ -\sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Note that the diagram  $\Gamma$  does not have to be connected. For example,



can perfectly well be regarded as a single diagram, the space  $V_\Gamma$  being 6-dimensional. In situations like this, if the  $i$ -th and  $j$ -th vertices of  $\Gamma$  belong to different components, then  $m_{ij} = 2$ ; so if  $v_i$  and  $v_j$  are the corresponding elements of the canonical basis of  $V_\Gamma$ , then  $f_\Gamma(v_i, v_j) = 0$ . Thus if  $\Gamma$  has  $k$  connected components altogether, the canonical basis splits into  $k$  mutually disjoint subsets such that  $f_\Gamma(v_i, v_j) = 0$  whenever  $v_i$  and  $v_j$  are from different subsets. This yields a decomposition of  $V_\Gamma$  as  $V_{\Gamma_1} \oplus V_{\Gamma_2} \oplus \cdots \oplus V_{\Gamma_k}$ , where the  $\Gamma_i$  are the connected components of  $\Gamma$ , and

$$f_\Gamma(x_1 + x_2 + \cdots + x_k, x'_1 + x'_2 + \cdots + x'_k) = f_{\Gamma_1}(x_1, x'_1) + \cdots + f_{\Gamma_k}(x_k, x'_k)$$

whenever  $x_i, x'_i \in V_{\Gamma_i}$  (since the “cross terms”, like  $f_\Gamma(x_4, x'_2)$ , are all zero). Now if  $\Gamma$  is admissible, then  $f_\Gamma$  is positive definite, and so

$$f_\Gamma(x_1 + x_2 + \cdots + x_k, x_1 + x_2 + \cdots + x_k) > 0$$

whenever  $x_1 + x_2 + \cdots + x_k \neq 0$ . In particular, if  $0 \neq x_i \in V_{\Gamma_i}$ , then (putting  $x_j = 0$  for all  $j \neq i$ ) it follows that  $f_\Gamma(x_i, x_i) > 0$ . Hence the component  $\Gamma_i$  is also an admissible diagram. Conversely, if each  $\Gamma_i$  is admissible then  $\Gamma$  must

be admissible also, since if  $0 \neq x \in V_\Gamma$ , then  $x = x_1 + x_2 + \cdots + x_k$  for some  $x_i \in V_{\Gamma_i}$ , and since  $x_i \neq 0$  for at least one  $i$ , we conclude that

$$\begin{aligned} f_\Gamma(x_1 + x_2 + \cdots + x_k, x_1 + x_2 + \cdots + x_k) &= f_\Gamma(x_1, x_1) + \cdots + f_\Gamma(x_k, x_k) \\ &= \sum_{x_i \neq 0} f_\Gamma(x_i, x_i) \\ &> 0. \end{aligned}$$

We have now proved the following proposition.


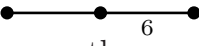
**6.12 PROPOSITION** *A disconnected diagram is admissible if all its connected components are admissible.*

Let us give another, slightly different, proof that if the connected components of  $\Gamma$  are admissible then so is  $\Gamma$  itself.

**Proof.** Observe first that the set of all  $v \in \mathbb{R}^{n+m}$  whose last  $m$  coordinates are all zero constitutes an  $n$ -dimensional subspace  $V$  which is isomorphic to  $\mathbb{R}^n$  in an obvious way. In other words, there is a vector space isomorphism  $\phi: \mathbb{R}^n \rightarrow V$  which preserves angles between vectors. Similarly, the set of all  $w \in \mathbb{R}^{n+m}$  whose first  $n$  coordinates are zero constitutes an  $m$ -dimensional subspace  $W$  which is isomorphic to  $\mathbb{R}^m$ ; this gives a vector space isomorphism  $\psi: \mathbb{R}^m \rightarrow W$  which preserves angles. Furthermore, if  $v \in V$  and  $w \in W$  then  $v \cdot w = 0$ : the spaces  $V$  and  $W$  are perpendicular to each other. Note also that  $\mathbb{R}^{n+m}$  is the direct sum of  $V$  and  $W$ .

Let  $\Gamma_1$  and  $\Gamma_2$  be admissible diagrams, and let  $\Gamma$  be the disconnected diagram consisting of  $\Gamma_1$  alongside  $\Gamma_2$ . Let  $\text{Vert}(\Gamma_1) = \{1, 2, \dots, n\}$  and  $\text{Vert}(\Gamma_2) = \{n+1, n+2, \dots, n+m\}$ , and let  $m$  be the labelling function. Since  $\Gamma_1$  is admissible there exists a basis  $u_1, u_2, \dots, u_n$  of  $\mathbb{R}^n$  such that the angle between  $u_i$  and  $u_j$  is  $\frac{m(i,j)-1}{m(i,j)}\pi$  for all  $i, j \in \{1, 2, \dots, n\}$ . Similarly, since  $\Gamma_2$  is admissible there exists a basis  $w_{n+1}, w_{n+2}, \dots, w_{n+m}$  of  $\mathbb{R}^m$  such that the angle between  $w_i$  and  $w_j$  is  $\frac{m(i,j)-1}{m(i,j)}\pi$  for all  $i, j \in \{n+1, n+2, \dots, n+m\}$ . Now define  $v_i = \phi(u_i)$  for all  $i \in \{1, 2, \dots, n\}$ , and  $v_i = \psi(w_i)$  for all  $i \in \{n+1, n+2, \dots, n+m\}$ . If  $i, j \in \{1, 2, \dots, n\}$  then the angle between  $v_i$  and  $v_j$  equals the angle between  $u_i$  and  $u_j$ ; if  $i, j \in \{n+1, n+2, \dots, n+m\}$  then the angle between  $v_i$  and  $v_j$  equals the angle between  $w_i$  and  $w_j$ ; if  $i \leq n$  and  $j \geq n+1$  then the angle between  $v_i$  and  $v_j$  is  $\pi/2$ . Hence the angle between  $v_i$  and  $v_j$  is  $\frac{m(i,j)-1}{m(i,j)}\pi$  in all cases. Since  $v_1, v_2, \dots, v_{n+m}$  is clearly a basis of  $\mathbb{R}^{n+m}$ , this proves that  $\Gamma$  is admissible.  $\square$

