The above logical identities are 'rules of logic' which are used all the time in mathematical reasoning. For instance, in a mathematical argument we might replace the statement 'It is not true that both $x \geq 0$ and $y \geq 1$' by the statement 'Either $x < 0$ or $y < 1$'. This is an example of the first of De Morgan's Laws above. We recognise (with a bit of thought) that the above two statements are logically equivalent.

For another example, if x is a real variable then we can replace the condition '$x > -1$ and $x^2 \geq 0$' by the condition '$x > -1$' because we know that, for real numbers, $x^2 \geq 0$ is always true. This is the same 'law of thought' or 'rule of logic' as the first property of T above.

Example As an example of the process of deducing new identities from a (small) basic set, we will show that $((p \wedge q) \to r) \leftrightarrow (p \to (q \to r))$ is a tautology, that is, we will show that $(p \wedge q) \to r$ and $p \to (q \to r)$ are logically equivalent. To do this, we will show that both terms are logically equivalent to the Boolean term $\neg p \vee (\neg q \vee r)$.

The following terms are equivalent:

$$(p \wedge q) \to r \qquad \neg(p \wedge q) \vee r \qquad (\neg p \vee \neg q) \vee r;$$

the first pair since $(a \to b) \leftrightarrow (\neg a \vee b)$ is a tautology and the second pair since $\neg(a \wedge b) \leftrightarrow (\neg a \vee \neg b)$ is a tautology.

Also the following terms are logically equivalent:

$$p \to (q \to r) \qquad p \to (\neg q \vee r) \qquad \neg p \vee (\neg q \vee r);$$

the first pair since $(a \to b) \leftrightarrow (\neg a \vee b)$ is a tautology and the second pair for the same reason.

Therefore the required identity follows since, by associativity, $(\neg p \vee \neg q) \vee r$ is logically equivalent to $\neg p \vee (\neg q \vee r)$.

The above list of logical identities may well seem familiar: if the reader has not already done so, then he or she should compare this list with that given as Theorem 2.1.1. Surely the similarity between these lists is no coincidence! Indeed it is not, and we will explain this in two ways.

The first is that the rules of logic which were used to establish Theorem 2.1.1 are simply the rules appearing in the above list. More precisely, the properties of '$\cap$', '$\cup$' and complementation '^{c}' are precisely analogous to those of '$\wedge$', '$\vee$' and '$\neg$' (look even at the words used in defining the set-theoretic operations).

As illustration, suppose that we are given sets X and Y. Let p be the proposition 'x is an element of X' and let q be 'x is an element of Y'. Then the statement 'x is an element of $X \cap Y$' means that x is in X and x is in Y, so is represented by

the proposition $p \wedge q$. Similarly, 'x is an element of $X \cup Y$' is represented by $p \vee q$ and 'x is not an element of X' by $\neg p$. If X is a subset of Y, we have that if x is in X then x is in Y. We therefore represent $X \subseteq Y$ by $p \rightarrow q$. Also, equality between sets, $X = Y$, is represented by logical equivalence: $p \leftrightarrow q$. Using this, we may translate any logical identity into a theorem about sets (and vice versa). For instance, $p \wedge q \leftrightarrow q \wedge p$ in Theorem 3.1.1 translates into $X \cap Y = Y \cap X$ in Theorem 2.1.1 (and $X \cap X^{\mathrm{c}} = \emptyset$ in Theorem 2.1.1 translates to $p \wedge \neg p \leftrightarrow F$ in Theorem 3.1.1 since $\emptyset$ and F correspond, as do U and T).

Another way to regard the similarity is to say that in each case we have an example of a Boolean algebra, as will be defined in Section 4.4, and that the properties expressed in 2.1.1 and 3.1.1 are just special cases of the defining properties of Boolean algebras.

The notion that the laws of reasoning might be amenable to an algebraic treatment, the idea of a 'logical calculus', seems to have appeared first in the work of Leibniz, a many-talented individual who, along with Newton, was one of the inventors of the integral and differential calculus. Leibniz' ideas on a logical calculus were not taken very seriously at the time and, indeed, for some time afterwards. It was only with Augustus De Morgan and, especially, George Boole, around the middle of the nineteenth century, that an algebraic treatment of logic was formalised. Boole noted that the 'logical operations', usually expressed using words such as 'and', 'or' and 'not', obey certain algebraic laws. He extracted those laws and came up with what is now termed 'Boolean algebra'.

Exercises 3.1

1. Here are a few examples of English-language propositions for you to write down in terms of simpler (constituent) propositions, as defined below.
   Let p be 'It is raining on Venus'.
   Let q be 'The Margrave of Brandenburg carries his umbrella'.
   Let r be 'The umbrella will dissolve'.
   Let s be 'X loves Y'.
   Let t be 'Y loves Z'.
   (a) Write down propositions in terms of p, q, r, s and t for the following.
      (i) 'If it is raining on Venus and the Margrave of Brandenburg carries his umbrella then the umbrella will dissolve'.
      (ii) 'If Y does not love Z and if it is raining on Venus then either X loves Y or the Margrave of Brandenburg carries his umbrella but not both'.

(b) Render into reasonable English each of the propositions expressed by the following:
(i) $(p \wedge q) \vee r$; (ii) $p \wedge (q \vee r)$;
(iii) $\neg p \rightarrow (s \wedge (r \rightarrow \neg t))$; (iv) $\neg(\neg s \vee \neg t) \rightarrow p$.

2. Write down the truth tables for each of the following Boolean terms and so decide which are tautologies and which are contradictions:
(i) $p \wedge (\neg q \vee p)$; (ii) $(p \wedge q) \vee r$;
(iii) $p \wedge \neg p$; (iv) $p \vee \neg p$;
(v) $(p \vee q) \rightarrow p$; (vi) $(p \wedge q) \rightarrow p$.

3. Which among the following Boolean terms are logically equivalent to each other?
$p \wedge (p \rightarrow q), q, (p \wedge q) \leftrightarrow p, p \rightarrow q, p \wedge q$.

4. Use the properties listed in Theorem 3.1.1 to establish the following:
(i) $(\neg p \leftrightarrow q) \leftrightarrow ((\neg q) \leftrightarrow p)$;
(ii) $((p \rightarrow \neg q) \wedge (p \rightarrow \neg r)) \leftrightarrow (\neg(p \wedge (q \vee r)))$;
(iii) $(p \rightarrow (q \vee r)) \leftrightarrow (\neg q \rightarrow (\neg p \vee r))$.

5. Suppose that X is a subset of Y. Let p be the proposition 'x is an element of X' and let q be the proposition 'x is an element of Y'. Write down a propositional term which represents the statement 'x is an element of $Y \backslash X$'. Hence or otherwise, establish the following identities for sets:
(i) $A \cap B = A \backslash B^{c}$;
(ii) $A \cup (B \backslash A) = A \cup B$;
(iii) $A \backslash (B \cup C) = (A \backslash B) \cap (A \backslash C)$;
(iv) $A \backslash (B \cap C) = (A \backslash B) \cup (A \backslash C)$.

3.2 Quantifiers

The logic of propositions that we discussed in the previous section captures only a small part of mathematical reasoning. It is merely a way of handling the logic of already formed propositions: it says nothing about how those propositions can be formed in the first place. If you look at various of the mathematical statements that we make in this book you will see that they typically involve functions (like that which squares a number or that which adds together any two numbers), relations (like that of one number being less than another or that of congruence modulo n) and constants (such as 0 and 1). They also involve quantifiers, the use of which is signalled by phrases such as 'for all', 'for every', 'there exists', 'we can find'. Mathematical logics much richer than the propositional logic of the previous section and containing all the above ingredients can be constructed and are rich enough for the expression of essentially all mathematics. We do

not discuss these here (for more see [Enderton, *Logic*], for example) but we do give a brief discussion of quantifiers, the use of which pervades mathematical reasoning.

Consider the following two statements:

(i) for every real number a there is a real number b such that $a \leq b$;
(ii) there is a real number b such that for every real number a we have $a \leq b$.

Think about these statements: we hope that you agree that the first is true (given a we can take $b = a + 1$ for example) and that the second is false (since it says that b is the largest real number and there is no such thing). We can introduce quantifiers as a mathematical shorthand which allows us to write assertions such as those above in a very compact (and unambiguous) way.

The **universal quantifier** is usually written as $\forall$ and read as 'for all' (also 'for every', 'for any', etc.): so the first phrase of (i) above can be abbreviated as '$\forall a$'. The use of an upside down A here is to remind one of its connection with the word 'all'. The word universal refers to the universe over which the variable 'x' varies. Unless we state otherwise, in our examples we will always take this to be the 'universe' of real numbers. The **existential quantifier** is usually written as $\exists$ and read as 'there exists' (also 'there is', 'we have' etc.). The use of a backwards E reminds one of the word 'exist'. Now the first phrase of (ii) above can be abbreviated as '$\exists b$'.

As further examples, $\forall x(x^2 \geq 0)$ is read as 'for all x, $x^2 \geq 0$' and $\exists x(x < 0)$ is read as 'there is an x with $x < 0$'. Things get more interesting when we combine quantifiers. For instance, $\forall x \forall y((x > 2 \wedge y > 2) \rightarrow (x + y < xy))$ reads as 'for all x and for all y if $x > 2$ and $y > 2$ then $x + y < xy$', which can be shortened to 'for all $x > 2$ and $y > 2$ we have $x + y < xy$'.

We can abbreviate the statements (i) and (ii) above as follows:

(i) $\forall a \exists b(a \leq b)$;
(ii) $\exists b \forall a(a \leq b)$.

The point to take from this example is that 'for all' and 'there exists' do not commute! The only formal difference between these statements is that the quantifiers have been interchanged and we noted that one is true whereas the other is false, so they are certainly not logically equivalent statements. (Quantifiers of the same kind do commute: $\forall x \forall y \ldots$ is equivalent to $\forall y \forall x \ldots$ and similarly for $\exists$.) Students can and do make errors of logic, and unjustified interchange of quantifiers is a common one! Using the formal symbols $\forall$ and $\exists$ can clarify the logic of a statement or argument.

We give two examples of valid and frequently used deductions which may be made when dealing with quantifiers and then we give an example of the kind of mistake which can be made when dealing with them.

First, negation interchanges $\forall$ and $\exists$ when it 'moves past' one of these quantifiers. That is, for any statement p (involving the variable x, so we will sometimes write $p(x)$ for emphasis), $\neg\forall x p$ is equivalent to $\exists x \neg p$ and $\neg\exists x p$ is equivalent to $\forall x \neg p$.

To illustrate the first, if $p(x)$ is '$x > 0$' then $\neg\forall x p(x)$ reads as 'not (for all x, $x > 0$)' or, more naturally, 'it is not the case that every x is greater than 0.' The formula $\exists x \neg p(x)$ reads as 'there exists x such that not $(x > 0)$' or, more naturally, 'there is some x which is not greater than 0'. This is logically equivalent to the first statement and is an example of the general rule that $\neg\forall x p(x)$ is equivalent to $\exists x \neg p(x)$.

The equivalence of the formulae $\neg\exists x p$ and $\forall x \neg p$ is the equivalence of the statements 'there is no x which satisfies p' and 'every x satisfies not-p' (that is, 'every x fails to satisfy p'). This actually follows completely formally from the first rule. To see this, apply the first rule with $\neg p$ in place of p to obtain that $\neg\forall x \neg p$ is equivalent to $\exists x \neg\neg p$; then use that $\neg\neg p$ is equivalent to p to deduce that $\neg\forall x \neg p$ is equivalent to $\exists x p$. It follows (negate each statement) that $\neg\neg\forall x \neg p$ is equivalent to $\neg\exists x p$ and hence that $\forall x \neg p$ is equivalent to $\neg\exists x p$.

For a second example of a correct deduction, first we notice a simple fact: from $\forall x(r(x) \to p(x))$ we may deduce the statement $\forall x(r(x)) \to \forall x(p(x))$. To see this, notice that if we know that it is always the case (for all x) that the statement $r(x)$ implies the statement $p(x)$ then, if we know that $\forall x(r(x))$ is true ($r(x)$ holds for every x) then $\forall x(p(x))$ holds ($p(x)$ holds for every value of x).

Now, assuming $\forall x(r(x) \to p(x))$ holds we have, since $\neg p(x) \to \neg r(x)$ is logically equivalent to $r(x) \to p(x)$ (being the contrapositive) that $\forall x(\neg p(x) \to \neg r(x))$ holds and hence, from the fact above, that $\forall x(\neg p(x)) \to \forall x(\neg r(x))$ holds. That is, from $\forall x(r(x) \to p(x))$ we may deduce $\forall x(\neg p(x)) \to \forall x(\neg r(x))$.

Here is an example of a non-implication: if we assume the truth of $\forall x(r(x) \to p(x) \vee q(x))$ then it does *not* follow that either $\forall x(r(x) \to p(x))$ or $\forall x(r(x) \to q(x))$ is true, that is, $\forall x(r(x) \to p(x) \vee q(x))$ does not imply $(\forall x(r(x) \to p(x))) \vee (\forall x(r(x) \to p(x)))$. To show this, we can take $r(x)$ to be $x^2 > 0$, $p(x)$ to be $x > 0$ and $q(x)$ to be $x < 0$. Then certainly $\forall x(r(x) \to p(x) \vee q(x))$ is true (it says that if the square of an element is strictly greater

than 0 then either the element is strictly greater than 0 or the element is strictly less than 0). But neither $\forall x(r(x) \to p(x))$ nor $\forall x(r(x) \to q(x))$ is true (for instance, the first says that if $x^2 > 0$ then $x > 0$, which is false). In this example it is quite easy to see that the statements are not equivalent but it is not unusual to see students make a mistake in logic based on this.

There are a number of further rules of deduction involving quantifiers which are used continually in mathematical argument and we refer to [Enderton, *Logic*], for example, for a full list. There is a full list, in the sense that one may write down a (small) number of shapes of rules of deduction from which all other rules of deduction follow. In principle, a computer can be programmed to use these rules in order to generate all valid mathematical deductions. The existence of such a 'generating set' of rules of deduction is related to Gödel's celebrated Completeness Theorem in mathematical logic. To state that theorem properly we would have to expand (considerably) on what we mean by 'valid' above and we refer to books on mathematical logic for this.

Another theorem from logic says that there exists nothing like truth tables for statements with quantifiers. We have seen that the method of computing truth tables allows us to decide, given any propositional term (and given enough time), whether that statement is a tautology or not. So we can, in principle, check any implication between propositional terms. There is no such method for statements involving quantifiers. We are not saying that no such method has been found: rather that it has been proved that there can be no such method!

Of course, the correctness or otherwise of many implications has been or can be established but there is no general method or collection of methods which will apply in all cases. That is, given two mathematical statements p and q there is no general method which we can apply in order to decide whether the implication $p \to q$ is correct: in particular, there is no general method which will either provide us with a deduction which starts with p as assumption and ends with q as conclusion or tell us that no such deduction exists.

