Chapter 1

Numbers, sets, and functions

1.1. The natural numbers, integers, and rational numbers

We assume that you are familiar with the set of natural numbers

$$\mathbb{N} = \{1, 2, 3, \dots\},$$

the set of integers

$$\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\},$$

and the set of rational numbers

$$\mathbb{Q} = \{p/q : p, q \in \mathbb{Z}, q \neq 0\}.$$

We also assume that you are familiar with the important method of proof known as the *principle of induction*. It says that if we have a property $P(n)$ that each natural number n may or may not have, such that:

(a) $P(1)$ is true, and

(b) if $k \in \mathbb{N}$ and $P(k)$ is true, it follows that $P(k+1)$ is true,

then $P(n)$ is true for all $n \in \mathbb{N}$. There is another way to state the principle of induction that shows it to be a fundamental property of the natural numbers.

1.1. Theorem. The following are equivalent.

(1) The principle of induction.

(2) Every nonempty subset of $\mathbb{N}$ has a smallest element.

Property (2) is called the *well-ordering property* of $\mathbb{N}$. We say that $\mathbb{N}$ is *well ordered*.

Proof. To show that the two statements are equivalent, we must prove that each implies the other.

(1) $\Rightarrow$ (2): Let S be a subset of $\mathbb{N}$ with no smallest element. Let $P(n)$ be the property that $k \notin S$ for all $k \leq n$. Since S has no smallest element, $1 \notin S$, so $P(1)$ is true. Also, if $P(n)$ is true, $P(n+1)$ must be true as well, for otherwise $n+1$ would be the smallest element of S. Thus $P(n)$ satisfies (a) and (b), so by assumption, $P(n)$ holds for all $n \in \mathbb{N}$ and S is empty.

(2) $\Rightarrow$ (1): Let $P(n)$ be a property of natural numbers satisfying (a) and (b). Define S to be the set of those $n \in \mathbb{N}$ for which $P(n)$ is false. Then (a) says that $1 \notin S$, and (b) (or rather its contrapositive) says that if $k \in S$, $k > 1$, then $k - 1 \in S$. Therefore S has no smallest element, so by assumption S must be empty, which means that $P(n)$ is true for all $n \in \mathbb{N}$. □

1.2. Remark. The *contrapositive* of an implication $P \Rightarrow Q$ is the implication not-$Q \Rightarrow$ not-P. These two implications are logically equivalent. Thus, if we want to prove that P implies Q, then we can instead prove that not-Q implies not-P. This is sometimes convenient. Do not confuse the contrapositive with the *converse* of $P \Rightarrow Q$, which is the implication $Q \Rightarrow P$. An implication and its converse are in general *not* equivalent.

We can think of $\mathbb{Z}$ as an extension of $\mathbb{N}$ that allows us to do subtraction without any restrictions, and of $\mathbb{Q}$ as an extension of $\mathbb{Z}$ that allows us to do division with the sole restriction that division by zero cannot be reasonably defined. The set $\mathbb{Q}$ with addition and multiplication and all the familiar rules satisfied by these operations is a mathematical structure called a *field*.

1.3. Definition. A *field* is a set F with two operations, *addition*, denoted $+$, and *multiplication*, denoted $\cdot$, such that the following axioms are satisfied.

A1 *Associativity:* $a + (b + c) = (a + b) + c$, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for all $a, b, c \in F$.

A2 *Commutativity:* $a + b = b + a$, $a \cdot b = b \cdot a$ for all $a, b \in F$.

A3 *Distributivity:* $a \cdot (b + c) = a \cdot b + a \cdot c$ for all $a, b, c \in F$.

A4 *Additive identity.* There is an element called 0 in F such that $a + 0 = a$ for all $a \in F$.
Multiplicative identity. There is an element called 1 in F such that $1 \neq 0$ and $a \cdot 1 = a$ for all $a \in F$.

A5 *Additive inverses.* For every $a \in F$, there is an element called $-a$ in F such that $a + (-a) = 0$.
Multiplicative inverses. For every $a \in F$, $a \neq 0$, there is an element called a^{-1} in F such that $a \cdot a^{-1} = 1$.

