Chapter 2

Sets , First Steps in Algebra and Functions

2.1 Sets

The mathematics we study in this course can be expressed entirely in terms of *sets* and *functions* between sets. We therefore begin with a summary of the set-theoretical concepts and definitions used in the course. We also use the occasion to summarise further notational conventions we use throughout this course.

A *set* is almost any reasonable collection of things. We shall not even attempt a more formal definition in this course. The things in the collection are called the *elements* of the set in question. We write

$$x \in A$$

to denote that x is an element of the set A and

$$x \notin A$$

to denote that x is not an element of the set A. Note that we do not exclude the possibility that x be a set in its own right, except that x cannot be A:

We explicitly exclude $A \in A$.

Two sets are considered to be the same when they comprise precisely the same elements, in other words, when every element of the first set is also an element of the second and vice versa. Formally,

Definition 19. Given two sets A and B, we say that A is a *subset* of B if and only if $x \in B$ whenever $x \in A$. We write

$$A \subseteq B$$

whenever this is the case. We say that $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

B is a *proper* subset of A if B is a subset of A, but $B \neq A$. In such a case we write

$$B \subset A.$$

Using our notational conventions, given two sets A and B,

$$A = B :\Longleftrightarrow \big((x \in A) \Leftrightarrow (x \in B)\big).$$

When we wish to describe a set, we can do so by *listing* all of its elements. Thus, if the set A has precisely a, b and c as its elements, then we write

$$A = \{a, b, c\}.$$

Another way of describing a set is by prescribing a number of *conditions* for membership of the set. In this case we write

$$\{x \mid P(x), Q(x), \ldots\}$$

to denote that the set in question consists of all those x for which $P(x), Q(x), \ldots$ all hold.

Note that, by the above, $\{a, b\}$, $\{a, b, b, b\}$ and $\{a, a, a, a, a, a, b\}$ are all the same set.

There are several operations on sets.

Definition 20. The *union* of two sets A and B consists of all those objects which are elements of one, or other (or both) of the sets. It is denoted by

$$A \cup B.$$

Using the notation above,

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

[Here := has been used to signify that the expression on the left hand side is defined to be equal to the expression on the right hand side.]

Example 21. If $A = \{1, 2, 3\}$ and $B = \{2, 4\}$, then

$$A \cup B = \{1, 2, 3, 4\}$$

Definition 22. The *intersection* of the sets A and B consists of all those objects which are elements of both and is denoted by

$$A \cap B,$$

so that,

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

Example 23. If $A = \{1, 2, 3\}$ and $B = \{2, 4\}$, then

$$A \cap B = \{2\}.$$

Definition 24. The *(Cartesian) product* of the sets A and B consists of all ordered pairs of objects, the first of each being an element of A, and the second an element of B, and is denoted by

$$A \times B,$$

so that,

$$A \times B := \{(x, y) \mid x \in A \text{ and } y \in B\}.$$

Example 25. If $A = \{1, 2, 3\}$ and $B = \{2, 4\}$, then

$$A \times B = \{(1, 2), (1, 4), (2, 2), (2, 4), (3, 2), (3, 4)\}.$$

Definition 26. Those elements of A that are not also elements of B form a set in their own right, which we denote by

$$A \setminus B,$$

so that

$$A \setminus B := \{x \in A \mid x \notin B\}.$$

Example 27. If $A = \{1, 2, 3\}$ and $B = \{2, 4\}$, then

$$A \setminus B = \{1, 3\}.$$

Convention 28. When all sets, A, under consideration are subsets of a fixed set, X, it is customary to call $X \setminus A$ the *complement* of A, and we denote this by A'.

We write $\emptyset$ for the *empty set*, which is the (unique!) set with no elements. Note that it is a subset of every set, that is, if X is any set, then $\emptyset \subseteq X$.

Observation 29. It is common to at first confuse $\in$ and $\subseteq$, partly because it is common to say x is in X, when we mean that x is a element of the set X ($x \in X$), as well as to say A is in X, when we mean that the set A is a subset of the set X.

Be careful to avoid this mistake. Remember the relationship between $\in$ and $\subseteq$:

$$x \in X \text{ if and only if } \{x\} \subseteq X.$$

Relations on sets play a central role in mathematics. Of particular importance for this course is *ordering*.

Orderings. We begin mathematics by learning to count, when we learn that the counting numbers come in a particular order. Order relations generalise this.

Definition 30. Let $\preceq$ be a binary relation on the set X — that is, a relation between pairs of elements of X — and write $a \preceq b$ whenever a is related to b in the sense of $\preceq$. Then $\preceq$ is an *partial order* on X if and only if for all $a, b, c \in X$

(i)	$a \preceq a$	**Reflexivity**
(ii)	If $a \preceq b$ and $b \preceq a$, then $a = b$.	**Antisymmetry**
(iii)	If $a \preceq b$ and $b \preceq c$, then $a \preceq c$.	**Transitivity**

It is common to write $b \succeq a$ for $a \preceq b$.

It is customary to write $a \prec b$ (or $b \succ a$) to denote that $a \preceq b$, but $a \neq b$.

The partial ordering $\preceq$ on X is a *total order* on X if and only if, in addition,

(iv) Either $a \preceq b$ or $b \preceq a$.

Hence, for elements, a, b of a totally ordered set precisely one of the following holds.

$$a \prec b, \qquad a = b, \qquad a \succ b.$$

Example 31. The number systems familiar from school are totally ordered sets, with $\preceq$ being the familiar ordering less than or equal to, so that $a \preceq b$ if and only if $a \leq b$.

Example 32. An example of a partially ordered set, which is not totally ordered is obtained by taking as X the set of all subsets of a fixed set, Y, which has at least two elements, so that

$$A \in X \text{ if and only if } A \subseteq Y,$$

and defining $\preceq$ by

$$A \preceq B \text{ if and only if } A \subseteq B.$$

We verify that this defines a partial order, but not a total order, on X.

Let A, B, C be subsets of Y.

(i) Since every element of A is an element of A, $A \subseteq A$, or, equivalently, $A \preceq A$.

(ii) Suppose that $A \preceq B$ and $B \preceq A$. Then $A \subseteq B$ and $B \subseteq A$. By Definition 19, this is equivalent to $A = B$.

(iii) Suppose that $A \preceq B$ and $B \preceq C$, so that $A \subseteq B$ and $B \subseteq C$. To show that $A \preceq C$, we must show that $A \subseteq C$. To do this, take any $y \in A$.

Since $A \subseteq B$, we must have $y \in B$, Then, since $B \subseteq C$, we must have $y \in C$.

Since we have shown that any element of A is an element of C, we have proved that $A \subseteq C$, as required.

(iv) Finally, to see that our $\preceq$ is not a total order, take two distinct elements, of Y, $a \neq b$.

Then $a \notin \{b\}$, whence $\{a\} \not\subseteq \{b\}$, that is, $\{a\} \not\preceq \{b\}$.

Similarly $b \notin \{a\}$, whence $\{b\} \not\subseteq \{a\}$, that is, $\{b\} \not\preceq \{a\}$.

Certain subsets of ordered sets play a prominent part in mathematics.

Definition 33. Let X be (parially) ordered by $\preceq$. Take $a, b \in X$. Then $x \in$ *lies between* a and b if and only if either $a \preceq x \preceq b$ or $b \preceq x \preceq a$.

The subset I of X is an *interval* if and only if given $a, b \in I$ every element of X which lies between a and b is an element of I.

Given $a, b \in X$ with $a \prec b$,

(i) $]a, b[:= \{x \in X \mid a \prec x \prec b\}$ is the *open interval* from a to b,

(ii) $[a, b[:= \{x \in X \mid a \preceq x \prec b\}$ is the *interval* from a to b, *closed* at a and *open* at b.

(iii) $]a, b] := \{x \in X \mid a \prec x \preceq b\}$ is the *interval* from a to b, open at a and closed at b.

(iv) $[a, b] := \{x \in X \mid a \preceq x \preceq b\}$ is the *closed interval* from a to b.

It is obvious that these are all intervals.

A number of sets occur with such frequency that special notation has been introduced for them. These include the sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ consisting respectively of all *natural numbers*, all *integers*, all *rational numbers*, all *real numbers* and all *complex numbers*, which we introduce below. Observe that

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

Explicitly,

$$\begin{aligned}
\mathbb{N}: &= \{0,1,2,3,\dots\} \\
\mathbb{Z}: &= \{\dots -3,-2-1,0,1,2,3,\dots\} \\
\mathbb{Q}: &= \{\, x \in \mathbb{R} \mid x = \frac{p}{q} \text{ for some } p,q \in \mathbb{Z}, \text{ with } q \neq 0 \,\} \\
&= \{\, x \in \mathbb{R} \mid x = \frac{p}{q} \text{ with } p \in \mathbb{Z} \text{ and } q \in \mathbb{Z} \setminus \{0\} \,\}
\end{aligned}$$

We also write $\mathbb{N}^*$ for $\mathbb{N} \setminus \{0\} = \{1,2,3,\dots\}$, the set of *counting numbers*.

