Part II

Sets and functions

6

The language of set theory

The students' task in learning set theory is to steep themselves in unfamiliar and essentially shallow generalities till they become so familiar that they can be used with almost no conscious effort.

Paul R. Halmos, *Naive set theory* (adapted slightly).

The language of set theory is used throughout mathematics. Many general results involve 'an integer n' or 'a real number a' and, to start with, set theory notation provides a simple way of asserting for example that n is an integer. However, it turns out that this language is remarkably flexible and powerful and in much mathematics it is indispensable for a proper expression of the ideas involved.

In the second part of this book we introduce the basic vocabulary of the language of sets and functions. The third part of the book will then provide some experience in using this language as we use it to give a precise formulation of the idea of counting, one of the earliest mathematical concepts.

6.1 Sets

At this introductory level it is sufficient to define the notion of *set* as any well-defined collection of objects. We can think of a set as a box containing certain objects. In this section some ways of specifying sets are introduced and also some frequently used notation.

We frequently use a single letter to denote a set. This represents a further stage of mathematical abstraction. The reader will already have accepted the abstract notion of a positive integer, for example 'two' is abstracted from 'two apples' and 'two chairs'. Now we move on to consider the set of all positive integers as a single mathematical object. Each time there is further step of abstraction like this it takes time to

become familiar with new sorts of objects. At the end of this chapter we will make a further abstraction step when we consider sets which are collections of sets!

Particular sets which are often used have standard symbols to represent them. In this book the following will be used.

$\mathbb{Z}$ denotes the set of all integers.†
$\mathbb{Z}^+$ denotes the set of all positive integers (natural‡ or counting numbers): 1, 2, 3, etc.
$\mathbb{Z}^{\geqslant}$ denotes the set of all non-negative integers: 0, 1, 2, 3, etc.
$\mathbb{Q}$ denotes the set of all rational numbers (fractions).
$\mathbb{R}$ denotes the set of all real numbers (i.e. numbers expressible as infinite decimals).
$\mathbb{R}^+$ denotes the set of all positive real numbers.
$\mathbb{R}^{\geqslant}$ denotes the set of all non-negative real numbers.
$\mathbb{C}$ denotes the set of all complex numbers.

The objects in a set are called the *elements*, *members* or *points* of the set. We write

$$x \in E$$

to denote the fact that the object x is an element of the set E. Thus for example '$a \in \mathbb{R}$' is read 'a is an element of the set of real numbers' or more simply just 'a is a real number'. The symbol '$\in$', first used in this way by the Italian mathematician Giuseppe Peano towards the end of the nineteenth century, is a variant of the Greek letter epsilon and care should normally be taken to distinguish it from that letter which is usually written 'ε' or 'ϵ'. However, some books do not make this distinction and use epsilon in place of '$\in$'. It is common to use upper case letters to represent sets and lower case to represent elements but this is not always appropriate: in geometry it is usual to use upper case letters to represent points and lower case letters to represent lines (which are sets of points); another exception occurs when a set is itself considered as an element of another set.

The negation of the statement $x \in A$ is written $x \notin A$. Thus $\sqrt{2} \notin \mathbb{Q}$ is the statement that $\sqrt{2}$ is not a rational number.

There are basically three ways of specifying a set: we can list the

† This symbol comes from '*Zahlen*', the German word for 'numbers'.
‡ Some people include the number 0 in the set of natural numbers but this seems to me unnatural as we usually start counting at 1. Because of this ambiguity, in this book we will normally refer to 'the non-negative integers' or 'the positive integers' depending on whether or not the number 0 is included.

elements, specify a condition for membership, or give a formula or algorithm constructing the elements of the set.

(a) Listing the elements of a set

When we list the elements of a set we denote the set by enclosing the elements in curly brackets. Thus

$$A = \{1, 3, \pi, -14\}$$

is the statement that A is the set whose four elements are 1, 3, π and -14. In this case we have $1 \in A$ and $2 \notin A$. Notice that the order in which the elements are listed is unimportant and repeating an element makes no difference. Thus, for the above set A we also have

$$A = \{\pi, 3, -14, 1\} = \{\pi; 3, \pi, 1, -14, -14\}.$$

On the face of it the listing notation is only practical for sets with a small number of elements. However it can be extended to large or even infinite sets. For example we can write

$$\mathbb{Z}^+ = \{\, 1, 2, 3, \ldots \}$$

where as usual the dots are read 'and so on'.

(b) The conditional definition of a set

Alternatively, a set may be described by specifying some condition which determines whether or not an object is an element of the set. Let us illustrate this with an example. Consider the set B defined as follows:

$$B = \{\, n \in \mathbb{Z} \mid 0 < n < 6 \,\}.$$

Here we write $n \in \mathbb{Z}$ simply to indicate the sort of objects that we are considering. The statement $0 < n < 6$ is a predicate. The definition means that B is the set of integers which when substituted in this predicate give a true proposition, i.e. given an integer n,

$$n \in B \Leftrightarrow 0 < n < 6.$$

Of course we could have described this particular set by listing the elements,

$$B = \{1, 2, 3, 4, 5\},$$

but in some cases this is difficult or impossible.

In the conditional definition of a set the vertical line† '|' is read as 'such that' and so the above definition would be read as 'B is the set of integers n such that $0 < n < 6$' or just 'B is the set of integers between 0 and 6'. This last reading, which makes no reference to the variable 'n', demonstrates that the symbol n in this definition represents a *dummy variable*: its only rôle is to indicate the internal logic of the definition and it can be replaced by any other symbol (not already in use) without any change in meaning. Thus we could equally write

$$B = \{\alpha \in \mathbb{Z} \mid 0 < \alpha < 6\}.$$

Notice that, although $0 < n < 6$ is a predicate and so a mathematical statement in the sense of Chapter 1, $\{n \in \mathbb{Z} \mid 0 < n < 6\}$ is not a statement; it is simply a mathematical object and can be no more true or false than the number 2 can be true or false. However, we can make statements about this object, such as '$2 \in \{n \in \mathbb{Z} \mid 0 < n < 6\}$' which is true, or '$m \in \{n \in \mathbb{Z} \mid 0 < n < 6\}$' which is equivalent to '$m \in \mathbb{Z}$ and $0 < m < 6$'.

(c) The constructive definition of a set

The other systematic method of describing a set is to give a formula (or more generally an algorithm) constructing the elements of the set. For example

$$\{n^2 \mid n \in \mathbb{Z}\} = \{0, 1, 4, 9, 16, \dots\}$$

is the set of integer squares, which means that the formula n^2 constructs the elements of the set as n takes all possible integer values, i.e. an element is in the set if and only if it can be written as n^2 for some integer n. Notice that in this case all non-zero elements in the set arise twice. This makes no difference to the set so that for example $\{n^2 \mid n \in \mathbb{Z}^{\geqslant}\}$ defines the same set.

Similarly,

$$\{2q \mid q \in \mathbb{Z}\} = \{\dots, -4, -2, 0, 2, 4, 6, \dots\}$$

is the set of even integers, and

$$\{a/b \mid a, b \in \mathbb{Z},\ b \neq 0\}$$

is the set of rational numbers, $\mathbb{Q}$. This last example illustrates the

† Alternative notations are to use a colon ':' or a semi-colon ';' here instead of '|' as follows: $B = \{n \in \mathbb{Z} : 0 < n < 6\}$ or $\{n \in \mathbb{Z}; 0 < n < 6\}$.

convention that when more than one condition is listed after the symbol '|' this means that all the conditions must be satisfied for an object to be in the set so that the word 'and' is understood.

Equality of sets

Quite commonly problems in mathematics take the form of seeking to pass between different ways of specifying the elements of a set. For example when we learn how to solve quadratic equations of the form $ax^2 + bx + c = 0$, where a, b and c are given real numbers, we are learning how to list the elements of the set $\{ x \in \mathbb{R} \mid ax^2 + bx + c = 0 \}$, or possibly the set $\{ x \in \mathbb{C} \mid ax^2 + bx + c = 0 \}$ if we allow complex solutions.

Definition 6.1.1 *Two sets A and B are* equal, *written* $A = B$, *if they have precisely the same elements, i.e. $A = B$ means $x \in A \Leftrightarrow x \in B$.*

To put it another way: a set is determined by its elements. Notice that this means that to show that two sets A and B are equal it is necessary to prove two things (although they can often be done together in simple cases): every element of A is an element of B and conversely every element of B is an element of A.

Example 6.1.2 $\{ x \in \mathbb{R} \mid x^2 - x - 2 = 0 \} = \{-1, 2\}$.

Constructing a proof. This is another way of stating that, for x a real number,

$$x^2 - x - 2 = 0 \Leftrightarrow x = -1 \text{ or } x = 2.$$

This is easily proved by the factorization method.

Proof For x a real number, $x^2 - x - 2 = 0 \Leftrightarrow (x - 2)(x + 1) = 0 \Leftrightarrow x - 2 = 0$ or $x + 1 = 0$ (by Proposition 4.4.1) $\Leftrightarrow x = 2$ or $x = -1$. Hence $x^2 - x - 2 = 0$ if and only if $x = -1$ or $x = 2$, as required. □

At this stage it is convenient to introduce two other ideas from set theory.

Definition 6.1.3 *The* empty set *is the unique set which has no elements at all. It is denoted by the symbol* $\emptyset$.

Thus the statement that the quadratic equation $x^2 + 2x + 2 = 0$ has no real solutions may be written $\{ x \in \mathbb{R} \mid x^2 + 2x + 2 = 0 \} = \emptyset$.

Take care to distinguish '$\emptyset$' which is a variant of a Scandinavian letter from the Greek letter phi written 'ϕ'.

Definition 6.1.4 *Given sets A and B we say that* $\boxed{A \text{ is a subset of } B}$, *written* $\boxed{A \subseteq B}$, *or $B \supseteq A$, when every element of A is an element of B, i.e. $x \in A \Rightarrow x \in B$. If A and B are in addition unequal so that B contains some element not contained in A, then we say that A is a* $\boxed{\text{proper}}$ *subset of B and write* $\boxed{A \subset B}$.†

Thus

$$A = B \Leftrightarrow (A \subseteq B \text{ and } B \subseteq A).$$

It is important to distinguish between the symbols $\in$ and $\subseteq$ although they are closely related as follows:

$$a \in A \Leftrightarrow \{a\} \subseteq A.$$

If sets are defined by predicates then there is a correspondence between the notions of 'implication' and 'subset': the universal statement that $P(a) \Rightarrow Q(a)$ for all $a \in A$ is equivalent to the statement that the set $\{ a \in A \mid P(a) \}$ is a subset of $\{ a \in A \mid Q(a) \}$.

Notice that if $A \subseteq B$ and $B \subseteq C$ then $A \subseteq C$ since $x \in A \Rightarrow x \in B$ and $x \in B \Rightarrow x \in C$ together imply that $x \in A \Rightarrow x \in C$. Furthermore $\emptyset \subseteq A$ for all sets A whereas $A \subseteq \emptyset$ only if $A = \emptyset$.

6.2 Operations on sets

Definition 6.2.1 *Given two sets A and B we can form the set of elements which lie both in A and in B. This is called the* $\boxed{\text{intersection}}$ *of A and B and is denoted by* $\boxed{A \cap B}$. *Thus*

$$A \cap B = \{ x \mid x \in A \text{ and } x \in B \}.$$

† Some mathematicians use $\subset$ where $\subseteq$ is used in this book. In fact when Peano originally introduced this notation he used it the other way round, writing $A \supset B$ to indicate that A is a subset of B!

Two sets A and B are said to be disjoint *if $A \cap B = \emptyset$, i.e. A and B have no elements in common.*

Definition 6.2.2 *Given two sets A and B we can form the set of elements which lie in A or lie in B. This is called the* union *of A and B and is denoted by $A \cup B$. Thus*

$$A \cup B = \{ x \mid x \in A \text{ or } x \in B \}.$$

Definition 6.2.3 *Given two sets A and B we can form the set of elements which lie in A but not in B. This is called the* difference *of A and B and is denoted by† $A - B$. Thus*

$$A - B = \{ x \mid x \in A \text{ and } x \notin B \}.$$

Notice that $A \cap A = A = A \cup A$, $A \cap \emptyset = \emptyset$, $A \cup \emptyset = A$, $A - A = \emptyset$ and $A - \emptyset = A$.

Proposition 6.2.4 *Given any two sets A and B, the three sets $A \cap B$, $A - B$ and $B - A$ are pairwise disjoint (i.e. each pair of these sets is disjoint) and*

$$A \cup B = (A \cap B) \cup (A - B) \cup (B - A).$$

Probably the simplest way to prove a statement like this is by means of truth tables as follows.

Proof Consider the truth tables on the following page.

The fact that the final two columns of the second table are the same tells us that the above equality of sets holds. The fact that no row has more than one T in the third, fourth and fifth columns of the first table

† The reader should be aware that this notation is sometimes used in algebra to denote the set $\{ a - b \mid a \in A, b \in B \}$. In this case the difference of the two sets A and B is denoted by $A \setminus B$.

tells us that the three corresponding sets are disjoint.

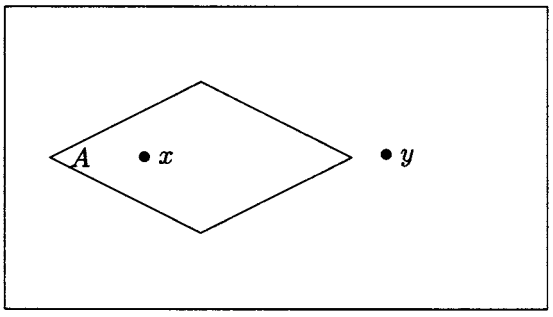
$x \in A$	$x \in B$	$x \in A \cap B$	$x \in A - B$	$x \in B - A$
T	T	T	F	F
T	F	F	T	F
F	T	F	F	T
F	F	F	F	F

$x \in A$	$x \in B$	$x \in (A \cap B) \cup (A - B) \cup (B - A)$	$x \in A \cup B$
T	T	T	T
T	F	T	T
F	T	T	T
F	F	F	F

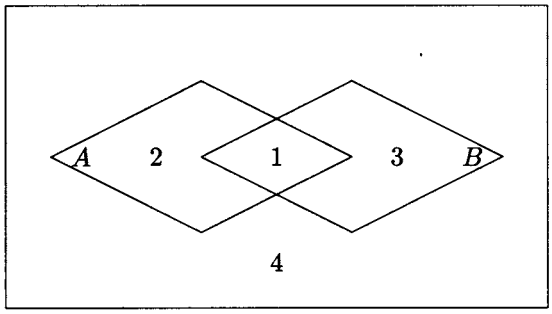
□

In the truth tables in the above proof the columns are headed by statements $x \in A$, $x \in B$, etc. which can be true or false, not simply by the names of the sets A, B, etc. which would not be statements.