We usually write $a \cdot b$ as ab, $a + (-b)$ as $a - b$, a^{-1} as $1/a$, and ab^{-1} as a/b.

From the field axioms we can derive many familiar properties of fields. It is a good exercise to work out careful proofs of some of these properties based only on the axioms. Here are a few examples. If you prefer, you can simply take $F = \mathbb{Q}$.

1.4. Example. From A2 and A4 we see that $0 + a = a$ and $1 \cdot a = a$ for all $a \in F$.

1.5. Example. The additive identity 0 is unique. Namely, assume $0'$ is another additive identity. By A4, $a + 0 = a$ for all $a \in F$. In particular, taking $a = 0'$, we see that $0' + 0 = 0'$, so by A2, $0 + 0' = 0'$. On the other hand, by assumption, $a + 0' = a$ for all $a \in F$, so taking $a = 0$, we see that $0 + 0' = 0$. We conclude that $0' = 0 + 0' = 0$. Similarly, the multiplicative identity is unique.

Exercise 1.1. Using only the axioms A1–A5, show that the additive inverse of $x \in F$ is unique, that is, if $x + y = 0$ and $x + z = 0$, then $y = z$ (so talking about *the* additive inverse of x is justified). Show also that the multiplicative inverse of $x \in F$, $x \neq 0$, is unique.

1.6. Example. From A2 and A5 we see that for $x \in F$, $(-x) + x = 0$. By Exercise 1.1, we conclude that the additive inverse of $-x$ must be x, that is, $-(-x) = x$. Similarly, for $x \neq 0$, $(x^{-1})^{-1} = x$.

1.7. Example. For every $x \in F$,

$$0 \cdot x \overset{\text{A4}}{=} (0 + 0) \cdot x \overset{\text{A2, A3}}{=} 0 \cdot x + 0 \cdot x.$$

Adding the additive inverse $-(0 \cdot x)$ of $0 \cdot x$ to both sides, we get $0 = 0 \cdot x$. By A2, $x \cdot 0 = 0$ as well.

Exercise 1.2. In A5, $-x$ was introduced as a symbol for the additive inverse of $x \in F$. Using Example 1.7, show that $-x$ is in fact the product of x and the additive inverse -1 of the multiplicative identity 1. In particular,

$$(-1)(-1) = -(-1) = 1.$$

If $x \in F$ and $n \in \mathbb{N}$, $n \geq 2$, we write x^n for the product of n factors of x. By A1, it does not matter how we bracket the product. For example, $x^3 = (x \cdot x) \cdot x = x \cdot (x \cdot x)$. We set $x^0 = 1$ and $x^1 = x$. If $x \neq 0$, we write x^{-n} for $(x^{-1})^n$, which equals $(x^n)^{-1}$. Then $x^{m+n} = x^m x^n$ and $(x^m)^n = x^{mn}$ for all $m, n \in \mathbb{Z}$.

There is more to the rationals than addition and multiplication. The rationals are also *ordered* in a way that interacts well with addition and multiplication. This structure is called an *ordered field*.

1.8. Definition. An *ordered field* is a field F with a relation $<$ (read 'less than') such that the following axioms are satisfied.

A6 For every $a, b \in F$, precisely one of the following holds: $a < b$, $b < a$, or $a = b$.

A7 If $a < b$ and $b < c$, then $a < c$ (the order relation is *transitive*).

A8 If $a < b$, then $a + c < b + c$ for all $c \in F$.

A9 If $a < b$ and $0 < c$, then $ac < bc$.

We take $a \leq b$ to mean that $a < b$ or $a = b$; $a > b$ to mean that $b < a$; and $a \geq b$ to mean that $b \leq a$. We say that a is *positive* if $a > 0$, and *negative* if $a < 0$.

Again, the axioms imply many further properties.

1.9. Example. We claim that 1 is positive. Note that if $1 < 0$, then adding -1 to both sides gives $0 < -1$ by A8, so multiplying both sides by -1 gives $0 = 0(-1) < (-1)(-1) = 1$ by A9, Example 1.7, and Exercise 1.2, but having both $1 < 0$ and $0 < 1$ contradicts A6.