It does follow, from what we said above, that one could programme a computer to start generating all correct mathematical implications: if an implication is correct it will eventually be output by the computer, and every implication output by the computer will be correct. But if you have a particular implication that you want to check (say, do the axioms for a group prove that such-and-such a formula holds in every group?) then, if it is correct, the computer will eventually output it (but you have no idea when) whereas, if it is false, you will never discover this just by waiting for the computer. If it is false it will never be output but you cannot discover this fact just by waiting to see whether it appears.

In the following exercises, you will be asked to determine whether certain statements involving quantifiers are true or false. This will not involve you

in waiting for an arbitrarily long time! Rather, the sentences will, like those discussed in the text, have a fairly clear truth value which can often be determined by rewriting the statement in English rather than in symbols. One of the main objects of the exercises is for you to become accustomed to 'translating' statements from symbols into words since this is such a common part of mathematical argument.

Exercises 3.2

1. Let $W(x)$ be the statement 'x likes whisky', let $S(x)$ be 'x is Scottish'.
   (a) Give English-language readings of the following statements with quantifiers.
      (i) $\forall x(S(x) \to W(x))$
      (ii) $\forall x(W(x) \to S(x))$
      (iii) $\exists x(S(x) \wedge \neg W(x))$
      (iv) $\neg\forall x(S(x) \to W(x))$
      (v) $\neg\forall x(S(x) \wedge W(x))$
      (vi) $\exists x\exists y(x \neq y \wedge W(x) \wedge W(y))$.
   (b) Write down formal statements which have the following meanings.
      (i) There is someone who is not Scottish and likes whisky.
      (ii) If there is someone who likes whisky then there is someone who is Scottish and likes whisky.
      (iii) Everyone who is not Scottish does not like whisky.
      (iv) There are at least two people who are not Scottish and who like whisky.
2. Decide which of the following are correct.
   (i) $(\forall x(r \to p)) \wedge (\forall x(p \to q))$ implies $\forall x(r \to q)$.
   (ii) $(\exists x(r \wedge p)) \wedge (\exists x(p \wedge q))$ implies $\exists x(r \wedge q)$.
   (iii) $\exists x\forall y(x < y)$ implies $\forall y\exists x(x < y)$.
   (iv) $\forall y\exists x(x < y)$ implies $\exists x\forall y(x < y)$.

3.3 Some proof strategies

In this section we gather together some of what we might call 'methods of proof' or 'proof strategies' that are used in the book. We do not claim that this list is complete or is in any way a classification of strategies. Our aim is simply to point out that, although the details of a proof depend on the specific situation, there are certain forms of argument that are used time and again in mathematical proofs. Longer proofs will use more than one of these strategies, possibly many times.

Certainly we are not giving recipes for constructing proofs: but the comments below might help you in understanding and even producing proofs since they make explicit some of their building blocks. After each 'strategy' we give some references to places where these are used in the text. We have left it for you to identify, within each argument that we reference, exactly where the strategy is used.

You should also look out for these strategies being used when you read proofs in this, and other, books.

Argument by contradiction We want to prove a statement so we prove that its negation leads to a contradiction. It is the law of the excluded middle (see the list in 3.1.1) which is the basis for the validity of this way of arguing. Examples: 1.1.1, 1.1.2, Example on p. 58, 4.1.3, analysis of groups of small order p. 225.

Argument by cases Sometimes it is easier or even necessary to treat different possibilites by different arguments, so we split into cases but it is necessary to make sure that all possibilities are covered! Examples: 1.3.2, 4.1.2, 4.2.7, 5.1.3 (i).

Sometimes this is combined with argument by contradiction: we split the range of all possibilities into two or more cases and show that all but one leads to a contradiction, so that case must hold. Example: the proof of 1.2.2.

Argument by contrapositive The idea here is that $p \to q$ is logically equivalent to $\neg q \to \neg p$. It may sometimes be easier to prove $\neg q \to \neg p$ than $p \to q$. Example: deductions with quantifiers on p. 139.

Choosing the least This method can be used when dealing with situations involving or indexed by positive integers. For example, we choose the least positive integer (or natural number), in some set, satisfying some condition and we want to establish some property of this integer. We show that if it did not have this property then we could produce a smaller integer in the set, contradicting the fact that our first choice was already supposed to be the smallest in the set. Examples: 1.1.1, 1.1.2, 4.2.3, 5.4.3.

Showing equality indirectly If they are sets, we can show $X = Y$ by showing that $X \subseteq Y$ and $Y \subseteq X$. If they are integers we can show $a = b$ by showing that $a \leq b$ and $b \leq a$. If they are positive integers we can show $a = b$ by showing that each divides the other. Examples: p. 81, 1.1.4, 5.4.3.

'Doing the same to both sides' This can be used for an equation or inequality for instance. Examples: 1.1.6, 1.3.3, 1.4.4, 5.1.1.

Mathematical induction Used to prove 'obvious' properties – they may seem obvious from what has been proved up to that point but, still, they should be proved (for instance extending a result from '$n = 2$' to general n). Examples: 1.3.2, 4.2.1 (iv).

Used where we do something to simplify (say by 'removing one term') and so reduce to the case covered by the induction hypothesis. Examples: 1.3.2, 1.3.3.

Used where we start with the induction hypothesis and 'add the next term to each side'. Example on p. 17.

Showing that a construction terminates If at each stage the construction produces, say, a natural number, and these numbers are strictly decreasing at each new stage then (by the well-ordering principle) the construction must stop. Examples: 1.1.5, 1.3.3 (though it is not explicitly written that way).

Use of key results Certain results are used time and time again in proofs of other results. Sometimes they are major theorems but sometimes they are just very useful lemmas. Examples: 1.1.3 used in 1.1.6, 1.4.3, 1.5.2; 5.2.4 used in 5.2.6 and Section 5.3; as well as more obvious examples like Fermat's and Euler's Theorems (1.6.3 and 1.6.7) and Lagrange's Theorem (5.2.3).

Use of definitions Many mathematical exercises are of the form 'prove that every set (or integer or ...) which satisfies property A also satisfies property B'. Before being able to attempt such an exercise, it is essential to have a clear idea of what properties A and B are! This may involve going back to an earlier chapter (or previous lecture notes) to find precise definitions.

Sometimes, one may be asked to explain why something is not true, in other words to give a 'disproof'. This can be done by giving a **counterexample**. To do this we give explicit values of the variables and show that, with these values, the result does not hold. For example if we are trying to show that some statement about integers is not true, we might be able to express our condition algebraically and arrive at something like 'if the condition is true then $ad - bc = a + d$'. If we are considering our variables as integers, we should now give explicit values (such as $a = 1$, $b = 2$, $c = 0$ and $d = 0$) to show that $ad - bc$ need not equal $a + d$ and so deduce that the original statement does not hold.

We emphasise that constructing a proof is quite different (and a lot harder!) than reading a proof. Of course, standard proofs may well combine many of the techniques above. In order to get some idea of which techniques might apply in any given case, it is best to try to write proofs as soon as possible in your

mathematical studies. Some proofs are straightforward to find in the sense that, if you understand the definitions, understand what is being assumed and can see where you are heading then it is rather obvious what steps to take in reaching that goal. But usually one needs some insight to guide one's efforts in finding a proof. Some proofs are based on a clever idea (for instance 1.3.4). Others, though they might not be so difficult to understand once found, require deep understanding of 'what is going on'. Fermat's, Euler's and Lagrange's Theorems surely come under this heading. A professional mathematician would not be likely to describe these as being, in the present-day context, deep theorems, simply because the ideas are now so familiar (to mathematicians). But in the contexts in which they were first proved, they required deep understanding of structure behind what is obvious and, indeed, they were instrumental in shaping some of the major concepts of mathematics which we can use so easily now.

To conclude this section, we will give some examples in the style of our end-of-section exercises, and consider what proof strategies might apply.

Example 1 Prove that, for any positive integer n the last digit of n (n written in base 10) is the same as the last digit of n^5.

As with many problems, the initial difficulty is in finding a mathematical formulation of the problem. In this case, we want to show that n and n^5 have the same last digit. This will happen if and only if $n^5 - n$ ends in a zero. Another way to say that a number ends in a zero is to say that that the number is divisible by 10. Thus, we can rephrase our original problem as: prove that, for any positive integer n, 10 divides $n^5 - n$.

As a general rule, the appearance of the words 'for any positive integer' suggests trying to use mathematical induction. We try this first: with the statement that 10 divides $1^5 - 1$ (the base case) being clear. So now suppose, for the inductive hypothesis, that 10 divides $n^5 - n$. We then consider $(n+1)^5 - (n+1)$. In order to proceed by induction, therefore, we need to be able to do something with $(n+1)^5$. An expansion of this expression requires the binomial theorem (1.2.1). This gives (after working out the binomial coefficients)

$$\begin{aligned}(n+1)^5 - (n+1) &= n^5 - n + 5n^4 + 10n^3 + 10n^2 + 5n + 1 - 1 \\ &= (n^5 - n) + 5(n^4 + n) + 10\,(n^3 + n^2).\end{aligned}$$

It is now almost clear that 10 divides the right-hand side. We know 10 divides $(n^5 - n)$ and (of course) $10\,(n^3 + n^2)$, so the proof will be complete once we show that 10 divides $5(n^3 + n^3)$. Clearly 5 divides this number so, by 1.1.6 (ii)

we are left with the problem of showing why 2 divides $n^3 + n^2 = n^2(n + 1)$. This is clear because (considering cases) either n is even so 2 divides n or n is odd (in which case 2 divides $n + 1$). This completes the proof by induction. However, that was a somewhat complicated proof of its type, so we just pause before proceeding to our next example to see if we could find alternatives to this proof.

As we saw, we considered $n^5 - n$. We could start by factorising this to get $n(n^4 - 1)$. By Fermat's Theorem, 1.6.3, $n^4 - 1$ is divisible by 5 (when n is not divisible by 5). Again if n is even, then 2 divides n, but if n is odd then 2 divides $n^4 - 1$ since then n^4 will be odd. Thus, using Fermat (a 'key result'), we see that 10 divides $n^5 - n$ except, possibly when five divides n. If 10 divides n, then n ends in a zero, and 10 divides n^5, so now we are left only with the case when 5 divides n, but 10 does not (so n is odd). In that case 5 divides n and 2 divides $n^4 - 1$, so 10 divides $n^5 - n$.

We have now seen two proofs of this fact, the second being a combination of cases and key results. There are many more proofs yet of this fact. The reader might try to find another using congruence classes modulo 10 and so a division into 10 cases. This illustrates the fact that it is worth thinking carefully about the possible strategies.

Example 2 As a second example, prove that if n^2 is odd then n is odd.

Again several proofs are possible, but before discussing these, the reader should perhaps pause and decide which seems the best to try.

In fact, the most straightforward one would be to consider the contrapositive statement: if n is even, then n^2 is even. It is clear that this holds, since if n is even 2 divides n and so 2 divides n^2 (in fact 4 would divide n^2 in that case). Thus if n^2 is odd then n is odd.

Example 3 Show that $\sqrt{2}$ cannot be written as a rational number (a quotient of two integers). It is worth checking our list of proof strategies to decide which one to try. The best possibility seems to be proof by contradiction. Accordingly, we suppose that $\sqrt{2}$ can be written in the form a/b for integers a/b, where we can suppose that all the common factors of a and b have been cancelled to give a reduced fraction. Then, squaring both sides would give $2 = a^2/b^2$ or $2a^2 = b^2$. This would mean that 2 divides b^2 and so b cannot be odd (otherwise b^2 would be odd). Since 2 therefore divides b, 4 divides b^2, that is $b^2 = 4c$ for some integer c. Then $2a^2 = 4c$ so $a^2 = 2c$ and a^2 is even so a is even. Thus 2 divides both a and b so the fraction was not reduced. This contradiction shows that $\sqrt{2}$ is not rational.

Exercises 3.3

In each of the following cases think about and discuss which proof strategies are likely to be helpful and write down at least one proof.

1. For any integer n, $n^2 + n + 1$ is not an even number.
2. If a, b are integers, then $a + b$ is odd precisely if one of a, b is odd.
3. If a, b are integers with $a + b$ an even integer, then $a - b$ is an even integer. Give a counterexample to show that if $a + b$ is even then ab need not be even.

Summary of Chapter 3

In this chapter, we discussed some ideas from elementary mathematical logic. The first section was concerned with propositions (statements which have a truth value) and ways to combine them using negation, conjunction, disjunction and implication. We also discussed a standard way to decide the truth values of a Boolean expression built from propositions and these operations, using truth tables. The second section introduced the idea of quantifiers: the universal quantifier ($\forall$), and the existential quantifier ($\exists$). Since these symbols are often used in mathematical arguments, our aim is for the reader to become familiar with them and to be able to use them freely. Finally, in Section 3.3 we considered some strategies of proof and illustrated them by examples.

4 Examples of groups

The mathematical concept of a group unifies many apparently disparate ideas. It is an abstraction of essential mathematical content from particular situations. Abstract group theory is the study of this essential content. There are several advantages to working at this level of generality. First, any result obtained at this level may be applied to many different situations, and so the result does not have to be worked out or rediscovered in each particular context. Futhermore, it is often easier to discover facts when working at this abstract level since one has shorn away details which, though perhaps pertinent at some level of analysis, are irrelevant to the broad picture.

Of course, to work effectively in the abstract one has to develop some intuition at this level. Although some people can develop this intuition by working only with abstract concepts, most people need to combine such work with the detailed study of particular examples, in order to build up an effective understanding.

That is why we have deferred the formal definition of a group until the third section of this fourth chapter. For you will see that you have already encountered examples of groups in Chapter 1, so, when you come to the definition of a group in Section 4.3, you will be able to interpret the various definitions and theorems which follow that in terms of the examples that you know. In Sections 4.1 and 4.2 we consider permutations: these provide further examples of groups and they have significantly different properties from the arithmetical groups of Chapter 1. A key section of this book is Section 4.3, in which the definition of a group is given. We illustrate this concept by many examples. Finally, in Section 4.4 we give examples of other kinds of algebraic structures.

4.1 Permutations

Definition Let X be a set. A **permutation** of X is a bijection from X to itself (in other words, a 'rearrangement' of the elements of X).

Thus, for example, the identity function, id_X, on any set X is a permutation of X (albeit a rather uninteresting one).

For finite sets X there are two notations available for expressing the action of a permutation of X. These are used in preference to the usual notation for functions.