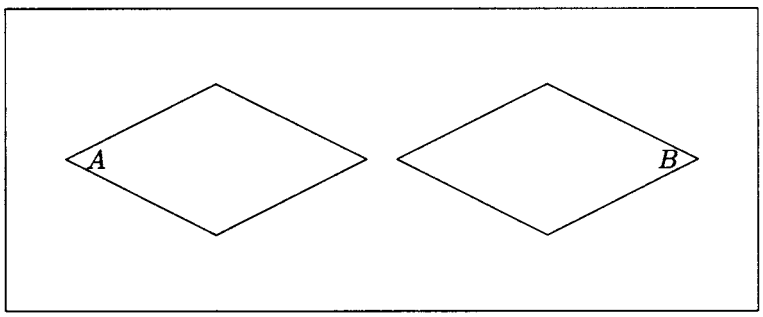
This proof illustrates the close relationship between the logical connectives introduced in Chapter 1 and the operations on sets defined above in terms of those connectives. Proofs of the above type are usually illustrated by a *Venn diagram.* We indicate a set A by the interior region of some curve drawn on the page so that the elements of the set correspond to points in the region. Thus for example in the following diagram $x \in A$ but $y \notin A$.



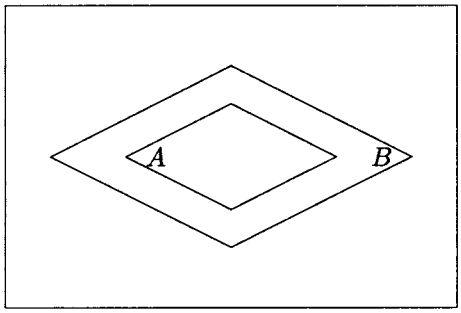
Two general sets A and B are described by two overlapping regions. Consider the following diagram.



Here the interior of the left-hand rhombus (regions 1 and 2) denotes A and the interior of the right-hand rhombus (regions 1 and 3) denotes B. Then the four rows of the above truth tables correspond to the four regions in the diagram. Thus region 1 corresponds to $A \cap B$, region 2 corresponds to $A - B$, region 3 corresponds to $B - A$ and the three regions 1, 2 and 3 together give $A \cup B$. The fact that regions 1, 2 and 3 are disjoint corresponds to the fact that the three sets on the right-hand side of Example 6.2.4 are disjoint and the fact that these three regions together make up $A \cup B$ corresponds to the equality of sets in Proposition 6.2.4. From this point of view the result is essentially obvious. Care is needed in using this sort of diagrammatic proof in more elaborate examples because unless proper care is taken it is sometimes possible to draw a diagram that does not include all possible regions or some feature of the diagram has nothing to do with the set theory (see for example the solution to Exercise 6.6). Notice that this diagram does not imply that $A \cap B$ is non-empty for it may be that there are no elements in the region 1. However, if we were given that $A \cap B = \emptyset$ then we could denote this by the following Venn diagram.



We can represent $A \subseteq B$ as follows.



6.3 The power set

Definition 6.3.1 *The* power set *of a set X, denoted by $\mathcal{P}(X)$, is the set of all subsets of the set X. Thus $A \in \mathcal{P}(X)$ is another way of writing $A \subseteq X$.*

Example 6.3.2 *If $X = \{a, b, c\}$ then*

$$\mathcal{P}(X) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, X\}.$$

Notice that the empty set $\emptyset$ is an element of the power set $\mathcal{P}(X)$ for any set X since $\emptyset \subseteq X$. A set like $\{a\}$ with a single element is called a *singleton*. It is important to distinguish the singleton set $\{a\}$ from the element a. In particular the singleton $\{\emptyset\}$ is to be distinguished from the empty set $\emptyset$: a box containing an empty box is not an empty box!

It is often the case that all the sets we are considering are subsets of some fixed set, say the set of real numbers. We then consider this to be the *universal set*.

Definition 6.3.3 *Once we have fixed a universal set U we can define the* complement *of any $A \in \mathcal{P}(U)$, denoted by A^c, to be the difference of U and A. Thus*

$$A^c = U - A = \{\, x \in U \mid x \notin A \,\}.$$

For example, if the universal set is $\mathbb{Z}$, the set of integers, and E is the set of even integers, then the complement E^c is the set of odd integers.

The intersection, union and complement of subsets of some universal set U correspond to the logical connectives 'and', 'or' and 'not'. The relationships between these operations may be summed up in the following theorem.

Theorem 6.3.4 *Let A, B and C be subsets of some universal set U (i.e. $A, B, C \in \mathcal{P}(U)$). Then we have the following identities.*

(i) associativity: $A \cup (B \cup C) = (A \cup B) \cup C$, $A \cap (B \cap C) = (A \cap B) \cap C$.
(ii) commutativity: $A \cup B = B \cup A$, $A \cap B = B \cap A$.
(iii) distributivity: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
(iv) De Morgan laws:† $(A \cup B)^c = A^c \cap B^c$, $(A \cap B)^c = A^c \cup B^c$.
(v) complementation: $A \cup A^c = U$, $A \cap A^c = \emptyset$.
(vi) double complement: $(A^c)^c = A$.

These can be proved by truth tables or Venn diagrams. Alternatively they can be proved by using logical argument from the definitions. Let us illustrate this by writing out the proof of one part.

Proof of $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Proof of '$\subseteq$': Suppose that $x \in A \cap (B \cup C)$. Then $x \in A$ and $x \in (B \cup C)$. Since $x \in B \cup C$, $x \in B$ or $x \in C$. If $x \in B$ then, since $x \in A$ as well, we have $x \in A \cap B$ and so $x \in (A \cap B) \cup (A \cap C)$ as required. On the other hand, if $x \notin B$, then we must have $x \in C$ and so, since also $x \in A$, we have $x \in A \cap C$ and so $x \in (A \cap B) \cup (A \cap C)$.

Proof of '$\supseteq$': Suppose now that $x \in (A \cap B) \cup (A \cap C)$. Then $x \in A \cap B$ or $x \in A \cap C$. If $x \in A \cap B$, then $x \in A$ and $x \in B$ so that $x \in A$ and $x \in B \cup C$ which gives $x \in A \cap (B \cup C)$ as required. On the other hand if $x \notin A \cap B$ then $x \in A \cap C$ and again we get $x \in A \cap (B \cup C)$. □

We can now use these results to derive other set identities by algebraic manipulation. Here is an example.

Proposition 6.3.5

$$(A \cup B) \cap (C \cup D) = (A \cap C) \cup (A \cap D) \cup (B \cap C) \cup (B \cap D).$$

† These results was observed independently in the nineteenth century by the British mathematician Augustus De Morgan and the United States mathematician Benjamin Peirce (see C.B. Boyer and U.C. Merzbach, *A history of mathematics*, Wiley, Second edition 1989). They correspond to the equivalence of the statements 'not (P and Q)' and '(not P) or (not Q)' commented on in the solution to Exercise 1.2.

Proof First notice that commutativity means that distributivity on one side implies distributivity on the other side so that Theorem 6.3.4(iii) implies that $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$.

$$\begin{aligned}
&(A \cup B) \cap (C \cup D) \\
&\quad = \big(A \cap (C \cup D)\big) \cup \big(B \cap (C \cup D)\big) \quad \text{by distributivity} \\
&\quad = (A \cap C) \cup (A \cap D) \cup (B \cap C) \cup (B \cap D) \quad \text{by distributivity.}
\end{aligned}$$

The associativity of the union operation means that we do not need any additional brackets here. □

Exercises

6.1 The following are standard subsets of the set of real numbers known as the real intervals with endpoints the real numbers a and b.

The *open interval:* $(a, b) = \{ x \in \mathbb{R} \mid a < x < b \}$.
The *closed interval:* $[a, b] = \{ x \in \mathbb{R} \mid a \leqslant x \leqslant b \}$.
The *right half-open interval:* $[a, b) = \{ x \in \mathbb{R} \mid a \leqslant x < b \}$.
The *left half-open interval:* $(a, b] = \{ x \in \mathbb{R} \mid a < x \leqslant b \}$.

(i) Prove that $0 \notin (0, 1)$, $0 \in [0, 1]$, $0 \in [0, 1)$ and $0 \notin (0, 1]$.
(ii) Find the elements of the set $[a, b] - (a, b)$.
(iii) Prove that $(a, b) = \emptyset$ if and only if $a \geqslant b$. [Hint: Prove the contrapositive.]
Find the corresponding results for the other real intervals with endpoints a and b.
(iv) Prove that, if $a \leqslant b$, then $[a, b] \subseteq (c, d)$ if and only if $c < a$ and $b < d$.

6.2 Prove that

(i) $\{ x \in \mathbb{R} \mid x^2 + x - 2 = 0 \} = \{1, -2\}$,
(ii) $\{ x \in \mathbb{R} \mid x^2 + x - 2 < 0 \} = (-2, 1)$,
(iii) $\{ x \in \mathbb{R} \mid x^2 + x - 2 > 0 \} = \{ x \in \mathbb{R} \mid x < -2 \} \cup \{ x \in \mathbb{R} \mid x > 1 \}$.

6.3 Find predicates which determine the following subsets of the set of integers $\mathbb{Z}$: (i) $\{3\}$, (ii) $\{1, 2, 3\}$, (iii) $\{1, 3\}$.

6.4 By using a truth table prove that $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$. Draw a Venn diagram to illustrate the proof.

6.5 Prove that

(i) $A \subseteq B \Leftrightarrow A \cup B = B$,
(ii) $A \subseteq B \Leftrightarrow A \cap B = A$.

6.6 Prove by contradiction that, if $A \cap B \subseteq C$ and $x \in B$, then $x \notin A - C$. [Work from the definitions of '$A \cap B$', '$A - C$', and '$\subseteq$'.]

6.7 Using the fact that an implication is equivalent to its contrapositive, prove that, for subsets of a universal set U, $A \subseteq B$ if and only if $B^c \subseteq A^c$.

7
Quantifiers

In Chapter 1 a predicate was described as an expression containing one or more free variables; it becomes a proposition, and so is true or false, when a specific value is assigned to each free variable. Of course whether this proposition is true or is false usually depends on the values selected.

However, we saw in the last chapter that a proposition can be created from a predicate in another way – by making a statement about the set of values of the free variables which make it true. Many results in mathematics take the form of listing the values of this set, for example when we solve an equation. But often results simply address the question of whether there is any choice of values of the free variables resulting in a true proposition and whether there is any choice resulting in a false proposition. Statements that such values exist are known as existential statements. Statements that they do not can be thought of as universal statements. We met examples of universal statements when discussing implications in Chapter 2.

In this chapter we discuss general universal and existential statements.

7.1 Universal statements

Suppose that $P(a)$ is a predicate with a single free variable a with possible values in a set A. Usually $P(a)$ is true for some elements of the set A and false for others, and in the last chapter we described how such a predicate could be used to describe a subset of A, denoted by $\{\, a \in A \mid P(a) \,\}$, the subset of elements of A for which the statement $P(a)$ is true. In certain cases this subset is the whole of A; the statement that this occurs is a *universal statement.*

Definition 7.1.1 *The notation* $\boxed{\forall a \in A,\ P(a)}$ *is an alternative way of writing*

$$\{\, a \in A \mid P(a) \,\} = A.$$

It is read: 'for each element a in the set A the proposition $P(a)$ is true' or '$P(a)$ is true for each a in the set A'.

In this notation, the comma between '$\forall a \in A$' and the predicate '$P(a)$' is included simply to clarify the expression. Sometimes brackets are used. The symbol '$\forall$' is called the *universal quantifier symbol* and is read 'for each', 'for every', 'for all' or 'for any'. In this book we will adopt the practice of putting quantifiers before the predicates to which they refer. However, we will usually write these statements in words and in this case show more flexibility about the order.

As an example consider Proposition 3.1.4 which asserts that $a^2 > 0$ for any non-zero real number a. This is a statement about the set of non-zero real numbers, $\mathbb{R} - \{0\}$. The predicate is '$a^2 > 0$'. We may write the statement in symbols as

$$\forall a \in \mathbb{R} - \{0\},\ a^2 > 0$$

or equivalently

$$\{\, a \in \mathbb{R} - \{0\} \mid a^2 > 0 \,\} = \mathbb{R} - \{0\}.$$

A third way of writing the statement is as the universal implication

$$a \in \mathbb{R} - \{0\} \Rightarrow a^2 > 0$$

and if you look back to Chapter 3 you will see that this is how we proved the statement.

Notice that in all three of these expressions a is a 'dummy' or 'bound' variable and can be changed to any other symbol without changing the meaning. For example

$$\forall x \in \mathbb{R} - \{0\},\ x^2 > 0$$

is equivalent to the above statements.

7.2 Existential statements

We now turn to the negation of a universal statement: the statement that it is false. In Chapter 2 we considered the statement $x > 0 \not\Rightarrow x \geqslant 1$ for real numbers x and explained that this meant that the universal

statement $x > 0 \Rightarrow x \geqslant 1$, which is shorthand for the statement $\forall x \in \mathbb{R}\,(x > 0 \Rightarrow x \geqslant 1)$, is false. Table 2.1.2 described the behaviour for different values of x and of course, for any x such that $0 < x < 1$, $x > 0$ is true and $x \geqslant 1$ is false so that $x > 0 \Rightarrow x \geqslant 1$ is false. However, to demonstrate that $\forall x \in \mathbb{R}\,(x > 0 \Rightarrow x \geqslant 1)$ is false we simply have to show that there is a single value of $x \in \mathbb{R}$ for which $x > 0 \Rightarrow x \geqslant 1$ is false, i.e. $x > 0$ and $x \ngeqslant 1$, in other words the set $\{\, x \in \mathbb{R} \mid x > 0 \text{ and } x \ngeqslant 1 \,\}$ is non-empty. The simplest way to do this is to give a specific element in the set, such as $1/2$: $1/2 > 0$ and $1/2 \ngeqslant 1$. Such a statement, that the subset defined by a predicate is non-empty, is called an *existential statement.*

Definition 7.2.1 *The notation* $\boxed{\exists a \in A,\ P(a)}$ *is an alternative way of writing*

$$\{\, a \in A \mid P(a) \,\} \neq \emptyset.$$

It is read: 'for some element a in the set A the proposition $P(a)$ is true' or '$P(a)$ is true for some a in the set A'.

The symbol '$\exists$' is called the *existential quantifier symbol* and is read 'for some', 'for at least one' or sometimes 'there exists ... such that'.

Remarks 7.2.2 Notice that the word 'any' sometimes indicates a universal statement and sometimes an existential statement.

The normal meaning of 'any' is 'every' as in '$a^2 \geqslant 0$ for any real number a'. This is a universal statement which can be written symbolically as '$\forall a \in \mathbb{R},\ a^2 \geqslant 0$'. However, in negative or interrogative statements 'any' is used idiomatically to mean 'some'. For example, 'There is not any real real number a such that $a^2 < 0$' is asserting that the existential statement '$\exists a \in \mathbb{R},\ a^2 < 0$' is false. And 'Is there any real number a such that $a^2 = 2$?' is asking whether the existential statement '$\exists a \in \mathbb{R},\ a^2 = 2$' is true.