Having derived a contradiction from the assumption that $1 < 0$, we must reject the assumption as false. Since $0 \neq 1$ by A4, the one remaining possibility by A6 is $0 < 1$.

Exercise 1.3. (a) Show that if $x > 0$, then $-x < 0$. Likewise, if $x < 0$, then $-x > 0$. In particular, by Example 1.9, $-1 < 0$.

(b) Show that if $x > 0$, then $x^{-1} > 0$. Show that if $x > 1$, then $x^{-1} < 1$.

1.10. Definition. An *interval* in an ordered field F is a subset of F of one of the following types, where $a, b \in F$.

$$\begin{aligned}
(a, b) &= \{x : a < x < b\} \\
[a, b] &= \{x : a \leq x \leq b\} \\
(a, b] &= \{x : a < x \leq b\} \\
[a, b) &= \{x : a \leq x < b\} \\
(a, \infty) &= \{x : x > a\} \\
(-\infty, a) &= \{x : x < a\} \\
[a, \infty) &= \{x : x \geq a\} \\
(-\infty, a] &= \{x : x \leq a\} \\
(-\infty, \infty) &= F
\end{aligned}$$

The intervals (a, b), (a, ∞), $(-\infty, a)$, and F itself are said to be *open*. The intervals $[a, b]$, $[a, \infty)$, $(-\infty, a]$, and F itself are said to be *closed*. Taking $a > b$, we see that the empty set is an interval which is both open and closed. One-point sets $[a, a]$ and the empty set are called *degenerate* intervals. Thus an interval is *nondegenerate* if it contains at least two points.

Exercise 1.4. Show that a nondegenerate interval contains infinitely many points.

1.11. Remark. By A7, if I is an interval, $x < y < z$, and $x, z \in I$, then $y \in I$. In other words, along with any two of its points, an interval contains all the points in between. Conversely, when F is the field of real numbers, a set satisfying this property is an interval (Exercise 2.12).

1.12. Definition. If a and b are elements of an ordered field and $a \leq b$, then we write $\min\{a, b\} = a$ for the *minimum* of a and b, and $\max\{a, b\} = b$ for the *maximum*.

1.13. Definition. The *absolute value* of an element a in an ordered field is the nonnegative element

$$|a| = \max\{a, -a\} = \begin{cases} a & \text{if } a \geq 0, \\ -a & \text{if } a < 0. \end{cases}$$

1.14. Theorem (triangle inequality)**.** For all elements a and b in an ordered field,

$$|a + b| \leq |a| + |b|.$$

For all elements x, y, z in an ordered field,

$$|x - z| \leq |x - y| + |y - z|.$$

Proof. Three cases need to be considered: $a, b \geq 0$; $a \geq 0$ and $b < 0$ (the case when $a < 0$ and $b \geq 0$ is analogous and does not need to be written out in detail); and $a, b < 0$. Let us treat the second case and leave the others as an exercise.

Since $a \geq 0$, we have $-a \leq 0 \leq a$, so, adding $-b$, we get $-(a + b) \leq a - b = |a| + |b|$. Since $b < 0$, we have $b < 0 < -b$, so, adding a, we get $a + b < a - b = |a| + |b|$. These two inequalities together give

$$|a + b| = \max\{a + b, -(a + b)\} \leq |a| + |b|.$$

To get the second inequality, take $a = x - y$ and $b = y - z$. □

Although the rational numbers have a rich structure, they suffer from limitations that call for a larger number system. The following result is attributed to Pythagoras and his associates some 2500 years ago.

1.15. Theorem. There is no rational number with square 2.

Proof. Suppose there are $p, q \in \mathbb{N}$ with $(p/q)^2 = 2$. Choose q to be as small as possible. Now $q < p < 2q$, so $0 < p - q < q$ and $2q - p > 0$. It is easily computed that $\left(\dfrac{2q - p}{p - q}\right)^2 = 2$, contradicting the minimality of q. □

1.16. Remark. Theorem 1.15 has many different proofs. Here is another one. Suppose there was $r \in \mathbb{Q}$ with $r^2 = 2$. We can write $r = p/q$, where p and q are integers with no common factors. We will derive a contradiction from this assumption.