The first of these, known as two-row notation, was introduced by Cauchy in a paper of 1815. To use this for a permutation π, list the elements of X in some fixed order $a,b,c,\ldots$, then write down a matrix with 2 rows and n columns which has $a,b,c,\ldots$ along the top row and has $\pi(a)$, $\pi(b)$, $\pi(c),\ldots$ along the second row (thus underneath each element x of X appears its image $\pi(x)$):

$$\begin{pmatrix} a & b & c & \ldots \\ \pi(a) & \pi(b) & \pi(c) & \ldots \end{pmatrix}.$$

Example Suppose that X is the set $\{a,b,c,d\}$ and that π is the permutation on X given by $\pi(a) = d$, $\pi(b) = c$, $\pi(c) = a$ and $\pi(d) = b$. Then the two-row notation for π is

$$\begin{pmatrix} a & b & c & d \\ d & c & a & b \end{pmatrix}.$$

If X is a finite set with, say, n elements then there is a bijection from the set of integers $\{1,2,\ldots,n\}$ to X. If we write x_i for the image of $i \in \{1,\ldots,n\}$, then we may think of such a bijection as being just a way of listing the elements of X as $\{x_1, x_2, \ldots, x_n\}$. When we use two-row notation to express permutations it saves time to write not x_1 but just 1, not x_2 but 2, ..., and so on. It even makes sense to 'identify' the elements of X with the integers $\{1,\ldots,n\}$. Hence all our discussion of permutations may be placed within the context of permutations of sets of the form $\{1,\ldots,n\}$ (thus we permute not the elements but the labels for the elements).

The fact that the function π is a bijection means that in this two-row notation no integer occurs more than once in the second row (since π is injective) and each integer in the set $\{1,\ldots,n\}$ occurs at least once in the second row (since π is surjective). Thus the second row is indeed a rearrangement, or permutation, of the first row.

Example Take $X = \{1,2,3\}$; there are $3! = 6$ permutations of this set:

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

Since permutations are functions from a set to itself, we may compose them. There are many examples in the following pages.

Definition Let n be a positive integer. Denote by $S(n)$ the set of all permutations of the set $\{1, \ldots, n\}$, equipped with the operation of composition (of functions). $S(n)$ is called the **symmetric group** on n symbols (or elements).

We now consider some properties of this operation of composition of permutations. We will use Greek letters ρ ('rho'), σ ('sigma') and τ ('tau') for permutations as well as π (with which we assume you are familiar).

Theorem 4.1.1 *Let n be a positive integer. Then $S(n)$ satisfies the following conditions:*

(Cl) *if π, σ are members of $S(n)$ then so is the composition $\pi\sigma$;*
(Id) *the identity function $\mathrm{id} = \mathrm{id}_{\{1,\ldots,n\}}$ is in $S(n)$;*
(In) *if π is in $S(n)$ then the inverse function π^{-1} is in $S(n)$.*

Also $S(n)$ has $n!$ elements.

Proof The three conditions (Cl), (Id), (In) (short for 'closure', 'identity' and 'inverse') may be rephrased as

(Cl) the composition of any two bijections is a bijection,
(Id) the indentity function is a bijection,
(In) the inverse of a bijection (exists and) is a bijection.

Each of these has already been established in Corollary 2.2.4.

To see that $S(n)$ has $n!$ elements, note that, in terms of the two-row notation, there are n choices for the entry in the second row under 1, for each such choice there are $n-1$ choices left for the entry under 2 (thus there are $n(n-1)$ choices for the first two entries of the second row), and so on. □

The notation $S(n)$ is sometimes used for the set of permutations (with the operation of composition) of any set with n elements. Of course this will not be the 'same' structure as that we have defined above but it is 'essentially' the same structure (refer to 'isomorphism of groups' in Section 5.3 below).

We will regard the operation of composition of permutations as a kind of 'multiplication'. Suppose that we have two permutations π, σ in $S(n)$ given in the two-row notation; how do we calculate the 'product' $\pi\sigma$? Since this is composition of functions, $\pi\sigma$ means: do σ, then do π. The result may be computed using two-row notation. An easy way to do this is to write (the two-row notation for) π underneath (that for) σ, and then to reorder the columns of π so that they occur in the order given by the second row of σ. This gives us four rows in which the second and third rows are identical. The two-row notation for the composition is obtained by deleting these identical rows, and writing only the first and fourth.

Example Consider the permutations in $S(5)$

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

The four-row array for computing $\pi\sigma$ is

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

Reordering the third and fourth rows together in the order determined by the second row gives

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

and so the composition is

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

This method is a little cumbersome to write down and so it is usually abbreviated as follows. The entry which will come below '1' (say) in the two-row notation for $\pi\sigma$ is found by looking at the entry below '1' in the two-row notation for σ – say that entry is k – and then looking below 'k' in the two-row notation for π: that entry (m say) is the one to place below '1' in the two-row notation for $\pi\sigma$. Proceed in the same way for $2, \ldots, n$.

It should be clear why this works: the first function, σ, takes 1 to k (since 'k' occurs below '1' in the notation for σ), and then the second function, π, takes

k to m – therefore the composition takes 1 to m, and so 'm' is placed below '1' in the notation for $\pi\sigma$.

Example In $S(3)$ we have

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

Notice that

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

Hence the operation of composition is **non-commutative** in the sense that $\pi\sigma$ need not equal $\sigma\pi$. Therefore, it is important to remember that we are using the convention that $\pi\sigma$ is the function obtained by applying σ and then applying π.

Example Consider $S(5)$: write id for the identity function, and take

$$\pi=\begin{pmatrix}1&2&3&4&5\\2&3&1&5&4\end{pmatrix},\sigma=\begin{pmatrix}1&2&3&4&5\\1&2&4&5&3\end{pmatrix},\tau=\begin{pmatrix}1&2&3&4&5\\2&1&5&4&3\end{pmatrix}$$

then (as the reader should check)

$$\sigma\pi=\begin{pmatrix}1&2&3&4&5\\2&4&1&3&5\end{pmatrix},\pi\sigma=\begin{pmatrix}1&2&3&4&5\\2&3&5&4&1\end{pmatrix},\tau^2=\tau\tau=\begin{pmatrix}1&2&3&4&5\\1&2&3&4&5\end{pmatrix}=\mathrm{id},$$

$$\sigma^2=\begin{pmatrix}1&2&3&4&5\\1&2&5&3&4\end{pmatrix},\quad\sigma^3=\sigma^2\sigma=\begin{pmatrix}1&2&3&4&5\\1&2&3&4&5\end{pmatrix}=\mathrm{id},$$

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

$$\pi^5=\begin{pmatrix}1&2&3&4&5\\3&1&2&5&4\end{pmatrix},\ \pi^6=\begin{pmatrix}1&2&3&4&5\\1&2&3&4&5\end{pmatrix}=\mathrm{id}.$$

It must be stressed that the reader probably will understand little of what follows in this and the next section, unless complete confidence in multiplication of permutations has been acquired. With this in mind, a large selection of calculations is provided in the exercises at the end of this section.

The two-row notation is also very useful when calculating the inverse of a permutation. The inverse is calculated by exchanging the upper and lower rows, and then reordering the columns so that the entries on the upper row occur in the natural order.

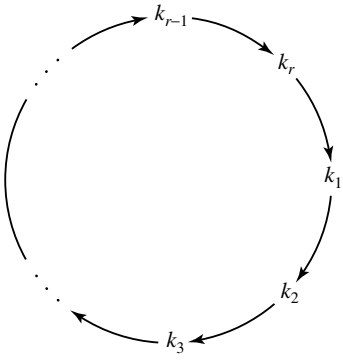

Fig. 4.1

Example In $S(7)$ the inverse of

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

is

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

We now consider the other notation for permutations.

Definition A permutation $\pi \in S(n)$ is **cyclic** or a **cycle** if the elements $1, \ldots, n$ may be rearranged, as say $k_1, \ldots, k_r, k_{r+1}, \ldots, k_n$ (we allow the possibilities that $r + 1 = 1$ or $r = n$), in such a way that π fixes each of $k_{r+1}, \ldots, k_n$ and 'cycles' the remainder, sending k_1 to k_2 sending k_2 to $k_3, \ldots$, sending k_{r-1} to k_r and finally sending k_r back to k_1. The integer r (that is, the number of elements in, or the length of, the cycling part) is called the **length** of π. (The algebraic significance of this integer will be explained later.) We say that the length of the identity permutation is 1. A cycle of length 2 is called a **transposition**. A cycle of length r is termed an ***r*-cycle**.

There is a special notation for cycles: write down, between parentheses, the integers which are moved by the cycle, in the order in which they are moved. Thus the cycle above could be denoted by $(k_1 k_2 \ldots k_r)$. See Fig. 4.1.

The point of the cycle at which to start may be chosen arbitrarily, so for any given cycle there will be a number of ways (equal to its length) of writing such a notation for it. For example, if π is the member of $S(5)$ which sends 1 to 3,

3 to 4, 4 to 1, and fixes 2 and 5 (so π is a cycle of length 3), then π may be written using this notation as (1 3 4), or as (3 4 1), or as (4 1 3).

Example

$$\begin{pmatrix}1&2&3&4&5\\2&3&1&4&5\end{pmatrix}, \begin{pmatrix}1&2\\2&1\end{pmatrix}, \begin{pmatrix}1&2&3&4\\4&1&3&2\end{pmatrix}, \begin{pmatrix}1&2&3&4&5&6&7\\4&7&6&2&1&5&3\end{pmatrix}$$

are cycles of lengths 3, 2, 3 and 7 respectively, but

$$\begin{pmatrix}1&2&3&4&5\\2&1&4&5&3\end{pmatrix}, \begin{pmatrix}1&2&3&4\\2&1&4&3\end{pmatrix}, \begin{pmatrix}1&2&3&4&5&6&7\\2&5&1&7&3&4&6\end{pmatrix}$$

are not cycles. Notations for the four cycles listed are (1 2 3), (1 2) (a transposition), (1 4 2) and (1 4 2 7 3 6 5).

Note that in the definition of cyclic permutation the case $r + 1 = 1$ corresponds to the permutation which does not move anything – in other words to the identity permutation id (which is therefore a cycle: its cycle notation would be empty, so we continue to write it as 'id'). The case $r = n$ is the case of a cycle which moves every element (the fourth, but not the first, cycle in the above example is of this kind).

Definition Let π and σ be elements of $S(n)$. Then π and σ are **disjoint** if every integer in $\{1, \ldots, n\}$ which is moved by π is fixed by σ and every integer moved by σ is fixed by π (we say that π **moves** $k \in \{1, \ldots, n\}$ if $\pi(k) \neq k$, otherwise π **fixes** k).

Theorem 4.1.2 *If π and σ are disjoint permutations in $S(n)$, then π and σ commute, that is, $\pi\sigma = \sigma\pi$.*

Proof For any permutation ρ in $S(n)$, let $\mathrm{Mov}(\rho)$ be the set of integers in $\{1, 2, \ldots, n\}$ which are moved by ρ. More formally

$$\mathrm{Mov}(\rho) = \{m : 1 \leq m \leq n \quad \text{and} \quad \rho(m) \neq m\}.$$

To say that π and σ are disjoint is just to say that the intersection of $\mathrm{Mov}(\pi)$ with $\mathrm{Mov}(\sigma)$ is empty.

We have the following possibilities for $m \in \{1, \ldots, n\}$:

$m \in \mathrm{Mov}(\pi)$;
$m \in \mathrm{Mov}(\sigma)$;
m is in neither $\mathrm{Mov}(\pi)$ nor $\mathrm{Mov}(\sigma)$.

In the first case m is sent to $\pi(m)$ by both $\pi\sigma$ and $\sigma\pi$. For we have $\pi\sigma(m) = \pi(\sigma(m)) = \pi(m)$ (since m is moved by π it is fixed by σ): on the other hand we have $\sigma\pi(m) = \sigma(\pi(m)) = \pi(m)$. The last equality follows since $\pi(m)$ is moved by π (so is fixed by σ): for otherwise we would have $\pi(\pi(m)) = \pi(m)$ and so, since π is 1-1, $\pi(m) = m$, a contradiction.

The other two cases are dealt with by similar arguments and we leave these to the reader. Thus $\pi\sigma = \sigma\pi$, since they have the same effect on the elements of $\{1, \ldots, n\}$. □

You might find it useful to go through the above proof with particular choices of disjoint π and σ.

Remark As we have already noted, the conclusion of 4.1.2 can fail for non-disjoint cycles. For another example, consider, in $S(3)$, $(1\ 2)(1\ 3) = (1\ 3\ 2)$, whereas $(1\ 3)(1\ 2) = (1\ 2\ 3) \neq (1\ 3\ 2)$.

The next result says that any permutation may be written as a product of disjoint cycles. But first we present an example illustrating how to 'decompose' a permutation in this way.

Example Let π be the following permutation in $S(14)$.

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \\ 4 & 9 & 10 & 7 & 5 & 2 & 6 & 13 & 1 & 3 & 11 & 12 & 14 & 8 \end{pmatrix}.$$

We begin by considering the repeated action of π on 1: π sends 1 to 4, which in turn is sent to 7, which is sent to 6, to 2, to 9, and then back to 1. So we find the 'circuit' to which 1 belongs, and write down the cycle in $S(14)$ that corresponds to this 'circuit', namely (1 4 7 6 2 9). Now we look for the first integer in $\{1, \ldots, 14\}$ that is not moved by this cycle: that is 3. The 'circuit' to which 3 belongs takes 3 to 10, which in turn goes back to 3. The cycle of $S(14)$ corresponding to this is (3 10); note that this cycle is disjoint from (1 4 7 6 2 9). The next integer which has not yet been encountered is 5: π fixes 5, so we do not need to write down a cycle for 5. The next integer not yet treated is 8: the cycle corresponding to the repeated action of π on 8 is (8 13 14); note that this is disjoint from each of the other two cycles found. Finally π fixes both 11 and 12. Thus we obtain an expression of π as a product of disjoint cycles:

$$\pi = (1\ 4\ 7\ 6\ 2\ 9)(3\ 10)(8\ 13\ 14).$$

By Theorem 4.1.2, we may rearrange the order of the cycles occurring, say as

$$(3\ 10)(1\ 4\ 7\ 6\ 2\ 9)(8\ 13\ 14)$$

but the actual cycles which occur are uniquely determined by π. See Fig. 4.2.