2.2 Arithmetic and First Steps in Algebra

The above sets, with the exception of the set of complex numbers, are familiar from school. A feature they share, also familiar from school, is that they allow *arithmetic*: we can add and multiply them. We briefly outline the path followed at school, leading from the counting numbers to the real numbers.

Both addition and multiplication were originally defined for counting numbers.

These satisfy several *Laws of Arithmetic*.

Given counting numbers, x, y, z

A1 $(x+y)+z = x+(y+z)$ **Associativity of an Addition**

A4 $x+y = y+x$ **Commutativity of Addition**

M1 $(x \times y) \times z = x \times (y \times z)$ **Associativity of an Addition**

M2 $1 \times x = x = x \times 1$ **Existence of a Multiplicative Neutral Element**

M4 $x \times y = y \times x$ **Commutativity of Multiplication**

D $x \times (y+z) = (x \times y) + (x \times z)$ and $(x+y) \times z = (x \times z) + (y \times z)$
Distributivity of Multiplication over Addition

We extend the counting numbers to the natural numbers by adjoining 0 in a manner which allows us to extend addition and multiplication. This introduces a new Law of Arithmetic.

Let x be a natural number.

A2 $x+0 = x = 0+x$ **Existence of an Additive Neutral Element**

This expanded number system has many uses, but some limitations. For example, if we are given natural numbers, a, b, how do we find natural numbers x, y so that

$$a + x = b \qquad (\diamondsuit)$$
$$a \times y = b \qquad (\diamondsuit\diamondsuit)$$

To allow us to solve all equations of the form $(\diamondsuit)$, we extend the natural numbers to the integers by adjoining *negative numbers* in a manner which allows us to extend addition and multiplication. This introduces a new Law of Arithmetic.

Let x be an integer.

A3 $(-x) + x = 0 = x + (-x)$ **Existence of Additive Inverses**

It also allows us to define *subtraction* by

$$a - b := a + (-b)$$

It is not possible to solve all equations of the form $(\diamondsuit\diamondsuit)$, because $0 \times x$ is always 0, which means that $(\diamondsuit\diamondsuit)$ cannot have a solution when $a = 0$ and $b \neq 0$. Fortunately, this is the only exceptional case and we extend the integers to the rational by adjoining *fractions* — expressions of the form $\frac{m}{n}$ with m an integer and m a counting number — in a manner which allows us to extend addition and multiplication. This introduces another new Law of Arithmetic.

Let x be an rational number.

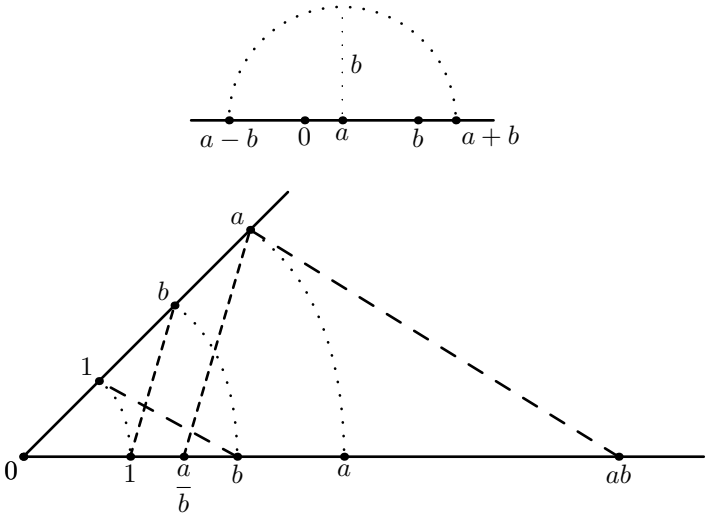
M3 $(\frac{1}{x}) \times x = 1 = x \times (\frac{1}{x})$ for $x \neq 0$ **Existence of Multiplicative Inverses**

This allows us to define *division by non-zero rational numbers* by

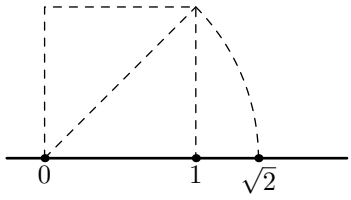
$$b \div a := b \times \frac{1}{a} \qquad \text{for } a \neq 0.$$

Convention 34. We adopt the convention of writing xy instead of $x \times y$ when there is no danger of confusion, as well as the usual conventions on omitting parentheses, so that we write $xy + z$ for $(x \times y) + z$.

The extension to real numbers is of a different nature. We can represent rational numbers by points on a straight line, by choosing one point to represent 0 and choosing another to represent 1. Then the arithmetic operations can be performed using compasses and rule.



Moreover, we can construct points on our line which do not represent any rational number. For example, we can construct a point representing the number $\sqrt{2}$ by constructing a unit square, and taking its diagonal.



Lemma 35. $\sqrt{2}$ *is not rational.*

Proof. Suppose that $\sqrt{2}$ is rational.

Since $\sqrt{2}$ is positive, there are counting numbers, m, n with no common factors, such that

$$\sqrt{2} = \frac{m}{n}.$$

Squaring and multiplying through by n^2 yields

$$m^2 = 2n^2$$

from which we see that m^2 must be even. But then m must be even, so that $m = 2k$, for some counting number, k. Thus

$$4k^2 = 2n^2,$$

or, equivalently,

$$n^2 = 2k^2,$$

whence we get (as above) that n must also be even, say $n = 2\ell$, for some counting number, ℓ.

Thus both m and n are divisible by 2, contradicting the choice of m and n. □

As the above indicates, we may regard the passage from the rational numbers to the real numbers as "filling the gaps" between rational numbers.

The study of the set of real numbers and functions defined on it forms the bulk of this course our main concern in this course. We gather together here the main properties and features of the set of real numbers.

The set of real numbers, $\mathbb{R}$, has two distinct structures, the first is the algebraic structure arising from arithmetic, and the second arising from the total ordering (cf Example 31). These, and their interplay, are characterised by the following axioms.

Algebraic Axioms

Take $x, y, z \in \mathbb{R}$.

A1 $x + (y + z) = (x + y) + z$ **Associativity of Addition**

A2 $x + 0 = x = 0 + x$ **Existence of an Additive Neutral Element**

A3 $(-x) + x = 0 = x + (-x)$ **Existence of Additive Inverses**

A4 $y + x = x + y$ **Commutativity of Addition**

M1 $x(yz) = (xy)z$ **Associativity of Multiplication**

M2 $x1 = x = 1x$ **Existence of a Multiplicative Neutral Element**

M3 $(\frac{1}{x})x = 1 = x(\frac{1}{x})$ for $x \neq 0$ **Existence of Multiplicative Inverses**

M4 $yx = xy$ **Commutativity of Multiplication**

D $x(y+z) = xy + xz$ and $(x+y)z = xz + yz$
Distributivity of Multiplication over Addition

Order Axioms

Take $x, y, z \in \mathbb{R}$.

O1 If $x < y$, then $x + z < y + z$.

O2 If $x < y$ and $z > 0$, then $xz < yz$.

O3 If $x > 0$, there is a $q \in \mathbb{Z}$, with $qx \leq y < (q+1)x$. **Archimidean Property**

The properties familiar from school mathematics can be deduced from the above axioms.

Theorem 36. *Take $x, y, z \in \mathbb{R}$.*

(i) There is only one additive neutral element, and there is only one multiplicative neutral element.

(ii) x has precisely one additive inverse, and, if $x \neq 0$, it has precisely one multiplicative inverse.

(iii) $(-(-x)) = x$ *and, if* $x \neq 0$, $\dfrac{1}{\frac{1}{x}} = x$.

(iv) $0x = 0$.

(v) $(-x) = (-1)x$.

(vi) $(-x)y = (-(xy))$.

(vii) If $xy = 0$, *then either* $x = 0$ *or* $y = 0$.

(viii) $x > y$ *if and only if* $x - y > 0$.

(ix) $x > 0$ *if and only if* $(-x) < 0$.

(x) $x^2 \geq 0$, *with equality if and only if* $x = 0$.

(xi) $x > y$ *if and only if* $(-x) < (-y)$

(xii) $x > 0$ *if and only if* $\frac{1}{x} > 0$.

(xiii) If $0 < x < y$, *then* $0 < \frac{1}{y} < \frac{1}{x}$.

These statements follow by direct application of the axioms. Similar proofs apply to similar algebraic settings we meet later, so it is important that the reader realise that many common properties of real numbers depend only on the algebraic structure and are therefore shared by any other set with a similar algebraic structure. Since the reader may not yet be comfortable with rigorous proofs from axioms, we present the proofs, leaving some parts as exercises, which the reader is urged to complete, in order to learn to master techniques of proof.