Now $2 = r^2 = p^2/q^2$, so $p^2 = 2q^2$ and p^2 is even. Hence p is even, say $p = 2k$, where k is an integer. Then $2q^2 = p^2 = (2k)^2 = 4k^2$, so $q^2 = 2k^2$ and q^2 is even. Hence q is even, so p and q are both divisible by 2, contrary to our assumption.

Exercise 1.5. Show that there is no rational number with square 3 by modifying the proof of Theorem 1.15 given in Remark 1.16. Where does the proof fail if you try to carry it out for 4? For which $n \in \mathbb{N}$ can you show by the same method that there is no rational number with square n?

This deficiency of $\mathbb{Q}$ leads us to a larger and more sophisticated number system. The real number system has a crucial property called *completeness* which implies, among many other consequences, that every positive real number has a real square root.

A small amount of set theory is essential for real analysis, so before turning to the real numbers we will review some basic concepts to do with sets and functions.

1.2. Sets

The notion of a *set* is a (many would say *the*) fundamental concept of modern mathematics. It cannot be defined in terms of anything more fundamental. Rather, the notion of a set is circumscribed by axioms (usually the so-called *Zermelo-Fraenkel axioms* along with the *axiom of choice*) from which virtually all of mathematics can be derived, at least in principle.

Our approach will be informal. We think of a set as any collection of objects. The objects are called the *elements* of the set. If x is an element of a set A, we write $x \in A$. A set is determined by its elements, that is, two sets are the same if and only if they have the same elements. Thus the most common way to show that sets A and B are equal is to prove, first, that if $x \in A$, then $x \in B$, and second, that if $x \in B$, then $x \in A$.

1.17. Definition. Let A and B be sets. We say that A is a *subset* of B and write $A \subset B$ (some write $A \subseteq B$) if every element of A is also an element of B. We say that A is a *proper subset* of B if $A \subset B$ and $A \neq B$. The *union* of A and B is the set

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

The *intersection* of A and B is the set

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

We say that A and B are *disjoint* if they have no elements in common. The *complement* of A in B is the set

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

Sometimes $B \setminus A$ is written as $B - A$, or as A^c if B is understood.

1.18. Remark. In mathematics, the conjunction *or* (as in the definition of the union $A \cup B$) is always understood in the inclusive sense: 'p or q' always means 'p or q *or both*'. If we want the exclusive *or*, then we must say so explicitly by adding the phrase 'but not both'.

1.19. Remark. The operations on sets in Definition 1.17 satisfy various identities reminiscent of the laws of arithmetic. There are the associative laws

$$A \cup (B \cup C) = (A \cup B) \cup C, \qquad A \cap (B \cap C) = (A \cap B) \cap C,$$

the commutative laws

$$A \cup B = B \cup A, \qquad A \cap B = B \cap A,$$

the distributive laws

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \qquad A \cup (B \cap C) = (A \cup B) \cap (A \cup C),$$

and *De Morgan's laws*

$$A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C), \qquad A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C).$$

Let us prove the second De Morgan's law. There are two implications to be proved: first, the implication that if $x \in A \setminus (B \cap C)$, then $x \in (A \setminus B) \cup (A \setminus C)$, and second, the converse implication. So suppose that $x \in A \setminus (B \cap C)$. This means that $x \in A$ but $x \notin B \cap C$. Now $x \notin B \cap C$ means that $x \notin B$ *or* $x \notin C$, so we conclude that either $x \in A$ and $x \notin B$, or $x \in A$ and $x \notin C$ (*either ... or* is still the inclusive *or*). Hence $x \in A \setminus B$ or $x \in A \setminus C$, that is, $x \in (A \setminus B) \cup (A \setminus C)$. We leave the converse implication to you.

Note that this proof required three things:

- knowing how to prove that two sets are equal,
- unravelling the definitions of the sets $A \setminus (B \cap C)$ and $(A \setminus B) \cup (A \setminus C)$,
- being able to *negate* the statement $x \in B \cap C$, that is, realising that $x \notin B \cap C$ means that $x \notin B$ or $x \notin C$.

1.20. Definition. The *empty set* is *the* set with no elements, denoted $\varnothing$.

1.21. Remark. To say that A is a subset of B is to say that if $x \in A$, then $x \in B$. Hence, to say that A is *not* a subset of B is to say that there is $x \in A$ with $x \notin B$. It follows that the empty set $\varnothing$ is a subset of *every* set B. Otherwise, there would be an element $x \in \varnothing$ with $x \notin B$, but $\varnothing$ has no elements at all.

Exercise 1.6. Prove that if $A \subset B$, then $A \setminus B = \varnothing$.

We can take unions and intersections not just of two sets, but of arbitrary collections of sets.

1.22. Definition. Let $(A_i)_{i \in I}$ be a *family* of sets, that is, we have a set I (called an *index set*), and associated to every $i \in I$, we have a set called A_i. The *union* of the family is the set

$$\bigcup_{i \in I} A_i = \{x : x \in A_i \text{ for some } i \in I\}.$$

The *intersection* of the family is the set

$$\bigcap_{i \in I} A_i = \{x : x \in A_i \text{ for all } i \in I\}.$$

1.23. Example. Define a family $(A_n)_{n \in \mathbb{N}}$ of sets by setting $A_1 = \mathbb{N}$, $A_2 = \{2, 3, 4, \dots\}$, $A_3 = \{3, 4, 5, \dots\}$, and so on, that is, $A_n = \{n, n+1, n+2, \dots\}$ for each $n \in \mathbb{N}$. Then $A_1 \supset A_2 \supset A_3 \supset \cdots$, so we have

$$\bigcup_{n \in \mathbb{N}} A_n = A_1 \cup A_2 \cup A_3 \cup \cdots = A_1 = \mathbb{N}.$$

Also,

$$\bigcap_{n \in \mathbb{N}} A_n = A_1 \cap A_2 \cap A_3 \cap \cdots = \varnothing,$$

because there is no natural number that belongs to A_n for all $n \in \mathbb{N}$. Indeed, if $k \in A_1 = \mathbb{N}$, then $k \notin A_{k+1}$.

1.24. Definition. The *product* of sets A and B, denoted $A \times B$, is the set of all *ordered pairs* (a, b) with $a \in A$ and $b \in B$.

What is an ordered pair, you may ask. All you need to know is that $(a_1, b_1) = (a_2, b_2)$ if and only if $a_1 = a_2$ and $b_1 = b_2$. But you may be interested to also know that we do not need to take an ordered pair as a new fundamental notion. If we define (a, b) to be the set $\{\{a\}, \{a, b\}\}$, then we can prove that $(a_1, b_1) = (a_2, b_2)$ if and only if $a_1 = a_2$ and $b_1 = b_2$.

It is unfortunate that the same notation is used for an ordered pair and an open interval, but the intended meaning should always be clear from the context.

1.3. Functions

1.25. Definition. A *function* (or a *map* or a *mapping*—these are synonyms) f consists of three things:

- a set A called the *source* or *domain* of f,
- a set B called the *target* or *codomain* of f,
- a *rule* that assigns to each element x of A a unique element of B. This element is called the *image* of x by f or the *value* of f at x, and denoted $f(x)$.

We write $f : A \to B$ to indicate that f is a function with source A and target B, that is, a function *from A to B*.

1.26. Remark. Note that the source and the target of the function must be specified for the function to be well defined. Also, the rule does not have to be a formula. Any unambiguous description will do.

1.27. Definition. The *identity function* of a set A is the function $\mathrm{id}_A : A \to A$ with $\mathrm{id}_A(x) = x$ for all $x \in A$.

1.28. Definition. Let $f : A \to B'$ and $g : B \to C$ be functions such that $B' \subset B$. The *composition* of f and g is the function $g \circ f : A \to C$ with $(g \circ f)(x) = g(f(x))$ for all $x \in A$ ('first apply f, then g').

1.29. Definition. Let $f : A \to B$ be a function. The *image* by f of a subset $C \subset A$ is the subset

$$f(C) = \{f(x) : x \in C\}$$

of B. The *image* or *range* of f is the set $f(A)$. The *preimage* or *inverse image* by f of a subset $D \subset B$ is the subset

$$f^{-1}(D) = \{x \in A : f(x) \in D\}$$

of A, that is, the set of elements of A that f maps into D. If D consists of only one element, say $D = \{y\}$ for some $y \in B$, then, for simplicity, we write $f^{-1}(y)$ for $f^{-1}(\{y\})$, and call $f^{-1}(y)$ the *fibre* of f over y.

1.30. Example. Assuming for the purposes of this example that we know about the real numbers, consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by the formula $f(x) = x^2$. Instead of $f(x) = x^2$, we can write $f : x \mapsto x^2$ (the arrow $\mapsto$ is read 'maps to'). The range of f consists of all the nonnegative real numbers, that is, $f(\mathbb{R}) = [0, \infty)$. We have

$$f^{-1}(0) = \{0\}, \qquad f^{-1}(1) = \{1, -1\}, \qquad f^{-1}(\{1, 4\}) = \{1, -1, 2, -2\}.$$

The function $g : \mathbb{R} \to [0, \infty)$, $x \mapsto x^2$, is not the same function as f because its target is different. And the function $h : [0, \infty) \to [0, \infty)$, $x \mapsto x^2$, is different still, because its source is different. All three functions are defined by the same formula and have the same range $[0, \infty)$.

Images and preimages interact with unions, intersections, and complements to a certain extent. Note that preimages are better behaved than images.

1.31. Theorem. Let $f : A \to B$ be a function. For subsets $K, L \subset A$ and $M, N \subset B$, the following hold.

(1) $f(K \cup L) = f(K) \cup f(L)$.
(2) $f^{-1}(M \cup N) = f^{-1}(M) \cup f^{-1}(N)$.
(3) $f^{-1}(M \cap N) = f^{-1}(M) \cap f^{-1}(N)$.
(4) $f^{-1}(M \setminus N) = f^{-1}(M) \setminus f^{-1}(N)$.

Proof. We shall prove (4) and leave the other parts as an exercise. Normally we prove the equality of two sets as two separate implications, but here things are simple enough that we can prove both implications at the same time. Namely, we have $x \in f^{-1}(M \setminus N)$ if and only if $f(x) \in M \setminus N$ if and only if $f(x) \in M$ and $f(x) \notin N$ if and only if $x \in f^{-1}(M)$ and $x \notin f^{-1}(N)$ if and only if $x \in f^{-1}(M) \setminus f^{-1}(N)$. □

Exercise 1.7. Finish the proof of Theorem 1.31.

1.32. Remark. It is not true in general that $f(K \cap L) = f(K) \cap f(L)$ or $f(K \setminus L) = f(K) \setminus f(L)$. For example, take f as in Example 1.30, $K = \{1\}$, and $L = \{-1\}$. Then $f(K \cap L) = f(\varnothing) = \varnothing$, but $f(K) \cap f(L) = \{1\} \cap \{1\} = \{1\}$. Also, $f(K \setminus L) = f(\{1\}) = \{1\}$, but $f(K) \setminus f(L) = \{1\} \setminus \{1\} = \varnothing$.

1.33. Definition. A function $f : A \to B$ is called:

- *injective* (or *one-to-one*) if it takes distinct elements to distinct elements, that is, if $x, y \in A$ and $f(x) = f(y)$, then $x = y$;
- *surjective* (or *onto*) if $f(A) = B$, that is, every element of B is the image by f of some element of A;
- *bijective* if f is both injective and surjective.

An injective function is also called an *injection*, a surjective function is called a *surjection*, and a bijective function is called a *bijection*.

1.34. Remark. Note that a function $f : A \to B$ is:

- injective if and only if the fibre $f^{-1}(y)$ contains *at most* one element for every $y \in B$,
- surjective if and only if the fibre $f^{-1}(y)$ contains *at least* one element for every $y \in B$,
- bijective if and only if the fibre $f^{-1}(y)$ contains *precisely* one element for every $y \in B$.