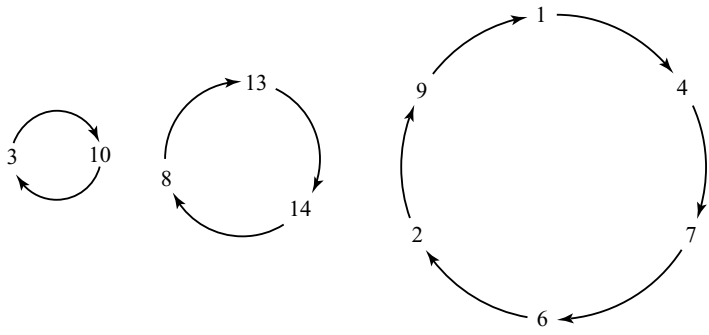

Fig. 4.2

Theorem 4.1.3 *Let π be an element of $S(n)$. Then π may be expressed as a product of disjoint cycles. This* ***cycle decomposition*** *of π is unique up to rearrangement of the cycles involved.*

Proof First look for the smallest integer which is not fixed by π: suppose that this is k. Apply π successively to k: let k_1 be k, k_2 be $\pi(k_1)$, k_3 be $\pi(k_2)$ and so on. Since the set $\{1,2,\ldots,n\}$ is finite, we will obtain repetitions after sufficiently many steps. Let r be the smallest integer such that k_r equals k_s for some s strictly less than r. If s were greater than 1 then we could write

$$\pi(k_{s-1}) = k_s = k_r = \pi(k_{r-1}).$$

Since π is injective, we deduce $k_{s-1} = k_{r-1}$, contrary to the minimality of r. It follows that s is 1, and so $(k_1 k_2 \ldots k_r)$ is an r-cycle.

Repeat the process for the smallest integer not fixed by π and not in the set $(k_1, k_2, \ldots, k_r)$ of integers already encountered. Continuing in this way, we obtain an expression of π as a product of disjoint cycles. From the construction it follows that the decomposition will be unique up to rearrangement of the cycles. □

You should note that the proof just given simply formalises the procedure used in the preceding example.

Two further examples of cycle notation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 7 & 2 & 10 & 12 & 5 & 4 & 8 & 1 & 6 & 3 & 9 & 11 \end{pmatrix} = (1\ \ 7\ \ 8)(3\ \ 10)(4\ \ 12\ \ 11\ \ 9\ \ 6),$$

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 4 & 7 & 8 & 6 & 2 & 10 & 3 & 9 & 5 & 1 \end{pmatrix} = (1\ \ 4\ \ 6\ \ 10)(2\ \ 7\ \ 3\ \ 8\ \ 9\ \ 5).$$

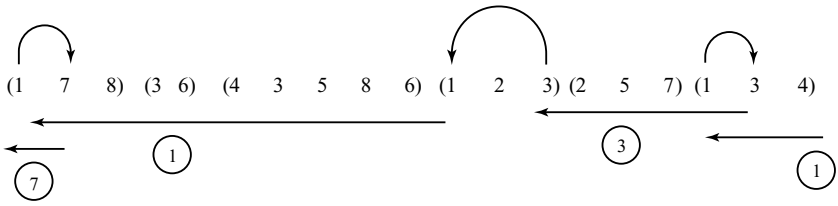

Fig. 4.3

In order to multiply together two permutations which are written using cycle notation, one can write down their two-row notations, multiply, and then write down the cycle notation for the result. But this is a cumbersome process, and the multiplication is best done directly. The basic manipulation involved is what we will call a switch. Suppose we are given a product, π, of cycles and we want to compute its cycle decomposition. We visualise the effect of π on an integer i moving from right to left, encountering the various cycles, possibly being switched to a new value at each encounter. To switch i, seek the first occurrence of i to the left of its present position. This lies in a cycle of π, and i is now switched to the number, k say, to which this cycle takes i. Now think of k continuing to move to the left, and repeat this switching process until the left-hand end is reached. The number, m say, which finally emerges at the left-hand end is $\pi(i)$.

The multiplication is carried out by repeating these switches, starting the process with each integer in $\{1,2,\ldots,n\}$ in sequence if the result is to be written in two-row notation, or in the order determined as the process continues otherwise. The method is illustrated by an example.

Example 1 Compute the cycle decomposition of the product π:

$$(1\ 7\ 8)(3\ 6)(4\ 3\ 5\ 8\ 6)(1\ 2\ 3)(2\ 5\ 7)(1\ 3\ 4).$$

Start with the integer 1 at the right-hand end of the above product. The first cycle encountered involves 1, switching it to 3. The number 3 continues to move to the left, and is switched back to 1 by the third cycle from the right since this cycle takes 3 to 1. Now 1 continues to move to the left, and is switched to 7 by the last cycle encountered. Therefore the product sends 1 to 7. See Fig. 4.3.

If we want to write the result in cycle notation, then it is most convenient to repeat the process next starting with the integer $7 = \pi(1)$: 7 is switched to 2; 2 to 3; then 3 goes to 5. Therefore 7 is sent to 5.

Continuing in this way ($5 = \pi(7)$ is treated next), we obtain an 8-cycle (1 7 5 8 3 6 4 2). Now we look for the first integer which has not yet been

'fed into' the right-hand end: in this case there are none, so the answer is just the above 8-cycle.

Example 2 In order to give further examples of multiplication of cycles, and to illustrate Theorem 4.1.1, we present the complete multiplication table for $S(3)$. The entry at the intersection of the row labelled σ and the column labelled τ is $\sigma\tau$.

	id	(123)	(132)	(12)	(13)	(23)
id	id	(123)	(132)	(12)	(13)	(23)
(123)	(123)	(132)	id	(13)	(23)	(12)
(132)	(132)	id	(123)	(23)	(12)	(13)
(12)	(12)	(23)	(13)	id	(132)	(123)
(13)	(13)	(12)	(23)	(123)	id	(132)
(23)	(23)	(13)	(12)	(132)	(123)	id

Example 3 The following permutations in $S(4)$ have a multiplication table as shown:

id; (1 3 4 2); (1 4)(2 3); (1 2 4 3); (2 3); (1 4); (1 2)(3 4); (1 3)(2 4).

	id	(1342)	(14)(23)	(1243)	(23)	(14)	(12)(34)	(13)(24)
id	id	(1342)	(14)(23)	(1243)	(23)	(14)	(12)(34)	(13)(24)
(1342)	(1342)	(14)(23)	(1243)	id	(13)(24)	(12)(34)	(23)	(14)
(14)(23)	(14)(23)	(1243)	id	(1342)	(14)	(23)	(13)(24)	(12)(34)
(1243)	(1243)	id	(1342)	(14)(23)	(12)(34)	(13)(24)	(14)	(23)
(23)	(23)	(12)(34)	(14)	(13)(24)	id	(14)(23)	(1342)	(1243)
(14)	(14)	(13)(24)	(23)	(12)(34)	(14)(23)	id	(1243)	(1342)
(12)(34)	(12)(34)	(14)	(13)(24)	(23)	(1243)	(1342)	id	(14)(23)
(13)(24)	(13)(24)	(23)	(12)(34)	(14)	(1342)	(1243)	(14)(23)	id

We finish the section by describing how to write down the inverse of a cycle: one simply reverses the order of the terms which appear (and then, if one wishes to, rewrites the resulting cycle with the smallest integer first). For example, $(1\ 2\ 3\ 4\ 5)^{-1} = (5\ 4\ 3\ 2\ 1) = (1\ 5\ 4\ 3\ 2)$.

It follows that if a permutation is written as a product of disjoint (hence commuting) cycles then the inverse is found by applying this process to each of its component cycles. If a permutation is written as a product of not necessarily disjoint cycles then the order of the components must also be reversed, because $(\pi\sigma)^{-1} = \sigma^{-1}\pi^{-1}$ (by 2.2.4(i)).

Permutations were important in the development of group theory, in that permutation groups of the roots of a polynomial were a key feature of Galois' work on solvability of polynomial equations by radicals. They also figure in

the work of Lagrange, Cauchy and others, as actions on polynomials (see the proof of Theorem 4.2.8 below). For more on this, see the notes at the end of Section 4.3.

The reader is strongly advised to attempt the exercises that follow before continuing to the next section.

Exercises 4.1

Let π_1, π_2, π_3, π_4 and π_5 be the following permutations:

$$\pi_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 2 & 1 & 6 & 5 & 4 & 9 & 8 & 7 \end{pmatrix},$$

$$\pi_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \end{pmatrix},$$

$$\pi_3 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 4 & 5 & 6 & 7 & 8 & 9 & 1 & 2 & 3 \end{pmatrix},$$

$$\pi_4 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 9 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 11 & 12 & 10 \end{pmatrix},$$

$$\pi_5 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 12 & 7 & 2 & 8 & 4 & 6 & 3 & 9 & 5 & 1 & 11 & 10 \end{pmatrix}.$$

1. Calculate the following products:
   $\pi_1\pi_2$, $\pi_2\pi_3$, $\pi_3\pi_1$, $\pi_3\pi_2$, $\pi_2\pi_1\pi_3$, $\pi_2\pi_2\pi_2$, $\pi_4\pi_5$, $\pi_5\pi_4$, $\pi_1\pi_3$, $\pi_2\pi_2$, $\pi_2\pi_1$, $\pi_3\pi_3$, $\pi_2\pi_1\pi_2$, $\pi_2\pi_3\pi_2$, $\pi_4\pi_4$, $\pi_5\pi_5$.
2. Find the inverses of π_1, π_2, π_3, π_4 and π_5.
3. Write each permutation in Example 4.1.1 as a product of disjoint cycles.
4. Compute the following products, writing each as a product of disjoint cycles:
   (i) (1 2 3 4 5)(1 3 6 8)(6 5 4 3)(1 3 6 8);
   (ii) (1 12 10)(2 7 3)(4 6 9 5)(1 3)(4 6)(7 9);
   (iii) (1 4 7)(2 5 8)(3 6 9)(1 2 3 4 5 6 7 8 9)(10 11).
5. Write down the complete multiplication table for the following set of permutations in $S(4)$:

   id, (1 2 3 4), (1 3)(2 4), (1 4 3 2), (1 3), (2 4), (1 2)(3 4) and (1 4)(2 3).

6. The study of the symmetric group $S(52)$ has engaged the attention of many sharp minds. As an aid to their investigations, devotees of this pursuit make use of a practical device which provides a concrete realisation of $S(52)$. This device is technically termed a 'deck of playing cards'. We now give

some exercises based on the permutations of these objects. Since it is a time-consuming task even to write down a typical permutation of 52 objects, we will work with a restricted deck which contains only 10 cards – say the ace to ten of spades (denoted $1, \ldots, 10$) for definiteness.

Permutations of the deck are termed 'shuffles' and 'cuts': let us regard these as elements of $S(10)$.

Define s to be the 'interleaving' shuffle which hides the top card:

$$s = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 2 & 4 & 6 & 8 & 10 & 1 & 3 & 5 & 7 & 9 \end{pmatrix}.$$

Let t be the interleaving shuffle which leaves the top card unchanged:

$$t = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1 & 3 & 5 & 7 & 9 & 2 & 4 & 6 & 8 & 10 \end{pmatrix}.$$

Finally, let c be the cut:

$$c = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 6 & 7 & 8 & 9 & 10 & 1 & 2 & 3 & 4 & 5 \end{pmatrix}.$$

Show that cutting the deck according to c and then applying the shuffle s has the same effect as the single shuffle t.

Write s, t, c, cs and scs using cycle notation.

For each of these basic and combined shuffles s, t, c, cs and scs, how many times must the shuffle be repeated before the cards are returned to their original positions?

4.2 The order and sign of a permutation

Definition Let π be a permutation. The positive **powers**, π^n, of π are defined inductively by setting $\pi^1 = \pi$ and $\pi^{k+1} = \pi \cdot \pi^k$ (k a positive integer). We also define the negative powers: $\pi^{-k} = (\pi^{-1})^k$ where k is a positive integer, and finally set $\pi^0 = \mathrm{id}$.

The following index laws for powers are obtained using mathematical induction.

Theorem 4.2.1 *Let π be a permutation and let r, s be positive integers. Then*

(i) $\pi^r \pi^s = \pi^{r+s}$,
(ii) $(\pi^r)^s = \pi^{rs}$,
(iii) $\pi^{-r} = (\pi^r)^{-1}$,
(iv) *if π, σ are permutations such that $\pi\sigma = \sigma\pi$ then $(\pi\sigma)^r = \pi^r\sigma^r$.*

Proof (i) The proof is by induction on r. If $r = 1$, then $\pi \cdot \pi^s = \pi^{s+1}$ by definition. Now suppose that

$$\pi^r \pi^s = \pi^{r+s}.$$

Then

$$\begin{aligned} \pi^{r+1}\pi^s &= (\pi \cdot \pi^r)\pi^s \\ &= \pi(\pi^r \pi^s) \\ &= \pi(\pi^{r+s}) \\ &= \pi^{r+s+1} \end{aligned}$$

as required.

The proofs of (ii) and (iii) are also achieved using mathematical induction and are left as exercises (for (iii) use the fact (2.2.4(i)) that $(fg)^{-1} = g^{-1}f^{-1}$ if f and g are bijections from a set to itself).

The fourth part also is proved by induction. We actually need a slightly stronger statement: that $(\pi\sigma)^k = \pi^k\sigma^k$ and $\sigma\pi^k = \pi^k\sigma$ (we use the second equation within the proof). By assumption the result is true for $k = 1$. So suppose inductively that $(\pi\sigma)^k = \pi^k\sigma^k$ and $\sigma\pi^k = \pi^k\sigma$. Then

$$\begin{aligned} (\pi\sigma)^{k+1} &= \pi\sigma(\pi\sigma)^k \quad \text{(by definition)} \\ &= \pi\sigma\pi^k\sigma^k \quad \text{(by induction)} \\ &= \pi\pi^k\sigma\sigma^k \quad \text{(also by induction)} \\ &= \pi^{k+1}\sigma^{k+1} \quad \text{(by definition)} \end{aligned}$$

Also

$$\begin{aligned} \sigma\pi^{k+1} &= \sigma\pi\pi^k = \pi\sigma\pi^k \quad \text{(by assumption)} \\ &= \pi\pi^k\sigma \quad \text{(by induction)} \\ &= \pi^{k+1}\sigma \quad \text{(by definition).} \end{aligned}$$

So we have proved both parts of the induction hypothesis for $k + 1$ and the result therefore follows by induction. □

Theorem 4.2.2 *Let π be an element of $S(n)$. Then there is an integer m, greater than or equal to 1, such that $\pi^m = \mathrm{id}$.*

Proof Consider the successive powers of π: $\pi; \pi^2; \pi^3; \ldots$ Each of these powers is a bijection from $\{1, \ldots, n\}$ to itself. Since there are only finitely many such functions (4.1.1) there must be repetitions within the list: say $\pi^r = \pi^s$ with

$r < s$. Since π^{-1} exists, we may multiply each side by π^{-r} to obtain (using 4.2.1(iii)) $\mathrm{id} = \pi^{s-r}$. So m may be taken to be $s - r$. □

Definition The **order** of a permutation π, $o(\pi)$, is the least positive integer n such that π^n is the identity permutation. Note that the order of id is 1 and id is the only permutation of order 1.

Example The order of any transposition is 2.

Example The successive powers of the cycle (3 4 2 5) are (3 4 2 5), (3 2)(4 5), (3 5 2 4), id. Thus the order of (3 4 2 5) is 4.