Clearly, Proposition 6.12 reduces the problem of classifying admissible diagrams to the problem of classifying connected admissible diagrams.

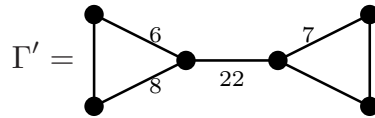
Our strategy for obtaining the complete list of admissible diagrams is as follows. We obtain a long list of inadmissible diagrams, proved inadmissible by methods similar to those we used above for the diagrams  and . The key step is to prove that if  $\Gamma$  is any inadmissible diagram, then any diagram which (in some sense) is more complicated than  $\Gamma$  is also inadmissible. It is a straightforward task to describe all diagrams which are not more complicated than any of the inadmissible diagrams on our list, and which are therefore the only possible admissible diagrams. Each of these possibly admissible diagrams  $\Gamma$  is proved to be admissible by explicitly finding linearly independent vectors  $b_i$  in Euclidean space such that the angle between  $b_i$  and  $b_j$  is  $\frac{m_{ij}-1}{m_{ij}}\pi$ .

The definition of “more complicated” is as follows:  $\Gamma'$  is more complicated than  $\Gamma$  if  $\Gamma'$  can be obtained from  $\Gamma$  by adding extra vertices and/or adding extra edges and/or increasing the labels on some edges.

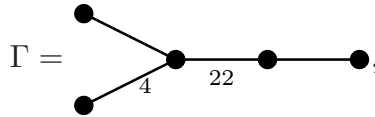
6.13 DEFINITION If  $\Gamma$  and  $\Gamma'$  are diagrams, we say that  $\Gamma'$  is more complicated than  $\Gamma$ , and we write  $\Gamma' \geq \Gamma$ , if the following conditions both hold.

- (i) If  $\Gamma, \Gamma'$  have  $n, n'$  vertices respectively, then  $n \leq n'$ .
- (ii) Given a numbering  $1, 2, \dots, n$  of the vertices of  $\Gamma$ , there exists a numbering  $1, 2, \dots, n'$  of the vertices of  $\Gamma'$ , such that  $m'(i, j) \geq m(i, j)$  for all  $i, j \in \{1, 2, \dots, n\}$ , where  $m'$  and  $m$  are the labelling functions for  $\Gamma'$  and  $\Gamma$  (respectively).

For example, if



and



then  $\Gamma' \geq \Gamma$ .

Let us prove the key proposition forthwith.

6.14 PROPOSITION If  $\Gamma, \Gamma'$  are diagrams with  $\Gamma' \geq \Gamma$ , and if  $\Gamma$  is inadmissible, then  $\Gamma'$  is also inadmissible.†

**Proof.** Let  $n = |\text{Vert}(\Gamma)|$  and  $n' = |\text{Vert}(\Gamma')|$ , and assume that, in accordance with Definition 6.13,

$$m'(i, j) \geq m(i, j) \quad \text{for all } i, j \in \{1, 2, \dots, n\}$$

where  $m$  and  $m'$  are the labelling functions for  $\Gamma$  and  $\Gamma'$  (for some numberings of the vertices of these diagrams). Let  $V$  and  $f$  be the space and form associated with  $\Gamma$ , so that  $V$  has basis  $v_1, v_2, \dots, v_n$  and  $f(v_i, v_j) = \cos\left(\frac{m(i, j) - 1}{m(i, j)}\pi\right)$ . Similarly, let  $V'$  and  $f'$  be the space and form associated with  $\Gamma'$ , the relevant basis of  $V'$  being  $v'_1, v'_2, \dots, v'_{n'}$ .