The proofs are not light reading or entertaining. The reader should read, but not dwell on, them, returning to them as needed.

Proof. (i) Take $a \in \mathbb{R}$ with the property that $a + x = x$ for every $x \in \mathbb{R}$. Then

$$\begin{aligned} a &= a + 0 && \text{by } \mathbf{A2} \\ &= 0 && \text{by the choice of } a \end{aligned}$$

The corresponding statement about the multiplicative neutral element is left to the reader.

(ii) Given $x \in \mathbb{R}$, take $y \in \mathbb{R}$ with $y + x = 0$. Then

$$\begin{aligned} y &= y + 0 && \text{by } \mathbf{A2} \\ &= y + (x + (-x)) && \text{by } \mathbf{A3} \\ &= (y + x) + (-x) && \text{by } \mathbf{A1} \\ &= 0 + (-x) && \text{by the choice of } y \\ &= (-x) && \text{by} \mathbf{A2} \end{aligned}$$

The corresponding statement about multiplicative inverses is left to the reader.

(iii) This follows from (ii), since, by **A3**, both $(-(-x))$ and x are additive inverses of $(-x)$.

The corresponding statement for the multiplicative cased is left to the reader.

(iv) Take $x \in \mathbb{R}$. Then

$$\begin{aligned} 0x &= 0x + 0 && \text{by } \mathbf{A2} \\ &= 0x + (0x + (-(0x)) && \text{by } \mathbf{A3} \\ &= (0x + 0x) + (-(0x)) && \text{by } \mathbf{A1} \\ &= ((0 + 0)x) + (-(0x)) && \text{by } \mathbf{D} \\ &= (0x) + (-(0x)) && \text{by } \mathbf{A2} \\ &= 0 && \text{by } \mathbf{A3} \end{aligned}$$

(v) Take $x \in \mathbb{R}$. Then

$$\begin{aligned} (-1)x + x &= (-1)x + 1x && \text{by } \mathbf{M2} \\ &= ((-1) + 1)x && \text{by } \mathbf{D} \\ &= 0x && \text{by } \mathbf{A3} \\ &= 0 && \text{by (iv)} \end{aligned}$$

Since both $(-1)x$ and $(-x)$ are additive inverses of x, by (ii), $(-1)x = (-x)$.

(vi) Take $x, y \in \mathbb{R}$. Then

$$\begin{aligned} (-x)y + xy &= ((-x) + x)y && \text{by } \mathbf{D} \\ &= 0y && \text{by } \mathbf{A3} \\ &= 0 && \text{by (iv)} \end{aligned}$$

Since both $(-x)y$ and $(-(xy))$ are additive inverses of xy, by (ii), $(-x)y = (-(xy))$.

(vii) Take $x, y \in \mathbb{R}$, with $xy = 0$ and $x \neq 0$. Then

$$\begin{aligned} y &= 1y && \text{by } \mathbf{M2} \\ &= ((\tfrac{1}{x})x)y && \text{by } \mathbf{M3}, \text{ as } x \neq 0 \\ &= (\tfrac{1}{x})(xy) && \text{by } \mathbf{M1} \\ &= (\tfrac{1}{x})0 && \text{as } xy = 0 \\ &= 0 && \text{by } \mathbf{M4} \text{ and (iv)} \end{aligned}$$

(viii) Take $x, y \in \mathbb{R}$.

(a) Suppose that $x > y$. Then

$$\begin{aligned} x - y &= x + (-y) && \text{by the definition of subtraction} \\ &> y + (-y) && \text{by } \mathbf{O1} \\ &= 0 && \text{by } \mathbf{A3} \end{aligned}$$

Thus $x > y$ only if $x - y > 0$.

(b) Suppose that $x - y > 0$. Then

$$\begin{aligned} x &= x + 0 && \text{by } \mathbf{A2} \\ &= x + ((-y) + y) && \text{by } \mathbf{A3} \\ &= (x + (-y)) + y && \text{by } \mathbf{A1} \\ &= (x - y) + y && \text{by the definition of subtraction} \\ &> 0 + y && \text{by } \mathbf{01} \text{ as } x - y > 0 \\ &= y && \text{by } \mathbf{A2} \end{aligned}$$

(ix) Take $x \in \mathbb{R}$. Suppose that $x > 0$. Then

$$\begin{aligned} 0 &= x + (-x) && \text{by } \mathbf{A3} \\ &> 0 + (-x) && \text{by } \mathbf{O1} \text{ as } x > 0 \\ &= (-x) && \text{by } \mathbf{A2} \end{aligned}$$

Thus $x > 0$ only if $(-x) < 0$.

The converse is left to the reader.

(x) Take $x \in \mathbb{R}$.

If $x > 0$, then,

$$\begin{aligned} x^2 &= xx \\ &> x0 && \text{by } \mathbf{02} \\ &= 0 && \text{by } \mathbf{M4} \text{ and (iv)} \end{aligned}$$

If $x < 0$, then

$$\begin{aligned} x^2 &= xx \\ &= (-(-x))(-(-x)) && \text{by (iii)} \\ &= ((-1)(-x))((-1)(-x) && \text{by (v)} \\ &= ((-1)(-1))((-x)(-x)) && \text{by } \mathbf{M1} \text{ and } \mathbf{M4} \\ &= ((-(-1))((-x)(-x)) && \text{by (v)} \\ &= 1((-x)(-x)) && \text{by (iii)} \\ &= (-x)(-x) && \text{by } \mathbf{M2} \\ &> 0 && \text{as, by (ix), } (-x) > 0 \end{aligned}$$

(xi) By (ix), $x > y$ if and only if $x + (-y) = x - y > 0$. But then

$$\begin{aligned} (-y) &= 0 + (-y) && \text{by } \mathbf{A2} \\ &= ((-x) + x) + (-y) && \text{by } \mathbf{A3} \\ &= (-x) + (x + (-y)) && \text{by } \mathbf{A1} \\ &> (-x) + 0 && \text{by } \mathbf{O1}. \text{ as } x + (-y) > 0 \\ &= (-x) && \text{by } \mathbf{A2} \end{aligned}$$

The converse is left to the reader.

(xii) Since $y > x > 0$, $0 < \frac{1}{x}, \frac{1}{y}$, by (xii). Moreover,

$$x(\tfrac{1}{x}) < y(\tfrac{1}{x}) \qquad \text{by } \mathbf{02}, \text{ as } \tfrac{1}{x} > 0,$$

whence

$$(x(\tfrac{1}{x}))(\tfrac{1}{y}) < (y(\tfrac{1}{x}))(\tfrac{1}{y}) \qquad \text{by } \mathbf{02}, \text{ as } \tfrac{1}{y} > 0.$$

Using **M3** and **M4**, we deduce that

$$1(\tfrac{1}{y}) < ((\frac{1}{x})y)(\tfrac{1}{y}).$$

By **M2** and **M1** we see that

$$\tfrac{1}{y} < (\tfrac{1}{x})(y(\tfrac{1}{y})),$$

from which we deduce, using **M3** and **M2**, that

$$\tfrac{1}{y} < \tfrac{1}{x}.$$

□

We presented the above proof in great detail for the benefit of those readers unaccustomed to rigorous proofs. While our proofs will remain rigorous throughout, we shall increasingly leave more purely routine details to the reader.

Before continuing, we show how the above explains one of matters which pupils typically find problematic, namely why we "change the signs when expanding something in parentheses if there is a negative sign outside".

Example 37. We show that for all real numbers a, b, we have $-(a-b) = (-a) + b$.

$$\begin{aligned}
-(a-b) &= -\big(a + (-b)\big) && \text{by the definition of subtraction}\\
&= (-1)\big(a + ((-1)b)\big) && \text{by Theorem 36 (v)}\\
&= \big((-1)a\big) + (-1)\big((-1)b\big) && \text{by } \mathbf{D}\\
&= \big((-1)a\big) + \big((-1)(-1)\big) && \text{by } \mathbf{M1}\\
&= (-a) + (-(-1)b && \text{by Theorem 36 (v)}\\
&= (-a) + 1b && \text{by Theorem 36 (iii)}\\
&= (-a) + b && \text{by } \mathbf{M2}
\end{aligned}$$

Observation 38. We note that if we replaced $\mathbb{R}$, the set of all real numbers, by $\mathbb{Q}$, the set of all rational numbers, Theorem 36 would still be true, and the same proofs would apply. So while we have captured axiomatically crucial features of the real numbers, we have not distinguished it from other sets with similar structure, such as the set of rational numbers. We turn to this next.

Definition 39. Let A be a subset of the totally ordered set X.

K is a *lower bound* for A if and only if $x \succeq K$ whenever $x \in A$. A is *bounded below* if and only if it has a lower bound.

L is an *upper bound* for A if and only if $x \preceq L$ whenever $x \in A$. A is *bounded above* if and only if it has an upper bound.

A is *bounded* if it is both bounded below and bounded above.

$i \in X$ is the *infimum*, or *greatest lower bound* of S if and only if

(i) $i \preceq x$ for all $x \in A$.

(ii) If $t \in X$ satisfies $t \preceq x$ for all $x \in A$, then $t \preceq i$.

In such a case we write $\inf A = i$.

If, in addition, $\inf A = i \in A$, then i is the *minimum* of A and we write $\min A = i$.

$s \in X$ is the *supremum*, or *least upper bound* of S if and only if

(i) $x \preceq s$ for all $x \in A$.

(ii) If $t \in X$ satisfies $x \preceq t$ for all $x \in A$, then $s \preceq t$.

In such a case we write $\sup A = s$.

If, in addition, $\sup A = s \in A$, then s is the *maximum* of A and we write $\max A = s$.

Example 40. We take $\mathbb{R}$ with its standard total order.

(i) $A = \{x \in \mathbb{R} \mid x < \sqrt{2}\}$ is bounded above, but not below. It has a supremum, namely, $\sqrt{2}$, but not a maximum.

(ii) $A = \{x \in \mathbb{Q} \mid x^2 \leq 2\}$ is bounded above, by, for example, 5, and below by, for example, -2. It has neither an infimum nor a supremum.

(iii) $A = \{x \in \mathbb{R} \mid x^2 \leq 2 \text{ and } x \in \mathbb{Q}\}$ is bounded above and below. It has $-\sqrt{2}$ as infimum and $\sqrt{2}$ as supremum. It has neither a minimum nor a maximum.

(iiv) $A = \{x \in \mathbb{R} \mid x^2 \leq 2\}$ is bounded above and below. It has $-\sqrt{2}$ as minimum and $\sqrt{2}$ as maximum.

Example 41. We take $\mathbb{Q}$ with its standard total order.

(i) $A = \{x \in \mathbb{Q} \mid x < \sqrt{2}\}$ is bounded above, but not below. It has no supremum, because $\sqrt{2}$ is not rational.

(ii) $A = \{\frac{n}{n+1} \mid n \in \mathbb{N}^*\} = \{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots\}$ is bounded above and below. It has 0 as infimum and 1 as supremum. It has neither a minimum nor a maximum.

We can now formulate the *Completeness Axiom*, which distinguishes the real numbers. Our version is not the usual one, but is equivalent to it.

Completeness Axiom

Every non-empty subset of $\mathbb{R}$ which is bound above, has a supremum.

The axioms we have listed — the Algebraic Axioms, the Order Axioms and the Completeness Axiom — determine $\mathbb{R}$ uniquely. The proof of this deep result is beyond this course.

We went to great trouble to prove several properties of the real numbers which followed without further ado from the axioms **A1**–**4**, **M1**–**4** and **D**. We did this because we did not need to use anything about the real numbers except that they satisfy the axioms. As a consequence, the results apply equally to any system satisfying the same axioms.

Definition 42. Any set, F, for which we can define two operations, α ("addition") and μ ("multiplication") satisfying Axioms **A1**–**4**, **M1**–**4** and **D** axioms is a *field*.

Example 43. The set of all rational numbers, $\mathbb{Q}$, with the usual addition as α and the usual multiplication as μ, forms a field.

The set of all real numbers, $\mathbb{R}$, with the usual addition as α and the usual multiplication as μ, forms a field.

Neither the set of all natural numbers, $\mathbb{N}$, nor the set of all integers, $\mathbb{Z}$, forms a field with the usual addition as α and the usual multiplication as μ.

One of the most important fields is the field of complex numbers, which we introduce next. Its importance is not immediately apparent, but it is crucial to much of mathematics, and indispensable for applications to theoretical physics, electronics, signal processing, statistics.

2.3 The Complex Numbers

As the reader knows, there is no real number, x, satisfying $x^2 = -1$. Yet there is a need for such a number in many situations. This was recognised in the 15^{th} century, when *Cardano's Formulæ* (due to Tartaglia!) appeared in print. Such a number was termed *imaginary*. We

now introduce and extension of the real number system which includes a square root of -1, denoted i, on which we define an "addition" and a "multiplication". We show that this new system is a field, that we can regard the real numbers as elements of this field, and that when we do, the two different additions and two different multiplications coincide.

Definition 44. The set, $\mathbb{C}$, of all complex numbers consists of all expressions of the form $x \dot{+} iy$, with x, y real numbers.

$$\mathbb{C} := \{x \dot{+} iy \mid x, y \in \mathbb{R}\}$$

The complex numbers $x \dot{+} iy$ and $u \dot{+} iv$ agree if and only if $x = u$ and $y = v$.

Given complex numbers, $x \dot{+} iy$ and $u \dot{+} iv$, we define their *sum* and *product* by

$$(x \dot{+} iy) \boxplus (u \dot{+} iv) := (x + u) \dot{+} i(y + v)$$
$$(x \dot{+} iy) \boxtimes (u \dot{+} iv) := (xu - yv) \dot{+} i(xv + yu)$$

Observation 45. We have used the unusual symbols $\dot{+}, \boxplus$ and $\boxtimes$ to keep separate the different additions and muliplications in use: $x + y$ and xy denote the customary addition and multiplication of real numbers, $\dot{+}$ represents the formal addition of the two parts of a complex numbers, and $z \boxplus w$ and $z \boxtimes w$ denote the addition and multiplication just introduced for complex numbers.

Once the properties of complex numbers have been established, we shall abandon the exotic notation, trusting the reader to know from the context which arithmetic operation is intended.

Observation 46. A convenient way to remember this is to think of i as if it were a real number, with the unusual property that $i^2 = -1$. Treating the new additions and multiplication as ordinary addition and multiplication, and replacing i^2 by -1 yields the formulæ used in the definition.

Lemma 47. $\mathbb{C}$ *forms a field with respect to* $\boxplus$ *and* $\boxtimes$.

Proof. The proof consists, for the most part, of routine verifications. We present some here, leaving the rest for the reader.

Take $a \dot{+} ib, r \dot{+} is, u \dot{+} iy \in \mathbb{C}$.

Associativity of Addition

$$\begin{aligned}
\big((a \dot{+} ib) \boxplus (r \dot{+} is)\big) \boxplus (u \dot{+} iv) &= \big((a + r) \dot{+} i(b + s)\big) \boxplus (u \dot{+} iv) && \text{by the definition of } \boxplus \\
&= \big((a + r) + u\big) \dot{+} i\big((b + s) + v\big) && \text{by the definition of } \boxplus \\
&= \big(a + (r + u)\big) \dot{+} i\big(b + (s + v)\big) && \text{by the associativity of } + \\
&= (a \dot{+} ib) \boxplus \big((r + u) \dot{+} i(s + v)\big) && \text{by the definition of } \boxplus \\
&= (a \dot{+} ib) \boxplus \big((r \dot{+} is) \boxplus (u \dot{+} iv)\big) && \text{by the definition of } \boxplus
\end{aligned}$$

Existence of a Multiplicative Neutral Element Take $a \dot{+} ib$ and $x \dot{+} iy \in \mathbb{C}$. Since $(a \dot{+} ib) \boxtimes (x \dot{+} iy) := (ax - by) \dot{+} i(ay + bx)$ in order for $a \dot{+} ib$ is to be the multiplicatise neutral element, we must have

$$ax - by = x$$
$$ay + bx = y$$

An obvious solution is $a = 1, b = 0$. Since, by Theorem 36(i), there cannot be more than one multiplicative neutral element, it is $1 \dot{+} i0$.

Existence of an Additive Neutral Element It follows immediately from the definition of $\boxplus$ and Theorem 36(i) that $0 \dot{+} i0$ is the additise neutral element.

Associativity of Multiplication

$$\begin{aligned}
&\big((a \dot{+} ib) \boxtimes (r \dot{+} is)\big) \boxtimes (u \dot{+} iv) \\
&= \big((ar - bs) \dot{+} i(as + br)\big) \boxtimes (u \dot{+} iv) && \text{by the definition of } \boxtimes \\
&= \big((ar - bs)u - (as + br)v\big) \dot{+} i\big((ar - bs)v + (as + br)u\big) && \text{by the definition of } \boxtimes \\
&= \big(aru - bsu - asv - brv\big) \dot{+} i\big(arv - bsv + asu + bru\big) && \text{by the arithmetic properties of } \mathbb{R} \\
&= \big(a(ru - sv) - b(rv + su)\big) \dot{+} i\big(a(rv + su) + b(ru - sv)\big) && \text{by the arithmetic properties of } \mathbb{R} \\
&= (a \dot{+} ib) \boxtimes \big((ru - sv) \dot{+} i(rv + su)\big) && \text{by the definition of } \boxtimes \\
&= (a \dot{+} ib) \boxtimes \big((r \dot{+} is) \boxtimes (u \dot{+} iv)\big) && \text{by the definition of } \boxtimes
\end{aligned}$$

Existence of Multiplicative Inverses A multiplicative inverse of $a \dot{+} ib \in \mathbb{C}$ is a complex number, $x \dot{+} iy$, satisfying

$$(a \dot{+} ib) \boxtimes (x \dot{+} iy) = (ax - by) \dot{+} (bx + ay) = 1 \dot{+} i0,$$

which is equivalent to finding $x, y \in \mathbb{R}$ such that

$$ax - by = 1 \tag{i}$$
$$bx + ay = 0. \tag{ii}$$

Multiplying (i) by a, (ii) by b and adding yields

$$(a^2 + b^2)x = a \tag{iii}$$

Subtracting b times (i) from a times (ii) yields

$$(a^2 + b^2)y = -b \tag{iv}$$

If $a^2 + b^2 \neq 0$, each of (iii) and (iv) has a unique solution, namely,

$$x = \frac{a}{a^2 + b^2}$$
$$y = \frac{-b}{a^2 + b^2}.$$

If, on the other hand, $a^2 + b^2 = 0$, we must have $a = b = 0$, in which case (i) has no solution.

In other words, if $a \dot{+} ib \neq 0 \dot{+} i0$ (the additive neutral element), then it has a multiplicative inverse. We then have the explicit formula

$$\frac{1}{a \dot{+} ib} = \frac{a}{a^2 + b^2} \dot{+} \frac{-b}{a^2 + b^2}.$$

□

Observation 48. Direct substitution verifies that for all real numbers a, u,

$$(a \dot{+} i0) \boxplus (u \dot{+} i0) = (a + u) \dot{+} i0$$
$$(a \dot{+} i0) \boxtimes (u \dot{+} i0) = (au) \dot{+} i0$$

This means that we may identify $\mathbb{R}$ with the subset $\{x \dot{+} i0 \mid x \in \mathbb{R}\}$ of $\mathbb{C}$.

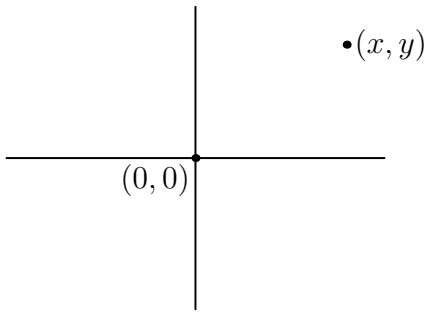
When we do so,we see that $\boxplus$ is just an *extension* of $+$.

Moreover, it is plain that $a \dot{+} ib = (a \dot{+} i0) \boxplus (0 \dot{+} ib)$, so we may regard $\dot{+}$ as a special case of $\boxplus$.

In light of these observations, we shall write $+$ instead of both $\dot{+}$ and $\boxplus$.

A Geometric Representation of Complex Numbers

Since each complex number, z, is of the form $x + iy$, with $x, y \in \mathbb{R}$, we can assign to each complex number the point (x, y) in the co-ordinate plane.



We call this the *plane of complex numbers* or the *Argand plane*.

In this representation, the horizontal axis represents the real numbers, which we have identified with the complex numbers of the form $a + i0$.

By contrast, the complex numbers of the form $0 + ia$ are the *purely imaginary* complex numbers.

We adopt the convention of omitting 0's, unless there is the danger of confusion, writing a for $a + i0$ and ib for $0 + ib$.

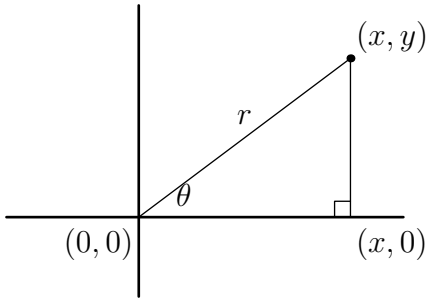
Definition 49. If $z = x + iy$, then x is the *real part* and y is the *imaginary part* of z. We write

$$\Re(z + iy) = x \qquad \text{and} \qquad \Im(x + iy) = y$$

The *complex conjugate* of $z = x + iy$ is the complex number $\overline{z} := x - iy$.

It is immediate that $\overline{\overline{z}} = z$ for any $z \in \mathbb{C}$.

The use of a co-ordinate plane to represent complex numbers suggests using *polar co-ordinates*.



Given the point (x, y) in the xy–plane, we can write $x = r\cos\theta$ and $y = r\sin\theta$, for some $r \geq$ and $0 \leq \theta < 2\pi$, with r and θ uniquely determined whenever $(x, y) \neq (0, 0)$ (that is $r > 0$). Indeed, $r = \sqrt{x^2 + y^2}$.

Definition 50. If $z = r\cos\theta + ir\sin\theta$, then r is the *modulus* and θ the *argument* of z. We write

$$|z| = r \qquad \text{and} \qquad \arg(z) = \theta$$

Lemma 51. $z\,\overline{z} = |z|^2$.

Proof. Let $z = x + iy$, with $x, y \in \mathbb{R}$. Then, since $i^2 = -1$,

$$z\,\overline{z} = (x + iy)(x - iy) = x^2 - i^2y^2 = x^2 + y^2.$$

□

Corollary 52. *$z = 0$ if and only if $|z| = 0$, and, if $z \neq 0$, then*

$$\frac{1}{z} = \frac{\overline{z}}{z\,\overline{z}} = \frac{\overline{z}}{|z|^2}.$$

Let the complex numbers z, w be $r\cos\theta + ir\sin\theta$ and $s\cos\psi + is\sin\psi$ respectively. Then

$$\begin{aligned} wz &= (r\cos\theta + ir\sin\theta)(s\cos\psi + is\sin\psi) \\ &= rs\big((\cos\theta\cos\psi - \sin\theta\sin\psi) + i(\cos\theta\sin\psi + \sin\theta\cos\psi)\big) \\ &= rs(\cos(\theta + \psi) + i\sin(\theta + \psi)), \end{aligned}$$

which establishes the next lemma.

Lemma 53. *Given complex numbers, w, z,*

$$|wz| = |w|\,|z| \qquad \textit{and} \qquad \arg(wz) = \arg(w) + \arg(z),$$

where this sum is taken modulo 2π.

Corollary 54. *If $z = r(\cos\theta + i\sin\theta)$ and n is any counting number, then*

$$z^n = r^n\big(\cos(n\theta) + i\sin(n\theta)\big)$$

Proof. We prove this by induction on n.

The case $n = 1$: When $n = 1$, we have

$$z^1 = z = r(\cos\theta + i\sin\theta) = r^1\big(\cos(1.\theta) + i\sin(1.\theta)\big),$$

showing that the proposition is true for $n = 1$.

The case $n \geq 1$: We make the inductive hypothesis that

$$z^n = r^n\big(\cos(n\theta) + i\sin(n\theta)\big) \tag{IH}$$

Then

$$\begin{aligned} z^{n+1} &= zz^n \\ &= \big(r(\cos\theta + i\sin\theta)\big)\Big(r^n\big(\cos(n\theta) + i\sin(n\theta)\big)\Big) && \text{by (IH)} \\ &= rr^n\big(\cos\theta\cos(n\theta) - \sin\theta\sin(n\theta)\big) + i\big(\cos\theta\sin(n\theta) + \sin\theta\cos(n\theta)\big) \\ &= r^{n+1}\Big(\cos\big((n+1)\theta\big) + i\sin\big((n+1)\theta\big)\Big), \end{aligned}$$

completing the inductive step. □

Corollary 55. *Let w be a non-zero complex number and n a counting number. Then the equation*

$$z^n = w \tag{$*$}$$

has the n distinct solutions

$$s^{\frac{1}{n}}\left(\cos\left(\tfrac{\alpha+2\pi k}{n}\right) + i\sin\left(\tfrac{\alpha+2\pi k}{n}\right)\right) \qquad (k = 0, 1, \ldots, n-1),$$

where $s := |w|$ and $\alpha = \arg(w)$).

Proof. Express w and z in modulus-argument form as

$$w = s(\cos\alpha + i\sin\alpha) \qquad \text{and} \qquad z = r(\cos\theta + i\sin\theta),$$

respectively, with $r, s > 0$ and $0 \leq \alpha, \theta < 2\pi$.

By Lemma 54, $z^n = r^n\big(\cos(n\theta) + i\sin(n\theta)\big)$ and $0 \leq n\theta < 2n\pi$

Thus $z^n = w$ if and only if

$$\begin{aligned} r^n &= s \\ \cos(n\theta) &= \cos\alpha \\ \sin(n\theta) &= \sin\alpha \end{aligned}$$

with $0 \leq n\theta < 2n\pi$, whence

$$\begin{aligned} r &= s^{\frac{1}{n}} \big(= \sqrt[n]{s}\big) \\ n\theta &\equiv \alpha \mod 2\pi. \end{aligned}$$

Since $0 \leq n\theta < 2n\pi$, we must have

$$n\theta = \alpha + 2k\pi$$

with $0 \leq k < n$ an integer. In other words,

$$\theta \in \{\alpha + k\tfrac{2\pi}{n} \mid k = 0, 1, \ldots, n-1\}.$$

Thus, the equation $z^n = w$ has n solutions

$$s^{\frac{1}{n}} \left(\cos\left(\tfrac{\alpha+2\pi k}{n}\right) + i \sin\left(\tfrac{\alpha+2\pi k}{n}\right)\right) \qquad (k = 0, 1, \ldots, n-1),$$

where $s := |w|$ and $\alpha = \arg(w)$).

These are distinct since, for $0 \leq \beta, \gamma < 2\pi$, $\cos(\alpha + \beta) = \cos(\alpha + \gamma)$ and $\sin(\alpha + \beta) = \sin(\alpha + \gamma)$ if and only if $\beta = \gamma$. □

2.4 Functions

To compare sets, we have the notion of a *function* or *map* or *mapping*.

Definition 56. A *function*, *map*, or *mapping* consists of three separate data, namely

(i) a *domain* that is, a set on which the function is defined,

(ii) a *codomain*, that is, a set in which the function takes its values, and

(iii) the assignment to each element of the domain of definition of a uniquely determined element from the set in which the function takes it values.

This is conveniently depicted diagrammatically by

$$f : X \longrightarrow Y,$$

or

$$X \xrightarrow{\ f\ } Y.$$

Here X is the domain of definition, Y is the set in which the function takes its values and f is the name of the function. (Note that the function f need not be given in terms of a mathematical formula.) We write $X = \mathrm{dom}(f)$ and $Y = \mathrm{codom}(f)$ to indicate that X is the domain and Y the codomain of f.

Nevertheless, we do often denote the function by f alone, but only when there is no danger of confusion. If we wish to express explicitly that the function, $f : X \longrightarrow Y$, assigns the element $y \in Y$ to the element $x \in X$, then we write $f : x \longmapsto y$ or, equivalently, $y = f(x)$. (This latter form is certainly familiar to the reader.) Sometimes the two parts are combined as

$$f : X \longrightarrow Y, \quad x \longmapsto y$$

or as

$$\begin{aligned} f : \ & X \ \longrightarrow Y \\ & x \ \longmapsto y. \end{aligned}$$

Example 57. Let X and Y both be the set of all human beings.

(i) $f\colon X \to Y,\ x \mapsto y$, where y is the (biological) father of x, is a function.

(ii) $g\colon X \to Y,\ x \mapsto y$, where y is the (biological) son of x, is not a function, since not everyone has a son, and some people have more than one son.

Example 58. We can express the algebraic operations on numbers in terms of functions. The fact that to each pair of numbers we assign their (uniquely determined) sum and product means that we have two functions

$$\alpha\colon \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}, \quad (x, y) \longmapsto x + y$$
$$\beta\colon \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}, \quad (x, y) \longmapsto xy$$

Such functions are *binary operations*, because they assign to each ordered pair of elements of a set a uniquely determined element of that set.

Definition 59. If f assigns $y \in Y$ to $x \in X$, then we say that y is the *image of x under f* or just the *image of x*.

Two functions f and g are *equal*, that is $f = g$ if and only if

(i) $\mathrm{dom}(f) = \mathrm{dom}(g)$

(ii) $\mathrm{codom}(f) = \mathrm{codom}(g)$

(iii) $f(x) = g(x)$ for every $x \in \mathrm{dom}(f)$.

In other words, to be the same, two functions must share both domain and codomain as well as agreeing everywhere.

Note that a function (or map, or mapping) is **not** just a formula.

As Example 57 shows, not every function can be expressed by a formula.

Even when a function is given by a formula, that formula need not be unique. Proving that two different formulæ define the same function can be a significant theorem.

Example 60. Pythagoras' Theorem states, in effect, that the function

$$\mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto 1$$

is the same function as

$$\mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \cos^2 x + \sin^2 x,$$

despite their being given by different formulæ.

Furthermore, there are distinct functions whose domains agree, which agree at every point (and therefore have the same range). Thus the only difference between them is that they have different codomains: *They only differ in the values they do* **not** *take!*. At this stage, it may seem peculiarly pedantic to distinguish such functions, but there are important algebraic and geometric examples, whose detailed study lies beyond the scope of these notes.

Finally, we can define functions *piecewise*, so that its values are determined differently in different parts of its domain.

Lemma 61. *Given functions $g\colon A \to Y$ and $h\colon B \to Y$ such that $g(x) = h(x)$ whenever $x \in A \cap B$, there is a unique function $f\colon A \cup B \to Y$ such that $f(a) = g(a)$ for all $a \in A$ and $f(b) = h(b)$ for all $b \in B$.*

Proof. Put $X := A \cup B$ and define f by

$$f\colon X \longrightarrow Y, \quad x \longmapsto \begin{cases} g(x) & \text{if } x \in A \\ f(x) & \text{if } x \in B \end{cases}.$$

This definition is forced by the requirement that $f(a) = g(a)$ for $a \in A$ and $f(b) = h(b)$ for $b \in B$. This means that the only possible definition of f is the one we have just given. In other words, there cannot be more than one function meeting our requirements.

The only question remaining is whether there is any such function at all, or, equivalently, whether our f is, in fact, a function.

(i) Since $X = A \cup B$ is the union of two sets, it is, itself, a set.

(ii) Y is, by hypothesis, also a set.

(iii) Take $x \in X$. Since $X = A \cup B$, either $x \in A$ or $x \in B$ (or possibly both).

If $x \in A$, then f assigns $g(x) \in Y$ to x, and $g(x)$ is uniquely determined, since $g\colon A \to Y$ is a function.

If $x \in B$, then f assigns $h(x) \in Y$ to x, and $h(x)$ is uniquely determined, since $h\colon B \to Y$ is a function.

Hence, $f\colon X \to Y$ is a function unless it happens to assign two different elements of Y to some element of X . This can only occur when $x \in A \cap B$, for then f assigns both $g(x)$ and $h(x)$ to x. But, by assumption, $g(x) = h(x)$ whenever $x \in A \cap B$, so that $f\colon X \to Y$ is, indeed, a function.[1]

□

Observation 62. In Lemma 61, the fact that $X = A \cup B$ ensures that there cannot be more than one function meeting our requirements, and the fact that g and h agree on $A \cap B$ ensure that there must be at least one such function.

Example 63. Consider the definition

$$|\ |\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \begin{cases} -x & \text{if } x \leq 0 \\ x & \text{if } x \geq 0 \end{cases}.$$

To see that $|\ |$ is a function, we define $\mathbb{R}_0^- := \{x \in \mathbb{R} \mid |x \leq 0\}$ and $\mathbb{R}_0^+ := \{x \in \mathbb{R} \mid x \geq 0\}$. Then

(i) $g\colon \mathbb{R}_0^- \to \mathbb{R},\ x \mapsto -x$ and $h\colon \mathbb{R}_0^+ \to \mathbb{R},\ x \mapsto x$ are functions;

(ii) $\mathbb{R} = \mathbb{R}_0^- \cup \mathbb{R}_0^+$;

(iii) $\mathbb{R}_0^- \cap \mathbb{R}_0^+ = \{0\}$ and $g(0) = -0 = 0 = h(0)$.

Hence, by Lemma 61, $|\ |$ is a function.

[1] We also say that f is "well-defined".

We adhere to the practice of specifying functions in a formally correct manner, stating its domain and co-domain rather than just a formula, so that it can become ordinary matter of course for the reader do so as well.

A function $f : X \longrightarrow Y$ can be represented by means of its *graph*.

Definition 64. The *graph* , $Gr(f)$, of the function $f : X \to Y$ is

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

This representation should be familiar from school.

Definition 65. The *range* or *image*, $\mathrm{im}(f)$, of the function $f\colon X \to Y$ is the subset of Y defined by

$$\mathrm{im}(f) := \{y \in Y \mid y = f(x) \text{ for some } x \in X\}.$$

Note that $\mathrm{im}(f) \subseteq \mathrm{codom}(f)$ always holds, with equality holding only sometimes. For example if $f : \mathbb{R} \longrightarrow \mathbb{R}$ is defined by $f(x) := 1$ for every $x \in \mathbb{R}$, then $\mathrm{im}(f) = \{1\} \neq \mathbb{R} = \mathrm{codom}(f)$.

Definition 66. Given a function $f : X \to Y$ and subsets A of X and B of Y, we define

$$f(A) := \{y \in Y \mid y = f(x) \text{ for some } x \in A\}$$
$$f^{-1}(B) := \{x \in X \mid f(x) \in B\}.$$

Then $f(A)$ is called the *image of A under f* and $f^{-1}(B)$ is called the *inverse image of B under f*, or the *pre-image of B under f*.

Definition 67. The *identity function*, on the set X, denoted id_X, is the function

$$id_X : X \longrightarrow X,\ x \longmapsto x,$$

which we may also write as $id_X(x) = x$.

Notice that both the domain and codomain must be precisely X for this definition to specify the identity function.

We can characterise subsets purely in terms of functions.

Definition 68. Suppose that $X \subseteq Y$. The *inclusion function* is

$$i_X^Y\colon X \longrightarrow Y, \quad x \longmapsto x$$

On the left x is considered as element of X . On the right, it is considered as element of Y.

Lemma 69. *Let X, Y be sets. Then $X \subseteq Y$ if and only if*

$$X \longrightarrow Y, \quad x \longmapsto x \qquad (*)$$

is a function.

Proof. The proof is essentially no more than a restatement of the definition of a function.

Suppose that $X \subseteq Y$. Then, by definition, every element of the set X is also an element of the set Y. Since every x trivially determines itself uniquely. Hence $(*)$ meets all three requirements in the definition of a function.

For the converse, suppose that $(*)$ defines a function, then since each $x \in X$ is mapped to itself, but now as element of Y, each element of X must be an element of Y, that is to say, $X \subseteq Y$. □

Definition 70. Given a function $f : X \to Y$ and a subset A of X we can use f to define a function $f\,|_A \colon A \longrightarrow Y$, called the *restriction of f to A* by means of

$$f\,|_A\,(x) = f(x) \quad \text{for every } x \in A.$$

Note that unless of course $A = X$, this is **not** the same function as f, even though the two functions agree everywhere they are both defined.

Functions can sometimes be *composed*.

Definition 71. Given functions $f : X \to Y$ and $g : Y \to Z$ their *composition*, denoted by $g \circ f$, is the function defined by

$$g \circ f : X \longrightarrow Z, \quad x \longmapsto g\big(f(x)\big),$$

as long as $\mathrm{im}(f) = \mathrm{dom}(g) = Y$.

In such a case,

$$\begin{aligned} \mathrm{dom}(g \circ f) &= \mathrm{dom}(f) \\ \mathrm{codom}(g \circ f) &= \mathrm{codom}(g) \\ \mathrm{im}(g \circ f) &\subseteq \mathrm{im}(g). \end{aligned}$$

The first two of these statements are true by definition and the last is an immediate consequence.

Equality need not hold in the last of these statements. To see this consider the functions $f : \mathbb{R} \longrightarrow \mathbb{R},\ x \longmapsto 1$ and $g : \mathbb{R} \longrightarrow \mathbb{R},\ y \longmapsto y$. Clearly $\mathrm{im}(g \circ f) = \{1\} \neq \mathbb{R} = \mathrm{im}(g)$.

Example 72. The function $f \colon \mathbb{R} \to \mathbb{R},\ x \mapsto 2x^2 + 1$ can be written as a composition $f = j \circ h \circ g$ where g, h, j are the functions

$$\begin{aligned} g &\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto x^2 \\ h &\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad y \longmapsto 2y \\ j &\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad z \longmapsto z + 1 \end{aligned}$$

so that

$$f(x) = j\Big(h\big(g(x)\big)\Big) = j\big(h(x^2)\big) = j(2x^2) = 2x^2 + 1.$$

Observation 73. The composition in Example 72 describes, step-by-step, how we actually evaluate the function. This illustrates one of the important practical applications of the composition of functions: we decompose complicated functions as the composition of simpler ones in order to facilitate computation.

Even more importantly, we shall use the decomposition of functions into compositions for theoretical purposes. This will allow us to provide techniques and formulæ for calculating limits and derivatives.

It also provides explanations for common constructions, such as the restriction of functions.

Given $A \subseteq X$ and a function $f : X \to A$, the restriction $f\mid_A : A \to Y$ is, in fact, the composition of f and the inclusion of A into X:

$$f\mid_A = f \circ i_A^X.$$

Observation 74. It is common to compose functions f and g even when $\operatorname{codom}(f) \neq \operatorname{dom}(g)$, as long as $\operatorname{im}(f) \subseteq \operatorname{dom}(g)$. While this is not strictly correct, we can justify it within our formal framework.

Take $f\colon X \to Y$ and $g\colon A \to Z$, with $B := \operatorname{im}(f) \subseteq A \cap Y$.

We may then replace f by

$$\tilde{f} : X \longrightarrow B, \quad x \longmapsto f(x),$$

and define $g \circ f : X \to Z$ to be the composite $g \circ i_B^A \circ \tilde{f}$.

The composition of functions is *associative*.

Lemma 75. *Take functions $h\colon W \to X$, $g\colon X \to Y$ and $f\colon Y \to Z$. Then the compositions $(f \circ g) \circ h\colon W \to Z$ and $f \circ (g \circ h)\colon W \to Z$ are the same function.*

Proof. Since $\operatorname{dom}\big((f \circ g) \circ h\big) = \operatorname{dom} h = \operatorname{dom}(g \circ h) = \operatorname{dom}\big(f \circ (\circ g \circ h)\big) = W$ and since $\operatorname{codom}\big((f \circ g) \circ h\big) = \operatorname{codom}(f \circ g) = \operatorname{codom} f = \operatorname{codom}\big(f \circ (g \circ h)\big) = Z$, it only remains to show that for each $w \in W$, $\big((f \circ g) \circ h\big)(w) = \big(f \circ (g \circ h)\big)(w)$. But, for $w \in W$,

$$\begin{aligned}
\big((f \circ g) \circ h\big)(w) &:= (f \circ g)\big(h(w)\big) \\
&:= f\Big(g\big(h(w)\big)\Big) \\
&=: f\big((g \circ h)(w)\big) \\
&=: \big(f \circ (g \circ h)\big)(w)
\end{aligned}$$

□

The identity functions act as *neutral elements* with respect to composition.

Lemma 76. *Let $f : X \to Y$ be a function, then*

$$\begin{aligned}
id_Y \circ f &= f \\
f \circ id_X &= f.
\end{aligned}$$

Proof. Take $x \in X$. Then

$$\begin{aligned}
(id_Y \circ f)(x) &:= id_Y\big(f(x)\big) := f(x) \\
(f \circ id_X)(x) &:= f\big(id_X(x)\big) := f(x)
\end{aligned}$$

□

Sometimes the effect of one function can be "undone" by another, if the first assigns y to x, the second allows us to determine x from y. Composition of functions allows us to formulate this precisely.

Definition 77. The functions $f\colon X \to Y$ and $g\colon Y \to X$ are *inverse functions* if and only if $g \circ f = id_X$ and $f \circ g = id_Y$.

A function cannot have more than one inverse.

Definition 78. Let $e, g\colon Y \to X$ be functions inverse to $f\colon X \to Y$. Then $e = g$.

Proof. Since $\mathrm{dom}(e) = \mathrm{dom}(g) = Y$ and $\mathrm{codom}(e) = \mathrm{codom}(g) = X$, we only need to verify that for all $y \in Y$, $e(y) = g(y)$.

Take $y \in Y$. Then

$$\begin{aligned} e(y) &= (e \circ (id_Y)(y) && \text{by Lemma 76} \\ &= \big(e \circ (f \circ g)\big)(y) && \text{as } f \circ g = id_Y \\ &= \big((e \circ f) \circ g\big)(y) && \text{as composition of functions is associative} \\ &= (id_X \circ g)(y) && \text{as } e \circ f = id_X \\ &= g(y) && \text{by Lemma 76} \end{aligned}$$

□

Observation 79. The fact that a function, $f\colon X \to Y$, cannot have more than one inverse justifies the notation f^{-1} usually used to denote the function $Y \to X$ inverse to f, for it is uniquely determined by f whenever f is invertible.

Finally, we introduce some important properties of functions.

Definition 80. The function $f : X \longrightarrow Y$ is said to be

(i) *1–1* or *injective* or *mono* if and only if for all x, x' it follows from $f(x) = f(x')$ that $x = x'$;

(ii) *onto* or *surjective* or *epi* if and only if given any $y \in Y$ there is an $x \in X$ with $f(x) = y$ — in other words $\mathrm{im}(f) = \mathrm{codom}(f)$;

(iii) *1–1 and onto* or *bijective* or *iso* if and only if it is both 1–1 and onto.

Thus a function is injective if and only if it distinguishes different elements of its domain: different elements of its domain are mapped to different elements of its codomain.

Similarly, a function is surjective if and only if its range coincides with its codomain.

Example 81. We write $\mathbb{R}_0^+$ for $\{x \in \mathbb{R} \mid x \geq 0\}$.

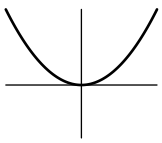
(i) $f : \mathbb{R} \to \mathbb{R},\ x \mapsto x^2$ is neither injective nor surjective, as $f(1) = f(-1)$ and there is no $x \in \mathbb{R}$ with $f(x) = -4$.

(ii) $g : \mathbb{R} \to \mathbb{R}_0^+,\ x \mapsto x^2$ is not injective, but it is surjective, as $f(1) = f(-1)$ and every non-negative real number can be written as the square of a real number.

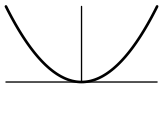
(iii) $h : \mathbb{R}_0^+ \to \mathbb{R},\ x \mapsto x^2$ is injective, but it not surjective, as $f(x) = f(x')$ if and only if $x^2 = x'^2$ if and only if $x' = \pm x$ if and only if $x' = x$ as, by definition, $x, x' \geq 0$. On the other hand, there is no $x \in \mathbb{R}_0^+$ with $f(x) = -4$.

(iv) $k : \mathbb{R}_0^+ \to \mathbb{R}_0^+,\ x \mapsto x^2$ is both injective, and surjective, as should be clear from parts (ii) and (iii).

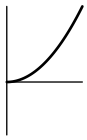
The differences between these functions is illustrated by their respective graphs



Graph of f



Graph of g



Graph of h

Graph of k

Observation 82. The notions of injectivity, surjectivity and bijectivity can also be expressed in terms of equations.

Take sets X and Y, and suppose we have a relation between elements of X and elements of Y, which we express by writing

$$y = f(x)$$

whenever $y \in Y$ is related to $x \in X$.

Then this f is a function if and only if for each $x \in X$, the equation $y = f(x)$ has one and only one solution $y \in Y$.

If we restrict attention to relations which are functions, then the function f is injective if and only if for each $y \in Y$, the equation $y = f(x)$ has *at most one solution* $x \in X$, and it is surjective if and only if for each $y \in Y$, the equation $y = f(x)$ has at least one solution $x \in X$.

This formulation in terms of equations suggests that a function has an inverse if and only if it is bijective (1 –1 and onto). This is indeed the case.

Theorem 83. *The function $f\colon X \to Y$ has an inverse if and only if it is bijective.*

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

Suppose that $g\colon Y \to X$ is inverse to f, so that $g \circ f = id_X$ and $f \circ g = id_Y$.

To see that f must be surjective, take $y \in Y$. Put x := g(y). Then

$$\begin{aligned} y &= id_Y(y) \\ &= (f \circ g)(y) && \text{as } g \text{ is the inverse function of } f \\ &= f\big(g(y)\big) && \text{by the definition of composition} \\ &= f(x) && \text{where } x := g(y) \in X, \text{ since } g : Y \to X \text{ is a function.} \end{aligned}$$

To see that f must be injective, take $x, u \in X$, with $f(x) = f(u)$. Then

$$\begin{aligned} u &= id_X(u) \\ &= (g \circ f)(u) && \text{as } g \text{ is the inverse function of } f \\ &= g\big(f(u)\big) && \text{by the definition of composition} \\ &= g\big(f(x)\big) && \text{as } f(u) = f(x) \\ &= (g \circ f)x) && \text{by the definition of composition} \\ &= x \end{aligned}$$

For the converse, suppose that f is bijective.

Define $g\colon Y \to X$ by $g\colon y \mapsto x$ if and only if $f\colon x \mapsto y$.

It follows immediately that $f\big(g(y)\big) = y$ for every $y \in Y$ and $g\big(f(x)\big) = x$ for every $x \in X$.

Thus, g is the inverse of f as long is g is, in fact, a function.

Since the domain of g is Y and its co-domain is X, and both are sets, it only remains to verify that g assigns to each $y \in Y$ a uniquely determined $x \in X$.

But f is bijective. The surjectiveness of f ensures that for each $y \in Y$, there is some $x \in X$ with $f\colon x \mapsto y$, and the inectiveness shows there cannot be two such elements in X. □

We now consider the case where the set, Y, is ordered, by say $\leq$.

Definition 84. The function $: X \to Y$ is *bounded above* if and only if there is a $K \in Y$ such that for all $x \in X$, $f(x) \leq K$.

f is *bounded below* if and only if there is a $B \in Y$ such that for all $x \in X$, $B \leq f(x)$.

f is *bounded* if and only if it is both bounded below and bounded above.

If the set X is also ordered, by say, $\preceq$, we can ask whether the function $f\colon X \to Y$ respects the orders.

Definition 85. The function $f\colon X \to Y$ is

(i) *(monotonically) non-decreasing* if and only if for all $a, b \in X$ $f(a) \leq f(b)$ whenever $a \preceq b$,

(ii) *(monotonically) increasing* if and only if for all $a, b \in X$ $f(a) < f(b)$ whenever $a \prec b$,

(iii) *(monotonically) non-increasing* if and only if for all $a, b \in X$ $f(b) \leq f(a)$ whenever $a \preceq b$,

(iv) *(monotonically) decreasing* if and only if for all $a, b \in X$ $f(b) < f(a)$ whenever $a \prec b$,

(v) *monotonic* if and only if it satisfies one of the four previous conditions and *strictly monotonic* if and only if either (ii) or (iv) holds.

Example 86. Take $\mathbb{R}$ with its usual ordering.

(i) The function

$$f\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases}$$

is monotonically non-decreasing, without being increasing.

(ii) The function

$$f : \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto -x$$

is monotonically decreasing.

(iii) The function

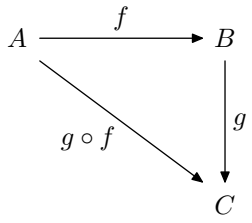
$$f \colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto x^2$$

is not monotonic because $f(-1) > f(0)$ and $f(0) < f(1)$.

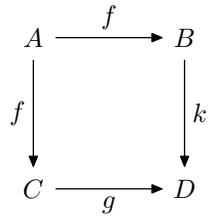
Lemma 87. *Every strictly monotonic function defined on a totally ordered set is injective..*

Proof. Suppose that X is totally ordered and that $f \colon X \to Y$ is monotonically increasing.
Take $a, b \in X, a \neq b$.
Then either $a < b$ or $b < a$.
In the former case, $f(a) < f(b)$ and in the latter $f(b) < f(a)$.
Thus, in both cases $f(a) \neq f(b)$.
The case when f is decreasing can be handled similarly. □

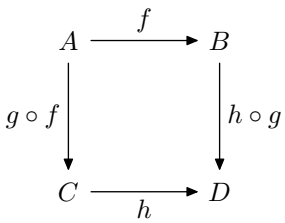
We often depict functions using diagrams. We say that the diagram



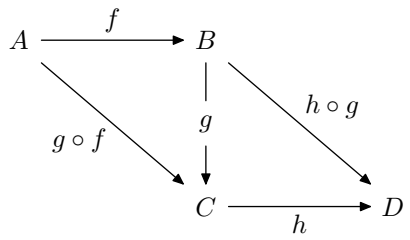
commutes when $h = g \circ f$, in other words, when the composition $g \circ f$ coincides with h. Similarly, the diagram



commutes when $k \circ j = g \circ f$, in other words, when the compositions $g \circ f$ and $k \circ j$ coincide. We can express the fact that the composition of functions is associative — that is to say, if $f : W \to X$, $g : X \to Y$ and $h : Y \to Z$ are functions then $(h \circ g) \circ f = h \circ (g \circ h)$ — by means of commutative diagrams, namely by stating that



commutes, or, equivalently, that the following diagram commutes.



2.5 Exercises

2.5.1. Let A, B and C be sets. Show that

(i) $A \cap B \subseteq A$ and $A \cap B \subseteq B$

(ii) If $C \subseteq A$ and $C \subseteq B$, then $C \subseteq A \cap B$

(iii) $A \subseteq A \cup B$ and $B \subseteq A \cup B$

(iv) If $A \subseteq C$ and $B \subseteq C$, then $A \cup B \subseteq C$.

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

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

(vii) $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$

(viii) $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$

2.5.2. Let A and B be sets. Show that the following are equivalent.

(i) $A \subseteq B$

(ii) $A \cap B = A$

(iii) $A \cup B = B$

2.5.3. Determine which of the following sets

(a) is bounded below,

(b) has an infimum (greatest lower bound),

(c) has a minimum,

(d) is bounded above,

(e) has a supremum (least upper bound),

(f) has a maximum.

(i) $\{x \in \mathbb{Z} \mid x \leq \sqrt{2}\}$

(ii) $\{x \in \mathbb{Q} \mid x \leq \sqrt{2}\}$

(iii) $\{x \in \mathbb{R} \mid x \leq \sqrt{2}\}$

(iv) $\{x \in \mathbb{R} \mid x \leq \sqrt{2} \text{ and } x \in \mathbb{Q}\}$

(v) $\{x \in \mathbb{Z} \mid x^2 < 2\}$

(vi) $\{x \in \mathbb{Q} \mid x^2 < 2\}$

(vii) $\{x \in \mathbb{R} \mid x^2 < 2\}$

(viii) $]0, 1] := \{x \in \mathbb{R} \mid 0 < x \leq 1]$.

Explain your answer.

2.5.4. Let a be a real number. Show that the equation $x^2 = a$ has

(i) no real solution if $a < 0$,

(ii) one real solution if $a = 0$.

(iii) two real solutions if $a > 0$.