Example The successive powers of the permutation (1 3)(2 5 4) are (1 3)(2 5 4), (2 4 5), (1 3), (2 5 4), (1 3)(2 4 5), id, (1 3)(2 5 4), and so on. In particular, the order of (1 3)(2 5 4) is six.

Theorem 4.2.3 *Let π be a permutation of order n. Then $\pi^r = \pi^s$ if and only if r is congruent to s modulo n.*

Proof From the proof of 4.2.2 it follows that if $\pi^r = \pi^s$ then $\pi^{s-r} = \mathrm{id}$. If, conversely, $\pi^{s-r} = \mathrm{id}$ then, multiplying each side by π^r and using 4.2.1, we obtain $\pi^s = \pi^r$. We will therefore have proved the result if we show that $\pi^k = \mathrm{id} = (\pi^0)$ precisely if k is congruent to 0 modulo n, that is, precisely if k is divisible by n.

To see this, observe first that if k is a multiple of n, say $k = nt$, then, using 4.2.1(ii),

$$\pi^k = \pi^{nt} = (\pi^n)^t = (\mathrm{id})^t = \mathrm{id}.$$

Suppose conversely that $\pi^k = \mathrm{id}$. Apply the division algorithm (1.1.1) to write k in the form $nq + r$ with $0 \leq r < n$. Then, again using 4.2.1, we have

$$\mathrm{id} = \pi^k = \pi^{nq+r} = \pi^{nq}\pi^r = (\pi^n)^q\pi^r = (\mathrm{id})^q\pi^r = \mathrm{id} \cdot \pi^r = \pi^r.$$

The definition of n (as giving the least positive power of π equal to id) now forces r to be zero: that is, n divides k. □

How can we quickly find the order of a permutation? For cycles the order turns out to be just the length of the cycle.

Theorem 4.2.4 *Let π be a cycle in S(n). Then $o(\pi)$ is the length of the cycle π.*

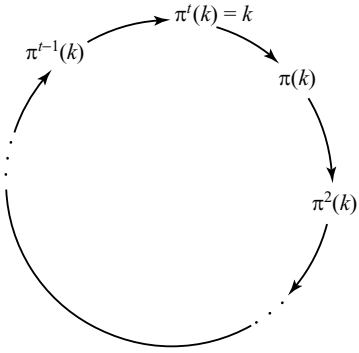

Fig. 4.4

Proof Think of the elements which are moved by π arranged in a circle, so that π^n just has the effect of moving each element n steps forward in the circuit (Fig. 4.4). From this picture it should be clear that if t is the length of π then the least positive integer n for which π^n equals the identity is t.

We can argue more formally as follows. If $\pi = \mathrm{id}$ then the result is clear; so we may suppose that there is $k \in \{1,\ldots,\mathrm{n}\}$ with $\pi(k) \neq k$. Since π is a cycle, the set of integers moved by π is precisely $\mathrm{Mov}(\pi) = \{k, \pi(k), \pi^2(k), \ldots, \pi^{t-1}(k)\}$, where t is the length of π. There are no repetitions in the above list, so the order of π is at least t.

On the other hand, $\pi^t(k) = k$, and hence for every value of r

$$\pi^t(\pi^r(k)) = \pi^{t+r}(k) = \pi^{r+t}(k) = \pi^r(\pi^t(k)) = \pi^r(k).$$

Therefore π^t fixes every element of the set $\mathrm{Mov}(\pi)$. Since π fixes all other elements of $\{1, \ldots, n\}$ so does π^t. Thus, $\pi^t = \mathrm{id}$.

Therefore the least positive power of π equal to the identity permutation is the tth, so $o(\pi) = t$, as required. □

Next we consider those permutations that are products of two disjoint permutations.

Lemma 4.2.5 *If π, σ are disjoint permutations in $S(n)$ then the order of $\pi\sigma$ is the least common multiple,* $\mathrm{lcm}(o(\pi), o(\sigma))$*, of the orders of π and σ.*

Proof Suppose that $o(\pi) = r$ and $o(\sigma) = s$, and let $d = ra$, $d = sb$ where $d = \mathrm{lcm}(r,s)$. Then certainly we have

$$(\pi\sigma)^d = \pi^d\sigma^d = \pi^{ra}\sigma^{sb} = (\pi^r)^a(\sigma^s)^b = \mathrm{id}$$

(the first equality by Theorem 4.2.1 since π and σ commute). So it remains to show that d is the least positive integer for which $\pi\sigma$ raised to that power is the identity permutation.

So suppose that $(\pi\sigma)^e = \mathrm{id}$. Since π and σ commute it follows, by Theorem 4.2.1, that $\pi^e\sigma^e = \mathrm{id}$. Let $k \in \{1, \ldots, n\}$. If k is moved by π then it is fixed by σ, and hence by σ^e: so $k = \mathrm{id}(k) = \pi^e\sigma^e(k) = \pi^e(k)$. On the other hand if k is fixed by π then certainly it is fixed by π^e. Therefore $\pi^e = \mathrm{id}$. Since $\pi^e\sigma^e = \mathrm{id}$, it then follows that also $\sigma^e = \mathrm{id}$. So by Theorem 4.2.3 it follows that r divides e and s divides e: hence d divides e (by definition of least common multiple), as required. □

Example The permutation

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

may be written as the product $(1\ 5\ 4)(2\ 7)$, of disjoint permutations. Therefore the order of π is the lcm of 3 and 2: that is $o(\pi) = 6$. (It is an instructive exercise, which illustrates the proof of 4.2.5, to compute the powers of π, and their cycle decompositions, up to the sixth.)

Example The permutation $\pi = (1\ 6)(3\ 7\ 4\ 2)$ is already expressed as the product of disjoint cycles, one of length 4 and the other of length 2. The order of π is therefore the lcm of 4 and 2 (which is 4): $o(\pi) = 4$. Note in particular that the order is not in this case the product of the orders of the separate cycles. You should compute the powers of π, and their cycle decompositions (up to the fourth), to see why this is so.

Theorem 4.2.6 *Let π be an element of $S(n)$, and suppose that $\pi = \tau_1\tau_2\ldots\tau_k$ is a decomposition of π as a product of disjoint cycles. Then the order of π is the least common multiple of the lengths of the cycles $\tau_1, \ldots, \tau_k$.*

Proof The proof is by induction on k. When k is 1, the result holds by Theorem 4.2.4. Now suppose, inductively, that the theorem is true if π can be written as a product of $k-1$ disjoint cycles. If π is a permutation which is of the form

$$\tau_1\tau_2\ldots\tau_k,$$

with the τ_j disjoint, then apply the induction hypothesis to the product $\tau_1\tau_2\ldots\tau_{k-1}$ to deduce that

$$o(\tau_1\tau_2\ldots\tau_{k-1}) = \mathrm{lcm}(o(\tau_1), \ldots, o(\tau_{k-1})),$$

and then apply Lemma 4.2.5 to the product $(\tau_1\tau_2\ldots\tau_{k-1})\tau_k$ to obtain the result (since $\mathrm{lcm}(\mathrm{lcm}(o(\tau_1),\ldots,o(\tau_{k-1})),o(\tau_k)) = \mathrm{lcm}(o(\tau_1),\ldots,o(\tau_k))$). □

In passing, we say a little about the shape of a permutation. By the 'shape' of a permutation π we mean the sequence of integers (in non-descending order) giving the lengths of the disjoint cyclic components of π. Thus if π has shape (2,2,5) then π is a product of three disjoint cycles, two of length 2 and one of length 5; the permutation (1 3 4)(2 5 8 6) has shape (3,4). We say that permutations π and σ are **conjugate** if there exists some permutation τ such that $\sigma = \tau^{-1}\pi\tau$. Then it may be shown that two permutations have the same shape if and only if they are conjugate. This is proved for transpositions – permutations of shape (2) – below (see the proof of Theorem 4.2.9(iv)) but, since we do not need the general result, we simply refer the reader to [Ledermann, Proposition 21] for a proof of the general result, which is due to Cauchy.

Finally in this section, we consider the sign of a permutation. There are a number of (equivalent) ways to define this notion: here we take the following route.

Definition Let $n \geq 2$ be an integer. Define the polynomial $\Delta = \Delta(x_1,\ldots x_n)$ in the indeterminates $x_1,\ldots,x_n$ to be

$$\Delta(x_1,\ldots,x_n) = \Pi\{(x_i - x_j)\colon i, j \in \{1,\ldots,n\}, i < j\}$$

the product of all terms of the form $(x_i - x_j)$ where $i < j$. For instance:

$\Delta(x_1,x_2) = (x_1 - x_2)$;
$\Delta(x_1,x_2,x_3) = (x_1 - x_2)(x_1 - x_3)(x_2 - x_3)$;
$\Delta(x_1,x_2,x_3,x_4) = (x_1 - x_2)(x_1 - x_3)(x_1 - x_4)(x_2 - x_3)(x_2 - x_4)(x_3 - x_4)$.

One may, of course, multiply out the terms but there is no need to do so: it will be most convenient to handle such polynomials in this factorised form.

Now let $n \geq 2$ and let $\pi \in S(n)$. We define a new polynomial, denoted $\pi\Delta$, from $\Delta = \Delta(x_1,\ldots,x_n)$ and π by the following rule: wherever Δ has a factor $x_i - x_j$, $\pi\Delta$ has the factor $x_{\pi(i)} - x_{\pi(j)}$. It is important to observe that $\pi\Delta$ is as Δ but with x_i replaced throughout by $x_{\pi(i)}$ for each i. More formally, we define $\pi\Delta$ by

$$\pi\Delta(x_1,\ldots,x_n) = \Pi\{(x_{\pi(i)} - x_{\pi(j)})\colon i, j \in \{1,\ldots,n\}, i < j\}.$$

For example, suppose that $n = 3$ and that π is the transposition (2 3). Then $\pi\Delta$ is obtained from Δ by replacing x_2 by x_3 and x_3 by x_2:

$$\pi\Delta(x_1,x_2,x_3) = (x_1 - x_3)(x_1 - x_2)(x_3 - x_2).$$

Now we note that this is just $-\Delta(x_1, x_2, x_3)$. To see this, just interchange the first two factors of $\pi\Delta$ and also write $(x_3 - x_2)$ as $-(x_2 - x_3)$.

The general case is similar: the only effect of applying a permutation to Δ in this way is to interchange the order of the factors and to replace some factors $x_i - x_j$ by $x_j - x_i$. We state this as our next result.

Lemma 4.2.7 *Let $\pi \in S(n)$ and let $\Delta(x_1, \ldots, x_n)$ be the polynomial as defined above. Then either $\pi\Delta = \Delta$ or $\pi\Delta = -\Delta$.*

Proof Consider a single factor $x_i - x_j$ of Δ. Since π is a bijection there are unique values k and l in $\{1, \ldots, n\}$ such that $\pi(k) = i$ and $\pi(l) = j$: also $k \neq l$ since $i \neq j$. There are two possibilities.

If $k < l$ then the factor $x_k - x_l$ occurs in Δ, and it is transformed in $\pi\Delta$ into $x_i - x_j$.

If $k > l$ then the factor $x_l - x_k$ occurs in Δ, and it is transformed in $\pi\Delta$ into $x_j - x_i = -(x_i - x_j)$.

Thus for every factor $x_i - x_j$ of Δ, either it or minus it occurs as a factor of $\pi\Delta$. Clearly (by the construction of $\pi\Delta$) Δ and $\pi\Delta$ have the same number of factors. It follows therefore (on collecting all the minus signs together) that $\pi\Delta$ is either Δ or $-\Delta$. □

We advise you to work through the above proof with some particular example(s) of π is $S(3)$ and $S(4)$.

Definition Let $\pi \in S(n)$. Define the **sign** of π, $\mathrm{sgn}(\pi)$, to be 1 or -1 according as $\pi\Delta = \Delta$ or $-\Delta$. Thus $\pi\Delta = \mathrm{sgn}(\pi) \cdot \Delta$. If $\mathrm{sgn}(\pi)$ is 1 then π is said to be an **even** permutation: if $\mathrm{sgn}(\pi) = -1$ then π is an **odd** permutation.

Theorem 4.2.8 *Let $\pi, \sigma \in S(n)$. Then* $\mathrm{sgn}(\sigma\pi) = \mathrm{sgn}(\sigma) \cdot \mathrm{sgn}(\pi)$.

Proof We compute, in two slightly different ways, the effect of applying the composite permutation $\sigma\pi$ to $\Delta = \Delta(x_1, \ldots, x_n)$. First we apply π to Δ, to get $\pi\Delta$: the effect is to replace, for each i, x_i by $x_{\pi(i)}$ throughout. Without rearranging, we immediately apply the permutation σ: this results in each $x_{\pi(i)}$ being replaced throughout by $x_{\sigma(\pi(i))} = x_{\sigma\pi(i)}$. The net result is that for each i, x_i has been replaced throughout by $x_{\sigma\pi(i)}$. So the resulting polynomial is, by definition, $(\sigma\pi)\Delta$ and, by definition, $(\sigma\pi)\Delta = \mathrm{sgn}(\sigma\pi) \cdot \Delta$.

Now we also have, by definition, that $\pi\Delta = \mathrm{sgn}(\pi) \cdot \Delta$. So, when we apply σ to $\pi\Delta$, we are just applying σ to $\mathrm{sgn}(\pi) \cdot \Delta$ (which is either Δ or $-\Delta$). The result of that is therefore equal to $\mathrm{sgn}(\pi) \cdot \sigma\Delta$, which equals $\mathrm{sgn}(\pi) \cdot \mathrm{sgn}(\sigma) \cdot \Delta$.

So the net result of applying $\sigma\pi$ to Δ may be expressed in two ways, as $\operatorname{sgn}(\sigma\pi)\cdot\Delta$ and as $\operatorname{sgn}(\pi)\operatorname{sgn}(\sigma)\cdot\Delta$. Equating these expressions, we obtain that the polynomials $\operatorname{sgn}(\sigma\pi)\cdot\Delta$ and $\operatorname{sgn}(\pi)\cdot\operatorname{sgn}(\sigma)\cdot\Delta$ are identical. Hence it must be that $\operatorname{sgn}(\sigma\pi)=\operatorname{sgn}(\pi)\cdot\operatorname{sgn}(\sigma)$, as required. □

You may observe that what we are using in the proof above is an 'action' of the symmetric group $S(n)$ on the set of polynomials $p(x_1,\ldots,x_n)$ in the variables $x_1,\ldots,x_n$. Given $\pi\in S(n)$ and $p(x_1,\ldots,x_n)$, we define the polynomial πp to be as p but with each x_i replaced by $x_{\pi(i)}$. What we used was, in essence, that if $\pi,\sigma\in S(n)$ then $(\sigma\pi)p=\sigma(\pi(p))$.

There are other routes to defining the sign of a permutation (see, for example, [Fraleigh, Chapter 5] and [MacLane and Birkhoff, Chapter III, Section 6]).