Let  $i, j \in \{1, 2, \dots, n\}$  with  $i \neq j$ . Then  $2 \leq m(i, j) \leq m'(i, j)$ , whence

$$\pi/2 \geq \pi/m(i, j) \geq \pi/m'(i, j),$$

and since  $\cos$  is decreasing on the interval  $[0, \pi/2]$ ,

$$0 \leq \cos(\pi/m(i, j)) \leq \cos(\pi/m'(i, j)).$$

Since  $\cos(\pi - \theta) = -\cos \theta$  for all  $\theta$ ,

$$\begin{aligned} 0 \geq \cos\left(\frac{m(i, j) - 1}{m(i, j)}\pi\right) &= -\cos(\pi/m(i, j)) \\ &\geq -\cos(\pi/m'(i, j)) = \cos\left(\frac{m'(i, j) - 1}{m'(i, j)}\pi\right), \end{aligned}$$

and so, by the definitions of  $f$  and  $f'$ ,

$$0 \geq f(v_i, v_j) \geq f'(v'_i, v'_j).$$

We have shown that this holds for all  $i, j \in \{1, 2, \dots, n\}$  with  $i \neq j$ .

---

† Note that a disconnected diagram is more complicated than any of its connected components; so 6.14 provides another proof that every connected component is admissible if the whole diagram is.

Given that  $\Gamma$  is inadmissible, there must exist  $v \in V$  with  $f(v, v) < 0$ . Since  $v$  must be expressible as a linear combination of the basis vectors  $v_1, v_2, \dots, v_n$ , let  $\lambda_1, \lambda_2, \dots, \lambda_n$  be scalars such that

$$v = \sum_{i=1}^n \lambda_i v_i = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n.$$

Now define  $v' \in V'$  by

$$v' = \sum_{i=1}^n |\lambda_i| v'_i = |\lambda_1| v'_1 + |\lambda_2| v'_2 + \dots + |\lambda_n| v'_n.$$

(Recall that  $n' = \dim V' \geq n = \dim V$ . The coefficient of  $v'_i$  in the expression for  $v'$  is zero if  $i > n$ .)

Bilinearity of  $f$  yields that

$$0 > f(v, v) = f\left(\sum_{i=1}^n \lambda_i v_i, \sum_{j=1}^n \lambda_j v_j\right) = \sum_{i=1}^n \lambda_i^2 f(v_i, v_i) + \sum_{i \neq j} \lambda_i \lambda_j f(v_i, v_j).$$

But  $f(v_i, v_i) = 1$  for all  $i$ , and since  $f(v_i, v_j) \leq 0$  whenever  $i \neq j$  it follows that  $\lambda_i \lambda_j f(v_i, v_j) \geq |\lambda_i| |\lambda_j| f(v_i, v_j)$ . Moreover, since  $f'(v_i, v_j) \geq f'(v'_i, v'_j)$ , we have  $|\lambda_i| |\lambda_j| f(v_i, v_j) \geq |\lambda_i| |\lambda_j| f'(v'_i, v'_j)$  whenever  $i \neq j$ . Hence

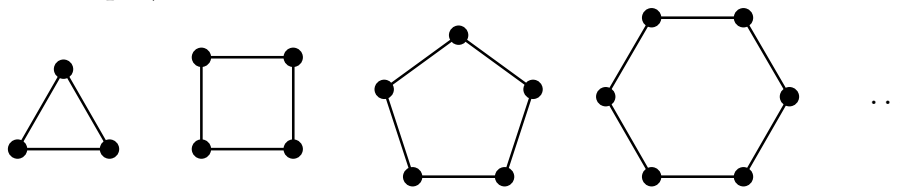
$$0 > \sum_{i=1}^n \lambda_i^2 + \sum_{i \neq j} |\lambda_i| |\lambda_j| f(v_i, v_j) \geq \sum_{i=1}^n |\lambda_i|^2 f'(v'_i, v'_i) + \sum_{i \neq j} |\lambda_i| |\lambda_j| f'(v'_i, v'_j).$$

But this last expression equals  $f'(\sum_{i=1}^n |\lambda_i| v'_i, \sum_{j=1}^n |\lambda_j| v'_j) = f'(v', v')$ , and we have shown that the element  $v' \in V'$  has the property that  $f'(v', v') < 0$ . So  $f'$  is not positive definite, and so  $\Gamma'$  is inadmissible, as required.  $\square$

We now need a long list of inadmissible diagrams. We defer the proofs for the time being, but here is the list.

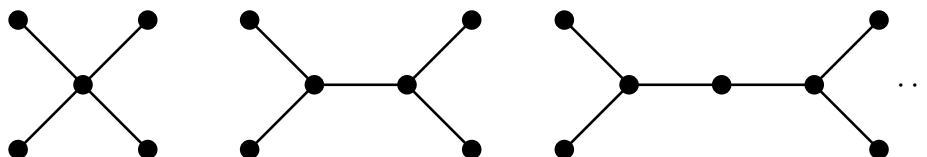


(i) The simple† circuits

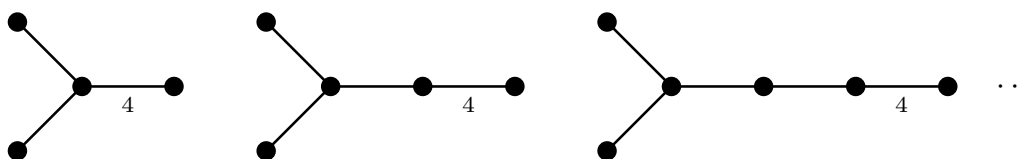


are all inadmissible.