Fowler† gives some non-mathematical examples: 'Have you any bananas?' with the possible answers 'No we haven't any bananas' and 'Yes we have some bananas'.

Real care is required with questions involving 'any'. 'Is there any integer a such that $a \geqslant 1$?' seems clear enough and is asking whether '$\exists a \in \mathbb{Z},\ a \geqslant 1$' is true. But 'Is $a \geqslant 1$ for any integer a?' seems less clear

† H.W. Fowler, *A dictionary of modern English usage* (revised by Ernest Gowers), Oxford University Press, Second edition 1968.

to me and might be taken to asking about the same statement as the first question, '$\exists a \in \mathbb{Z}$, $a \geqslant 1$' (which is true) but might also be taken to be asking about '$\forall a \in \mathbb{Z}$, $a \geqslant 1$' (which is false). Great care is needed in using 'for any' in interrogative statements.

7.3 Proving statements involving quantifiers

An enormous number of results in advanced mathematics take the form of asserting the truth or falsehood of some universal or existential statement; this is one of the factors which distinguishes advanced from elementary mathematics and many of the results in this book take this form. This section provides an overview of the main methods of proof but in effect much of the whole book is about proving such results.

(a) Proving statements of the form $\forall a \in A$, $P(a)$

We usually prove statements of this form by rewriting them in the form

$$a \in A \Rightarrow P(a).$$

An example of this is the proof of Proposition 3.1.4 which we have already discussed.

(b) Proving statements of the form $\exists a \in A$, $P(a)$

We often prove statements of this form by simply exhibiting a particular element $a \in A$ for which $P(a)$ is true. This is *proof by example.*

Example 7.3.1 To prove $\exists n \in \mathbb{Z}$, $n^2 = 9$.

Solution Observe that $3 \in \mathbb{Z}$ and $3^2 = 9$ and so $n = 3$ provides an example proving this statement. □

There are, however, less direct methods of proving existential statements such as the use of the counting arguments which will be considered in Chapter 11.

(c) Proving statements involving both quantifiers

Very many statements involve both quantifiers. Consider the result of Exercise 3.3.

Proposition 7.3.2 *For integers n, if n is even then n^2 is even.*

This is a universal implication: $\forall n \in \mathbb{Z}\,(n \text{ is even} \Rightarrow n^2 \text{ is even})$. However, the hypothesis that n is even is an existence statement, which may be written $\exists q \in \mathbb{Z},\ n = 2q$. We can use this by making use of some specific integer q such that $n = 2q$. Thus we begin the proof of this result, by the direct method, as follows.

Suppose that n is an even integer. Then $n = 2q$ for some integer q.

The conclusion which we are aiming for is the statement that n^2 is even, which may be written, again spelling out the definition, $\exists q \in \mathbb{Z},\ n^2 = 2q$. Recall that in such statements the symbol 'q' is a dummy variable and its only rôle is to glue the notation together: in words we could write this as 'n^2 is twice some integer.' We could replace q by any other symbol not already in use, for example $\exists p \in \mathbb{Z},\ n^2 = 2p$. Indeed in this case we ought to do this since the symbol q is already in use: when we wrote 'Then $n = 2q$ for some integer q' we were using q to denote some specific integer with the property that $n = 2q$. It is only by doing this that we can make use of the existence statement. We can now proceed to find an integer p such that $n^2 = 2p$ and complete the proof as follows.

Therefore $n^2 = (2q)^2 = 4q^2 = 2(2q^2)$ and so, since $2q^2$ is an integer, n^2 is even.

Hence, if n is even, then n^2 is even.

This is actually written out without reference to 'p' although we could have said '$n^2 = 2p$ where $p = 2q^2$ is an integer and so n^2 is even.'

There is a genuine ambiguity about the statement '$n = 2q$ for some $q \in \mathbb{Z}$'. It may be the existential statement that the integer q exists and this is what is meant by the statement $\exists q \in \mathbb{Z},\ n = 2q$. Alternatively, it may be the statement that q is a specific integer such that $n = 2q$ as occurs in this proof. The statement with the second meaning is possible because of the statement with the first meaning, and we could make this clear by writing out the proof as follows.

Suppose that n is even. Then $\exists q \in \mathbb{Z},\ a = 2q$. So let q_1 be an integer such that $n = 2q_1$. Then $n^2 = (2q_1)^2 = 2(2q_1^2)$. Hence, since $2q_1^2$ is an integer, $\exists p \in \mathbb{Z},\ n^2 = 2p$. Thus n^2 is even.

Hence, if n is even then n^2 is even.

In practice this distinction is blurred and I would encourage the reader not to worry about it – it is a case where ambiguity is better than

pedantry. Most problems are avoided so long as a different dummy variable is used each time a definition involving a quantifier is used in the proof.

7.4 Disproving statements involving quantifiers

The idea of disproving statements can appear a little strange at first, but to some extent this is a matter of presentation: disproving 'P' is the same as proving 'not P'.

(a) Disproving statements of the form $\forall a \in A$, $P(a)$

We have already observed that the negation of this statement is the statement

$$\exists a \in A, \text{ not } P(a)$$

and so we can disprove it by giving a single example for which it is false. This is called *disproof by counterexample* to $P(a)$.

Example 7.4.1 To disprove the statement $\forall x \in \mathbb{R}$, $x^2 > 2$.

Solution A counterexample is provided by $x = 1$ since $1 \in \mathbb{R}$ and $1^2 = 1 \leqslant 2$. □

Great care is required in interpreting negatives of universal statements in everyday speech. For example consider the statement 'All the members of the class are not here' which would normally be taken to be the negative of the universal statement 'All the members of the class are here', in other words 'Some member of the class is not here.' This differs from our careful usage. Consider a mathematical statement of the same structure: 'All the numbers in the set are not even.' This must mean the same as 'All the numbers in the set are odd' since 'odd' means the same as 'not even'. But the everyday usage indicated above would give 'Some number in the set is odd.' Take care to use language forms which are not open to misunderstanding: if we wish to indicate an absence from a class then we should say 'Not all the members of the class are here.' But then, away from mathematics, one can have a lot of fun with ambiguity!

(b) Disproving statements of the form $\exists a \in A,\ P(a)$

The negation of this statement, often written

$$\nexists a \in A,\ P(a),$$

is the statement

$$\forall a \in A, \text{ not } P(a)$$

and this gives one way of disproving the statement. We made use of this fact in the proof of Proposition 2.2.4. Here is another very simple and familiar example.

Proposition 7.4.2 *There does not exist a real number x such that $x^2 = -1$.*

Proof We know that, for all $x \in \mathbb{R}$, we have the inequality $x^2 \geqslant 0$ and so $x^2 \neq -1$. Hence there does not exist $x \in \mathbb{R}$ such that $x^2 = -1$. □

The other way of disproving an existence statement is by contradiction. Here we show that the statement $P(a)$ where $a \in A$ necessarily leads to a contradiction. Proposition 4.1.1 which is a non-existence statement was proved using this method.

7.5 Proof by induction

We can reformulate the method of proof by induction using the language of set theory. Recall from Chapter 5 that induction is used to prove statements of the form $\forall n \in \mathbb{Z}^+,\ P(n)$. Such a statement can be rewritten as $\{\, n \in \mathbb{Z}^+ \mid P(n) \,\} = \mathbb{Z}^+$. Thus induction can be thought of as a method for proving that certain subsets of $\mathbb{Z}^+$, the set of positive integers, are in fact the whole set. From this point of view we can express the induction principle (Axiom 5.1.1) as follows.

Axiom 7.5.1 (The induction principle reformulated) *Suppose that A is a subset of $\mathbb{Z}^+$, the set of positive integers. Then $A = \mathbb{Z}^+$ if*

(i) $1 \in A$, *and*
(ii) $\forall k \in \mathbb{Z}^+ (k \in A \Rightarrow k + 1 \in A)$.

This statement reduces to Axiom 5.1.1 if we put $A = \{\, n \in \mathbb{Z}^+ \mid P(n) \,\}$. It can be deduced from Axiom 5.1.1 if we write $P(n)$ for the predicate $n \in A$.

Induction is quite often formulated in this more formal way.

7.6 Predicates involving more that one free variable

We have already met a number of universal and existential statements involving more than one variable. If $P(a, b)$ is a predicate involving two free variables $a \in A$ and $b \in B$ then we can form propositions involving quantifiers as follows.

(i) $\forall a \in A, \forall b \in B, P(a, b)$.
(ii) $\exists a \in A, \exists b \in B, P(a, b)$.
(iii) $\forall a \in A, \exists b \in B, P(a, b)$.
(iv) $\exists b \in B, \forall a \in A, P(a, b)$.
(v) $\forall b \in B, \exists a \in A, P(a, b)$.
(vi) $\exists a \in A, \forall b \in B, P(a, b)$.

The meaning of the first two of these is fairly clear. Examples of these which we have aready met are as follows.

Proposition 3.1.1 $\forall a, b \in \mathbb{R}^+, a < b \Rightarrow a^2 < b^2$.

Proposition 3.2.1 $\forall a, b \in \mathbb{R}, a < b \Rightarrow 4ab < (a + b)^2$.

Proposition 4.1.1 *It is not true that* $\exists m, n \in \mathbb{Z}, 14m + 20n = 101$.

Notice in these examples that '$\forall a, b \in A$' is a shorthand for '$\forall a \in A$, $\forall b \in A$' and similarly for '$\exists$'.

Statements involving both quantifiers require some care in particular regarding the order of the quantifiers. Consider for example the predicate '$m < n$' involving positive integers m and n.

Example 7.6.1 $\forall m \in \mathbb{Z}^+, \exists n \in \mathbb{Z}^+, m < n$.

This is the statement that $\{ m \in \mathbb{Z}^+ \mid \exists n \in \mathbb{Z}^+, m < n \} = \mathbb{Z}^+$ or that $m \in \mathbb{Z}^+ \Rightarrow (\exists n \in \mathbb{Z}^+, m < n)$. Notice how the use of the single quantifier '$\exists n \in \mathbb{Z}^+$' leads to a predicate '$\exists n \in \mathbb{Z}^+, m < n$' with a single free variable m. We can then consider for which values of m this predicate is true. In this case we are considering the assertion that it holds for all positive integers m. This means that for each positive integer m, there exists a greater integer n. This is clearly the case and we can prove it by example: take $n = m + 1$. We can write out a formal proof as follows.

Proof This result is true because, given a positive integer m, if we put $n = m + 1$ then n is a positive integer and $m < n$. □

Example 7.6.2 $\exists n \in \mathbb{Z}^+, \forall m \in \mathbb{Z}^+, m < n.$

This is the statement that the set $\{ n \in \mathbb{Z}^+ \mid \forall m \in \mathbb{Z}^+, m < n \}$ is non-empty. For a positive integer n to be in this set we must have $\forall m \in \mathbb{Z}^+, m < n$, in other words it must be greater than all positive integers. But it certainly isn't greater than n itself and so we can disprove the statement $\forall m \in \mathbb{Z}^+, m < n$ by the counterexample $m = n$. Writing this out formally gives the following.

Proof This result is false because, for each positive integer n, if we put $m = n$ then m is a positive integer and $m \not< n$ so that $m = n$ provides a counterexample to the statement $\forall m \in \mathbb{Z}^+, m < n$ which is therefore false. □

Alternatively, the proof might be written more briefly as follows leaving the reader to sort out the quantifiers.

Proof This result is false because, for each positive integer n, if we put $m = n$ then m is a positive integer and $m \not< n$. □

Example 7.6.3 $\forall n \in \mathbb{Z}^+, \exists m \in \mathbb{Z}^+, m < n.$

Let us consider the set $\{ n \in \mathbb{Z}^+ \mid \exists m \in \mathbb{Z}^+, m < n \}$. This is the set of positive integers which are strictly greater than some other positive integer. If $n > 1$ then n does lie in this set since we can take $m = n - 1$ but on the other hand 1 is not an element of this set since $\forall m \in \mathbb{Z}^+, m \geqslant 1$. Hence $\{ n \in \mathbb{Z}^+ \mid \exists m \in \mathbb{Z}^+, m < n \} = \mathbb{Z}^+ - \{1\}$. Since 1 does not lie in this set the universal statement is false. All we need say is the following.

Proof This statement is false and a counterexample is $n = 1$ since $m \not< 1$ for all positive integers m. □

Example 7.6.4 $\exists m \in \mathbb{Z}^+, \forall n \in \mathbb{Z}^+, m < n.$

This is the statement that the set $\{ m \in \mathbb{Z}^+ \mid \forall n \in \mathbb{Z}^+, m < n \}$ is non-empty. For a positive integer m to be in this set we must have $\forall n \in \mathbb{Z}^+, m < n$, in other words it must be smaller than all positive integers. But it certainly isn't smaller than itself and so we can disprove the statement $\forall n \in \mathbb{Z}^+, m < n$ by the counterexample $n = m$. Writing this out formally gives the following.

Proof This result is false because, for each positive integer m, if we put $n = m$ then n is a positive integer and $m \not< n$ so that $n = m$ provides a counterexample to the statement $\forall n \in \mathbb{Z}^+$, $m < n$ which is therefore false. □

The alternative more usual briefer form is as follows.

Proof This result is false because, for each positive integer m, if we put $n = m$ then n is a positive integer and $m \not< n$. □

The reader should carefully compare the proofs in Example 7.6.2 and Example 7.6.4. Although in both cases the implication used is that if $m = n$ then $m \not< n$, in the first example we start from a general positive integer n and then define m by $m = n$, whereas in the second example we start from a general positive integer m and then define n by $n = m$. The distinction may be clarified when the reader does Exercise 7.2.

7.7 The Cartesian product of two sets

At the beginning of this chapter we introduced quantifiers in terms of properties of the subset defined by a predicate involving one free variable. Predicates involving more than one free variable also define subsets – of a set known as the Cartesian product. For simplicity we restrict attention to the case of two free variables.

Definition 7.7.1 *Given sets X and Y, the* Cartesian product *of X and Y, denoted by* $X \times Y$, *is the set of all* ordered pairs (x, y) *where $x \in X$ and $y \in Y$. Thus*

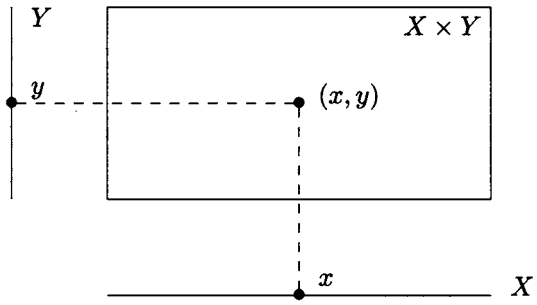
$$X \times Y = \{\, (x, y) \mid x \in X \text{ and } y \in Y \,\}.$$

In referring to an ordered pair *in this definition we mean that two such pairs, (x_1, y_1) and (x_2, y_2), are equal, $(x_1, y_1) = (x_2, y_2)$, if and only if $x_1 = x_2$ and $y_1 = y_2$. We say that the ordered pair (x, y) has* coordinates *x and y.*

When $Y = X$ we write $X \times X = X^2$.