Theorem 4.2.9 *Let π and σ be in $S(n)$. Then*

(i) $\operatorname{sgn}(\mathrm{id})=1$,
(ii) $\operatorname{sgn}(\pi)=\operatorname{sgn}(\pi^{-1})$,
(iii) $\operatorname{sgn}(\pi^{-1}\sigma\pi)=\operatorname{sgn}(\sigma)$,
(iv) *if τ is a transposition then* $\operatorname{sgn}(\tau)=-1$.

Proof (i) This is immediate from the definition of sign.

(ii) By Theorem 4.2.8 we have

$$\operatorname{sgn}(\pi^{-1})\operatorname{sgn}(\pi)=\operatorname{sgn}(\pi^{-1}\pi)=\operatorname{sgn}(\mathrm{id})=1$$

using (i). So either both π^{-1} and π are even or both are odd, as required.

(iii) This is immediate by Theorem 4.2.8 and (ii).

(iv) The proof proceeds by showing this for increasingly more general transpositions.

First notice that the result is obviously true for $\tau=(1\ 2)$ since the only factor of Δ whose sign is changed by interchanging 1 and 2 is x_1-x_2.

Secondly, note that any transposition involving '1' is a conjugate of $(1\ 2)$:

$$(1\ \ k)=(2\ \ k)(1\ \ 2)(2\ \ k)=(2\ \ k)^{-1}(1\ \ 2)(2\ \ k).$$

So by (iii) $\operatorname{sgn}(1\ k)=\operatorname{sgn}(1\ 2)=-1$.

Finally we notice that every transposition is conjugate to one involving '1':

$$(m\,k)=(1\,k)(1\,m)(1\,k)=(1\,k)^{-1}(1\,m)(1\,k).$$

So, by another application of (iii), we obtain $\operatorname{sgn}(m\ k)=-1$, as required. □

Note, by the way, that we have illustrated the remark after 4.2.6 by showing that every two transpositions are conjugate.

Example For any positive integer n, let $A(n)$ denote the set of all even permutations (permutatations with sign $+1$) in $S(n)$. We notice, using Theorems 4.2.8 and 4.2.9, that the product of any two elements of $A(n)$ is in $A(n)$, that id is in $A(n)$ and that the inverse of any element of $A(n)$ is in $A(n)$. Also, provided $n \geq 2$, $(1\ 2)$ is in $S(n)$ but is not in $A(n)$, and so not every permutation is even.

Since $(1\ 2)$ is odd, multiplying an even permutation by $(1\ 2)$ gives an odd permutation, and multiplying an odd permutation by $(1\ 2)$ gives an even permutation. The map f from the set of even permutations to the set of odd permutations defined by $f(\pi) = (1\ 2)\pi$ is a bijection, so it follows that half the elements of $S(n)$ are even and the other half are odd. Hence $A(n)$ has $n!/2$ elements. You can think of the map f more concretely by imagining the elements of $A(n)$ written out in a row; then, beneath each such element π, write its image $(1\ 2)\pi$. It is easy to show that the second row contains no repetitions and contains all odd permutations, so it is clear that $A(n)$ contains exactly half of the $n!$ elements of $S(n)$.

We finish the section by showing that every permutation may be written (in many ways) as a product of transpositions (not disjoint in general of course!). It will follow by Theorems 4.2.8 and 4.2.9 that a permutation is even or odd according as the number of transpositions in such a product is even or odd (hence the terminology).

Theorem 4.2.10 *Every cycle is a product of transpositions. If π is a cycle then* $\operatorname{sgn}(\pi) = (-1)^{\text{length}(\pi)-1}$.

Proof To see that a cycle $(x_1\ x_2 \ldots x_k)$ can be written as a product of transpositions, we just check:

$$(x_1\ x_2 \ldots x_k) = (x_1\ x_k) \ldots (x_1\ x_3)(x_1\ x_2).$$

There are $k - 1 = \text{length}(\pi) - 1$ terms on the right-hand sign each with sign -1, by Theorem 4.2.9(iv). By Theorem 4.2.8 it follows that

$$\operatorname{sgn}(\pi) = \operatorname{sgn}((x_1\ x_k) \cdots (x_1\ x_3)(x_1\ x_2)) = (-1)^{\text{length}(\pi)-1}. \quad \square$$

Next we extend this result to arbitrary permutations.

Theorem 4.2.11 *Suppose $n \geq 2$. Every permutation in $S(n)$ is a product of transpositions. Although there are many ways of writing a given permutation π as a product of transpositions, the number of terms occurring will always be either even or odd according as π is even or odd.*

Proof It is immediate from Theorems 4.1.3 and 4.2.10 that every permutation may be written as a product of transpositions. Suppose that we write π as a product of transpositions. Then, by the multiplicative property of sign (Theorem 4.2.8) and Theorem 4.2.9(iv), we have that $\text{sgn}(\pi)$ is -1 raised to the number of terms in the decomposition. Thus the statement follows. □

Exercises 4.2

1. Determine the order and sign of each of the following permutations:
   (i) (1 2 3 4 5)(8 7 6)(10 11);
   (ii) (1 3 5 7 9 11)(2 4 6 8 10);
   (iii) (1 2)(3 4)(5 6 7 8)(9 10);
   (iv) (1 2 3 4 5 6 7 8)(1 8 7 6 5 4 3 2).
2. Give an example of two cycles of lengths r and s respectively whose product does not have order $\text{lcm}(r,s)$.
3. Give an example of a permutation of order 2 which is not a transposition.
4. Show that if π and σ are permutations such that $(\pi\sigma)^2 = \pi^2\sigma^2$ then $\pi\sigma = \sigma\pi$.
5. Find permutations π, σ such that $(\pi\sigma)^2 \neq \pi^2\sigma^2$.
6. Compute the orders of the permutations

   $$(2\ 1\ 4\ 6\ 3), (1\ 2)(3\ 4\ 5) \text{ and } (1\ 2)(3\ 4).$$

7. Compute the orders of the following products of *non-disjoint* cycles:

   $$(1\ 2\ 3)(2\ 3\ 4); (1\ 2\ 3)(3\ 2\ 4); (1\ 2\ 3)(3\ 4\ 5).$$

8. Complete the proof of Theorem 4.2.1.
9. List the elements of $A(4)$ and give the order of each of them.
10. Show that every element of $S(n)$ $(n \geq 2)$ is a product of transpositions of the form $(k\ k+1)$.
    [Hint: $(k\ k+2) = (k\ k+1)(k+1\ k+2)(k\ k+1)$.]
11. What is the highest possible order of an element in
    (i) $S(8)$, (ii) $S(12)$, (iii) $S(15)$?
    You may be interested to learn that there is no formula known for the highest order of an element of $S(n)$.

Start

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	

End?

15	14	13	12
11	10	9	8
7	6	5	4
3	2	1	

Fig. 4.5

12. Refer back to Exercise 4.1.6 for notation and terminology. Compute the orders and signs of *s, t, c, cs,* and *scs*. You should find that the order of $t = cs$ is 6.
    Suppose that someone shuffles the cards according to the interleaving *s*, having attempted to make the cut *c* but, in making the cut, failed to pick up the bottom card, so that the first permutation actually performed was

$$x = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 5 & 6 & 7 & 8 & 9 & 1 & 2 & 3 & 4 & 10 \end{pmatrix}.$$

    Believing that the composite permutation *sc* (=*t*) has been made, and having read the part of this section on orders, this person repeats the shuffle *sc* five more times, is somewhat surprised to discover that the cards have not returned to their original order, but then continues to make *sc* shuffles, hoping that the cards will eventually return to their original order. Show that this will not happen.
    [Hint: use what you have learned about the sign of a permutation.]
13. A well known children's puzzle has 15 numbered pieces arranged inside a square as shown (Fig. 4.5). A move is made by sliding a piece into the empty position. Consider the empty position as occupied by the number 16, so that every move is a transposition involving 16. Show that the order of the pieces can never be reversed. [Hint: show that if a product of transpositions each involving the number 16 moves 16 to an even-numbered position on the 4×4 board then the number of transpositions must be even. Also consider the sign of the permutation which takes the 'start' board to the 'end?' board.]

4.3 Definition and examples of groups

We are now ready to abstract the properties which several of our structures share. We make the following general definition.

Definition A **group** is a set G, together with an operation $*$, which satisfies the following properties:

(G1) for all elements g and h of G, $g * h$ is an element of G (closure);
(G2) for all elements, g, h and k of G,

$$(g * h) * k = g * (h * k) \quad \text{(associativity)};$$

(G3) there exists an element e of G, called the **identity** (or **unit**) of G, such that for all g in G we have

$$e * g = g * e = g \quad \text{(existence of identity)};$$

(G4) for every g in G there exists an element g^{-1} called the **inverse** of g, such that

$$g * g^{-1} = g^{-1} * g = e \quad \text{(existence of inverse)}.$$

Definition The group $(G, *)$ is said to be **commutative** or **Abelian** (after Niels Henrik Abel (1802–29)) if the operation $*$ satisfies the commutative law, that is, if for all g and h in G we have $g * h = h * g$.

Normally we will use multiplicative notation instead of '$*$', and so write 'gh' instead of '$g * h$': for that reason some books use '1' instead of 'e' for the identity element of G. Occasionally (and especially if the group is Abelian) we will use additive notation, writing '$g + h$' instead of '$g * h$' and $-g$ for g^{-1}, in which case the symbol '0' is normally used for the identity element of G.

Note that the condition (G3) ensures that the set G is non-empty. Associativity means that $(gh)k = g(hk)$, and so we may write ghk without ambiguity. It follows, by induction, that this is true for any product of group elements: different bracketing does not change the value of the resulting element.

Use of the terms '*the* identity' and '*the* inverse' presupposes that the objects named are uniquely defined. We now justify this usage.

Theorem 4.3.1 *Let G be any group. Then there is just one element e of G satisfying the condition for being an identity of G. Also, for each element g in G*

there is just one element g^{-1} in G satisfying the condition for being an inverse of g.

Proof Suppose that both e and f satisfy the condition for being an identity element of G. Then we have

$$f = ef = e:$$

the first equality holds since e acts as an identity, the second equality holds since f acts as an identity. So there is just one identity element.

Given g in G, suppose that both h and k satisfy the condition for being an inverse of g. Then we have

$$h = he = h(gk),$$

since k is an inverse for g. But, by associativity, this equals $(hg)k$ and then, since h is an inverse for g, this in turn equals $ek = k$. Thus $h = k$, and the inverse of g is indeed unique. □

Let us now consider some examples of groups.

4.3.1 Groups of numbers

Example 1 The integers $(\mathbb{Z}, +)$ with addition as the operation, form a group. The closure and associativity properties are part of the unwritten assumptions we have made about $\mathbb{Z}$. The identity element for addition is 0. The inverse of n is $-n$. This group has an infinite number of elements.

In contrast, the set of natural numbers $\mathbb{N}$ equipped with addition is not a group, since not all its elements have additive inverses (*within* the set $\mathbb{N}$).

Note also that the integers with multiplication as the operation do not form a group since, for instance, 2 does not have a multiplicative inverse within the set $\mathbb{Z}$.

Example 2 The integers modulo n, $(\mathbb{Z}_n, +)$ (i.e. the set of congruence classes modulo n), equipped with addition modulo n (i.e. addition of congruence classes) as the operation, form a group. The identity element is the congruence class $[0]_n$ of 0 modulo n. The inverse of $[k]_n$ is $[-k]_n$. This example was discussed in Chapter 1, and it is an example of a group with a finite number of elements.

Notice again that if multiplication is taken as the operation then the set of congruence classes modulo n is not a group since not all elements have inverses (for example, $[0]_n$ has no multiplicative inverse).

Example 3 Consider $(G_n, \cdot)$, the set of invertible congruence classes modulo n under multiplication. This is a group. The identity element is $[1]_n$. The inverse of each element of G_n exists by definition of G_n (and is found as in Theorem 1.4.3). The number of elements in this group is $\phi(n)$ (the Euler ϕ-function was defined in Section 1.6).

Example 4 Other familiar number systems – the real numbers $\mathbb{R}$, the complex numbers $\mathbb{C}$, the rational numbers $\mathbb{Q}$ – are groups under addition. In each case, in order to obtain a group under the operation of multiplication, we must remove zero from the set.

Example 5 An interesting example of a finite non-Abelian group associated with a number system is provided by the **quaternions** $\mathbb{H}$ discovered by William Rowan Hamilton (1805–65). These can be regarded as 'hyper-complex' numbers, being of the form $a1 + b\mathrm{i} + c\mathrm{j} + d\mathrm{k}$ where *a,b,c,d* are real numbers. The product of any two of these numbers can be computed by using the following rules for multiplying i, j and k:

$$\mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = -1, \quad \mathrm{ij} = \mathrm{k}, \quad \mathrm{ji} = -\mathrm{k}, \quad \mathrm{jk} = \mathrm{i}, \quad \mathrm{kj} = -\mathrm{i}, \quad \mathrm{ki} = \mathrm{j}, \quad \mathrm{ik} = -\mathrm{j}.$$

Let $\mathbb{H}_0 = \{\pm 1, \pm \mathrm{i}, \pm \mathrm{j}, \pm \mathrm{k}\}$. Since $\mathbb{H}_0$ has only finitely many elements (eight), the closure under multiplication and associativity of multiplication are properties which can be checked (although the direct checking of associativity is very tedious). The set $\mathbb{H}_0$, under the multiplication defined above, is a group with 'multiplication table' as shown.

	1	−1	i	−i	j	−j	k	−k
1	1	−1	i	−i	j	−j	k	−k
−1	−1	1	−i	i	−j	j	−k	k
i	i	−i	−1	1	k	−k	−j	j
−i	−i	i	1	−1	−k	k	j	−j
j	j	−j	−k	k	−1	1	i	−i
−j	−j	j	k	−k	1	−1	−i	i
k	k	−k	j	−j	−i	i	−1	1
−k	−k	k	−j	j	i	−i	1	−1

We should perhaps point out the convention for forming a table, such as that above, which enables us to see the effect of an operation '*' on a set G. The table has the elements of the set G in some definite order as a heading row, and the elements, in the same order, as a leading column. The entry at the intersection of the row labelled by g and the column labelled by h is $g * h$.