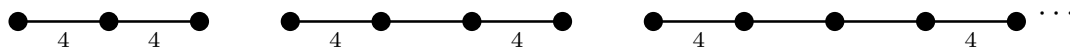
(ii) The following diagrams are all inadmissible.



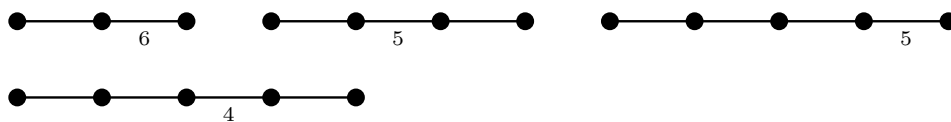
(iii) So are these.



(iv) And these.



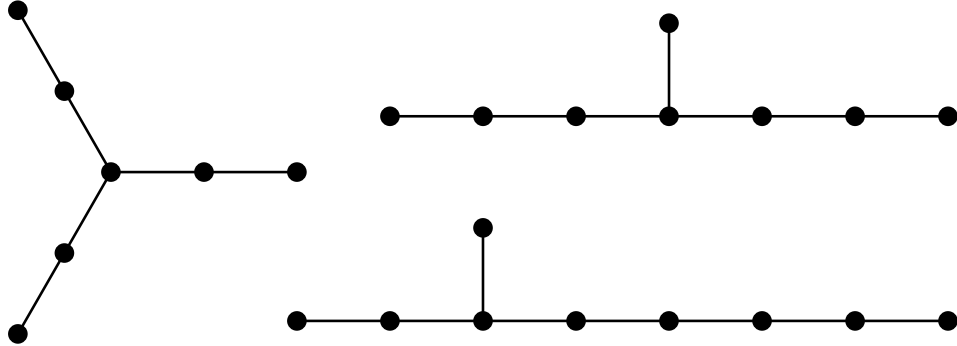
(v) Here are four more inadmissible diagrams.




---

† By *simple* we mean that the edge labels are all equal to three.

(vi) Finally, the following three diagrams are also inadmissible.



Accepting that it is true that the diagrams just listed are all inadmissible, let us determine exactly which diagrams might be admissible.

6.15 THEOREM *Let  $\Gamma$  be a connected admissible diagram. Then  $\Gamma$  is one of the following types.*

Type  $A_n$ : ( $n$  vertices, for any  $n \geq 1$ )

Type  $B_n$ : ( $n$  vertices, for any  $n \geq 2$ )

Type  $D_n$ : ( $n$  vertices, for any  $n \geq 4$ )

Type  $I_2(p)$ : (any  $p > 4$ )

Type  $H_3$ : 5

Type  $H_4$ : 5

Type  $F_4$ : 4

Type  $E_6$ : 6

Type  $E_7$ : 7

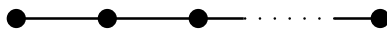
Type  $E_8$ : 8

**Proof.** If  $\Gamma$  has exactly one vertex then it is of type  $A_1$ , and if it has exactly two vertices then it is either of type  $A_2$  or  $B_2$ , or type  $I_2(p)$  for some  $p > 4$ . So we may assume that  $\Gamma$  has at least three vertices.

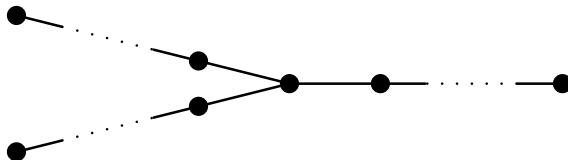
If  $\Gamma$  has an edge label which is 6 or greater, then  $\Gamma$  is more complicated than the first diagram in (iv) of our list of inadmissible diagrams, and by

Proposition 6.14 it follows that  $\Gamma$  is inadmissible, contradiction. So 3, 4 and 5 are the only labels that occur.

If  $\Gamma$  had a circuit then it would be more complicated than one of the simple circuits in (i) of our list of inadmissible diagrams, and if  $\Gamma$  had a vertex of valency 4 or more, then it would be more complicated than the first diagram in (ii) of our list of inadmissible diagrams. By 6.14, this is impossible. Similarly, it is impossible for  $\Gamma$  to have two or more vertices of valency 3, for otherwise  $\Gamma$  would be more complicated than one of the other diagrams in (ii) of our list of inadmissibles. These facts combine to tell us that  $\Gamma$  is either a string



with various labels 3, 4 or 5 on the edges, or else consists of three branches of various lengths emanating from the only vertex of valency 3,



again with variously labelled edges.

The diagrams in (iii) of the inadmissibles list show that  $\Gamma$  cannot have a vertex of valency 3 and an edge labelled 4 or more; so in the three branch cases  $\Gamma$  can have only simple edges (labelled 3). The first diagram in (vi) of the inadmissibles list shows that the three branches cannot all have length two or more; that is, at least one of the branches has length one. If two of the branches have length one then  $\Gamma$  is of type  $D_n$  for some  $n$ , and this is listed as a possibility in the theorem statement. So we can assume that exactly one of the branches has length one. Now if both the other branches had length three or more, then  $\Gamma$  would be more complicated than the second diagram in (vi) of the inadmissibles, which is impossible. So at least one of the other branches has length exactly two. The third branch can have length two, three or four, corresponding to types  $E_6$ ,  $E_7$  and  $E_8$ , but no more than that, or else  $\Gamma$  would be more complicated than the third diagram in (vi). So all the three branch possibilities are covered.