We can picture points in the Cartesian product as follows using lines to represent the sets X and Y and a rectangle to represent $X \times Y$. It is

customary to use a horizontal line to denote the first set and a vertical line to denote the second set.



Examples 7.7.2 (a) For $X = \{a, b, c\}$ and $Y = \{a, b\}$,

$$X \times Y = \{\,(a,a),(a,b),(b,a),(b,b),(c,a),(c,b)\,\}$$

and

$$Y \times X = \{\,(a,a),(a,b),(a,c),(b,a),(b,b),(b,c)\,\}.$$

Notice that these two sets are different.

(b) $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ is the familiar 2-dimensional Euclidean plane.

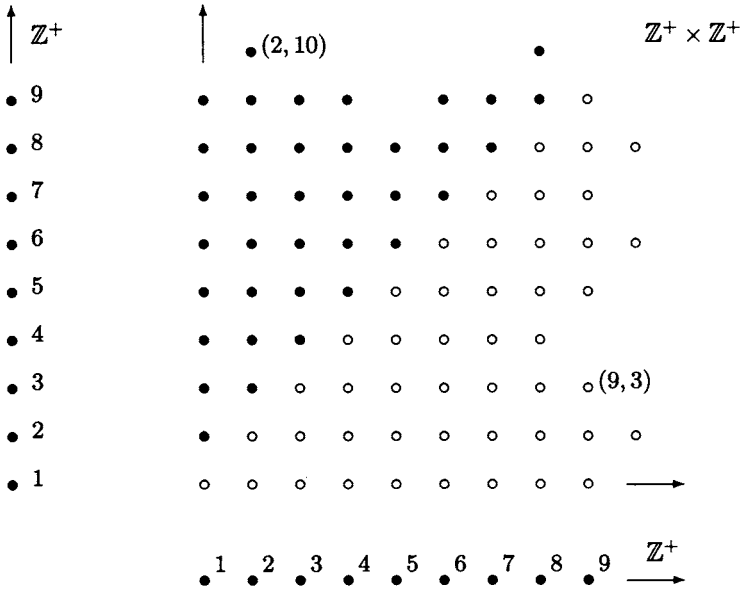
Now given a predicate $P(a, b)$ involving free variables $a \in A$ and $b \in B$, we can define the subset $\{\,(a,b) \in A \times B \mid P(a,b)\,\}$ of the Cartesian product. This is familiar in $\mathbb{R}^2$ from the study of plane curves using the methods of Cartesian geometry; we can describe certain subsets of $\mathbb{R}^2$ by giving an *equation*. For example we say that the circle with centre $(0, 0)$ and radius 1 has equation $x^2 + y^2 = 1$ to mean that the set of points in $\mathbb{R}^2$ which lie on this circle is given by $\{\,(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\,\}$.

Example 7.7.3 To illustrate the relationship between universal and existential statements involving a predicate with two free variables and the subset defined by the predicate we look again at the examples considered in the previous section. We can draw part of the picture of the subset

$$\{\,(m,n) \in \mathbb{Z}^+ \times \mathbb{Z}^+ \mid m < n\,\} \subseteq \mathbb{Z}^+ \times \mathbb{Z}^+$$

as in the diagram on the next page.

Here the solid dots ($\bullet$) indicate the elements of the Cartesian product for which the predicate $m < n$ is true and the circles ($\circ$) those for which



it is false. Thus $(2, 10)$ is in the subset determined by the predicate whereas $(9, 3)$ is not.

We now consider each example in turn.

Example 7.6.1 $\forall m \in \mathbb{Z}^+, \exists n \in \mathbb{Z}^+, m < n$. This is true because each vertical line of elements in the Cartesian product contains an element in the subset, for example $\{m\} \times \mathbb{Z}^+$, the vertical line of points whose first coordinate is m, contains the point $(m, m + 1)$.

Example 7.6.2 $\exists n \in \mathbb{Z}^+, \forall m \in \mathbb{Z}^+, m < n$. This is false because no horizontal line lies entirely in the subset, for example $\mathbb{Z}^+ \times \{n\}$, the horizontal line of points whose second coordinate is n, contains the point (n, n) which does not lie in the subset.

Example 7.6.3 $\forall n \in \mathbb{Z}^+, \exists m \in \mathbb{Z}^+, m < n$. This is false because there is a horizontal line containing no elements of the subset, namely the line $\mathbb{Z}^+ \times \{1\}$.

Example 7.6.4 $\exists m \in \mathbb{Z}^+, \forall n \in \mathbb{Z}^+, m < n$. This is false because each vertical line contains a point not in the subset, for example $\{m\} \times \mathbb{Z}^+$ contains the point (m, m).

To conclude this chapter here are some standard results relating the Cartesian product to union and intersection.

Proposition 7.7.4 *For all sets A, B, C and D the following hold:*

(i) $A \times (B \cup C) = (A \times B) \cup (A \times C)$;
(ii) $A \times (B \cap C) = (A \times B) \cap (A \times C)$;
(iii) $(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)$;
(iv) $(A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D)$.

These statements may be proved directly from the definitions. As an illustration here is a proof of part (ii). (Part (iv) appears as Exercise 7.7 and part (iii) appears in Problems II, Question 13.)

Proof of part (ii) A proof that two sets are equal requires us to prove two set inclusions. In this case we can do them together as follows.

$$\begin{aligned}(x,y) \in A \times (B \cap C) &\Leftrightarrow x \in A \text{ and } y \in B \cap C \\ &\Leftrightarrow x \in A \text{ and } y \in B \text{ and } y \in C \\ &\Leftrightarrow (x,y) \in A \times B \text{ and } (x,y) \in A \times C \\ &\Leftrightarrow (x,y) \in (A \times B) \cap (A \times C).\end{aligned}$$

Thus $(x,y) \in A \times (B \cap C) \Leftrightarrow (x,y) \in (A \times B) \cap (A \times C)$ as required. □

Exercises

7.1 Determine the following sets:

(i) $\{m \in \mathbb{Z}^+ \mid \exists n \in \mathbb{Z}^+,\ m \leqslant n\}$,
(ii) $\{m \in \mathbb{Z}^+ \mid \forall n \in \mathbb{Z}^+,\ m \leqslant n\}$,
(iii) $\{n \in \mathbb{Z}^+ \mid \exists m \in \mathbb{Z}^+,\ m \leqslant n\}$,
(iv) $\{n \in \mathbb{Z}^+ \mid \forall m \in \mathbb{Z}^+,\ m \leqslant n\}$.

7.2 Prove or disprove the following statements.

(i) $\forall m, n \in \mathbb{Z}^+,\ m \leqslant n$.
(ii) $\exists m, n \in \mathbb{Z}^+,\ m \leqslant n$.
(iii) $\forall m \in \mathbb{Z}^+, \exists n \in \mathbb{Z}^+,\ m \leqslant n$.
(iv) $\exists m \in \mathbb{Z}^+, \forall n \in \mathbb{Z}^+,\ m \leqslant n$.
(v) $\forall n \in \mathbb{Z}^+, \exists m \in \mathbb{Z}^+,\ m \leqslant n$.
(vi) $\exists n \in \mathbb{Z}^+, \forall m \in \mathbb{Z}^+,\ m \leqslant n$.

7.3 Prove or disprove the following statements.

(i) $\forall m, n \in \mathbb{Z},\ m \leqslant n.$
(ii) $\exists m, n \in \mathbb{Z},\ m \leqslant n.$
(iii) $\forall m \in \mathbb{Z},\ \exists n \in \mathbb{Z},\ m \leqslant n.$
(iv) $\exists m \in \mathbb{Z},\ \forall n \in \mathbb{Z},\ m \leqslant n.$
(v) $\forall n \in \mathbb{Z},\ \exists m \in \mathbb{Z},\ m \leqslant n.$
(vi) $\exists n \in \mathbb{Z},\ \forall m \in \mathbb{Z},\ m \leqslant n.$

7.4 Prove or disprove each of the following statements.

(i) $\forall x \in \mathbb{R},\ \exists y \in \mathbb{R},\ x + y = 0.$
(ii) $\exists y \in \mathbb{R},\ \forall x \in \mathbb{R},\ x + y = 0.$
(iii) $\forall x \in \mathbb{R},\ \exists y \in \mathbb{R},\ xy = 0.$
(iv) $\exists y \in \mathbb{R},\ \forall x \in \mathbb{R},\ xy = 0.$
(v) $\forall x \in \mathbb{R},\ \exists y \in \mathbb{R},\ xy = 1.$
(vi) $\exists y \in \mathbb{R},\ \forall x \in \mathbb{R},\ xy = 1.$
(vii) $\forall n \in \mathbb{Z}^+$, ($n$ is even or n is odd).
(viii) ($\forall n \in \mathbb{Z}^+$, n is even) or ($\forall n \in \mathbb{Z}^+$, n is odd).

7.5 Prove the following:

$$(\exists q \in \mathbb{Z},\ n = 2q + 1) \Rightarrow (\exists p \in \mathbb{Z},\ n^2 = 2p + 1).$$

7.6 Write the following universal statement in terms of quantifiers and prove it.

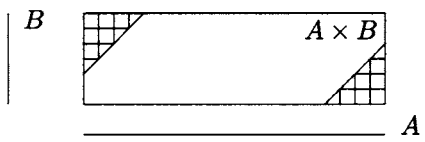
For integers a and b, if a and b are even then so is $a + b$.

7.7 For sets A, B, C and D prove that

$$(A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D),$$

and give an example to show that these sets are not always equal.

7.8 Suppose that the set $\{\,(a, b) \mid P(a, b)\,\}$ is given by the triangular regions in the diagram below.



Decide whether each of the following statements is true or false.

(i) $\forall a \in A,\ \exists b \in B,\ P(a,b)$.
(ii) $\exists b \in B,\ \forall a \in A,\ P(a,b)$.
(iii) $\forall b \in B,\ \exists a \in A,\ P(a,b)$.
(iv) $\exists a \in A,\ \forall b \in B,\ P(a,b)$.

7.9 Find a reformulation of Axiom 5.4.1 (the strong induction principle) as a method of proving that subsets of $\mathbb{Z}^+$ are the whole set (similar to Axiom 7.5.1).

8
Functions

The notion of *function* is one of the most fundamental in mathematics. It is probably familiar to the reader in the context of calculus, and it is here that the concept was first clarified and the difference between a function and a formula properly understood in the early years of the nineteenth century. Today, the language of functions is used throughout mathematics.

In this chapter the function concept is introduced. We discuss various ways of defining a function and consider the graph of a function.

8.1 Functions and formulae

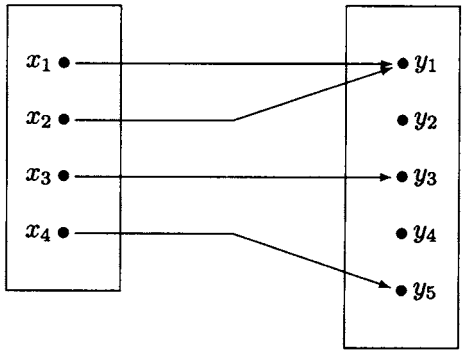
Definition 8.1.1 *Suppose that X and Y are sets. A* function, map *or* mapping *from the set X to the set Y is the assignment of a unique element of Y to each element of X. If f is a function from X to Y we write $f: X \to Y$ and denote the element of Y assigned to an element $x \in X$ by $f(x)$ writing $x \mapsto f(x)$. The element $f(x) \in Y$ is called the* value *of f at $x \in X$ or the* image *of x under f. The set X is called the* domain *of the function f and the set Y is called the* codomain.

Example 8.1.2 The most basic way of describing a function is by listing the values. For example, given $X = \{x_1, x_2, x_3, x_4\}$ and $Y = \{y_1, y_2, y_3, y_4, y_5\}$, the following table determines a function $f: X \to Y$.

x	x_1	x_2	x_3	x_4
$f(x)$	y_1	y_1	y_3	y_5

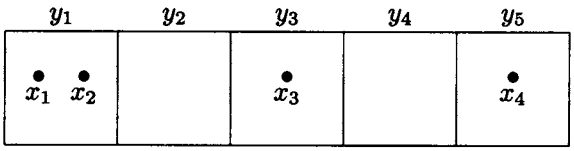
We determine the value of the function at an element of X by locating that element in the first row and reading down to the second row. Notice that each element of the domain X occurs precisely once in the first row so that this is a well-defined procedure. On the other hand elements of the codomain Y may occur once in the second row, but they may also occur more than once or not at all: an element of the codomain may be the value of the function at several elements of the domain or may not be a value at all.

We can picture this function as follows.



Here we find the value of f at an element of the domain X by following the arrow which begins at that element. Each element of the domain lies at the beginning of just one arrow so this is a well-defined procedure. However, elements in the codomain may lie at the end of one arrow, several arrows or no arrow at all.

Yet another way of describing a function is as follows. Think of the elements of the codomain set as boxes. The function describes how to place the objects of the domain set in these boxes. Thus the function under consideration would be pictured as follows.



Again notice that each element of the domain set occurs precisely once

in the picture. However, any given box, corresponding to an element of the codomain, may contain one object, several objects or none at all.

The reader may wonder at this variety of ways of thinking about a function. It is important to realize that mathematicians develop a variety of ways of thinking about the rather abstract concepts they deal with. Any given way may suit one person better than another and each leads to different insights: it can be useful to be aware of several.

Example 8.1.3 Suppose that $X = \{a, b, c\}$ and $Y = \{d, e\}$. Then there are precisely eight functions from the set X to the set Y given as follows.

x	$f_1(x)$	$f_2(x)$	$f_3(x)$	$f_4(x)$	$f_5(x)$	$f_6(x)$	$f_7(x)$	$f_8(x)$
a	d	d	d	d	e	e	e	e
b	d	d	e	e	d	d	e	e
c	d	e	d	e	d	e	d	e

Here the functions f_1 and f_8 are examples of *constant functions.* Given sets X and Y and any element $y_0 \in Y$ there is a *constant function* $c_{y_0}: X \to Y$ given by $c_{y_0}(x) = y_0$ for all $x \in X$.

Of course a function can have the same set as domain and codomain as in the next example.

Example 8.1.4 Suppose that $Z = \{a, b\}$. Then there are precisely four functions from Z to Z as follows.

x	$g_1(x)$	$g_2(x)$	$g_3(x)$	$g_4(x)$
a	a	a	b	b
b	a	b	a	b

In this case the functions g_1 and g_4 are constant functions. The function g_2 is an example of an *identity function.* Given a set X, the *identity function* on X, $I_X: X \to X$, is given by $I_X(x) = x$ for all $x \in X$.

Describing a function by listing the values is only practical when the domain set is small and impossible if the domain set is infinite. The most common way of describing a function f is by giving a *formula* for f which provides a procedure for finding the value of the function at each element of the domain. When defining a function using a formula it is important to be clear about which sets are the domain and the codomain of the function.

Example 8.1.5 Recall that $\mathbb{R}^{\geqslant}$ denotes the subset of $\mathbb{R}$, the set of real numbers, given by $\{ x \in \mathbb{R} \mid x \geqslant 0 \}$. Consider the following four functions:

(i) $f_1 : \mathbb{R} \to \mathbb{R}$ given by $f_1(x) = x^2$;
(ii) $f_2 : \mathbb{R}^{\geqslant} \to \mathbb{R}$ given by $f_2(x) = x^2$;
(iii) $f_3 : \mathbb{R} \to \mathbb{R}^{\geqslant}$ given by $f_3(x) = x^2$;
(iv) $f_4 : \mathbb{R}^{\geqslant} \to \mathbb{R}^{\geqslant}$ given by $f_4(x) = x^2$.

Notice that, although the *formula* $f_i(x) = x^2$ is the same for each of these four functions, they are considered to be four distinct functions since the domain and codomain are part of the definition of the function. We will see later that these four functions have different properties even though they are given by the same formula.

We need to take care that the formula makes sense for each element of the domain.

Example 8.1.6 The formula $f_1(x) = \dfrac{x^2 + x - 2}{x - 1}$ does not define a function $\mathbb{R} \to \mathbb{R}$ since it gives no value for $x = 1$. When working with the set of real numbers it is quite common not to specify the domain and codomain of a function given by a formula. The convention is that, if these are not specified, then we take as the domain the subset of $\mathbb{R}$ consisting of the numbers for which the formula makes sense and we take $\mathbb{R}$ as the codomain. Using this convention the above formula for f_1 defines a function $f_1 : \mathbb{R} - \{1\} \to \mathbb{R}$.

If we really want a function with domain $\mathbb{R}$ then in an example like this there are two basic techniques for extending f_1.

Rewriting the formula: We can rewrite the formula in such a way that it makes sense for all real numbers x using

$$(x^2 + x - 2)/(x - 1) = (x - 1)(x + 2)/(x - 1) = x + 2$$

for $x \neq 1$. Then $f_2(x) = x + 2$ does define a function on $\mathbb{R}$, i.e. with domain $\mathbb{R}$, extending the function f_1.

Explicit definition: Alternatively we can explicitly specify the value of the function at the element 1 where the formula for f_1 does not work. Thus

$$f_3(x) = \begin{cases} \dfrac{x^2 + x - 2}{x - 1} & \text{if } x \neq 1, \\ 3 & \text{if } x = 1, \end{cases}$$

defines a function $f_3 : \mathbb{R} \to \mathbb{R}$.

Notice that we may specify the values at individual points any way we like. It doesn't have to relate in a sensible way to the values elsewhere. In defining $f_3(1)$ we selected that value given by the formula $x + 2$ so that f_2 and f_3 are in fact the same function. However, we could extend the function f_1 to the whole of $\mathbb{R}$ in a different way. For example

$$f_4(x) = \begin{cases} \dfrac{x^2 + x - 2}{x - 1} & \text{if } x \neq 1, \\ 42 & \text{if } x = 1, \end{cases}$$

is another way of extending the function f_1 to the whole of $\mathbb{R}$.

Example 8.1.7 In some cases, for example $x \mapsto 1/x$, there appears to be no way of rewriting the formula to include in the domain of definition points where it gives no value (in this case $x = 0$). But in these cases we can still use the second method. For example

$$g(x) = \begin{cases} 1/x & \text{if } x \neq 0, \\ 73 & \text{if } x = 0, \end{cases}$$

is a perfectly respectable function $g : \mathbb{R} \to \mathbb{R}$.

We can use different formulae for different parts of the domain as in the following example.

Example 8.1.8 The *modulus function* $x \mapsto |x|$ is defined on $\mathbb{R}$ by

$$|x| = \begin{cases} x & \text{if } x \geqslant 0, \\ -x & \text{if } x \leqslant 0. \end{cases}$$

Notice here that the value of the function at 0 is given twice, since $0 \geqslant 0$ and $0 \leqslant 0$. This isn't a problem since the two formulae ($x \mapsto x$ and $x \mapsto -x$) give the same value (namely 0) for $x = 0$. But when specifying a function in this way we do need to be careful to check that it is *well-defined* (i.e. there is a uniquely specified value) at each point.

Definition 8.1.9 *Two functions $f : X \to Y$ and $g : X \to Y$ are* equal*, written* $f = g$*, when they have the same value at each point of the domain X, i.e. $f(x) = g(x)$ for all $x \in X$. Notice that it is implicit in this definition that two equal functions have the same domain and the same codomain.*

Example 8.1.10 The functions f_2 and f_3 of Example 8.1.6 are equal even though they are defined in different ways. Thus it is not the process leading to the values which determines the function but what the values are. Of course, in defining f_3 the value of $f_3(1)$ was selected precisely so that they would be equal.

We have seen in Example 8.1.6 how to extend the domain of a function. We can also make the domain smaller.

Definition 8.1.11 *Suppose that $f: X \to Y$ is a function and A is a subset of X, i.e. $A \subseteq X$. Then we can define a function $g: A \to Y$ by $g(a) = f(a)$ for all $a \in A$. This function is called the* restriction *of f to A and is denoted by $f|A$.*

Examples 8.1.12 (a) In Example 8.1.3 $f_1|\{a,b\} = f_2|\{a,b\}$ and is the constant function taking the value d.

(b) In Example 8.1.5 $f_2 = f_1|\mathbb{R}^{\geqslant}$ and $f_4 = f_3|\mathbb{R}^{\geqslant}$.

(c) In Example 8.1.6 $f_2|(\mathbb{R} - \{1\}) = f_4|(\mathbb{R} - \{1\}) = f_1$.

8.2 Composition of functions

Suppose that $f: X \to Y$ and $g: Y \to Z$. Then, given an element $x \in X$, the function f assigns to it an element $y = f(x) \in Y$ and now the function g assigns to this an element $g(y) = g(f(x)) \in Z$. Thus using f and g an element of Z has been assigned to x. This process defines a function with domain X and codomain Z called the composite of f and g.

Definition 8.2.1 *Given two functions $f: X \to Y$ and $g: Y \to Z$ the* composite *of f and g, denoted by $g \circ f: X \to Z$ or simply $gf: X \to Z$, is defined by*

$$g \circ f(x) = g(f(x)) \quad \text{for all } x \in X.$$

The order of the f and g in this definition should be carefully noted. Since we have adopted the convention (usual† in most areas of mathematics) of writing the symbol for the function on the left of the element

† But not used universally. In some branches of algebra it is quite common to write the value $f(x)$ as xf or even x^f.

where it is being evaluated (viz. $f(x)$) it is natural to denote the function obtained by applying f and then applying g by the symbol $g \circ f$ which presents them (from left to right) in the opposite order to their application. We can write the following to indicate this:

$$g \circ f: X \xrightarrow{f} Y \xrightarrow{g} Z.$$

Examples 8.2.2 (a) The function $x \mapsto (x+1)^2$ from $\mathbb{R}$ to $\mathbb{R}$ is the composite of the function $f: \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x + 1$ and the function $g: \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x^2$ because

$$g \circ f(x) = g(x+1) = (x+1)^2 \quad \text{for all } x \in X.$$

Notice that the order does matter here because

$$f \circ g(x) = f(x^2) = x^2 + 1;$$

composition of functions is not in general *commutative*. It is not usually a good idea to refer simply to 'the composite of f and g' when both orders are possible since there is no well-established convention about which order these words indicate. It is best to use the symbolism $g \circ f$ or $f \circ g$ in order to be quite clear.

(b) Suppose that A is a subset of a set X. Then we may define a function $i: A \to X$ by $i(a) = a$ for all $a \in A$. This function is called the *inclusion function* of A into X. If now $f: X \to Y$ is a function, then the composite $f \circ i: A \to Y$ is equal to the restriction $f|A: A \to Y$ of f to A.

(c) Suppose that $f: \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = x^2+1$ and $g: \mathbb{R}-\{0\} \to \mathbb{R}$ is defined by $g(x) = 1/x$. Then strictly speaking the composite $g \circ f$ is not defined since the codomain of f is not the same as the domain of g.

However, every value of f lies in the domain of g since $x^2 + 1 \geqslant 1 > 0$ for all real numbers x so that g can be evaluated at each value of f. In a case like this we can still attach a meaning to the composite $g \circ f$ as the function given by $g \circ f(x) = g(f(x))$. This gives the function $\mathbb{R} \to \mathbb{R}$ determined by $g \circ f(x) = 1/(x^2 + 1)$.

Proposition 8.2.3 *Suppose that $f: X \to Y$, $g: Y \to Z$ and $h: Z \to W$ are functions. Then*

(i) $(h \circ g) \circ f = h \circ (g \circ f): X \to Z$,

(ii) $f \circ I_X = f = I_Y \circ f: X \to Y$.

Proof These results are proved simply by evaluating the functions. For example, given $x \in X$, both the functions in (i) assign $h(g(f(x)))$ to this element and so the functions are equal. □

Notice that the first part of Proposition 8.2.3 means that we can write $h \circ g \circ f$ without ambiguity. Composition of functions is *associative.*

8.3 Sequences

Definition 8.3.1 *A function $f : \mathbb{Z}^+ \to A$ is called a* sequence *in the set A.*

The study of sequences in $\mathbb{R}$ or $\mathbb{C}$ is of great importance in advanced calculus and numerical analysis. Sequences may be defined by a formula, such as $n \mapsto n^2$, $n \mapsto 1/n$, $n \mapsto 2^n$, $n \mapsto (1 + 1/n)^n$ and $n \mapsto 1/n!$. But more interesting are those sequences defined inductively. We saw in Chapter 5 that strictly speaking formulae involving the exponent n or the factorial $n!$ really have to be defined inductively. Another example of an inductively defined sequence considered there was the Fibonacci sequence.

It is common in advanced calculus to solve numerical problems by generating a sequence of numbers x_n which provide increasingly good approximations to the solution as n increases – the sequence *converges* to a *limit* which is the required solution.†

One important reason for becoming adept at handling quantifiers is to be able to handle the definition of the limit of a sequence. To illustrate the ideas we introduce the idea of a null sequence, a sequence with limit 0.

Definition 8.3.2 *Given a sequence $f: \mathbb{Z}^+ \to \mathbb{R}$ of real numbers, we say that the sequence is* null*, written* $\lim f = 0$ *or* $\lim_{n\to\infty} f(n) = 0$*, when*

$$\forall \varepsilon \in \mathbb{R}^+, \exists N \in \mathbb{Z}^+, \forall n \in \mathbb{Z}^+ \, (n \geqslant N \Rightarrow |f(n)| < \varepsilon).$$

† A good example is the Newton–Raphson method for solving an equation, described in most books on advanced calculus (see for example G.B. Thomas and R.S. Finney, *Calculus and analytic geometry*, Addison-Wesley, Eighth edition 1992).

The use of ε for a general positive real number in this definition is traditional. This is not the place to investigate the full implications of this definition which involves *three* quantifiers! Here is a simple example of its use.

Example 8.3.3 The sequence $n \mapsto 1/\sqrt{n}$ is null.

Constructing a proof. We are required to prove that

$$\forall \varepsilon \in \mathbb{R}^+, \exists N \in \mathbb{Z}^+, \forall n \in \mathbb{Z}^+ \, (n \geqslant N \Rightarrow 1/\sqrt{n} < \varepsilon).$$

[Notice that $|1/\sqrt{n}| = 1/\sqrt{n}$.] So suppose we are given $\varepsilon \in \mathbb{R}^+$. Then to demonstrate the existence of the integer N we examine when $1/\sqrt{n} < \varepsilon$.

$$1/\sqrt{n} < \varepsilon \Leftrightarrow 1/n < \varepsilon^2 \Leftrightarrow n > 1/\varepsilon^2.$$

Hence so long as we choose a positive integer N such that $N > 1/\varepsilon^2$ we will have $n \geqslant N \Rightarrow n > 1/\varepsilon^2 \Rightarrow 1/\sqrt{n} < \varepsilon$ as required.

So for example, if $\varepsilon = 1$ then we must choose $N > 1$, if $\varepsilon = 1/2$ then we must choose $N > 4$, and if $\varepsilon = 1/100$ then $N > 10000$. The smaller the number ε the greater the number N is required to be. But whatever the positive real number ε is the condition $N > 1/\varepsilon^2$ tells us how large N has to be.

Proof Given a positive real number ε, $1/\sqrt{n} < \varepsilon$ if and only if $n > 1/\varepsilon^2$. Hence if we choose $N > 1/\varepsilon^2$ then $n \geqslant N$ implies that $n > 1/\varepsilon^2$ so that $1/\sqrt{n} < \varepsilon$ is required. □

This type of proof involving multiple quantifiers is quite elaborate and this is in fact an extremely simple example. More complicated examples are not considered in this book but are dealt with in detail in books on infinite sequences and series, advanced calculus and analysis.†

8.4 The image of a function

Given a function $f: X \to Y$, it is not necessary that every element of Y is a value of the function. For example the function $\mathbb{R} \to \mathbb{R}$ given by $x \mapsto x^2$ does not have -1 as a value. Thus we can obtain a subset of Y by considering those elements which are values.

† See, for example, R. Haggerty, *Fundamentals of mathematical analysis*, Addison-Wesley, Second edition 1993.

Definition 8.4.1 *Given a function $f: X \to Y$, the subset of the codomain Y consisting of those elements which are values of f is called the* image *of f and is denoted by* $\mathrm{Im}f$. *Thus*

$$\mathrm{Im}f = \{ f(x) \mid x \in X \}.$$

This means that the function f provides a constructive definition of the set $\mathrm{Im}f$, as discussed in Chapter 6.

8.5 The graph of a function

Definition 8.5.1 *Suppose that $f: X \to Y$ is a function. Then we define the* graph *of f to be the subset of the Cartesian product $X \times Y$ given by*

$$G_f = \{ (x, y) \in X \times Y \mid y = f(x) \} = \{ \big(x, f(x)\big) \mid x \in X \}.$$

Remarks 8.5.2 In the case that X and Y are subsets of $\mathbb{R}$ this is the usual idea of the graph of a function. Work in calculus makes it clear that the graph gives a great deal of information about the function and that it is useful to develop skill in drawing graphs.

Example 8.5.3 The graphs of the functions f_1 and f_2 of Example 8.1.3 are as follows.

Here we indicate the elements of the graph by solid dots (•).

The graph of the function $f: X \to Y$ is a particular example of a subset of the Cartesian product $X \times Y$ determined by a predicate: the predicate is $y = f(x)$. Not every subset of $X \times Y$ arises as the graph of a function. Each column $\{x_0\} \times Y$ must contain a single element, namely

$(x_0, f(x_0))$. However, the function f is determined by its graph G_f, for, given $x_0 \in X$, there is a unique element $(x, y) = (x_0, y_0) \in G_f$ such that $x = x_0$, and then $f(x_0) = y_0$.†

Exercises

8.1 Define functions f and $g: \mathbb{R}^2 \to \mathbb{R}$ by:

$$f(x,y) = \frac{x+y}{2} + \frac{|x-y|}{2};$$
$$g(x,y) = \begin{cases} x & \text{if } x \geqslant y, \\ y & \text{if } x \leqslant y. \end{cases}$$

Prove that the function g is well-defined. Prove that $f = g$.

8.2 Define functions f and $g: \mathbb{R} \to \mathbb{R}$ by $f(x) = x^3$ and $g(x) = 1 - x$. Find the functions (i) $f \circ f$, (ii) $f \circ g$, (iii) $g \circ f$, (iv) $g \circ g$.

List the elements of the set $\{ x \in \mathbb{R} \mid fg(x) = gf(x) \}$.

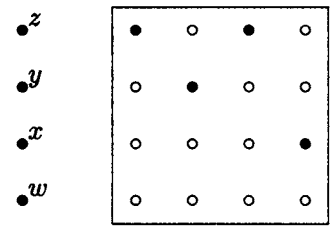
8.3 Find functions $f_i: \mathbb{R} \to \mathbb{R}$ with images as follows:

(i) $\mathrm{Im} f_1 = \mathbb{R}$;
(ii) $\mathrm{Im} f_2 = \mathbb{R}^+$;
(iii) $\mathrm{Im} f_3 = \mathbb{R} - \mathbb{Z}$;
(iv) $\mathrm{Im} f_4 = \mathbb{Z}$.

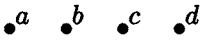
8.4 Prove that the sequence $n \mapsto 1/n$ is null.

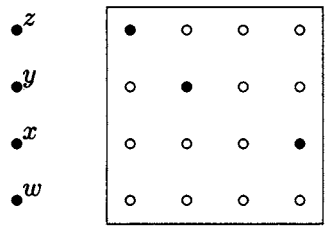
8.5 Let $X = \{ a, b, c, d \}$ and $Y = \{ w, x, y, z \}$. Which of the following subsets of $X \times Y$ is the graph of a function $f: X \to Y$? For those which are write down a table for the corresponding function as in Example 8.1.3. Explain why the others are not graphs of functions.

† Our definition of a function (Definition 8.1.1) is usually considered rather informal in more advanced mathematics. There a function is defined to *be* a graph: in other words a function $X \to Y$ is defined to be a subset of $X \times Y$ in which each element of X occurs as the first coordinate of an element precisely once (see for example Daniel J. Velleman, *How to prove it, a structured approach,* Cambridge University Press, 1994).

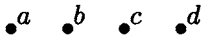


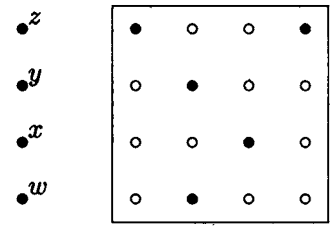
(i)



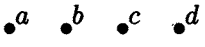


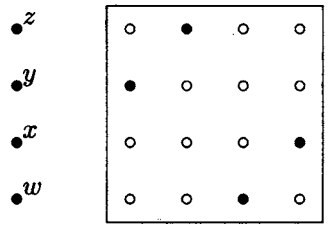
(ii)



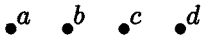


(iii)





(iv)



9

Injections, surjections and bijections

In defining a function $f: X \to Y$ we insist that a unique element of Y is assigned to each element of X. However, we do not require that each element of Y is assigned to some element of X nor do we prevent the possibility of the same element of Y being assigned to several (or even all the) elements of X. By imposing additional conditions concerning the number of elements of X to which elements of Y are assigned we get functions with particular properties. In this chapter we consider functions with particularly good properties and in particular functions which are *bijections* for which we can define an inverse function.

9.1 Properties of functions

Definition 9.1.1 *Suppose that $f: X \to Y$ is a function.*

(i) *If no element of Y is assigned to more than one element of X, i.e. the function takes a different value for each point of the domain, then we say that the function f is an* injection *(or that it is* injective *or* one-to-one*). In symbols we can write this*

$$\forall x_1, x_2 \in X, (x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2))$$

or equivalently using the contrapositive

$$\forall x_1, x_2 \in X, (f(x_1) = f(x_2) \Rightarrow x_1 = x_2).$$

(ii) *If each element of Y is assigned to some element of X, i.e. each point of the codomain is a value of the function, then we say that the*

function f is a surjection *(or that it is* surjective *or* onto*). In symbols we can write this*

$$\forall y \in Y, \exists x \in X, y = f(x).$$

(iii) *If f is both an injection and a surjection then we say that it is a* bijection *(or that it is* bijective *or* one-to-one and onto*).*

We can reformulate these definitions using the notion of a *pre-image.*

Definition 9.1.2 *For a function $f: X \to Y$, given an element $y \in Y$ a* pre-image *of y (under f) is an element $x \in X$ such that $y = f(x)$.*

So f is an injection if and only if every element of Y has at most one pre-image; f is a surjection if and only if every element of Y has at least one pre-image; and f is a bijection if and only if every element of Y has precisely one pre-image.

These ideas are simply expressed in terms of the various models for a function introduced in the previous chapter. If a function is described by listing the values of the function as in Example 8.1.2 then we find the pre-images of an element of the codomain by locating its occurrences in the list of values and reading up to the corresponding elements of the first row. Thus a function is injective when no element occurs more than once in the list of values, it is surjective when every element of the codomain does occur in the list of values and bijective when every element of the codomain occurs precisely once in the list of values.

In terms of the picture representing the function by arrows from the domain to the codomain (see page 90) we find pre-images by following arrows backwards. Thus, for example, a function is bijective if every point of the codomain is at the end of precisely one arrow.

Finally, if we think of a function as a procedure for placing the elements of the set X into a set of boxes Y (see page 90) then a pre-image of a particular box $y \in Y$ is simply an element which is placed in that box. From this point of view, f is injective when no two elements are placed in the same box, it is surjective when no box is left empty, and it is bijective when each box contains precisely one element.

Examples 9.1.3 (a) In Example 8.1.3 all the functions apart from the constant functions f_1 and f_8 are surjections. None of the functions are

injections.

(b) In Example 8.1.4 the non-constant functions f_2 and f_3 are bijections. Notice that, for any set X, the identity function I_X is a bijection.

(c) In Example 8.1.5 the function f_1 is neither an injection (since $(-1)^2 = 1^2 = 1$) nor a surjection (since there is no real number x such that $x^2 = -1$); the function f_2 is an injection (since $0 \leqslant x_1 < x_2 \Rightarrow x_1^2 < x_2^2$) but not a surjection; the function f_3 is a surjection but not an injection; and the function f_4 is a bijection. This example illustrates the importance of the domain and the codomain. For example, if a result has been proved for injections then it can be applied to f_2 and to f_4 but not to f_1 and f_3.

Remarks 9.1.4 Notice that we can easily convert any function into a surjection by changing the codomain. Recall that, given a function $f: X \to Y$, the *image* of f, denoted by $\mathrm{Im}f$, is the set of values of f. The assignment determining the function f also determines a function $f^+: X \to \mathrm{Im}f$ (i.e. $f^+(x) = f(x)$) which is a surjection.

Since the definitions of injectivity and surjectivity involve universal and existential statements, proofs and disproofs of these properties usually follow the format discussed in Chapter 7.

Example 9.1.5 To determine whether the function $f_1: \mathbb{R} \to \mathbb{R}$ given by $f_1(x) = x + 1$ is an injection, a surjection or a bijection.

For injectivity, it is usually easiest to use the second formulation:

$$f_1(x_1) = f_1(x_2) \Rightarrow x_1 = x_2.$$

Filling in the definition of the function this gives

$$x_1 + 1 = x_2 + 1 \Rightarrow x_1 = x_2$$

which is certainly true so that f_1 is injective.

For surjectivity, we rewrite the definition as usual as a universal implication:

$$y \in \mathbb{R} \Rightarrow \exists x \in \mathbb{R},\ y = f_1(x).$$

If we put the definition of the function into this then it easy to prove the existential statement by example. For

$$y = f_1(x) \Leftrightarrow y = x + 1 \Leftrightarrow x = y - 1$$

which tells us that, given $y \in \mathbb{R}$, then $y = f_1(x)$ if $x = y - 1$ so that y does have a pre-image under f_1. Hence f_1 is surjective and so a bijection.

In fact, what is written above is rather inefficient; we can often prove bijectivity in one fell swoop. The best way to discover whether a function is injective or surjective can be to investigate the pre-images of a general element in the codomain using the approach used above for surjectivity. In this case if y is a general element of the codomain $\mathbb{R}$ then

$$\begin{aligned} x \text{ is a pre-image of } y &\Leftrightarrow y = f_1(x) \\ &\Leftrightarrow y = x + 1 \\ &\Leftrightarrow x = y - 1. \end{aligned}$$

This shows that each element y of $\mathbb{R}$ has precisely one pre-image, namely $y - 1$, so that the function is a bijection.

We can write out the solution to this problem quite briefly as follows.

Solution Since $y = f_1(x) \Leftrightarrow y = x + 1 \Leftrightarrow x = y - 1$, for $x, y \in \mathbb{R}$, we see that each element y of $\mathbb{R}$ has precisely one pre-image under f_1. Hence f_1 is a bijection. □

Example 9.1.6 To determine whether the function $f_2 : \mathbb{R}^+ \to \mathbb{R}^+$ given by $f_2(x) = x + 1$ is injective, surjective or bijective.

This function is given by the same formula as the previous example. If we try the pre-image approach we get almost the same argument:

$$\begin{aligned} x \text{ is a pre-image of } y &\Leftrightarrow y = f_2(x) \\ &\Leftrightarrow y = x + 1 \\ &\Leftrightarrow x = y - 1. \end{aligned}$$

The difference becomes clear when we ask whether the formula for the pre-image makes sense. Remember that if x is to be a pre-image then it must certainly lie in the domain $\mathbb{R}^+$ (otherwise $f_2(x)$ isn't even defined). But, given $y \in \mathbb{R}^+$, $y - 1 \in \mathbb{R}^+$ if and only if $y - 1 > 0$ (by the definition of $\mathbb{R}^+$), i.e. $y > 1$. So we see that y has a pre-image if and only if $y > 1$. This proves that the image of f_2 is the set $\{ y \in \mathbb{R}^+ \mid y > 1 \}$ and, since this is not the whole of $\mathbb{R}^+$, f_2 is not surjective. As usual, in the formal proof we give an explicit counterexample, an element of $\mathbb{R}^+$ which is not a value of the function.

However, the function is injective and the proof is just the same as in the previous example.

Solution The function is injective since, for $x_1, x_2 \in \mathbb{R}^+$, $f_2(x_1) = f_2(x_2) \Rightarrow x_1 + 1 = x_2 + 1 \Rightarrow x_1 = x_2$. However, if $x \in \mathbb{R}^+$ then $x > 0$ so that $f_2(x) = x + 1 > 1$. Thus $1/2$ is not a value of f_2 which is not then surjective. □

Example 9.1.7 To determine whether the function $f_3: \mathbb{R} \to \mathbb{R}$ given by $f_3(x) = 4x^2 - 4x + 2$ is injective, surjective or bijective.

This time the pre-image approach gives

$$\begin{aligned} y = f_3(x) &\Leftrightarrow y = 4x^2 - 4x + 2 \\ &\Leftrightarrow y = (2x-1)^2 + 1 \\ &\Leftrightarrow x = (1 \pm \sqrt{y-1})/2. \end{aligned}$$

Again we must check whether our formula for the pre-image makes sense. Now if $y < 1$ then $y - 1$ has no (real) square root and so in that case the formula has no meaning. Looking back through the steps above we see what the problem is: $y = (2x-1)^2 + 1$ tells us that if y is a value then $y \geqslant 1$. This tells us that f_3 is not surjective.

The other thing we notice in the formula for the pre-image is the symbol $\pm$. This means that x is a pre-image for y if and only if $x = (1 + \sqrt{y-1})/2$ or $x = (1 - \sqrt{y-1})/2$. In other words if $\sqrt{y-1}$ exists and is non-zero, then y has two pre-images. For example, taking $y = 2$ for which $\sqrt{y-1} = 1$,† we obtain $f_3(x) = 2 \Leftrightarrow x = (1 \pm 1)/2$, i.e. $x = 1$ or $x = 0$. So f_3 is not injective.

Solution Because $f_3(0) = f_3(1) = 2$, the function is not injective.

For a real number x, $f_3(x) = 4x^2 - 4x + 2 = (2x-1)^2 + 1 \geqslant 1$. Thus 0 is not a value of f_3 and so it is not surjective. □

Notice that in presenting this proof no indication is given of how we found the counterexample to injectivity. Formally this is not required for the proof. In a case like this it is not difficult to spot such a counterexample (for example from the graph of the function or from rewriting the formula as $(2x-1)^2 + 1$) but it is often a good idea to help the reader by indicating how you found the counterexample.

† We adopt the usual convention that if y is a non-negative real number then $\sqrt{y}$ represents the non-negative square root of y, i.e. $x = \sqrt{y}$ if and only if $y = x^2$ and $x \geqslant 0$.

Example 9.1.8 To determine whether the function $f_4: \mathbb{R}^+ \to \{x \in \mathbb{R}^+ \mid x \geqslant 1\}$ given by $f_4(x) = 4x^2 - 4x + 2$ is injective, surjective or bijective.

The pre-image approach gives just the same formula as in the previous example. However, if y is in the codomain then $y \geqslant 1$ and so the square root in the formula for a pre-image element does determine a real number. Furthermore $(1 + \sqrt{y-1})/2 > 1/2$ and so this gives one pre-image for y showing that f_4 is surjective. For $y > 1$, we have a second pre-image precisely when $(1 - \sqrt{y-1})/2 > 0$ which occurs when $\sqrt{y-1} < 1$, i.e. $y < 2$. This proves that f_4 is not injective since elements y in the codomain have two pre-images if $y < 2$. However, it is probably more satisfying to give two specific elements of the domain where the function takes the same value; we can get these by taking $y = 5/4$ which has $\sqrt{y-1} = 1/2$ and so the pre-images $(1 \pm 1/2)/2$, i.e. $3/4$ and $1/4$.

Sketching the graph of this function is a great aid to seeing what is going on. The reader should do this.

Solution Notice that $y = f_4(x) \Leftrightarrow x = (1 \pm \sqrt{y-1})/2$. The function is surjective since given $y \geqslant 1$, $y = f_4(x)$ for $x = (1 + \sqrt{y-1})/2 \in \mathbb{R}^+$. The function is not injective since $f_4(1/4) = f_4(3/4) = 5/4$. □

9.2 Bijections and inverses

Definition 9.2.1 *A function $f: X \to Y$ is called* invertible *if*† *there exists a function $g: Y \to X$ such that*

$$y = f(x) \Leftrightarrow x = g(y) \quad \textit{for all } x \in X \textit{ and for all } y \in Y.$$

In this case we call g an inverse (function) *of f and write* $g = f^{-1}$.

The symmetry of the definition shows that in this case g is also invertible and f is an inverse of g.

This means then that corresponding to the assignment of an element of Y to each element of X there is a reverse assignment of an element of X to each element of Y.

† The reader should note the way that the word 'if' is used in this definition. It really means 'if and only if' since we are defining the meaning of the word 'invertible'. Saying that f is invertible means precisely the same as saying that there is a function g with the properties given. This usage in definitions is very common although it has on the whole been avoided so far in this book.

Example 9.2.2 Suppose that $f: \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = 2x + 1$ and $g: \mathbb{R} \to \mathbb{R}$ is defined by $g(x) = (x - 1)/2$. Then, for $x, y \in \mathbb{R}$,

$$y = f(x) \Leftrightarrow y = 2x + 1 \Leftrightarrow x = (y - 1)/2 \Leftrightarrow x = g(y).$$

Thus f and g are invertible and each is the inverse of the other.

Theorem 9.2.3 *Let $f: X \to Y$. Then f is invertible if and only if it is a bijection. Furthermore, if it is invertible, then its inverse function is unique.*

Constructing a proof. Intuitively this result is pretty clear. For example if we think of the model of the function given by placing elements from the set X in the set of boxes Y then we have observed that if the function is bijective there is precisely one element of X in each box. In this case the inverse function is given by associating to each box the element which it contains. Furthermore this procedure is only possible if there is only one element in each box which means that f is bijective.

The formal proof makes these ideas a bit more precise in terms of the formal definitions of bijectivity and inverse and is constructed as so often by simply spelling out these definitions. It may seem quite long, but if you are clear about what you are trying to do there is really only one possible way forward at each stage and so the proof more or less writes itself.

The last part of the theorem asserts that f has a unique inverse. This means that f has an inverse (for this is what is meant by stating that it is invertible) but only one inverse. Uniqueness statements are very common. The usual approach to proving them is illustrated in this case: we demonstrate that if g_1 and g_2 are inverses of f then $g_1 = g_2$. Arguments of this type have already been met in proofs of the injectivity of a function f, for this asserts that each value of the function has a unique pre-image: this is proved by demonstrating that if x_1 and x_2 are pre-images of the same element then $x_1 = x_2$ (see for example Example 9.1.5).

Proof For the first statement two things have to be proved.

(a) 'f invertible $\Rightarrow$ f bijective': Suppose that f is invertible. Then, by definition, there is an inverse function $g: Y \to X$ so that

$$y = f(x) \Leftrightarrow x = g(y) \quad \text{for all } x \in X \text{ and for all } y \in Y.$$

Now to prove that f is bijective we must prove that f is surjective and that it is injective.

For injectivity, suppose that $x_1, x_2 \in X$ are such that $f(x_1) = f(x_2)$; then we must prove that $x_1 = x_2$. Put $y_0 = f(x_1) = f(x_2)$. Then $y_0 = f(x_1) \Rightarrow x_1 = g(y_0)$ (from the definition of the inverse) but on the other hand $y_0 = g(x_2) \Rightarrow x_2 = g(y_0)$. Hence $x_1 = g(y_0) = x_2$ as required.

For surjectivity suppose that $y_0 \in Y$. We must demonstrate that there exists an element $x_0 \in X$ such that $y_0 = f(x_0)$. Put $x_0 = g(y_0)$. Then it is immediate from the definition of an inverse function that $y_0 = f(x_0)$ as required.

(b) 'f bijective $\Rightarrow$ f invertible': Suppose that f is bijective. To show that it is invertible we must construct an inverse function $g: Y \to X$. Suppose that $y_0 \in Y$. Since f is a bijection, y_0 has precisely one pre-image, say $x_0 \in X$, such that $f(x_0) = y_0$. So we can define a function by the rule that, for each $y \in Y$, $g(y)$ is the unique element $x \in X$ such that $f(x) = y$. It follows from this definition that

$$y = f(x) \Leftrightarrow x = g(y) \quad \text{for all } x \in X \text{ and for all } y \in Y$$

and so g is an inverse of f.

(c) 'f invertible $\Rightarrow$ it has a unique inverse': Suppose that f is invertible and $g_1: Y \to X$ and $g_2: Y \to X$ are inverse functions for f. To demonstrate that $g_1 = g_2$ we must show that $g_1(y) = g_2(y)$ for all $y \in Y$. Let $y_0 \in Y$. Then put $x_1 = g_1(y_0)$ and $x_2 = g_2(y_0)$ so that we are required to prove that $x_1 = x_2$. This is easy because $x_1 = g_1(y_0) \Rightarrow y_0 = f(x_1)$ and $x_2 = g_2(y_0) \Rightarrow y_0 = f(x_2)$ so that $f(x_1) = f(x_2)$. But now, since f is invertible it is bijective (by (a)) and so in particular injective. It follows that $x_1 = x_2$, i.e. $g_(y_0) = g_2(y_0)$ as required. □

Because of this theorem the words 'bijective' and 'invertible' are used interchangeably when referring to functions. Many standard functions are not actually bijections but by suitable restrictions of the domain and codomain can be converted into bijections. Here are some examples from calculus.

Examples 9.2.4 (a) The function $\sin: \mathbb{R} \to \mathbb{R}$ is neither injective (since, for example, $\sin 0 = \sin \pi$) nor surjective (since, for example, there is no real number x such that $\sin x = 2$). To define an inverse function $\sin^{-1}$ we restrict the domain and the codomain of sin so that it is bijective. Surjectivity is easy: we simply change the codomain to the image of sin,

i.e. Im(sin), which is the set $[-1,1] = \{ x \in \mathbb{R} \mid -1 \leqslant x \leqslant 1 \}$ (since we know that each number in this set arises as a value of the sine function on $\mathbb{R}$). Injectivity is not much harder for we know that as x varies from $-\pi/2$ to $\pi/2$ the value of $\sin x$ increases steadily from -1 to 1 (sketch the graph). So we take as our domain $[-\pi/2, \pi/2] = \{ x \in \mathbb{R} \mid -\pi/2 \leqslant x \leqslant \pi/2 \}$. To sum up, the function†

$$\sin: [-\pi/2, \pi/2] \to [-1, 1]$$

is a bijection. By the above theorem it has an inverse

$$\sin^{-1}: [-1, 1] \to [-\pi/2, \pi/2].$$

In making this definition we had to choose a subset of $\mathbb{R}$ on which the sine function was injective. There are many possibilities, for example $[\pi/2, 3\pi/2]$, and each has its own inverse function. The above choice is the most natural and the function it leads to is called the *principal value* of the inverse.

In the same way we can restrict the cosine and tangent functions to obtain inverse functions

$$\cos^{-1}: [-1, 1] \to [0, \pi]$$

and

$$\tan^{-1}: \mathbb{R} \to (-\pi/2, \pi/2).$$

(b) Given $n \in \mathbb{Z}^+$, the function $x \mapsto x^n$ is a bijection $\mathbb{R} \to \mathbb{R}$ for n odd and so in this case we do have an inverse function $\mathbb{R} \to \mathbb{R}$, the *nth root function*, denoted by‡ $x \mapsto x^{1/n}$ or $x \mapsto \sqrt[n]{x}$.

† Strictly speaking this function should have a different name in order to indicate the change of domain and codomain. However, it is usual to use the same name.

‡ The notation $x^{1/n}$ is selected so that the law of exponents holds: $(x^{1/n})^n = (x^n)^{1/n} = x^1 = x$. With this definition it is possible to give some indication of the problem in defining 0^0 referred to in Definition 5.3.3. Notice that $0^{1/n}$ is defined to be 0. Furthermore, we have observed (Exercise 8.4) that the sequence $1/n$ is null, $\lim 1/n = 0$, which suggests that we ought to define $0^0 = 0$. However, if the law of indices is to hold then we must have $x^0 = 1$ for non-zero x (see the solution to Exercise 5.7) and this suggests that $0^0 = 1$, the convention adopted in this book. It is not possible to satisfy these conflicting demands. For $(x, y) \in \mathbb{R}^{\geqslant} \times \mathbb{R} - \{(0,0)\}$ it is possible to define x^y extending the definition for integer exponenents so that the laws of exponents are satisfied and so that the function of two variables $(x, y) \mapsto x^y$ is 'continuous'. It is not possible to extend this 'continuous' function to the point $(0, 0)$. A full explanation of this, and the definition of 'continuous', is beyond the scope of this book (see for example K.G. Binmore, *Mathematical analysis, a straightforward approach,* Cambridge University Press, Second edition 1982).

On the other hand it is neither an injection nor a surjection when n is even. [The case $n = 2$ is typical and is discussed in Example 9.1.3(c).] In this case we restrict the domain and the codomain to $\mathbb{R}^{\geqslant}$ so that we have a bijection. Thus the inverse function $x \mapsto x^{1/n} = \sqrt[n]{x}$ is a function $\mathbb{R}^{\geqslant} \to \mathbb{R}^{\geqslant}$. The convention is that for even n the expressions $x^{1/n}$ and $\sqrt[n]{x}$ represent the non-negative root of a non-negative real number x.

The following equivalent formulation of Definition 9.2.1 is often useful.

Proposition 9.2.5 *The functions $f: X \to Y$ and $g: Y \to X$ are inverses of each other if and only if $g \circ f = I_X$ and $f \circ g = I_Y$.*

Proof '$\Rightarrow$': Suppose that f and g are inverse to each other. Then

$$y = f(x) \Leftrightarrow x = g(y) \quad \text{for all } x \in X \text{ and all } y \in Y.$$

Now, given $x_0 \in X$, put $y_0 = f(x_0)$ so that $x_0 = g(y_0)$. Then $g \circ f(x_0) = g(y_0) = x_0$, i.e. $g \circ f = I_X$.

Similarly $f \circ g = I_Y$.

'$\Leftarrow$': Suppose that $g \circ f = I_X$ and $f \circ g = I_Y$. Suppose that $y_0 = f(x_0)$. Then $g(y_0) = g \circ f(x_0) = I_X(x_0) = x_0$. Thus $y = f(x) \Rightarrow x = g(y)$ for all $x \in X$, $y \in Y$.

Similarly, $x = g(y) \Rightarrow y = f(x)$ and so we have proved that f and g are inverses of each other. □

9.3 Functions and subsets

In the previous section it has been demonstrated that a function f has an inverse f^{-1} if and only if it is a bijection. However, the reader will discover that the notation f^{-1} which has been used for the inverse function is used even when the function f is not a bijection. This chapter ends with a brief description of this more general use.

To begin with recall the definition of the power set (Definition 6.3.1): the power set $\mathcal{P}(X)$ of a set X is the set of subsets of X so that $A \in \mathcal{P}(X)$ means simply $A \subseteq X$. We now describe how given a function $f: X \to Y$ we can associate to it two functions between the power sets of X and Y (one in each direction) each of which captures the function f in a different way.

Definition 9.3.1 *Suppose that $f: X \to Y$ is a function.*

(i) *The function* $\overrightarrow{f}: \mathcal{P}(X) \to \mathcal{P}(Y)$ *is defined by*

$$\overrightarrow{f}(A) = \{ f(x) \mid x \in A \}$$

for $A \in \mathcal{P}(X)$.

(ii) *The function* $\overleftarrow{f}: \mathcal{P}(Y) \to \mathcal{P}(X)$ *is defined by*

$$\overleftarrow{f}(B) = \{ x \in X \mid f(x) \in B \}$$

for $B \in \mathcal{P}(Y)$.

Remarks 9.3.2 The notations $\overrightarrow{f}$ and $\overleftarrow{f}$ are non-standard. Most writers denote the function $\overrightarrow{f}$ simply by f and denote the function $\overleftarrow{f}$ by f^{-1}. With experience this does not lead to confusion (and is another example of how ambiguity can be better than pedantry). However, in this book the notations $\overrightarrow{f}$ and $\overleftarrow{f}$ will be used since it seems potentially confusing to have two different functions with the same name when meeting these ideas for the first time.

In the case of $\overrightarrow{f}$ notice that each element x_0 of X corresponds to an element of $\mathcal{P}(X)$, namely the singleton subset $\{x_0\}$, and similarly for Y. So we can think of $\overrightarrow{f}$ as an extension of f in the sense that

$$\overrightarrow{f}(\{x_0\}) = \{ f(x) \mid x \in \{x_0\} \} = \{ f(x) \mid x = x_0 \} = \{f(x_0)\}.$$

Notice that $\overrightarrow{f}(X)$ is the image of f, $\mathrm{Im}(f)$ (see Definition 8.4.1).

In the case of $\overleftarrow{f}$,

$$\overleftarrow{f}(\{y_0\}) = \{ x \in X \mid f(x) \in \{y_0\} \} = \{ x \in X \mid f(x) = y_0 \}$$

for each $y_0 \in Y$. Thus $\overleftarrow{f}(\{y_0\})$ gives the set of pre-images of y_0. In the box model of a function (page 90) this gives the set of elements in the box y_0. Now, if f is a bijection with inverse f^{-1}, then $f(x) = y_0$ if and only if $x = f^{-1}(y_0)$ so that $\overleftarrow{f}(\{y_0\}) = \{f^{-1}(y_0)\}$ and we can think of $\overleftarrow{f}$ as an extension of f^{-1}.

If f is not a bijection then $\overleftarrow{f}(\{y\})$ will not be a singleton subset for some elements of Y: if it is not a surjection then it will be empty for elements y not in the image, and if not an injection then it will contain more than one element for some y.

9.4 Peano's axioms for the natural numbers

The notion of the successor of an integer n was introduced in Section 5.1. Chapter 5 was headed by a quotation from Richard Dedekind suggesting that this idea captured the essence of the natural numbers (or positive integers). The necessary language is now available so that some indication of the ideas behind the Dedekind quotation can be given.

Definition 9.4.1 *The* successor function, $s: \mathbb{Z}^+ \to \mathbb{Z}^+$, *is defined by* $s(n) = n + 1$ *for* $n \in \mathbb{Z}^+$.

Dedekind observed that the existence of the successor function together with the number 1 completely captures the properties of the natural numbers. These ideas are now usually associated with the name of the Italian mathematician Giuseppe Peano.†

Axioms 9.4.2 (Peano) *The set of positive integers* $\mathbb{Z}^+$ *is a set with a function* $s : \mathbb{Z}^+ \to \mathbb{Z}^+$ *and an element* $1 \in \mathbb{Z}^+$ *such that*

(i) s *is an injection,*
(ii) 1 *is not in the image of* s, *and*
(iii) *for* $A \subseteq \mathbb{Z}^+$, *if* $1 \in A$, *and* $n \in A \Rightarrow s(n) \in A$, *then* $A = \mathbb{Z}^+$.

Notice that (iii) is just the induction axiom as reformulated in Axiom 7.5.1.

† The first exposition of these ideas was by Richard Dedekind in his book *Was sind und was sollen die Zahlen?* published in 1888. Peano formulated his axioms for the natural numbers in his book *Arithmetices principia, nova methodo exposita* published in 1889. Peano did not see Dedekind's work until his own book was in press and Peano's exposition using the language of set theory is considered to have been much more influential. It was in this book that he introduced notations for set membership (the modern symbol $\in$) and set inclusion (he used an inverted 'C' which he also used for implication – recall that it was observed in Section 6.1 that there is a strong connection between set inclusion and implication) and he seems to have been the first to clarify the distinction between set membership and set inclusion. Peano was a remarkable mathematician. It is curious that his 1889 book was written in Latin rather than the Italian or French he usually used. In his book about Peano (Reidel, 1980), Hubert Kennedy describes it as 'the unique romantic act of his scientific career' and suggests that it reflected Peano's awareness that he was doing something historic – after all, the great classics such as Isaac Newton's *Principia* had been written in Latin. The following year brought Peano's most striking achievement, his construction of a space-filling curve; this is a curve in the plane $\mathbb{R}^2$ given by continuous functions $x = f(t)$, $y = g(t)$ such that, as t varies over the interval $[0, 1]$, (x, y) passes through every point of the unit square $[0, 1] \times [0, 1]$. This was perhaps the first indication of the subtlety of the notion of dimension, confounding intuition, whose exploration has now led to the modern study of fractal sets. (See for example Donald M. Davis, *The nature and power of mathematics*, Princeton University Press, 1993.)

It is not difficult to prove (by induction) that, given any set X with a function $s : X \to X$ and a distinguished element $1 \in X$ satisfying these axioms, there is a bijection $X \to \mathbb{Z}^+$ under which the distinguished elements 1 and the successor functions correspond.† This is what is meant by saying that these statements provide axioms for the positive integers.

We can use s to define algebraic operations on $\mathbb{Z}^+$ inductively as follows.

Definition 9.4.3 (Peano) *The sum $m + n$ of positive integers m and n may be defined by induction on n by*

(i) $m + 1 = s(m)$, *and*
(ii) *for* $k \in \mathbb{Z}^+$, $m + s(k) = s(m + k)$.

The product of positive integers mn or $m \times n$ is now defined (making use of addition) by induction on n by

(i) $m \times 1 = m$, *and*
(ii) *for* $k \in \mathbb{Z}^+$, $m \times s(k) = m \times k + m$.

It is an interesting exercise to prove the basic algebraic properties of addition and multiplication starting from these definitions, for example the commutativity of addition $m + n = n + m$.

Exercises

9.1 Determine whether each of the following functions $\mathbb{R} \to \mathbb{R}$ is injective, surjective or bijective.

(i) $f_1(x) = 2x + 5$.
(ii) $f_2(x) = x^2 + 2x + 1$.
(iii) $f_3(x) = x^2 - 2x$.
(iv) $f_4(x) = \begin{cases} 1/x & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$

9.2 In each of the above examples consider the effect of changing the domain and the codomain to $\mathbb{R}^+$.

† See for example G. Birkhoff and S. Mac Lane, *A survey of modern algebra*, Macmillan, Fourth edition 1977.

9.3 Find inverses for the following functions:

(i) $f_1: \mathbb{R} \to \mathbb{R}$ given by $f_1(x) = 3x + 2$;
(ii) $f_2: \mathbb{R} \to \mathbb{R}$ given by $f_2(x) = x^3 + 1$.

9.4 Suppose that $f: X \to Y$ and $g: Y \to Z$ are injections. Prove that $g \circ f: X \to Z$ is an injection.

9.5 Suppose that $f: X \to Y$ and $g: Y \to Z$ are bijections of sets. Prove that the composite $g \circ f: X \to Z$ is also a bijection and that

$$(g \circ f)^{-1} = f^{-1} \circ g^{-1}: Z \to Y \to X.$$

9.6 Let $f: X \to Y$ be a function with graph $G_f \subseteq X \times Y$. Prove that f is surjective if and only if $\forall y \in Y$, $(X \times \{y\} \cap G_f) \neq \emptyset$.

9.7 Let $f: X \to Y$ be a function and $B_1, B_2 \in \mathcal{P}(Y)$. Prove that

(i) $B_1 \subseteq B_2 \Rightarrow \overleftarrow{f}(B_1) \subseteq \overleftarrow{f}(B_2)$,
(ii) $\overleftarrow{f}(B_1 \cap B_2) = \overleftarrow{f}(B_1) \cap \overleftarrow{f}(B_2)$,
(iii) $\overleftarrow{f}(B_1 \cup B_2) = \overleftarrow{f}(B_1) \cup \overleftarrow{f}(B_2)$.

Prove that the converse of the first of these statements is not universally true (by constructing a counterexample).

Problems II: Sets and functions

1 Prove the following statements:

(i) $\{x \in \mathbb{R} \mid x^2 - 3x + 2 = 0\} = \{x \in \mathbb{Z} \mid 0 < x < 3\}$;
(ii) $\{x \in \mathbb{R} \mid x^2 - 3x + 2 < 0\} = \{x \in \mathbb{R} \mid x < 2\} \cap \{x \in \mathbb{R} \mid x > 1\}$;
(iii) $\{x \in \mathbb{R} \mid x^2 - 3x + 2 > 0\} = \{x \in \mathbb{R} \mid x > 2\} \cup \{x \in \mathbb{R} \mid x < 1\}$.

2. By using a truth table prove that, for sets A, B and C,

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

Draw a Venn diagram to illustrate the proof.

3. Prove the absorption laws:

(i) $A \cap (A \cup B) = A$;
(ii) $A \cup (A \cap B) = A$.

4. Prove by contradiction or otherwise that $A \cap B = A \cap C$ and $A \cup B = A \cup C$ if and only if $B = C$.

5. Using truth tables, prove that for sets A, B and C,

(i) $(A \cup C) - B \subseteq (A - B) \cup C$,
(ii) $(A \cap C) - B = (A - B) \cap C$.

Draw Venn diagrams to illustrate the proofs.

Prove that there is equality in the first of these results if and only if $B \cap C = \emptyset$.

Deduce from the second of these results that

$$(A - B) \cap C = \emptyset \text{ if and only if } A \cap C \subseteq B.$$

6. Use the distributivity law to prove that

$$(A \cap B) \cup (B \cap C) \cup (C \cap A) = (A \cup B) \cap (B \cup C) \cap (C \cup A).$$

7. For subsets of a universal set U prove that $B \subseteq A^c$ if and only if $A \cap B = \emptyset$. By taking complements deduce that $A^c \subseteq B$ if and only if $A \cup B = U$. Deduce that $B = A^c$ if and only if $A \cap B = \emptyset$ and $A \cup B = U$.

8. Given sets $A, B \in \mathcal{P}(X)$, their *symmetric difference* is defined by

$$A \Delta B = (A - B) \cup (B - A) = (A \cup B) - (A \cap B).$$

Prove that

(i) the symmetric difference is associative, $(A \Delta B) \Delta C = A \Delta (B \Delta C)$ for all $A, B, C \in \mathcal{P}(X)$,

(ii) there exists a unique set $N \in \mathcal{P}(X)$ such that $A \Delta N = A$ for all $A \in \mathcal{P}(X)$
[Hint: Guess what N is!],

(iii) for each $A \in \mathcal{P}(X)$, there exists a unique $A' \in \mathcal{P}(X)$ such that $A \Delta A' = N$,

(iv) for each $A, B \in \mathcal{P}(X)$, there exists a unique set C such that $A \Delta C = B$.

9. Using the notation of the previous problem, prove that for sets A, B, C, $D \in \mathcal{P}(X)$

$$A \Delta B = C \Delta D \Leftrightarrow A \Delta C = B \Delta D.$$

10. We define half-infinite† intervals as follows:

$$\begin{aligned} (a, \infty) &= \{ x \in \mathbb{R} \mid x > a \}; \\ [a, \infty) &= \{ x \in \mathbb{R} \mid x \geqslant a \}. \end{aligned}$$

Prove that

(i) $(a, \infty) \subseteq [b, \infty) \Leftrightarrow a \geqslant b$,

(ii) $[a, \infty) \subseteq (b, \infty) \Leftrightarrow a > b$.

† The symbol '∞' used in the notation in this question does not represent a number 'infinity'. The expression (a, ∞) is used by analogy with (a, b), where b is a real number, but the definitions are different (cf. Exercise 6.1). In this book no meaning is attached to the symbol '∞' on its own and it is only used in the notations in this question and in Definition 8.3.2. Further discussion of this use of the symbol can be found in books on analysis, for example R. Haggerty, *Fundamentals of mathematical analysis*, Addison-Wesley, Second edition 1993.

11. Give a proof or a counterexample for each of the following statements:

(i) $\forall x \in \mathbb{R}, \exists y \in \mathbb{R}, x + y > 0$;
(ii) $\forall x \in \mathbb{R}, \exists y \in \mathbb{R}, x - y > 0$;
(iii) $\exists x \in \mathbb{R}, \forall y \in \mathbb{R}, x + y > 0$;
(iv) $\forall x \in \mathbb{R}, \exists y \in \mathbb{R}, xy > 0$;
(v) $\exists x \in \mathbb{R}, \forall y \in \mathbb{R}, xy > 0$;
(vi) $\forall x \in \mathbb{R}, \exists y \in \mathbb{R}, xy \geqslant 0$;
(vii) $\exists x \in \mathbb{R}, \forall y \in \mathbb{R}, xy \geqslant 0$;
(viii) $\forall x \in \mathbb{R}, \exists y \in \mathbb{R}, (x + y > 0$ or $x + y = 0)$;
(ix) $\forall x \in \mathbb{R}, \exists y \in \mathbb{R}, (x + y > 0$ and $x + y = 0)$;
(x) $(\forall x \in \mathbb{R}, \exists y \in \mathbb{R}, x + y > 0)$ and $(\forall x \in \mathbb{R}, \exists y \in \mathbb{R}, x + y = 0)$.

12. Suppose that $A \subseteq \mathbb{Z}$. Write the following statement entirely in symbols using the quantifiers $\forall$ and $\exists$. Write out the negative of this statement in symbols.

There is a greatest number in the set A.

Give an example of a set A for which this statement is true. Give another example for which it is false.

13. Prove that, for sets A, B, C and D,

(i) $A \times (B \cup C) = (A \times B) \cup (A \times C)$,
(ii) $(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)$,

14. Define functions f and $g: \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$ and $g(x) = x^2 - 1$. Find the functions $f \circ f$, $f \circ g$, $g \circ f$, $g \circ g$.

List the elements of the set $\{ x \in \mathbb{R} \mid fg(x) = gf(x) \}$.

15. Given $A \in \mathcal{P}(X)$ define the *characteristic function* $\chi_A: X \to \{0, 1\}$ by

$$\chi_A(x) = \begin{cases} 0 & \text{if } x \notin A, \\ 1 & \text{if } x \in A. \end{cases}$$

Suppose that A and B are subsets of X.

(i) Prove that the function $x \mapsto \chi_A(x)\chi_B(x)$ (multiplication of integers) is the characteristic function of the intersection $A \cap B$.
(ii) Find the subset C whose characteristic function is given by

$$\chi_C(x) = \chi_A(x) + \chi_B(x) - \chi_A(x)\chi_B(x).$$

16. Determine which of the following functions $f_i: \mathbb{R} \to \mathbb{R}$ are injective, which are surjective and which are bijective. Write down an inverse function of each of the bijections.

(i) $f_1(x) = x - 1$;
(ii) $f_2(x) = x^3$;
(iii) $f_3(x) = x^3 - x$;
(iv) $f_4(x) = x^3 - 3x^2 + 3x - 1$;
(v) $f_5(x) = e^x$;
(vi) $f_6(x) = \begin{cases} x^2 & \text{if } x \geqslant 0, \\ -x^2 & \text{if } x \leqslant 0. \end{cases}$

17. Functions $f: \mathbb{R} \to \mathbb{R}$ and $g: \mathbb{R} \to \mathbb{R}$ are defined as follows.

$$f(x) = \begin{cases} x+2 & \text{if } x < -1, \\ -x & \text{if } -1 \leqslant x \leqslant 1, \\ x-2 & \text{if } x > 1. \end{cases}$$

$$g(x) = \begin{cases} x-2 & \text{if } x < -1, \\ -x & \text{if } -1 \leqslant x \leqslant 1, \\ x+2 & \text{if } x > 1. \end{cases}$$

Find the functions $f \circ g$ and $g \circ f$. Is g the inverse of the function f? Is f injective or surjective? How about g? Sketch and compare the graphs of these functions.

18. Suppose that $f: X \to Y$ and $g: Y \to Z$ are surjections. Prove that the composite $g \circ f: X \to Z$ is a surjection.

19. Let $f: X \to Y$ be a function. Prove that there exists a function $g: Y \to X$ such that $f \circ g = I_Y$ if and only if f is a surjection. [g is called a *right inverse* of f.]

20. Let $f: X \to Y$ be a function and $A_1, A_2 \in \mathcal{P}(X)$.

(i) Prove that $A_1 \subseteq A_2 \Rightarrow \overrightarrow{f}(A_1) \subseteq \overrightarrow{f}(A_2)$. Prove that the converse is not universally true. Give a simple condition on f which is equivalent to the converse.
(ii) Prove that $\overrightarrow{f}(A_1 \cap A_2) \subseteq \overrightarrow{f}(A_1) \cap \overrightarrow{f}(A_2)$. Prove that equality is not universally true.
(iii) Prove that $\overrightarrow{f}(A_1 \cup A_2) = \overrightarrow{f}(A_1) \cup \overrightarrow{f}(A_2)$.

21. Let $f: X \to Y$ be a function. Prove that

(i) f is injective $\Leftrightarrow$ $\overrightarrow{f}$ is injective $\Leftrightarrow$ $\overleftarrow{f}$ is surjective,

(ii) f is surjective $\Leftrightarrow$ $\overrightarrow{f}$ is surjective $\Leftrightarrow$ $\overleftarrow{f}$ is injective.

22. Starting from Peano's axioms prove that if $n \in \mathbb{Z}^+$ and $n \neq 1$ then n is a successor, i.e. $s(a) = n$ for some $a \in \mathbb{Z}^+$.

[Let $A = \mathrm{Im}(s) \cup \{1\}$ and prove that $A = \mathbb{Z}^+$.]

23. Starting from Definition 9.4.3 prove that addition of positive integers is

(i) associative, i.e. $(a+b)+c = a+(b+c)$ for positive integers a, b, c,

(ii) commutative, i.e. $a+b = b+a$ for positive integers a, b.

[For (ii), prove first of all that $a+1 = 1+a$ by induction on a.]