We have by now encountered a number of situations in which tables of the above type have arisen (see Section 1.4 and the end of Section 4.1). One reason

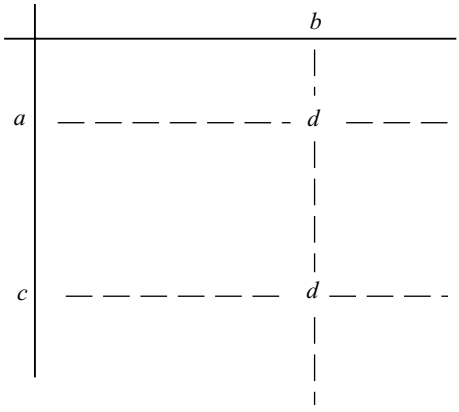

Fig. 4.6

for the use of such a table, which completely describes the operation, is that it helps one to determine whether or not the set under the operation is a group. The closure property of the set G under the operation is reduced to the question of whether or not every entry in the table is in the set G. The existence of an identity element can be determined by seeking an element of G such that the row labelled by that element is the same as the heading row of the table (and similarly for the columns). The inverse of an element g of G can be read off from the table by looking for the identity element in the row containing g and noting the heading for the column in which it occurs: the element heading that column is the inverse of g. For instance, in the example $\mathbb{H}_0$ above, to find the inverse of j we look along the row labelled 'j' until we find the identity element 1, then look to the head of that column: we conclude that $j^{-1} = -j$.

It is also possible to tell from its table whether or not G is Abelian: G will be Abelian exactly if the table is symmetric about its main diagonal (as in Example 6 below, but not as in Example 5 above). The only property which the table fails to give directly is associativity. In Example 5 above, in order to check this from the table, we would need to work out the products $(gh)k$ and $g(hk)$ for each of the eight choices for g, h and k: a total of 1024 calculations! Clearly some other method of checking associativity is usually sought (see Example 4 on p. 176).

When one is constructing such a group table it is useful to bear in mind that every row or column must contain each element of the group exactly once. For if one had an entry occurring twice in, say, the same column (as shown in Fig. 4.6) then one would have $ab = cb$. Multiplying on the right by b^{-1} would give the contradiction $a = c$.

In a paper of 1854, Cayley describes how to construct the group table. He also emphasises that each row and column contains each element exactly once.

Example 6 As an example of using a table to define a group, let G be the set $\{e, a, b, c\}$ with operation given by the table

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

Associativity can be checked case by case, and so one may verify that this is an Abelian group with 4 elements.

Example 7 The set, S, of all complex numbers (see the Appendix for these) of the form e^{ir} with r a real number, under multiplication ($e^{ir}e^{is} = e^{i(r+s)}$) is a group. The identity element is $e^{i0} = 1$, and the inverse of the element e^{ir} is $e^{i(-r)}$. This is an infinite Abelian group.

4.3.2 Groups of permutations

Example 1 We can now see Theorem 4.1.1 as saying that the set $S(n)$, of permutations of the set $\{1, 2, \ldots, n\}$, is a group under composition of functions. This is a group with $n!$ elements, and it is non-Abelian if $n \geq 3$. Notice that associativity follows by Theorem 2.2.1.

Example 2 We also saw in Section 4.2 (Example on p. 167) that the set $A(n)$ of even permutations is a group under composition of functions. This is a group with $n!/2$ elements (assuming $n \geq 2$), the **alternating group** on n elements. However, the set of all odd permutations in $S(n)$, with composition as the operation, fails to be a group since the product of two odd permutations is not odd, and so the set is not closed under the operation (if $n = 1$ it fails to be a group since it is empty!).

Example 3 Let G consist of the following four elements of $S(4)$:

$$G = \{\mathrm{id},\ (1\ \ 2)(3\ \ 4),\ (1\ \ 3)(2\ \ 4),\ (1\ \ 4)(2\ \ 3)\}.$$

Equipped with the usual product for permutations, the operation on this set is associative (since composition of any permutations is). To check the other group properties, it is easiest to work out the table.

	id	(12)(34)	(13)(24)	(14)(23)
id	id	(12)(34)	(13)(24)	(14)(23)
(12)(34)	(12)(34)	id	(14)(23)	(13)(24)
(13)(24)	(13)(24)	(14)(23)	id	(12)(34)
(14)(23)	(14)(23)	(13)(24)	(12)(34)	id

Notice that this is essentially the same table as that given in Example 6 in Section 4.3.1. For if we write e for id, a for (1 2)(3 4), b for (1 3)(2 4) and c for (1 4)(2 3), then we transform this table into that in the other example. This allows us to conclude that the operation in that example is indeed associative, because the tables for the two operations are identical (up to the relabelling mentioned above); so, since multiplication of permutations is associative, the other operation must also be associative. This is an example of a 'faithful permutation representation' of a group – where a set of permutations is found with the 'same' multiplication table as the group.

4.3.3 Groups of matrices

Example 1 Let $\mathrm{GL}(2,\mathbb{R})$ be the set of all invertible 2×2 matrices with real entries, equipped with matrix multiplication as the operation (a matrix is said to be **invertible** if it has an inverse with respect to multiplication).

This operation is associative: even if you have not seen this proved before, you may verify it quite easily for 2×2 matrices: you need to check that.

$$\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} e & f \\ g & h \end{pmatrix}\right)\begin{pmatrix} i & j \\ k & l \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\left(\begin{pmatrix} e & f \\ g & h \end{pmatrix}\begin{pmatrix} i & j \\ k & l \end{pmatrix}\right).$$

The identity for the operation is the 2×2 identity matrix

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

and the other conditions are easily seen to be satisfied. This example may be generalised by replacing '2' by 'n' so as to get the **general linear** group, $\mathrm{GL}(n, \mathbb{R})$, of all invertible $n \times n$ matrices with real entries.

Example 2 Let G be the set of all upper-triangular 2×2 matrices with both diagonal elements non-zero. Equivalently (as you may check), G is the set of

all invertible upper-triangular 2×2 matrices: those of the form

$$\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}$$

where a, b and c are real numbers with both a and c non-zero. Equipped with the operation of matrix multiplication, this set is closed (you should check this), contains the identity matrix, and the inverse of any matrix in G is also in G, as you should verify.

Example 3 Let G be the set of all 2×2 invertible diagonal matrices with real entries: that is, matrices of the form

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$$

where a and b are both non-zero. Then G is a group under matrix multiplication. The verification of this claim is left to the reader.

Example 4 Let X and Y be the matrices

$$X = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \qquad Y = \begin{pmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{pmatrix}$$

where i is a square root of -1. It can be seen that if $\mathbf{I}$ denotes the 2×2 identity matrix then

$$X^2 = Y^2 = -\mathbf{I} \quad \text{and} \quad XY = -YX.$$

Putting $Z = XY$, we deduce, or check, that

$$Z^2 = -\mathbf{I}, \quad YZ = X, \quad ZY = -X, \quad ZX = Y, \quad XZ = -Y.$$

It follows that the eight matrices $\mathbf{I}$, $-\mathbf{I}$, X, $-X$, Y, $-Y$, Z and $-Z$ have the same multiplication table as the quaternion group $\mathbb{H}_0$ (Example 5 on p. 172) since that table was constructed using essentially the same equations.

This is a matrix representation of $\mathbb{H}_0$, and it provides a nice proof of the fact that the operation on $\mathbb{H}_0$ is associative since matrix multiplication is associative. (Notice that the set of matrices of the form $a\mathbf{I} + bX + cY + dZ$ where a, b, c, d are real numbers gives a representation of the quaternions as 2×2 matrices with complex entries.)

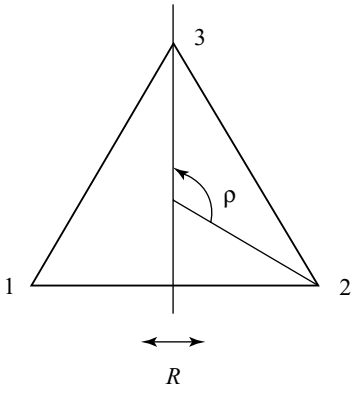

Fig. 4.7

4.3.4 Groups of symmetries of geometric figures

Now we turn to a rather different source of examples. Groups may arise in the form of groups of symmetries of geometric figures. By a symmetry of a geometric figure we mean an orthogonal affine transformation of the plane (or 3-space, if appropriate) which leaves the figure invariant. If the terms just used are unfamiliar, no matter: the meaning of 'symmetry' should become intuitively clear when you look at the following examples. We may say, roughly, that a symmetry of a geometric figure is a rigid movement of it which leaves it looking as it was before the movement was made.

Example 1 Consider an equilateral triangle such as that shown in Fig. 4.7. (The triangle itself is unlabelled, but we assign an arbitrary numbering to the vertices so as to be able to keep track of the movements made.)

There are a number of ways of 'picking up the triangle and then setting it down again' so that it looks the same as when we started, although the vertices may have been moved. In particular, we could rotate it anti-clockwise about its mid-point by an angle of $2\pi/3$: let us denote that operation (or 'symmetry') by ρ. Or we could reflect the triangle in the vertical line shown: let us denote that symmetry by R. Of course there are other symmetries, including that which leaves everything as it was (we denote that by e), but it will turn out that all the other symmetries may be described in terms of ρ and R.

We may define a group operation on this set of symmetries: if σ and τ are symmetries then define $\sigma\tau$ to be the symmetry 'do τ then apply σ to the result'. That this gives us a group is not difficult to see: we have just noted that closure holds; e is the identity element; the inverse of any symmetry is clearly a symmetry ('reverse the action of the symmetry'); and associativity follows because

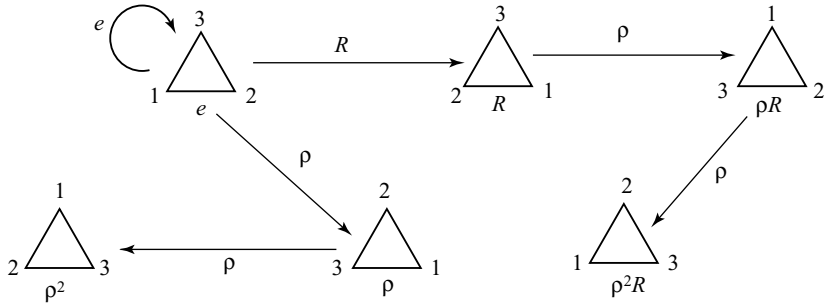

Fig. 4.8

symmetries are transformations (so functions). For the equilateral triangle there are six different symmetries, namely e, ρ, ρ^2, R, ρR and $\rho^2 R$ (there can be at most six symmetries since there are only six permutations of three vertices). See Fig. 4.8. We note some 'relations': $\rho^3 = e$; $R^2 = e$; $\rho^2 R = R\rho$. You may observe that there are others, but it turns out that they can all be derived from the three we have written down.

This group is often denoted as $D(3)$ and we may write down its group table, either by making use of the relations above or by calculating the effect of each product of symmetries on the triangle with labelled vertices. For instance, Fig. 4.9 gives us the relation $\rho^2 R = R\rho$. To compute, for example, the product $\rho R \cdot \rho^2 R$ we may either compute the effect of this symmetry on the triangle, using a sequence of pictures such as that in the figures, or we may use the relations above as follows:

$$\rho R \cdot \rho^2 R = \rho \cdot R\rho \cdot \rho R = \rho \cdot \rho^2 R \cdot \rho R = \rho^3 \cdot R\rho \cdot R = \rho^3 \cdot \rho^2 R \cdot R$$
$$= e\rho^2 R^2 = \rho^2 e = \rho^2.$$

Note in particular that this group is not Abelian

	e	ρ	ρ^2	R	ρR	$\rho^2 R$
e	e	ρ	ρ^2	R	ρR	$\rho^2 R$
ρ	ρ	ρ^2	e	ρR	$\rho^2 R$	R
ρ^2	ρ^2	e	ρ	$\rho^2 R$	R	ρR
R	R	$\rho^2 R$	ρR	e	ρ^2	ρ
ρR	ρR	R	$\rho^2 R$	ρ	e	ρ^2
$\rho^2 R$	$\rho^2 R$	ρR	R	ρ^2	ρ	e

We may obtain a permutation representation of this group by replacing each symmetry by the permutation of the three vertices which it induces. Thus ρ is replaced by (1 2 3), ρ^2 by (1 3 2), R by (1 2), ρR by (1 3) and $\rho^2 R$ by (2 3).

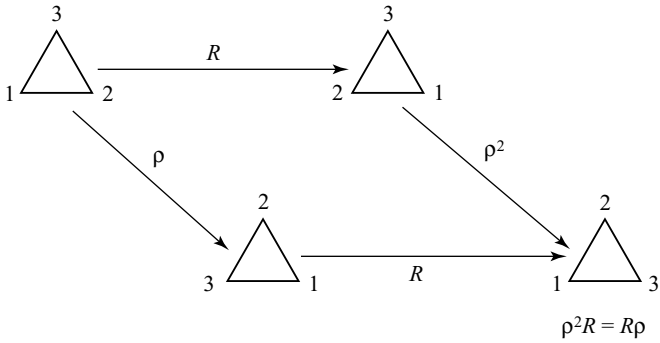

Fig. 4.9

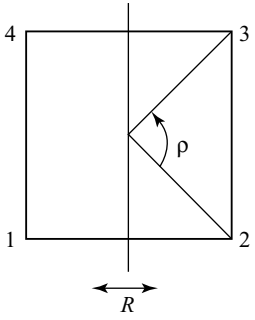

Fig. 4.10

With this relabelling, the above table becomes the table of the symmetric group $S(3)$ given in Section 4.1, p. 157.

Example 2 If we replace the triangle in Example 1 with a square, then we have a similar situation (Fig. 4.10). We take ρ to be rotation about the centre by $2\pi/4$ and R to be reflection in the perpendicular bisector of (say) the side joining vertices 1 and 2. This time we get a group with 8 elements (see Exercise 4.3.6), in which all the relations are consequences of the relations $\rho^4 = e$, $R^2 = e$, $\rho^3 R = R\rho$. This group is denoted by $D(4)$. (We can use the numbering of the vertices to obtain a faithful permutation representation of this group in $S(4)$.)

Examples 1 and 2 suggest a whole class of groups: the **dihedral group** $D(n)$ is the group of symmetries of a regular n-sided polygon. It has $2n$ elements and is generated by the rotation, ρ, anti-clockwise about the centre, by $2\pi/n$, together with the reflection, R, in the perpendicular bisector of any one of the sides, and these are subject to the relations $\rho^n = e$, $R^2 = e$, $\rho^{n-1} R = R\rho$.

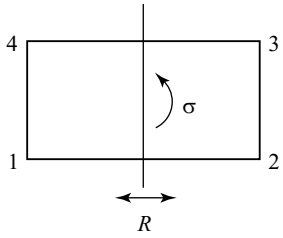

Fig. 4.11

Example 3 Let our geometric figure be a rectangle that is not a square (Fig. 4.11).

Then rotation by $2\pi/4$ is no longer a symmetry, although the rotation σ about the centre by $2\pi/2$ is. If, as before, we let R be reflection in the perpendicular bisector of the line joining vertices 1 and 2, then we see that the group has 4 elements e, σ, R, τ, where τ is σR and the table is as shown below.

	e	σ	R	τ
e	e	σ	R	τ
σ	σ	e	τ	R
R	R	τ	e	σ
τ	τ	R	σ	e

To conclude this section, we make some historical remarks.