Suppose, on the other hand, that  $\Gamma$  is a string, so that there are exactly two vertices of valency 1, the rest having valency 2. The diagrams in (iv) show that  $\Gamma$  cannot have two edges labelled 4 or more. That is, there is at most one non-simple edge. If all the edges are simple then  $\Gamma$  is of type  $A_n$ ; so we may assume that there is exactly one non-simple edge. If the label on

this edge is 5, then its endpoints cannot both have valency 2, or  $\Gamma$  would be more complicated than the second diagram in (v) of the list of inadmissible diagrams. In other words, if the label is 5 then the non-simple edge is one of the two end edges. And  $\Gamma$  must be either of type  $H_3$  or  $H_4$ , since if it had five or more vertices it would be more complicated than the third diagram in (v). So it remains to deal with the cases when the non-simple edge is labelled 4. If the non-simple edge is an end edge then  $\Gamma$  is of type  $B_n$ . If not, then  $\Gamma$  must be of type  $F_4$ , since if it had five or more vertices it would be more complicated than the fourth diagram in (v). So all the string possibilities are covered too.  $\square$

### §6d Existence and inadmissibility proofs

Overlooking the fact that a few proofs have been skipped, and a technicality to be mentioned in a moment, we have now completely classified the finite groups of transformations of Euclidean space which are generated by reflections. For, if  $G$  is such a group, it must have a root system, and the root system must have a base, and the base must correspond to an admissible diagram. The technicality is that, for all we have proved so far, several different groups generated by reflections might give the same diagram. In fact, it is not too difficult to prove (although we will not do it) that if  $G$  is generated by reflections, then it is also generated by the reflections corresponding to the roots in any base for its root system. This means that the diagram does determine  $G$  up to isomorphism. So the classification theorem for Euclidean reflection groups—which we have not quite proved—is as follows.

**6.16 THEOREM** *There is a one-to-one correspondence between isomorphism classes of finite Euclidean reflection groups and diagrams whose connected components come from the list in Theorem 6.15.*

We have also yet to prove that all the types listed in Theorem 6.15 are actually admissible, and correspond to finite reflection groups. To show that the diagrams are admissible simply requires finding, in each case,  $n$  vectors in Euclidean space with the right configuration of angles. Proving that there is a corresponding finite reflection group is more difficult, and requires constructing the entire root system (so that Proposition 6.2 can be applied). Finally, we have yet to prove the inadmissibility of all those diagrams.

The proofs of inadmissibility all use the same method, and so we will leave most of them as exercises. They depend on the following lemma.

6.17 LEMMA Suppose that  $f$  is a bilinear form on a vector space  $V$  over the field  $\mathbb{R}$ , and suppose that  $v_1, v_2, \dots, v_n$  is a basis of  $V$ . If there exist scalars  $\lambda_i \geq 0$  that are not all zero and have the property that  $\sum_{i=1}^n \lambda_i f(v_i, v_j) \leq 0$  for all  $j \in \{1, 2, \dots, n\}$ , then  $f$  is not positive definite.

**Proof.** Suppose that there exist such  $\lambda_i$ , and put  $v = \sum_{i=1}^n \lambda_i v_i$ . Then clearly  $v \neq 0$ , since the  $\lambda_i$  are not all zero and the  $v_i$  are linearly independent. However,

$$\begin{aligned} f(v, v) &= f\left(v, \sum_{j=1}^n \lambda_j v_j\right) = \sum_{j=1}^n \lambda_j f(v, v_j) \\ &= \sum_{j=1}^n \lambda_j f\left(\sum_{i=1}^n \lambda_i v_i, v_j\right) = \sum_{j=1}^n \lambda_j \left(\sum_{i=1}^n \lambda_i f(v_i, v_j)\right) \leq 0 \end{aligned}$$

since  $\lambda_j (\sum_{i=1}^n \lambda_i f(v_i, v_j)) \leq 0$  for all  $j$ . Hence  $f$  is not positive definite.  $\square$

To apply this lemma in practice, one should attempt to find  $\lambda_i$  such that  $\sum_{i=1}^n \lambda_i f(v_i, v_j) = 0$  for all but one value of  $j$ . Since the values  $f(v_i, v_j)$  are known, this involves solving a system of  $n - 1$  homogeneous linear equations in the  $n$  unknowns  $\lambda_i$ . The solution will probably be unique up to a scalar multiple. Take any nonzero solution and see whether  $\sum_{i=1}^n \lambda_i f(v_i, v_j) \leq 0$  for the remaining value of  $j$ . It will be, in every case we need.

For example, let  $\Gamma$  be the third diagram in (v), and let  $v_1, v_2, \dots, v_9$  be the canonical basis of  $V_\Gamma$ . Choose the numbering so that vertex  $i$  is adjacent to vertex  $i + 1$  for all  $i$  from 2 to 8, and vertex 1 is adjacent to vertex 4. Then the values of  $f_\Gamma$  on the vectors of the canonical basis are as follows: firstly,  $f_\Gamma(v_i, v_i) = 1$  for all  $i$ , then

$$\begin{aligned} f_\Gamma(v_1, v_4) &= f_\Gamma(v_2, v_3) = f_\Gamma(v_3, v_4) = f_\Gamma(v_4, v_5) = -1/2 \\ f_\Gamma(v_5, v_6) &= f_\Gamma(v_6, v_7) = f_\Gamma(v_7, v_8) = f_\Gamma(v_8, v_9) = -1/2 \end{aligned}$$

and  $f_\Gamma(v_i, v_j) = 0$  in all other cases. If we now consider the equations

$\sum_{i=1}^n \lambda_i f(v_i, v_j) = 0$  for all  $j \neq 2$ , we find the requirements to be

$$\begin{aligned} 0 &= -\frac{1}{2}\lambda_8 + \lambda_9 \\ 0 &= -\frac{1}{2}\lambda_7 + \lambda_8 - \frac{1}{2}\lambda_9 \\ 0 &= -\frac{1}{2}\lambda_6 + \lambda_7 - \frac{1}{2}\lambda_9 \\ 0 &= -\frac{1}{2}\lambda_5 + \lambda_6 - \frac{1}{2}\lambda_7 \\ 0 &= -\frac{1}{2}\lambda_4 + \lambda_5 - \frac{1}{2}\lambda_6 \\ 0 &= -\frac{1}{2}\lambda_3 - \frac{1}{2}\lambda_1 + \lambda_4 - \frac{1}{2}\lambda_9 \\ 0 &= -\frac{1}{2}\lambda_1 + \lambda_4 \\ 0 &= -\frac{1}{2}\lambda_2 + \lambda_3 - \frac{1}{2}\lambda_4 \end{aligned}$$

and if we put  $\lambda_9 = c$  we quickly find that  $\lambda_8 = 2c$ ,  $\lambda_7 = 3c$ ,  $\lambda_6 = 4c$ ,  $\lambda_5 = 5c$ ,  $\lambda_4 = 6c$ ,  $\lambda_1 = 3c$ ,  $\lambda_3 = 4c$  and  $\lambda_2 = 2c$ . Now, lo and behold!, we see that  $\sum_{i=1}^n \lambda_i f(v_i, v_2) = \lambda_2 - \frac{1}{2}\lambda_3 = 0$ . So the conditions of the lemma are satisfied, and the form  $f_\Gamma$  is not positive definite. Indeed, we have found that the nonzero vector

$$v = 2v_2 + 4v_3 + 6v_4 + 5v_5 + 4v_6 + 3v_7 + 2v_8 + v_9 + 3v_1$$

satisfies  $f_\Gamma(v, v) = 0$ .

It actually works just like this in most of the other cases, and we wind up with a nonzero vector  $v$  such that  $f_\Gamma(v, v_j) = 0$  for all  $j$  (which certainly gives  $f_\Gamma(v, v) = 0$ ). Only the diagrams with edge labels of 5 give slightly more complicated calculations.

As for the existence proofs, again we will content ourselves with one example: type  $E_8$ . This is, in fact, the most difficult. We start with an orthogonal basis  $e_1, e_2, \dots, e_8$  of 8-dimensional Euclidean space such that each  $e_i$  has length  $1/\sqrt{2}$ . For example, in  $\mathbb{R}^8$ , we could choose  $e_i$  to be the 8-tuple whose  $i$ -th component is  $1/\sqrt{2}$  and whose other components are all 0. Define

$$\begin{aligned} S = \{ &\pm e_i \pm e_j \mid i \neq j \mid i, j \in \{1, 2, \dots, 8\}, i \neq j \} \\ &\cup \{ \frac{1}{2} \sum_{i=1}^8 \varepsilon_i e_i \mid \varepsilon_i = \pm 1 \text{ and } \prod_{i=1}^8 \varepsilon_i = 1 \}. \end{aligned}$$

There are 240 vectors altogether in  $S$ , since in the first piece there are 112 (since there are  $\binom{8}{2} = 28$  choices for the pair  $\{i, j\}$  and 4 choices for the signs)

and 128 in the other piece (since the first seven signs can be chosen arbitrarily, giving  $2^7$  possibilities, and then the last sign is determined uniquely).

It is not hard to check that  $u \cdot u = 1$  and  $u \cdot v \in \{-\frac{1}{2}, 0, \frac{1}{2}\}$  for all  $u, v \in S$  with  $u \neq \pm v$ . Thus the angle between two vectors in  $S$  is always either  $\pi/3$ ,  $\pi/2$  or  $2\pi/3$ , and so we have that

$$\rho_u(v) = \begin{cases} v - u & \text{if the angle is } \pi/3 \\ v & \text{if the angle is } \pi/2 \\ v + u & \text{if the angle is } 2\pi/3 \end{cases}$$

It needs to be checked that  $\rho_u(v) \in S$  in all cases. Again, this is not hard. The point is that once it has been checked for one pair  $\{u, v\}$ , permuting the coordinates yields many other pairs which need not be checked separately.

Define  $a_1 = \frac{1}{2}(e_1 - e_2 - e_3 - e_4 - e_5 - e_6 - e_7 + e_8)$  and  $a_2 = e_1 + e_2$ , and, for  $3 \leq i \leq 8$ , define  $a_i = -e_{i-1} + e_{i-2}$ . Then if we take  $B = \{a_i \mid 1 \leq i \leq 8\}$  it can be checked that the inner products  $a_i \cdot a_j$  are as they should be for the diagram of type  $E_8$ . (That is,  $a_i \cdot a_j = \cos(2\pi/3) = -\frac{1}{2}$  if the vertices  $i$  and  $j$  are adjacent,  $a_i \cdot a_j = 0$  for nonadjacent vertices.) It is an interesting fact that when an arbitrary root is expressed as a linear combination of the simple roots  $a_i$ , all the coefficients turn out to be integers. A root  $\sum_{i=1}^8 \gamma_i e_i \in S$  is positive if the largest  $i$  with  $\gamma_i \neq 0$  has  $\gamma_i > 0$ .

It is possible to explicitly describe the linear transformations  $g$  in the reflection group  $G$  corresponding to this root system as matrices relative to the basis  $e_1, \dots, e_8$ . Firstly, there are the  $8!$  permutation matrices, and  $2^7$  diagonal matrices with diagonal entries  $\pm 1$  and determinant 1. These generate a group of order  $2^7 8! = 5160960$  which is a subgroup  $H$  of  $G$ . (In fact,  $H$  is itself a Euclidean reflection group: it is of type  $D_8$ .) The idea now is to investigate the cosets of  $H$  in  $G$ . If  $v_1, \dots, v_8$  is any orthonormal basis of the space of eight-component row vectors then there are  $2^7 8!$  orthogonal matrices of the form  $xg$  where  $x \in H$  and  $g$  is the matrix whose rows are  $v_1, \dots, v_8$ . (These matrices are obtained from  $g$  by permuting the rows and multiplying an even number of rows by  $-1$ .) We proceed to describe a large number of orthonormal bases which give rise to elements of  $G$ .

Let  $\eta_1, \eta_2, \dots, \eta_8$  be signs  $\eta_i = \pm 1$  with  $\prod_{i=1}^8 \eta_i = 1$ , and define

$$v_{ij} = \begin{cases} \frac{3}{4} & \text{if } i = j \\ -\frac{1}{4}\eta_i\eta_j & \text{if } i \neq j. \end{cases}$$

Let  $v_i$  be the row whose  $j$ th entry is  $v_{ij}$ . Then  $v_1, \dots, v_8$  form an orthonormal basis and give rise to a coset of  $H$  in  $G$ . In fact this gives 64 cosets, corresponding to the 64 choices for the signs  $\eta_i$ . These 64 cosets give us another  $64|H| = 330301440$  elements of  $G$ . (Who would have thought that there would be so many  $8 \times 8$  orthogonal matrices whose entries are all plus or minus three quarters or one quarter!)

Choose a division of the set  $\{1, 2, \dots, 8\}$  into two subsets  $J$  and  $J'$  of four elements each—there are thirty-five ways of doing this—and let  $\varepsilon = \pm 1$ . Let  $J = \{j_1, j_2, j_3, j_4\}$  and  $J' = \{j_5, j_6, j_7, j_8\}$ , and let  $\alpha_{ij}$  be the  $(i, j)$ -entry of the matrix

$$X = \begin{pmatrix} 1 & \varepsilon & \varepsilon & \varepsilon & 0 & 0 & 0 & 0 \\ \varepsilon & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ \varepsilon & -1 & 1 & -1 & 0 & 0 & 0 & 0 \\ \varepsilon & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \varepsilon & \varepsilon & \varepsilon \\ 0 & 0 & 0 & 0 & \varepsilon & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & \varepsilon & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 & \varepsilon & -1 & -1 & 1 \end{pmatrix}$$

Let  $v_i$  be the vector whose  $k$ th entry is  $\frac{1}{2}\alpha_{ijk}$ ; then the matrix  $g$  whose rows are the  $v_i$  is in  $G$ . Since there were 35 possible partions of  $\{1, 2, \dots, 8\}$  as  $J \cup J'$  and two choices for  $\varepsilon$ , we have in fact obtained another 70 cosets of  $H$ . This gives  $70|H| = 361267200$  elements of  $G$ .

We have now described  $1 + 64 + 70 = 135$  cosets of  $H$  in  $G$ , and in fact this is all of them. The group  $G$  has  $135|H| = 696729600$  elements altogether.



## Index of notation

Perp	4	$\text{Aut}(G)$	57
Plus	5	$Z(G)$	62
$(\text{Mult})_\lambda$	5	$\text{Inn}(G)$	62
$\text{Sym}(S)$	7	$C_G(g)$	66
$i$	8	$\rho_a$	76
$(\text{Dot})_\lambda$	10	$\text{plc}(\mathbf{B})$	89
$O(V)$	10	$V_\Gamma$	96
$[G : H]$	39	$f_\Gamma$	96
$ G $	39		

# Index

## — A —

Abelian groups . . . . .	16	automorphisms . . . . .	10
additive group of a field . . . . .	17	of cyclic groups . . . . .	58, 59
admissible diagram . . . . .	96	of Klein 4-group . . . . .	58
angle between two vectors . . . . .	71	axis of symmetry . . . . .	73
automorphism of a group . . . . .	57		

## — B —

base . . . . .	89	bilinear form . . . . .	72
----------------	----	-------------------------	----

## — C —

canonical homomorphism . . . . .	52	closure under an operation . . . . .	21
Cantor, Georg . . . . .	27	under inversion . . . . .	9, 22
cardinality . . . . .	26, 29	under multiplication . . . . .	8
Cartesian coordinates . . . . .	71	codomain . . . . .	3
Cartesian product . . . . .	4	$\pi$ -commensurable . . . . .	80
central quotient . . . . .	63	conjugate elements . . . . .	63
centralizer . . . . .	67	cosets of a subgroup . . . . .	23
centre . . . . .	62	of $C_G(g)$ . . . . .	67
of a $p$ -group . . . . .	69	countable set . . . . .	27
class equation . . . . .	69	Coxeter diagram . . . . .	94
classes of a group . . . . .	63	cycle notation . . . . .	8
of $\text{Sym}\{1, 2, 3\}$ . . . . .	64	cycle type . . . . .	64
of $\text{Sym}\{1, 2, \dots, n\}$ . . . . .	64	cyclic groups . . . . .	11
of $\text{Sym}\{1, 2, 3, 4, 5\}$ . . . . .	65		
of the dihedral group of order 8 . . . . .	66		
of $\text{GL}_3(\mathbb{C})$ . . . . .	66		

— D —

dense subset.....	27	dihedral group of order $2n$ .....	73
desmic tetrahedra.....	81	distance between two points.....	71
determinant homomorphism.....	40	domain.....	3
dihedral group of order 8.....	9	dot product.....	9, 71

— E —

endomorphism.....	60	equivalence class.....	31
enumerable set.....	27	Euclidean space.....	9
equivalence relation.....	30	even permutation.....	41

— F —

First Isomorphism Theorem.....	54	functions.....	3
foundations.....	3	fundamental roots.....	92

— G —

general linear group.....	18	greatest common divisor.....	60
generating a group.....	11	group of transformations.....	9
graph.....	94		

— H —

homomorphism.....	40, 45	Homomorphism Theorem.....	54
injective.....	54	hyperplane.....	75
kernel of.....	52		
natural.....	52		
Sym{1, 2, 3, 4} to Sym{1, 2, 3}.....	42		

— I —

identity permutation . . . . .	8	inner automorphism . . . . .	61
image of a homomorphism . . . . .	53	inner product space . . . . .	9
inadmissible diagram . . . . .	96	invariant subspace . . . . .	79
index of a subgroup . . . . .	39	isomorphic groups . . . . .	12
inherited operation . . . . .	21	isomorphism . . . . .	45
injectivity of a homomorphism . . . . .	54		

— K —

kernel of a homomorphism . . . . .	52	Klein's four group . . . . .	12
------------------------------------	----	------------------------------	----

— L —

labelling function . . . . .	95	Latin square . . . . .	20
------------------------------	----	------------------------	----

— M —

map, mapping . . . . .	4	multiplicative group of a field . . . . .	17
modulus homomorphism . . . . .	41		

— N —

natural homomorphism . . . . .	52	normal subgroup . . . . .	46
negative roots . . . . .	89	normalized root system . . . . .	87

— O —

odd permutation . . . . .	41	orthogonal transformation . . . . .	10
operations as relations . . . . .	5	orthogonal group . . . . .	10
operation on a set . . . . .	15	orthonormal basis . . . . .	71
orthogonal direct sum . . . . .	80		

— P —

parallelogram law . . . . .	74	polygonal $\pi$ -commensurable set . . . . .	80
parity of a permutation . . . . .	41	position vectors . . . . .	71
permutation . . . . .	7	positive definiteness . . . . .	9
permutation multiplication . . . . .	7	positive linear combination . . . . .	89
perpendicularity relation . . . . .	4	positive roots . . . . .	89
$\pi$ -commensurable . . . . .	80	Pythagorean triples . . . . .	2

— Q —

quotient by an equivalence relation . . . . .	32
---	----

— R —

rational number . . . . .	27	root system . . . . .	84
reflection formula . . . . .	75	rotation . . . . .	78
regular polygon . . . . .	72	rotation matrix . . . . .	79
relations . . . . .	4		
reflexive . . . . .	30		
symmetric . . . . .	30		
transitive . . . . .	30		

— S —

sets . . . . .	3	subset multiplication . . . . .	45
sign of a permutation . . . . .	41	surjectivity of a homomorphism . . . . .	54
simple edge . . . . .	101	symmetric group . . . . .	7
simple roots . . . . .	92	symmetries of a square . . . . .	8
squares as structured sets . . . . .	5	symmetry . . . . .	1
structured sets . . . . .	5	intuitive definition . . . . .	4
subgroup . . . . .	21	precise definition . . . . .	6
normal . . . . .	46		

— T —

tetrahedron . . . . .	81	transformation . . . . .	4
		relation-preserving . . . . .	6

— V —

vector spaces, as structured sets . . . . .	5
---	---