The emergence of group theory is one of the most thoroughly investigated developments in the history of mathematics. In [Wussing] three rather different sources for this development are distinguished: solution of polynomial equations and symmetric functions; number theory; geometry.

The best known source is the study of exact solutions for polynomial equations. The solution of a general quadratic equation

$$ax^2 + bx + c = 0$$

was possibly known to the Babylonians and certainly was known to the early Greeks, although the lack of algebraic symbolism and their over-reliance on geometric interpretations meant that their solution seems over-complicated today. The solution was also known to the Chinese. The Greeks considered only positive real solutions. Brahmagupta (c. 628) was quite happy to deal with negative numbers as solutions of equations (and in general) but it would be another thousand years before such solutions were accepted in Europe.

The Arab mathematician al-Khwarizmi (c. 800) presented in his *Al-jabr wa'l muqābalah* (hence 'algebra') the systematic solution of quadratic equations,

though he disallowed negative roots (and of course, complex roots). He also pointed out that a difficulty arises when b^2 is less than $4ac$: 'And know that, when in a question belonging to this case you have halved the number of roots and multiplied the half by itself, if the product be less than the number connected with the square, then the instance is impossible' (adapted from [Fauvel and Gray]).

The solution of the general cubic equation

$$ax^3 + bx^2 + cx + d = 0$$

was given by Cardano in his *Ars Magna* of 1545. He stated that the hint for the solution was given to him by Tartaglia. Scipione del Ferro had previously found the solution in some special cases.

Cardano's solution was split into a large number of cases because negative numbers were not then used as coefficients: for instance, we would think of $x^3 + 3x^2 + 5x + 2 = 0$ and $x^3 - 4x^2 - x + 1 = 0$ as both being instances of the general equation $ax^3 + bx^2 + cx + d = 0$ whereas, in those times, the second would necessarily have been presented as $x^3 + 1 = 4x^2 + x$. Still, Cardano had difficultly with the fact that the solutions need not always be real numbers. He did, however, have some inkling of the idea of 'imaginary' numbers: one of the problems he studied in *Ars Magna* is to divide 10 into two parts, the product of which is 40. The solutions of this problem are $5 + \sqrt{-15}$ and $5 - \sqrt{-15}$. Cardano says '... you will have to imagine $\sqrt{-15}$, ... putting aside the mental tortures involved, a solution is obtained which is truly sophisticated'. The solution of the general quartic equation

$$ax^4 + bx^3 + cx^2 + dx + e = 0$$

(as we would write!) was found by Ferrari at Cardano's request and was also included in his *Ars Magna*.

We turn now to the general quintic equation

$$ax^5 + bx^4 + cx^3 + dx^2 + ex + f = 0.$$

Several attempts were made to find an algebraic formula which would give the roots in terms of the coefficients a, b, c, d, e, f. In fact, there can be no such general formula. Ruffini gave a proof for this, but his argument contained gaps which he was unable to fill to the satisfaction of other mathematicians, and the first generally accepted proof was given in 1824 by Abel (1802–29). Of course there still remained the problem of deciding whether any particular polynomial equation has a solution in 'radicals' (one expressed in terms of the coefficients, using addition, subtraction, multiplication, division and the extraction of nth roots). This problem was solved in 1832 by Galois (1811–32),

and it is here that groups occur (some of the ideas already appeared in the work of Lagrange and Ruffini). The key idea of Galois was to associate a group to a given polynomial. The group is that of all permutations ('substitutions') of the roots of the polynomial which leave the polynomial invariant: the operation is just composition of permutations. Having translated the problem about equations into one about groups, Galois then solved the group theory problem and deduced the solution to the problem for polynomials. Galois was killed in a duel when he was 20, and one wonders what else he might have achieved had he lived.

It would be wrong to imply that Galois' work immediately revolutionised mathematics and produced a vast interest in 'group theory'. Both during his life, and initially after his death, the work of Galois went almost completely unappreciated. Partly this was due to a series of misfortunes which befell papers which he submitted for publication but also his work was very innovatory and difficult to understand at the time. Some of his main results were included in a letter written the night before the duel.

It should be emphasised that Galois did not understand the term 'group' in the (abstract, axiomatic) way in which we defined it at the beginning of this section. Galois used the term 'group' in a much more informal way and in a much more specific context – for Galois, groups were groups of substitutions. Indeed, although Cayley took the first steps towards defining an abstract group around 1850, his work was premature and groups were always groups of something, related to a definite context, until late in the century. It was Kronecker who, in 1870, first defined an abstract Abelian group. Then in 1882 von Dyck and Weber independently gave the definition of an abstract group.

In discussing the origins of group theory, one should also mention the work of Cauchy (1789–1857). In his work, groups occurred in a somewhat different way than they had in Galois'. Cauchy was interested in functions, $f(x_1, x_2, \ldots, x_n)$, of several variables, and in the permutations π in $S(n)$ which fix f (so that $f(x_1, x_2, \ldots, x_n) = f(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)})$). The set of permutations fixing a given function is a group under composition, and the group and the function f are closely related. Cauchy published an important paper in 1815 on groups and, between 1844 and 1846, developed and systematised the area significantly.

In 1846 Liouville published some of Galois' work. Serret lectured on it and gave a good exposition in his *Cours d'Algèbre supérieure* (1866). But it was only with the publication in 1870 of the book by Jordan – *Traité des substitutions et des équations algébriques* – that the subject came to the attention of a much wider audience.

There were other sources of groups. On the geometric side, Jordan had considered groups of transformations of a geometry. Number theory, of course, was another source, providing examples of the form $(\mathbb{Z}_n, +)$ and $(G_n, \cdot)$. Also

groups (represented as linear transformations of a vector space) appeared in the work of Bravais on the possible structures of crystals.

In the late nineteenth century, the ideas of group theory began to pervade mathematics. A particularly notable development was the Erlanger Programme of the geometer Klein (1849–1925): the development of geometry (and geometries) in terms of the group of transformations which leave the particular geometry invariant.

Exercises 4.3

1. Decide which of the following sets are groups under the given operations:
   (i) the set $\mathbb{Q}$ of rational numbers, under multiplication;
   (ii) the set of non-zero complex numbers, under multiplication;
   (iii) the non-zero integers, under multiplication;
   (iv) the set of all functions from $\{1, 2, 3\}$ to itself, under composition of functions;
   (v) the set of all real numbers of the form $a + b\sqrt{2}$ where a and b are integers, under addition;
   (vi) the set of all 3×3 matrices of the form
   $$\begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix}$$
   where a, b, c are real numbers, under matrix multiplication;
   (vii) the set of integers under subtraction;
   (viii) the set of real numbers under the operation $*$ defined by $a * b = a + b + 2$.
2. Let G be a group and let a, b be elements of G. Show that
   $$(ab)^{-1} = b^{-1}a^{-1},$$
   and give an example of a group G with elements a, b for which
   $$(ab)^{-1} \neq a^{-1}b^{-1}.$$
3. Let G be a group in which $a^2 = e$ for all elements a of G. Show that G is Abelian.
4. Let G be a group and let c be a fixed element of G. Define a new operation '$*$' on G by
   $$a * b = ac^{-1}b.$$
   Prove that the set G is a group under $*$.

5. Let G be the set of all 3×3 matrices of the form

$$\begin{pmatrix} a & a & a \\ a & a & a \\ a & a & a \end{pmatrix}$$

where a is a non-zero real number. Find a matrix A in G such that, for all X in G, $AX = X = XA$. Prove that G is a group.

6. It was seen in Example 1 of Section 4.3.4 above that each possible permutation of the labels of the three vertices of an equilateral triangle is induced by a symmetry of the figure. On the other hand there are $4! = 24$ permutations of the labels of the vertices of a square, but there turn out to be only 8 symmetries of the square. Can you find a (short) argument which shows why only one-third of the permutations in $S(4)$ are obtained?
7. Write down the multiplication table for the group $D(4)$ of symmetries of a square.
8. Fill in the remainder of the following group table (the identity element does not necessarily head the first column). When you have done this, find all solutions of the equations:

$$\text{(i)}\, ax = b;\ \text{(ii)}\, xa = b;\ \text{(iii)}\, x^2 = c;\ \text{(iv)}\, x^3 = d.$$

	a	b	c	d	f	g
a	c			f		
b		f			c	
c	a					
d				c		
f						
g		d				c

[Hint: when you are constructing the table, remember that each group element appears exactly once in each row and column.]

9. Consult the literature to find out how to solve cubic equations 'by radicals'.

4.4 Algebraic structures

In the previous section, we saw that a group is defined to be a set together with an operation satisfying certain conditions, or **axioms**. These axioms were not chosen arbitrarily so as to generate some kind of intellectual game. The axioms were chosen to reflect properties common to a number of mathematical structures and we were able to present many examples of groups. In this section we consider briefly some of the other commonly arising algebraic structures.

Definition A **semigroup** is a set S, together with an operation $*$, which satisfies the following two properties (closure and associativity):

(S1) for all elements x and y of S, $x * y$ is in S;
(S2) for all elements x, y and z of S, we have $x * (y * z) = (x * y) * z$.

Since these are two of the group axioms, it follows that every group is a semigroup. But a semigroup need not have an identity element, nor need it have an inverse for each of its elements.

Example 1 The set of integers under multiplication, $(\mathbb{Z}, \cdot)$, is a semigroup. This semigroup has an identity element 1 but not every element has an inverse, so it is not a group.

Example 2 For any set X, the set, $F(X)$, of all functions from X to itself is a semigroup under composition of functions, since function composition is associative. For a specific example, take X to be the set $\{1,2\}$ (we considered this example in Section 2.2, but gave the functions different names there). There are four functions from X to X; the identity function e (so, $e(1) = 1$ and $e(2) = 2$), the function a with $a(1) = 2$ and $a(2) = 1$, the function b with $b(1) = b(2) = 1$ and the function c with $c(1) = c(2) = 2$. Here is the multiplication table for these functions under composition.

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

You can see that this is not a group table: for example, b has no inverse. In fact, since $b^2 = b$, b is an example of an **idempotent** element: an element which satisfies the equation $x^2 = x$ (such elements figure prominently in Boole's *Laws of Thought* – part of the definition of a Boolean algebra, see below, is that every element is idempotent under '$\wedge$' and '$\vee$' – and the term was introduced by B. Peirce in 1870 in his *Linear Associative Algebras*). In a group G, if $g^2 = g$ then we can multiply each side by g^{-1} to obtain $g = e$; hence e is the only element in a group which is idempotent.

Example 3 As in Example 2, let X be a set and let $F(X)$ be the semigroup of all functions from X to itself, under composition. Let f, $g \in F(X)$. If f is a bijection hence, by Theorem 2.2.3, has an inverse, then, given an equation

$fg = fh$ we may compose with f^{-1} to get

$$f^{-1}(fg) = f^{-1}(fh);$$

hence

$$(f^{-1}f)g = (f^{-1}f)h;$$

that is

$$\mathrm{id}_X g = \mathrm{id}_X h \text{ and so } g = h.$$

But it is not necessary that f be invertible: in fact f is an injection if and only if whenever $fg = fh$ we must have $g = h$. For let $x \in X$. Then $fg = fh$ implies $(fg)(x) = (fh)(x)$. That is

$$f(g(x)) = f(h(x)).$$

So, since f is injective, it follows that $g(x) = h(x)$. This is true for every $x \in X$, so $g = h$.

Suppose, conversely, that $f \in F(X)$ is such that for all $g, h \in F(X)$ we have that $fg = fh$ implies $g = h$: we show that f is injective. For, if not, then there would be distinct $x_1, x_2 \in X$ such that $f(x_1) = f(x_2) = z$ (say). Take g to be the function on X that interchanges x_1 and x_2 and fixes all other elements: g is the permutation $(x_1\ x_2)$. Take h to be the identity function on X. Then $fg = fh$, yet $g \neq h$ – contradiction.

So we have shown that f is injective if and only if $fg = fh$ implies $g = h$ for all $g, h \in F(X)$. Similarly f is surjective if and only if $gf = hf$ implies $g = h$ for all $g, h \in F(X)$ (you are asked to prove this as an exercise at the end of the section).

Example 4 A further class of examples of semigroups is provided by the finite state machines we discussed in Section 2.4. We illustrate this by determining the semigroup associated with the machine M which has states $S = \{0, 1, 2\}$ and input alphabet $\{\mathrm{a}, \mathrm{b}\}$ and whose state diagram is as shown in Fig. 4.12.

For any word w in $\{\mathrm{a},\mathrm{b}\}$, we define a function $f_w\colon S \to S$ by

$f_w(0)$ is the state M would end up in, if it started in state 0 and read w,
$f_w(1)$ is the state M would end up in, if it started in state 1 and read w, and
$f_w(2)$ is the state M would end up in, if it started in state 2 and read w.

Since there are only a finite number (27) of possible functions from S to S, there are only a finite number of different functions f_w. These distinct functions are the elements of the semigroup of M. In our example, f_{a} is the identity

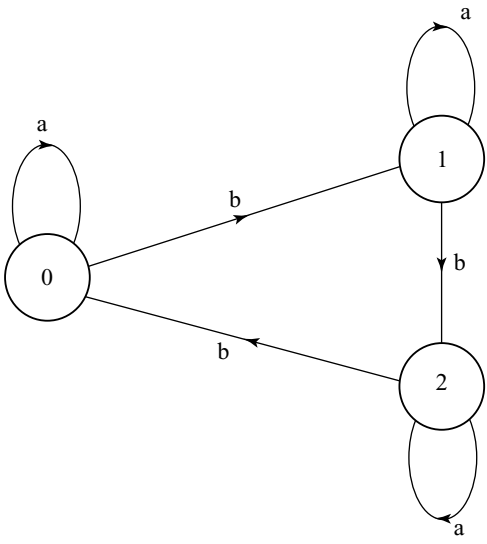

Fig. 4.12

function, taking each element of S to itself. The function f_{b} takes 0 to 1, 1 to 2 and 2 to 0. The function f_{bb} takes 0 to 2, 1 to 0 and 2 to 1. These are, in fact, the only distinct functions for M, so its semigroup has just three elements. The operation in the semigroup of M is that the 'product' of f_u and f_v is f_{uv}. We draw up the multiplication table for the semigroup, which is as shown.

	f_{a}	f_{b}	f_{bb}
f_{a}	f_{a}	f_{b}	f_{bb}
f_{b}	f_{b}	f_{bb}	f_{a}
f_{bb}	f_{bb}	f_{a}	f_{b}

(Note that, in our example, the final table is that of a group with three elements. That it is a group and not just a semigroup is just by chance.)

We now consider sets with two operations. These operations will be referred to as addition and multiplication although they need not be familiar 'additions' and 'multiplications': they need only satisfy the conditions listed below.

Definition A **ring** is a set R with two operations, called **addition** and **multiplication** and denoted in the usual way, satisfying the following properties: