Fundamentals of Pure Mathematics

Kenneth Falconer

Martinmas Semester 2010-11

About the Course

In this course we shall mostly talk and think about numbers, working carefully from definitions. The course will have the following components:

1. Proof and Mathematical Argument

2. Sets, Relations and Functions

3. Construction and Properties of Number Systems

4. Some Number Theory

5. Countability

If you wish to read further about various topics covered in this course and related to the course, the following books may be helpful (note that the texts are for further reading and it is certainly not compulsory to buy them).

I. Stewart and D. Tall, The Foundations of Mathematics, Oxford University Press, 1977; QA107.S8T2.

M. Liebeck, A Concise Introduction to Pure Mathematics, Chapman & Hall/ CRC, 2000; QA8.4L5.

H.-D. Ebbinghaus et al., Numbers, Springer-Verlag, New York, 1991; QA241.E3E8.

All three books are available on short loan in the Mathematics and Physics Library.
The assessment for the course will be by a 2 hour exam in January.

1 Proof and Mathematical Argument

Writing mathematics clearly and carefully becomes increasingly important as you get further into the subject. Advanced mathematical arguments can have a complicated logical structure, and this structure must be clear to any reader (and even more importantly to the writer!).

In particular, written mathematics should not just be a list of equations but should make clear their logical relationship to each other. The careful use of words to express such relationships is important. For example, writing "$x \geq 1, x^2 \geq x$" is ambiguous, but "For all $x \geq 1$ we have $x^2 \geq x$" makes clear what is intended.

Styles of mathematical writing vary considerably. You should develop your own style using words as well as symbols, to make your arguments as clear as you can.

1.1 Implication

The notion of implication is fundamental in any mathematical argument. If A and B are statements then "A implies B" (in symbols A ⇒ B) means that whenever A is true B must also be true. For example:

$$x = 2 \text{ implies } x^2 = 4.$$

Implication may be indicated in a variety of ways, such as:

A implies B, A ⇒ B, B is implied by A, If A then B, B if A.

It is crucial to realise that "If A then B" and "If B then A" mean very different things. "If $x = 2$ then $x^2 = 4$" is true, but "If $x^2 = 4$ then $x = 2$" is false (x might be -2).

However, sometimes two statements A and B are each implied by the other, in which case we say "A and B are equivalent" (in symbols A ⇔ B). Ways of writing this include:

A is equivalent to B, A ⇔ B, A implies and is implied by B, A if and only if B,
A iff B, A is a necessary and sufficient condition for B.

For example, "$x^2 = 4$ if and only if $x = 2$ or $x = -2$". Note that if you are asked to show that A holds if and only if B holds you have to do *two* things.

1.2 Proof

A *proof* is a careful argument that establishes a new fact or *theorem*, given certain *assumptions* or *hypotheses*. There are various kinds of proof, some of which we mention here.

Proof by deduction
A deductive proof consists of a sequence of *statements* or *sentences* each of which is deduced from previous ones or from hypotheses using standard mathematical properties. The final statement may be called a *theorem*. For example:

Theorem 1.1. *If* $x^2 - 3x + 1 < 0$ *then* $x > 0$.

Deductive proof. Assume that $x^2 - 3x + 1 < 0$. Then $3x > x^2 + 1$ (rearranging the inequality), which implies that $3x > 1$ (since $x^2 \geq 0$). It follows that $x > \frac{1}{3}$ (dividing), so $x > 0$ (by the order property).

[Note that in this argument each step may be deduced from the previous one by a standard mathematical fact. However, the steps are not all reversible.]

Proof by contradiction
Sometimes it is easier to argue by contradiction, i.e. to assume that the desired conclusion is false and derive a contradiction to some known fact.

Theorem 1.2. *If $x^2 - 3x + 1 < 0$ then $x > 0$.*

Proof by contradiction. Assume that $x^2 - 3x + 1 < 0$ and suppose that $x \leq 0$. Then $x^2 < 3x - 1 \leq 3 \times 0 - 1$ (rearranging and using $x \leq 0$), so $x^2 < -1$, which contradicts that the square of a real number is non-negative. We conclude that $x > 0$.

Counter-examples
To show that a statement is false it is enough to give a single instance for which it does not hold, called a *counter-example*.

For example, the statement "$x^2 - 4x + 1 > 0$ for all $x > 0$" is false. To see this we simply note that $2^2 - 4 \times 2 + 1 = -3 \leq 0$, so that $x = 2$ is a counter-example to the statement. In particular, to demonstrate falsity there is no need to 'solve the inequality'.

1.3 Mathematical Induction

Mathematical induction is a more sophisticated method of deductive proof used to derive formulae and facts throughtout mathematics. Induction is a method of proving statements involving the natural numbers $1, 2, 3, \ldots$. The idea is that (i) we prove the statement when $n = 1$ and (ii) show that if the statement is true for some integer n then it is true for the integer immediately above. From this we can conclude that the statement is true for all $n = 1, 2, 3, \ldots$. Formally:

The Principle of Mathematical Induction.
Let $P(n)$ be a statement depending on an arbitrary positive integer n. Suppose that we can do the following two steps:
(i) Verify that $P(1)$ is true,
(ii) for all positive integers n, show that if $P(n)$ is true then $P(n+1)$ is true.
Then the statement $P(n)$ is true for all positive integers n.

The statement $P(n)$ is called *the inductive hypothesis*, step (i) is called *starting the induction* or *anchor* and step (ii) is called *the inductive step*.

Note that the Principle of Induction is intuitively obvious: if $P(1)$ is true and $P(n) \Rightarrow P(n+1)$, that is the truth of $P(n)$ implies the truth of $P(n+1)$ for all $n = 1, 2, 3, \ldots$, then

$$P(1) \Rightarrow P(2) \Rightarrow P(3) \Rightarrow \ldots \Rightarrow P(n) \Rightarrow \ldots$$

by applying (ii) with $n = 1, 2, 3, \ldots$ in turn, so $P(n)$ is true for all n.

Example 1.3. (Summation of series). *For every positive integer n*

$$1 + 2 + \ldots + n = \frac{n(n+1)}{2}.$$

Proof by induction. Let $P(n)$ be the statement: $1+2+\ldots+n = \frac{n(n+1)}{2}$. Then $1 = \frac{1(1+1)}{2}$, so $P(1)$ holds, which starts the induction.

Now assume that $P(n)$ is true for some positive integer n. We relate the sum in $P(n+1)$ to that in $P(n)$:

$$\begin{aligned} 1+2+\ldots+n+(n+1) &= (1+2+\ldots+n)+(n+1) \\ &= \frac{n(n+1)}{2}+(n+1) \qquad \text{(using } P(n)) \\ &= \frac{n(n+1)+2(n+1)}{2} \\ &= \frac{(n+1)(n+2)}{2} \\ &= \frac{(n+1)((n+1)+1)}{2} \end{aligned}$$

which is the statement $P(n+1)$, completing the inductive step.

Thus by the Principle of Induction $P(n)$ is true for all positive integers n.

Example 1.4. *Use induction to show that $9^n - 2^n$ is divisible by 7 for all positive integers n.*

There are many variants on the Principle of Induction. For instance we may wish to start at an integer other than 1. Thus if for some integer n_0 we can show

(i) that $P(n_0)$ is true, and

(ii) for all $n \geq n_0$ that if $P(n)$ is true then $P(n+1)$ is true,

the Principle of Induction gives that $P(n)$ is true for all $n \geq n_0$.

Sometimes we need to use that $P(k)$ is true for all $k \leq n$ to deduce $P(n+1)$. Thus if we can show

(i) that $P(1)$ is true, and

(ii) that if $P(k)$ is true for all $k \leq n$ then $P(n+1)$ is true,

then $P(n)$ is true for all $n \geq 1$., This is called *strong* or *total mathematical induction*.

Recall that an integer $n \geq 2$ is a *prime number* if is cannot be expressed as a product of integers $n = rs$ with $r > 1$ and $s > 1$.

Theorem 1.5 (The Fundamental Theorem of Arithmetic)**.** *Every natural number $n \geq 2$ is a product of prime numbers, i.e.*

$$n = p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m},$$

for some distinct primes $p_1, \ldots, p_m$ and natural numbers $k_1, \ldots, k_m$. Moreover, up to the order of the p_i, this decomposition is unique.

Proof by induction. Let $P(n)$ be the statement: n may be expressed as a product of primes. Since 2 is prime, $P(2)$ is true, which starts the induction.

Now assume that for some $n \geq 2$, $P(k)$ is true for all integers $2 \leq k \leq n$. Consider the integer $n+1$. Either $n+1$ is prime (so a product of a single prime factor) or $n+1 = rs$ where $r > 1$ and

$s > 1$. In the latter case $2 \leq r, s \leq n$, so by $P(r)$ and $P(s)$ both r and s are products of primes, so $n+1 = rs$ is a product of primes. Thus $P(n+1)$ is true, completing the inductive step.

To show uniqueness we use the fact that if the integer a divides bc and a and b have no common prime factor, then a divides c; we omit the details here.

Induction is a general method that is used in virtually every area of mathematics. Further examples include complex numbers (e.g. for a proof of de Moivre's theorem), finding powers of matrices, terms of sequences, in number theory, graph theory, group theory, mathematical logic, ...

2 Sets, Relations and Functions

2.1 Sets

Today all of Mathematics is expressed in the language of sets – in a certain sense, the concept of a set is even more fundamental than that of a number.

For the main part of the course, we will simply take a *set* to be a collection of object which we term the *members* or *elements* of the set. Generally we use capital letters to denote sets and small letters to denote elements.

We write $a \in A$ to mean that an element a belongs to a set A and $a \notin A$ to mean it does not belong to A.

We sometimes write $\{\cdots\}$ for the set containing the elements $\cdots$.

Here are some of the standard sets used in mathematics:

$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ the *Integers*

$\mathbb{Q} = \{p/q$ such that $p, q \in \mathbb{Z}, q \neq 0\}$ the *Rationals*

$\mathbb{R} =$ the *Real Numbers*

$\mathbb{C} =$ the *Complex Numbers*

$\mathbb{N} = \mathbb{Z}^+ = \{1, 2, 3, \ldots\}$ the *Natural Numbers* or *Positive Integers*

$\mathbb{Q}^+ =$ the *Positive Rationals*, etc.

$\varnothing$ the *Empty Set* or *Null Set*, i.e. the set with no elements.

We write

$$\{x : \text{ Property involving } x\} \quad \text{or} \quad \{x | \text{ Property involving } x\}$$

to mean "the set of x such that the Property involving x holds".

For a common example:

$$[a,b] = \{x : a \leq x \leq b\} \quad \text{and} \quad (a,b) = \{x : a < x < b\}$$

denote the *closed* and *open intervals* of real numbers between a and b.

Two sets are said to be equal ($A = B$) if they consist of precisely the same elements. In other words $A = B$ if and only if every element of A is also an element of B and every element of B is an element of A. We say that A is a *subset* of B ($A \subseteq B$) if every element of A is an element of B (but the converse does not necessarily hold). Clearly

$$A = B \Longleftrightarrow A \subseteq B \ \& \ B \subseteq A.$$

Note that $2 \in \mathbb{Z}$ but $\{2\} \subseteq \mathbb{Z}$, etc.

If $A \subseteq B$ and $A \neq B$ we say A is a *proper* subset of B; note that some books distinguish this by writing $A \subset B$.

Often we work within a *universal set* U and confine attention to subsets of this, e.g. $\mathbb{Z}$ might be the universal set when working with the theory of prime numbers.

2.2 Set Operations

Here are some basic operations on sets:

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

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

Set difference $A \setminus B = \{x : x \in A \text{ and } x \notin B\}$;

Complement $A^c = \{x : x \in U \text{ and } x \notin A\}$ where U is the universal set.

More generally, if we have a collection of sets A_i ($i \in I$) indexed by another set I, we can form their intersection and union:

$$\bigcap_{i \in I} A_i = \{x : x \in A_i \text{ for all } i \in I\};$$
$$\bigcup_{i \in I} A_i = \{x : \text{ there exists } i \in I \text{ such that } x \in A_i)\}.$$

[Note that the symbol $\forall$ is often used as an abbreviation for ‘for all’; $\exists$ should be read ‘there exists’ or ‘for some’.]

Example 2.1. *For $i = 1, 2, 3, \ldots$ let $A_i = \{-i, -i+1, \ldots, i-1, i\} \subseteq \mathbb{Z}$. Determine $\bigcap_{i \in \mathbb{N}} A_i$ and $\bigcup_{i \in \mathbb{N}} A_i$.*

There are various standard set theoretic identities. These may be verified either by showing that every element in each side of the identity is also in the other side, or sometimes by using Venn diagrams.

Example 2.2. (The distributive laws for intersection and union) *For all sets A, B, C:*

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

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

More generally, for indexed unions and intersections:

$$A \cap (\bigcup_{i \in I} B_i) = \bigcup_{i \in I} (A \cap B_i),$$

$$A \cup (\bigcap_{i \in I} B_i) = \bigcap_{i \in I} (A \cup B_i).$$

Example 2.3. (De Morgan's laws) *For indexed unions and intersections:*

$$\big(\bigcup_{i \in I} A_i\big)^c = \bigcap_{i \in I} A_i^c,$$

$$\big(\bigcap_{i \in I} A_i\big)^c = \bigcup_{i \in I} A_i^c,$$

Power sets

The *power set* of A, written $\mathcal{P}(A)$, is the set of all subsets of A, that is $\mathcal{P}(A) = \{Y : Y \subseteq X\}$. For example, $\mathcal{P}(\{1,2\}) = \big\{\varnothing, \{1\}, \{2\}, \{1,2\}\big\}$.

Cartesian products

An *ordered pair* (a,b) should be thought of as a pair of elements in which the order of elements matters (and repetitions are allowed), so that $(a,b) \neq (b,a)$ unless $a = b$. For two sets A and B the set of ordered pairs

$$A \times B = \{(a,b) : a \in A,\ b \in B\}$$

is called their *Cartesian product* or *direct product*.

Example 2.4.

(1) *If* $A = \{1,2\}$ *and* $B = \{1,2,3\}$ *then*

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

(2) $\quad \mathbb{R} \times \mathbb{R} = \{(x,y) : x \in \mathbb{R}, y \in \mathbb{R}\} = \mathbb{R}^2 =$ the Euclidean plane.

2.3 Relations

Formally, a *relation* or *binary relation* R on a set X is any subset of the ordered pairs $X \times X$. However, instead of writing $(x,y) \in R$ we almost invariably write xRy, read as 'x is related to y'.

Example 2.5. $>$ *is a relation on* $\{1,2,3\}$*, formally given by* $\{(2,1),(3,1),(3,2)\}$*, that is* $2 > 1, 3 > 1$*, and* $3 > 2$*.*

Example 2.6. *The following are relations on $\{$ all people $\}$:*

$$\begin{array}{lll} xRy & \textit{if} & x \textit{ is a brother of } y \\ xRy & \textit{if} & x \textit{ is taller than } y \\ xRy & \textit{if} & x \textit{ is younger than } y. \end{array}$$

Example 2.7. *The following are relations on $\mathbb{Z}$:*

$$\begin{array}{llll} (a) & xRy & \textit{if} & x = y \\ (b) & xRy & \textit{if} & x|y \quad \textit{“x divides y”} \\ (c) & xRy & \textit{if} & x \leq y \\ (d) & xRy & \textit{if} & m|x-y \textit{ where } m \geq 2 \textit{ is a given integer} \\ & & – & \textit{this is “congruence modulo m”, written } x \equiv y(\mathrm{mod} m). \end{array}$$

Equivalence relations

A relation R on a set X is an *equivalence relation* if it has the following properties:

Reflexivity: xRx for all $x \in X$;

Symmetry: $xRy \Rightarrow yRx$ for all $x, y \in X$;

Transitivity: xRy and $yRz \Rightarrow xRz$ for all $x, y, z \in X$.

Example 2.8. *For the four relations in Example 2.7:*
Clearly, equality is an equivalence relation.

Divisibility is not; symmetry fails since $2 \mid 4$ *but* $4 \nmid 2$.

Congruence modulo m is a (very important) equivalence relation. Let us check that:

(R) $m \mid 0 = x - x \Rightarrow x \equiv x \pmod{m}$.

(S) $x \equiv y \pmod{m} \Rightarrow m \mid (x-y) \Rightarrow m \mid (y-x) \Rightarrow y \equiv x \pmod{m}$.

(T) $x \equiv y \;\&\; y \equiv z \pmod{m} \Rightarrow m \mid (x-y) \;\&\; m \mid (y-z) \Rightarrow m \mid (x-y)+(y-z) = x-z \Rightarrow x \equiv z \pmod{m}$.

Is $\leq$ an equivalence relation?

Example 2.9. *Give an example of a relation that is reflexive and symmetric but not transitive.*

Equivalence classes

The crucial property of an equivalence relation on a set X is that it splits X into well-defined equivalence classes.

Let R be an equivalence relation on X, and let $x \in X$. The *equivalence class* of x is

$$[x] = \{y \in X : xRy\},$$

that is the set of all y that are related to x.

Theorem 2.10. *The equivalence classes of an equivalence relation R on a set X partition X into disjoint subsets, that is:*

(i) for any $x, y \in X$ either $[x] \cap [y] = \emptyset$ or $[x] = [y]$;

(ii) $\bigcup_{x \in X} [x] = X$.

Proof. If $x \in X$ then xRx (R) so $x \in [x]$, giving that x lies in some equivalence class, for (ii).

We must show that two classes are disjoint or identical. Suppose that $[x]$ and $[y]$ have a common element, say $z \in [x] \cap [y]$. Then xRz and yRz, so zRx (S), so yRx (T).

Now if $u \in [x]$, then xRu, so yRu (T), that is $u \in [y]$. We conclude that $[x] \subseteq [y]$, and similarly $[y] \subseteq [x]$ so $[x] = [y]$ for (i). □

Example 2.11. *Let us describe the equivalence classes for congruence modulo m. First note that $x \equiv y \pmod{m}$ if and only if x and y give the same remainder when divided by m. So, the equivalence classes correspond to the different possible remainders modulo m. There are m different remainders, and so the equivalence classes are:*

$$C_r = \{qm + r : q \in \mathbb{Z}\} \ (r = 0, 1, \ldots, m-1).$$

Example 2.12. *What are the equivalence classes of equality on X?*

Order relations

Certain relations are particularly useful when comparing the size or position of set elements.

A relation R on a set X is an *order relation* if it has the following properties:

Anti-symmetric: For all $x, y \in X$, either xRy or yRx with both holding if and only if $x = y$;

Transitivity: xRy and $yRz \Rightarrow xRz$ for all $x, y, z \in X$.

Example 2.13.

(a) $\leq$ *is an order relation on $\mathbb{Z}$ or $\mathbb{R}$*

(b) *The relation $\leq_{\mathbb{C}}$ on $\mathbb{C}$ given by $z \leq_{\mathbb{C}} w$ if and only if $\mathrm{Re}(z) \leq \mathrm{Re}(w)$ and if $\mathrm{Re}(z) = \mathrm{Re}(w)$ then $\mathrm{Im}(z) \leq \mathrm{Im}(w)$ is an order relation.*

2.4 Functions

Functions relate elements of different sets. Informally, a function f from a set A to a set B is a rule or formula that associates to each element $a \in A$ exactly one element $f(a) \in B$.

Formally a *function, mapping, operation* or *transformation* $f : A \to B$ is a subset $f \subseteq A \times B$ with the property that for each $a \in A$ there exists a unique element $b \in B$ such that $(a,b) \in f$. The notation $f(a) = b$ is normally used instead of $(a,b) \in f$ [NB Some books, especially in algebra, write af instead of $f(a)$].

We call A the *domain* of f and B the *range* (or *codomain*) of f.

Note that two functions f and g are equal if and only if they have the same domain and the same range and $f(x) = g(x)$ for all x in the domain.

Example 2.14. *The following are functions:*

1. $f : \{a,b,c,d\} \to \{1,2,3\}$*;* $f(a) = 1$, $f(b) = 2$, $f(c) = 1$, $f(d) = 3$.
2. $f : \mathbb{Z} \to \mathbb{Z}$*;* $f(x) = x+1$ *for each* $x \in \mathbb{Z}$.
3. $f : \mathbb{Z} \to \mathbb{Z}$*;* $f(x) = 2x$ *for each* $x \in \mathbb{Z}$.
4. $f : \mathbb{Q} \to \mathbb{Q}$*;* $f(x) = 2x$ *for each* $x \in \mathbb{Q}$.
5. *For any given set A the mapping* $i_A : A \to A$ *given by* $f(x) = x$ *for each* $x \in A$ *is called the* identity function on A.

Example 2.15. *The operation of addition on, for example, the integers* $\mathbb{Z}$ *may be viewed as a function* $+ : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$. *In fact, a binary operation on a set X is formally defined as a mapping* $X \times X \to X$.

Example 2.16. *A sequence* $(a_1, a_2, a_3, \ldots)$ *of elements from a set A is simply a function* $\mathbb{N} \to A$.

Types of function

A function $f : A \to B$ is *surjective* or *onto*, if, for every $b \in B$, there exists a (not necessarily unique) $a \in A$ such that $b = f(a)$.

A mapping $f : A \to B$ is *injective* or *one-to-one*, if distinct elements of A always have distinct images in B, i.e. for $x, y \in A$, $x \neq y$, then $f(x) \neq f(y)$.

A mapping is *bijective* if it is both injective and surjective.

In Example 2.14 above, 1, 2, 4 and 5 are surjective; 3 is not. For example, to prove function 2 is surjective: given any $y \in \mathbb{Z}$, let $x = y - 1$, then $f(x) = y - 1 + 1 = y$. To show that 3 is not surjective, note that there is no element $x \in \mathbb{Z}$ such that $f(x) = 1$.

Normally, a proof of injectivity uses the contrapositive of the condition just stated – one shows that:

$$f(x) = f(y) \Rightarrow x = y.$$

Functions 2, 3, 4 and 5 are injective; 1 is not. For example, function 3 is injective because

$$f(x) = f(y) \implies 2x = 2y \implies x = y.$$

Function 1 is not injective since the distinct elements a and c both map to 1.

Thus functions 2, 4 and 5 are bijections; 1 and 2 are not.

Example 2.17. *The follow functions are all different:*

$$\begin{array}{lll}
\sin: & \mathbb{R} \to \mathbb{R} & \textit{is neither injective or surjective} \\
\sin: & \mathbb{R} \to [-1,1] & \textit{is surjective} \\
\sin: & [-\frac{\pi}{2}, \frac{\pi}{2}] \to \mathbb{R} & \textit{is injective} \\
\sin: & [-\frac{\pi}{2}, \frac{\pi}{2}] \to [-1,1] & \textit{is bijective}.
\end{array}$$

Composition and inverses of functions Given two functions $f : A \to B$ and $g : B \to C$, we often want to combine these functions to give us a new function from A to C whose effect on an element $a \in A$ is the same as first applying the f to a, and then applying g to the result:

For two functions $f : A \to B$ and $g : B \to C$, their *composition* is the function $g \circ f : A \to C$ defined by $(g \circ f)(a) = g(f(a))$ for all $a \in A$.

Theorem 2.18. *Composition is associative; that is if $f : A \to B$, $g : B \to C$ and $h : C \to D$ then $f \circ (g \circ h) : A \to D$ and $(f \circ g) \circ h : A \to D$ are equal.*

Proof. To prove that two functions are equal we simply check that they act in the same way on each element. For all $x \in X$

$$(f \circ (g \circ h))(x) = f((g \circ h)(x)) = f(g(h(x))$$

and

$$((f \circ g) \circ h)(x) = (f \circ g)(h(x)) = f(g(h(x)),$$

so $f \circ (g \circ h) = (f \circ g) \circ h$. □

Given a function $f : A \to B$, it is natural to ask whether there is a way of 'reversing' f, that is can we find a new function $g : B \to A$ that 'undoes' the effect of f?

Let $f : A \to B$. A function $g : B \to A$ is called an the *inverse* of f if $g \circ f = i_A$ and $f \circ g = i_B$. If such an inverse exists we say that f is is *invertible*. We write f^{-1} for the inverse of f, so that $f^{-1} \circ f = i_A$ and $f \circ f^{-1} = i_B$ (Note that an inverse must be unique).

It turns out that we can exactly characterize the invertible mappings:

Theorem 2.19. *Let $f : A \to B$. then f has an inverse if and only if it is a bijection.*

Proof. Suppose f has inverse f^{-1}. To show that f is injective, let $x, y \in A$ be such that $f(x) = f(y)$. Then $f^{-1}(f(x)) = f^{-1}(f(y))$, that is $x = y$.

To show f is surjective, let $z \in B$ and let $x = f^{-1}(z)$. Then $f(x) = f(f^{-1}(z)) = i_B(z) = z$. So f is surjective and thus bijective.

Now suppose f is a bijection. We define an inverse by 'reversing the arrows'. For each $y \in B$, the surjectivity of f implies the existence of an element $x \in A$ with $f(x) = y$. By the injectivity of f, this element is unique. So define $f^{-1}(y) = x$. Then $f(f^{-1}(y)) = f(x)$ where x is such that $y = f(x)$, so $f(f^{-1}(y)) = y$. Also $f^{-1}(f(x))$ equals the unique z such that $f(z) = f(x)$, so $z = x$ by injectivity and $f^{-1}(f(x)) = x$. □

Example 2.20. $\tan : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ *is a bijection and* $\tan^{-1} : \mathbb{R} \to (-\frac{\pi}{2}, \frac{\pi}{2})$ *is its inverse.*

Proposition 2.21. *Let $f : A \to B$, $g : B \to C$ be bijections. Then $g \circ f : A \to C$ is a bijection, and $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.*

Proof. That $g \circ f$ is a bijection follows easily from the definitions. Using associativity,

$$(g \circ f) \circ (f^{-1} \circ g^{-1}) = g \circ (f \circ (f^{-1} \circ g^{-1})) = g \circ ((f \circ f^{-1}) \circ g^{-1}) = g \circ (i_B \circ g^{-1}) = g \circ g^{-1} = i_C.$$

Similarly $(f^{-1} \circ g^{-1}) \circ (g \circ f) = i_A$. □

Induced functions

If a set is the domain of a function it is sometimes possible for a function to be defined on its equivalence classes in a natural way.

Suppose that X is a set and that R is an equivalence relation on X. Write X/R for the set of all equivalence classes of X under R (X/R is sometimes called the *quotient set*), Let $f : X \to Y$ be a function for some set Y. We will attempt to define a function

$$\tilde{f} : X/R \to Y \quad \text{by} \quad \tilde{f}([x]) = f(x)$$

where $[x]$ is the equivalence class containing x.

We need to be sure that $\tilde{f}$ is *well-defined*, i.e. that if $[x] = [y]$ then $\tilde{f}([x]) = \tilde{f}([y])$, that is $f(x) = f(y)$ whenever x and y are in the same equivalence class. We then call $\tilde{f}$ the *induced function* on X/R. [If you have studied any group theory, you will have met this notion with quotient groups.]

Example 2.22. *On $\mathbb{Z}$ let R be the equivalence relation xRy iff $2|x-y$, with equivalence classes E 'the evens' and O 'the odds'. Define $f : \mathbb{Z} \to \mathbb{Z}$ by $f(x) = (-1)^x$. Then if xRy we get*

$$f(x) = (-1)^x = (-1)^y(-1)^{x-y} = (-1)^y = f(y).$$

So the induced function $\tilde{f}(E) = 1, \tilde{f}(O) = -1$ is well-defined.

Example 2.23. Modular arithmetic *Start with the integers $\mathbb{Z}$ and the relation $\equiv \pmod{m}$. Denote by $\mathbb{Z}_m = \{[x] \,:\, x \in \mathbb{Z}\}$ the set of equivalence classes. We have seen that there are precisely m of them. Define addition and multiplication on $\mathbb{Z}_m$ by*

$$[x]+[y] = [x+y],\ [x][y] = [xy].$$

Then these operations are well defined.

To see this note that:

$$\begin{aligned}
[x]=[x_1]\ \&\ [y]=[y_1] &\Rightarrow x \equiv x_1 \pmod{m}\ \&\ y \equiv y_1 \pmod{m}\\
&\Rightarrow m \mid (x-x_1)\ \&\ m \mid (y-y_1)\\
&\Rightarrow m \mid (x-x_1)+(y-y_1) = (x+y)-(x_1+y_1)\\
&\Rightarrow x+y \equiv x_1+y_1 \pmod{m} \Rightarrow [x+y]=[x_1+y_1]\\
[x]=[x_1]\ \&\ [y]=[y_1] &\Rightarrow x \equiv x_1 \pmod{m}\ \&\ y \equiv y_1 \pmod{m}\\
&\Rightarrow m \mid (x-x_1)\ \&\ m \mid (y-y_1)\\
&\Rightarrow m \mid (x-x_1)y - x_1(y-y_1) = xy - x_1y_1\\
&\Rightarrow xy \equiv x_1y_1 \pmod{m} \Rightarrow [xy]=[x_1y_1].
\end{aligned}$$

This established, it is *very easy* to see that the basic laws of arithmetic hold in this new structure; for instance the associative law:

$$\begin{aligned}
([x]+[y])+[z] &= [x+y]+[z] = [(x+y)+z] = [x+(y+z)]\\
&= [x]+[y+z] = [x]+([y]+[z]).
\end{aligned}$$

But one has to be careful: for instance in $\mathbb{Z}_m$ we have

$$\underbrace{[x]+\ldots+[x]}_{m} = [0].$$

3 Number Systems

In this chapter we consider 'increasingly large' number systems: the integers, the rationals, the reals and the complex numbers. We will take the integers as our starting point and build up the other number systems from there.

We start by noting some properties are common to these number systems.

3.1 General Arithmetic Properties

Let R be a set (think of $\mathbb{R}$) and let $+$ and $\cdot$ be binary operations on R (called the *sum* and *product*), so that $x+y \in R$ and $xy \in R$ whenever $x,y \in R$.
$(R,+,\cdot)$ is called a *ring* if the following conditions hold for all $a,b,c \in R$:

(A1) $a+b=b+a$ (commutative law for +)

(A2) $a+(b+c)=(a+b)+c$ (associative law for +)

(A3) there exists $0 \in R$ such that $0+a=a$ for all $a \in R$ (zero)

(A4) for all $a \in R$ there exists $-a \in R$ such that $a+(-a)=0$ (negatives)

(M1) $ab=ba$ (commutative law for .)

(M2) $a(bc)=(ab)c$ (associative law for .)

(M3) there exists $1 \in R$ such that $1 \neq 0$ and $1a=a$ for all $a \in R$ (unity)

(D) $a(b+c)=ab+ac$ (distributive law).

$(R,+,\cdot)$ is called a *field* if, in addition,

(M4) for all $a \in R$ with $a \neq 0$ there exists $a^{-1} \in R$ such that $aa^{-1}=1$ (inverses).

[Jumping ahead, $\mathbb{Z}$ is a ring and $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are fields.]

Properties
From these properties we may show that in any ring:

(i) 0 is unique

(ii) 1 is unique

(iii) for all $a \in R$, $-(-a)=a$

(iv) $0a=a0=0$ for all $a \in R$.

Sample proofs (ii) Suppose 1 and q both satisfy (M3). Then $q=1q=q1=1$ using (M3), (M1) and that q satisfies (M3).

(iv) Note that $a0+a0=a(0+0)=a0$ by (D) and (A3). Hence

$$a0=a0+0=a0+(a0+-(a0))=(a0+a0)+-(a0)=a0+-(a0)=0,$$

using (A3), (A1), (A4), (A2) and (A4). □

Cancellation laws The usual cancellation laws also follow from the ring or field axioms.

(C1) In any ring $b+a=c+a$ implies $b=c$

(C2) In any field $ba=ca$ and $a \neq 0$ implies $b=c$.

Proof If $b+a=c+a$ then

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

using (A2), so by (A4), (A3),and (A1)

$$b=b+0=c+0=c.$$

(C2) is similar. □

Order properties
We will also be interested in the notion of 'order' on a ring. A ring R is *ordered* if there is a order relation $\leq$ on R such that for all $a,b,c,d \in R$ we have either $a \leq b$ or $b \leq a$ with both holding iff $a=b$ and:

(O1) $a \leq b, b \leq c$ implies $a \leq c$

(O2) $a \leq b, c \leq d$ implies $a+c \leq b+d$

(O3) $0 \leq a \leq b, 0 \leq c \leq d$ implies $ac \leq bd$.

Of course, once we have defined $\leq$ then the definitions of $\geq, <$ and $>$ follow by the usual conventions. E.g. $a<b$ means $a \leq b$ and $a \neq b$

3.2 Integers

One has to start from somewhere and in this course we will start from the natural numbers and build up other number systems from there. [It is possible to take a more basis starting point, namely the Peano Axioms, and develop the natural numbers and the integers from there - see Steward and Hall for details.]

Thus we will assume that the integers $\mathbb{Z}=\{\ldots,-3,-2,-1,0,1,2,3,\ldots\}$ are 'familiar' and 'well-understood'.

In particular, we assume the standard properties of addition, multiplication and order. Thus:

Proposition 3.1. *The integers $\mathbb{Z}$ form a ring under usual addition and multiplication, and this ring is ordered under $\leq$.*

The integers are adequate for many things, in particular for purposes of mathematical induction. However, they are in some ways limited, notably that integers do not have multiplicative inverses, in other words that they do not form a field.

3.3 Rational Numbers

Why do we want to *define* rational numbers? Why are we not satisfied with our intuition of them (as we pretended to be with $\mathbb{Z}$)? Well, we just about could. The conceptual difficulty arising with rational numbers is that different forms (such as, say, $3/5$ and $6/10$), actually represent the same rational number.

We will see how the language of equivalence relations can be used to overcome this difficulty. But also this will be a warm-up exercise for the construction of real numbers from the rationals – where our intuition is much more flaky.

A rational number, say 3/5, is just a fancy notation for an (ordered) pair of integers $(3,5)$ and we need to identify this with the pairs $(6,10),(9,15)$, etc, which are different representation of the same fraction.

Write $\mathbb{Z}^* = \mathbb{Z} \setminus 0$ for the set of non-zero integers. We define a relation on $\mathbb{Z} \times \mathbb{Z}^*$ by

$$(a,b)R(c,d) \iff ad = bc.$$

Lemma 3.2. *R is an equivalence relation.*

Proof. Using the properties of multiplication of integers:

(R) $ab = ba \Rightarrow (a,b)R(a,b)$.

(S) $(a,b)R(c,d) \Rightarrow ad = bc \Rightarrow cb = da \Rightarrow (c,d)R(a,b)$.

(T) $(a,b)R(c,d)$ & $(c,d)R(e,f) \Rightarrow ad = bc$ & $cf = de$
$\Rightarrow adf = bcf = bde \Rightarrow af = be$ (as $d \neq 0$) $\Rightarrow (a,b)R(e,f)$. □

Definition 3.3. *We call the equivalence classes of R the* rational numbers*:*

$$\mathbb{Q} = \{[(x,y)] : (x,y) \in X\}.$$

The equivalence class $[(a,b)]$ *is denoted by* $\frac{a}{b}$ *or* a/b*.*

Example 3.4. *Describe the equivalence classes of* $(0,1)$ *and* $(1,1)$*.*

We need to define the basic operations on $\mathbb{Q}$. We do this by 'mimicking' what we 'already know':

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd} \qquad [(a,b)] + [(c,d)] = [(ad+bc, bd)]$$

$$\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd} \qquad [(a,b)][(c,d)] = [(ac,bd)].$$

As with the modular arithmetic, we are defining operations for *equivalence classes* by using their *representatives*. We need to check that:

Proposition 3.5. *The operations of '+' and '·' on* $\mathbb{Q}$ *given above are well-defined.*

Proof. We prove this for addition; multiplication is left as an exercise. Suppose that

$$[(a_1,b_1)] = [(a_2,b_2)], \quad [(c_1,d_1)] = [(c_2,d_2)]. \tag{3.1}$$

Then

$$a_1b_2 = b_1a_2,\ c_1d_2 = d_1c_2.$$

Using the ring properties of + and $\cdot$ on $\mathbb{Z}$:

$$\begin{aligned}(a_1d_1+b_1c_1)b_2d_2 &= a_1d_1b_2d_2+b_1c_1b_2d_2 = \underline{a_1b_2}d_1d_2+b_1b_2\underline{c_1d_2} \\ &= b_1a_2d_1d_2+b_1b_2d_1c_2 = b_1d_1(a_2d_2+b_2c_2).\end{aligned}$$

This means that

$$(a_1d_1+b_1c_1, b_1d_1)R(a_2d_2+b_2c_2, b_2d_2),$$

and hence

$$[(a_1d_1+b_1c_1, b_1d_1)] = [(a_2d_2+b_2c_2, b_2d_2)],$$

as required. □

Example 3.6. *Try to define a 'funny addition' of rationals as follows:*

$$\frac{a}{b}\oplus\frac{c}{d}=\frac{a+c}{b+d}.$$

Show that this is not a well defined operation.

It is now straightforward to check the standard laws for addition and multiplication by using representatives. The laws are summed up in the following statement.

Theorem 3.7. *$\mathbb{Q}$ is a field under the operations of '$+$' and '$\cdot$' given above. The 'zero' is $[(0,1)]$, the'one' is $[(1,1)]$. The 'negative' of $[(a,b)]$ is $[(-a,b)]$ and the multiplicative inverse of $[(a,b)]$ (where $b\neq 0$) is $[(b,a)]$.*

Proof. We just check the distributive law; the other properties are very similar using representatives.

$$x=[(a,b)]=\frac{a}{b},\ y=[(c,d)]=\frac{c}{d},\ z=[(e,f)]=\frac{e}{f},$$

and then

$$\begin{aligned}x(y+z) &= \frac{a}{b}\left(\frac{c}{d}+\frac{e}{f}\right)=\frac{a}{b}\cdot\frac{cf+de}{df}=\frac{acf+ade}{bdf} \\ &= \frac{acbf+bdae}{bdbf}=\frac{ac}{bd}+\frac{ae}{bf}=\frac{a}{b}\cdot\frac{c}{d}+\frac{a}{b}+\frac{e}{f}=xy+xz.\end{aligned}$$

□

Recall that for two non-zero integers a,b their *greatest common divisor* $\gcd(a,b)$ is the largest natural number d which divides (without remainder) both a and b. We say that a and b are *co-prime* if $\gcd(a,b)=1$. It is clear from our definition of the equivalence R that

$$(a,b)R(ac,bc)\ (0\neq c\in\mathbb{Z}).$$

So we obtain the well-known representation 'in lowest terms' of a rational number:

Corollary 3.8. *For every non-zero rational number r there exist co-prime integers a and b such that $r = a/b$.*

Order properties of the rationals

We define an order $\leq$ on $\mathbb{Q}$ in terms of $\leq$ on $\mathbb{Z}$ by

$$\frac{a}{b} \leq \frac{c}{d} \quad (i.e.[(a,b)] \leq [(c,d)]) \quad \Leftrightarrow ad \leq bc$$

where we choose representatives of the equivalence classes with $b,d > 0$.

We can easily show that this condition does not depend on which representatives of the equivalence classes we choose.

Moreover, we can use representatives to show that it is an order relation and that the order properties (O1)-(O3) hold. For example for (O2):

$$\begin{aligned} \frac{a}{b} \leq \frac{c}{d}, \ \frac{e}{f} \leq \frac{g}{h} \ &\Rightarrow \ ad \leq bc, \ \ eh \leq gf \\ &\Rightarrow \ adfh + ehbd \leq bcfh + gfbd \\ \Rightarrow \ \frac{a}{b} + \frac{e}{f} = \frac{af+be}{bf} &\leq \frac{ch+gd}{dh} = \frac{c}{d} + \frac{g}{h}. \end{aligned}$$

The integers as a subset of the rationals

As things stand, integers are not rational numbers: integers are integers, and rational numbers are equivalence classes of pairs of integers, so we cannot say that $\mathbb{Z} \subseteq \mathbb{Q}$: We resolve this problem by showing that inside $\mathbb{Q}$ there is a set which behaves exactly like $\mathbb{Z}$.

Proposition 3.9. *The mapping*

$$f : \mathbb{Z} \to \mathbb{Q}, \ f(x) = \frac{x}{1}$$

is injective and respects the basic operations and the ordering on $\mathbb{Z}$, in particular for all $x, y \in \mathbb{Z}$ we have:

$$\begin{aligned} &f(x+y) = f(x) + f(y), \\ &f(xy) = f(x)f(y), \\ &x \leq y \Rightarrow f(x) \leq f(y). \end{aligned}$$

Proof. We have for $x, y \in \mathbb{Z}$:

$$\begin{aligned} &f(x) = f(y) \Rightarrow \frac{x}{1} = \frac{y}{1} \Rightarrow x \cdot 1 = y \cdot 1 \Rightarrow x = y, \\ &f(x) + f(y) = \frac{x}{1} + \frac{y}{1} = \frac{x \cdot 1 + y \cdot 1}{1 \cdot 1} = \frac{x+y}{1} = f(x+y), \\ &f(x)f(y) = \frac{x}{1} \cdot \frac{y}{1} = \frac{xy}{1} = f(xy), \\ &x \leq y \ \Rightarrow \ x1 \leq 1y \ \Rightarrow \ \frac{x}{1} \leq \frac{y}{1} \ \Rightarrow \ f(x) \leq f(y). \end{aligned}$$

□

Therefore, the image of $\mathbb{Z}$ under f

$$\text{Im}(f) = \{f(x) : x \in \mathbb{Z}\} = \{\frac{x}{1} : x \in \mathbb{Z}\}$$

behaves precisely as $\mathbb{Z}$, and we can identify the two sets.

Denseness of the rationals
As an ordered set, $\mathbb{Q}$ has a property which makes it very different from $\mathbb{Z}$.

Definition 3.10. *An ordered set X is called* dense *if for all $x, y \in X$ with $x < y$ there exists $z \in X$ such that $x < z < y$.*

Example 3.11. *The integers $\mathbb{Z}$ are not dense. For example, there is no integer z such that $2 < z < 3$.*

Proposition 3.12. *The rational numbers $\mathbb{Q}$ form a dense set.*

Proof. Let $x, y \in \mathbb{Q}$ with $x < y$. Let $z = (x+y)/2$. Then $x < z < y$. □

Incompleteness of the rationals
Although rational numbers are dense (see 3.12), it turns out that they still do not fit our intuitive notion of a geometrical line. If they did, they could be used for measuring lengths, i.e. every finite segment of a line would have the length which is a rational number. This, however, is not the case. For example, if the length of the diagonal of a unit square is c, then by Pythagoras' Theorem $c^2 = a^2 + b^2 = 1^2 + 1^2 = 2$. But what is c? Whatever it is, it isn't a rational:

Proposition 3.13. *There is no rational number c such that $c^2 = 2$.*

Proof. For a contradiction, suppose that such a rational number exists, we may write it as $c = a/b$, where a, b have no common factor. Then $a^2/b^2 = c^2 = 2$ and hence $a^2 = 2b^2$. It follows that a^2 is even, and so a itself must be even. Write $a = 2a_1$ and substitute into $a^2 = 2b^2$ to obtain $4a_1^2 = 2b^2$, and hence $2a_1^2 = b^2$. We now conclude that b^2 is even and so b itself is even. But both a and b are even which contradicts that they have no common factor. □

So it is as though there are numbers that are missing. Another way of viewing this is via the notion of *completeness*. For this we need the notions of 'maximum' and 'minimum' elements of a set, and introduce the concepts of upper and lower bounds:

Definition 3.14. *Let X be a set and $\leq$ an order relation on X. Let Y be a subset of X and m a member of X.*

m is an upper bound for Y if $m \geq y$ for all $y \in Y$

m is a least upper bound for Y if $m \geq y$ for all $y \in Y$ but if $m' < m$ there is $y \in Y$ with $y > m'$.

m is the maximum of Y, denoted by $\max Y$*, if $m \in Y$ and m is a upper bound for Y.*

m is a lower bound for Y if $m \leq y$ for all $y \in Y$

m is a greatest lower bound for Y if $m \leq y$ for all $y \in Y$ but if $m' > m$ there is $y \in Y$ with $y < m'$.

m is the minimum of Y, denoted by $\min Y$*, if $m \in Y$ and m is a lower bound for Y*

Example 3.15. *Let $X = \{0,1,2,3,\ldots\}$ and $Y = \{\ldots,-2,-1,0,1\}$ be subsets of $\mathbb{Z}$. Then*
X has a minimum 0 but no maximum
Y has a maximum 1 but no minimum
$X \cup Y$ has neither a minimum nor a maximum
$X \cap Y$ has a minimum 0 and a maximum 1.

Proposition 3.16. *The set*

$$A = \{q \in \mathbb{Q} : q^2 < 2\}$$

is bounded above (i.e. has an upper bound) but has no least upper bound.

Proof. It is clear that A is bounded above: 2 is one upper bound.

Suppose that A has the least upper bound $b \in \mathbb{Q}$. By Theorem 3.13 $b^2 \neq 2$, so that either $b^2 < 2$ or $b^2 > 2$. We will obtain a contradiction by showing that in the former case b is not an upper bound, while in the latter case it is not the *least* upper bound.

Case 1: $b^2 < 2$. It is enough to find a positive integer n such that $(b+1/n)^2 < 2$. But

$$2 - \left(b + \frac{1}{n}\right)^2 = 2 - b^2 - \frac{2b}{n} - \frac{1}{n^2} > (2-b^2) - \frac{2b+1}{n}.$$

Hence, taking $n > \frac{2b+1}{2-b^2} > 0$ we get $2-(b+1/n)^2 > 0$, with $b+1/n \in \mathbb{Q}$ as desired.

Case 2: $b^2 > 2$. This time we want to find a positive integer n such that $(b-1/n)^2 > 2$.

$$\left(b - \frac{1}{n}\right)^2 - 2 = b^2 - \frac{2b}{n} + \frac{1}{n^2} - 2 > (b^2-2) - \frac{2b}{n}.$$

Taking $n > \frac{2b}{b^2-2} > 0$ we get $(b-1/n)^2 - 2 > 0$, with $b - 1/n \in \mathbb{Q}$. □

Definition 3.17. *A totally ordered set X is said to be* complete *if every subset of X which is bounded above has a least upper bound.*

Corollary 3.18. *$\mathbb{Q}$ is not complete.*

3.4 Real Numbers

Real numbers are thought of as 'filling the holes' that we noted in $\mathbb{Q}$. That gives us a clue as to how define a real number: we will identify it with the set of all rational numbers lying below it (and for which it will act as the least upper bound).

Definition 3.19. *A* Dedekind cut *or* cut *is a non-empty subset $A \subseteq \mathbb{Q}$ with following properties:*

(C1) A is bounded above, i.e. has an upper bound;

(C2) A has no maximum;

(C3) A is closed downwards, that is if $x \in A$ and $y \leq x$ then $y \in A$.

The set of all Dedekind cuts is denoted by $\mathbb{R}$ and termed the real numbers.

Example 3.20. $\{x \in \mathbb{Q} : x < 1\}$ *is a Dedekind cut,* $\{x \in \mathbb{Q} : x \leq 1\}$*,* $\{x \in \mathbb{Z} : x < 1\}$ *and* $\mathbb{Q}$ *are not.*

In order to 'build' the theory of real numbers we need the following steps:

(1) Define the basic operations + and $\cdot$ on $\mathbb{R}$.

(2) Define $\leq$ on $\mathbb{R}$.

(3) Identify a copy of $\mathbb{Q}$ in $\mathbb{R}$.

(4) Establish the basic properties of the operations and $\leq$, that is show that $\mathbb{R}$ is an ordered field.

(5) Prove that $\mathbb{R}$ is complete.

Executing this whole programme in details is rather long and tedious, mostly because of a technical difficulty to do with multiplication of negative numbers (see below). We sketch the process, paying attention to the most important points.

The definition of addition presents no problems:

Definition 3.21. *For cuts A and B define their* sum *as follows:*

$$A + B = \{a + b : a \in A,\ b \in B\}.$$

Proposition 3.22. *If A and B are cuts then so is $A + B$.*

Proof. We need to check that the conditions (C1), (C2) and (C3) of Definition 3.19 are satisfied.

(C1) If m is an upper bound for A, and if n is an upper bound for B, then $m+n$ is an upper bound for $A+B$.

(C2) Suppose $A+B$ has a maximum element m, so $m = a+b$ where $a \in A$, $b \in B$. Since A satisfies (C2), there is $a_1 \in A$ such that $a_1 > a$. But then $a_1 + b \in A + B$ and $a_1 + b > a + b = m$, which is a contradiction.

(C3) Suppose that $x \in A+B$ and that $y \leq x$. Write $x = a+b$ with $a \in A$, $b \in B$. Note that $y = x + (y-x) = (a+y-x)+b$. Since $a+y-x \leq a$ we have $a+y-x \in A$, and hence $y \in A+B$. □

Example 3.23. *Why are the following definitions for multiplication, negatives and inverses are not good?*

$$\begin{aligned}
&AB = \{ab : a \in A,\ b \in B\};\\
&-A = \{-a : a \in A\};\\
&1/A = \{1/a : a \in A\}.
\end{aligned}$$

In order to define these operations properly, we need the distinction between positive and negative cuts. To do this, we will now introduce the ordering on cuts, and cuts corresponding to rational numbers:

Definition 3.24. *For cuts A and B define:*

$$A \leq B \Leftrightarrow A \subseteq B.$$

Proposition 3.25. *$\leq$ is an order relation on $\mathbb{R}$.*

Proof. It is clear that $\subseteq$ is transitive. Let $A, B \in \mathbb{R}$ be arbitrary, and suppose that $B \not\leq A$, i.e. $B \not\subseteq A$. This means that there exists $b \in B \setminus A$. Since A satisfies (C3), for every $a \in A$ we must have $a < b$. But since B satisfies (C3) as well, this implies that $a \in B$ for every $a \in A$. This proves that $A \subseteq B$, i.e. $A \leq B$. Since $A \leq B$ and $B \leq A$ imply that $A = B$, $\leq$ is antisymmetric. □

Definition 3.26. *For a rational number r define*

$$\bar{r} = \{x \in \mathbb{Q} : x < r\}.$$

The set $\overline{0}$ is denoted by O; the set $\overline{1}$ is denoted by I.

Proposition 3.27. *For every $r \in \mathbb{Q}$, the set $\bar{r}$ is a cut.*

Proof. Exercise. □

Thus, by identifying the rational number r with the cut $\bar{r}$ we can regard the rationals as a subset of the real numbers.

Now we can define the remaining operations:

Definition 3.28. *For a cut A define its* negative *as follows:*

$$-A = \{-x : x \in \mathbb{Q} \text{ is an upper bound for } A\text{, but not a least upper bound for } A\}.$$

Proposition 3.29. *If A is a cut, then so is $-A$ and $A + (-A) = 0$.*

Proof. Exercise. □

Definition 3.30. *A cut A is said to be* positive *if $A > O$.*

Example 3.31. *A is positive if and only if it contains some positive numbers.*

Definition 3.32. *For two cuts A and B define their* product *as follows:*

$$AB = \{ab : a \in A,\ b \in B,\ a > 0,\ b > 0\} \cup \{x \in \mathbb{Q} : x \le 0\}$$

if they are positive, and

$$AB = \begin{cases} -(A(-B)) & \textit{if } A > O,\ B < O; \\ -((-A)B) & \textit{if } A < O,\ B > O; \\ (-A)(-B) & \textit{if } A < O,\ B < O; \\ O & \textit{if } A = O \textit{ or } B = O. \end{cases}$$

Proposition 3.33. *If A and B are cuts, then so is AB.*

Proof. Messy, and not particularly exciting. We will omit it. □

Definition 3.34. *For a cut $A > O$ define*

$$\frac{1}{A} = \left\{ \frac{1}{x} : x \textit{ is an upper bound for } A \textit{ but not a least upper bound for } A \right\} \cup \{q \in \mathbb{Q} : q \le 0\},$$

and for $A < O$ define

$$\frac{1}{A} = -\frac{1}{-A}.$$

Proposition 3.35. *If A is a cut and $A \neq O$ then $1/A$ is a cut and $A(1/A) = 1$.*

Theorem 3.36. *With the definitions of addition, multiplication, negatives and inverses above, $\mathbb{R}$ identified with cuts of $\mathbb{Q}$ is a field with zero O and identity I. It is an ordered field under $\le$. The rationals $\mathbb{Q}$ may be regarded as a subset of $\mathbb{R}$ by identifying $r \in \mathbb{Q}$ with $\bar{r} \in \mathbb{Q}$, an identification that respects the operations and order.*

Proof. This requires a great deal of tedious checking of the field and order properties given in Section 3.1. A main difficulty is that for anything involving multiplication we have to consider all the possibilities of Definition 3.32 to allow for negative numbers.

The proof that $\mathbb{Q}$ may be regarded as a subset of $\mathbb{R}$ is along the lines of Theorem 3.9 and is straightforward. □

The final, and in many ways most important thing that we have to show is that $\mathbb{R}$ is complete – after all it is the lack of completeness of $\mathbb{Q}$ that we set out to 'mend'.

Theorem 3.37. *$\mathbb{R}$ is complete. In other words, every subset of $\mathbb{R}$ that is bounded above has a least upper bound.*

Proof. Let $X \subseteq \mathbb{R}$ be a set that has an upper bound. We can write $X = \{A_i : i \in J\}$, where each A_i is a cut. Let

$$M = \bigcup_{i \in J} A_i.$$

We claim that $M \in \mathbb{R}$, i.e. that M is a cut.

(C1) Let P be any upper bound for X. P itself is a cut, so it has an upper bound p. For all i we have $A_i \leq P$, i.e. $A_i \subseteq P$, so $M \subseteq P$. Hence p is an upper bound for M.

(C2) Suppose that M has maximum m. Then, for some i, we would have $m \in A_i$ and m would be the maximum of A_i, which is impossible as A_i is a cut.

(C3) Let $x \in M$ be arbitrary, and suppose $y \leq x$. Then $x \in A_i$ for some i, and, since A_i is a cut, we must have $y \in A_i$, implying $y \in M$.

Clearly M is an upper bound for X. If Q is any other upper bound for X, we would have $A_i \subseteq Q$, which implies $M = \bigcup_{i \in I} A_i \subseteq Q$, i.e. $M \leq Q$. We conclude that M is the least upper bound for X. □

Existence of Roots in $\mathbb{R}$

The definition of the real numbers resolves the question of finding numbers such as $\sqrt{2}$ and roots of other equations.

Proposition 3.38. *There exists a real number r such that $r^2 = 2$.*

Proof. Recall the proof of Theorem 3.16. We had the set

$$A = \{q \in \mathbb{Q} : q^2 < 2\},$$

which is bounded above but does not have a least upper bound. But in $\mathbb{R}$ it does have a least upper bound by Theorem 3.37; denote this least upper bound by b. Exactly as in the proof of Theorem 3.16 we can show that $b^2 < 2$ and $b^2 > 2$ cannot happen, so $b^2 = 2$. □

We want to do this in (much) more generality:

Theorem 3.39. *Let n be a positive integer, and let x be a positive real number. Then there exists exactly one positive real number y such that $y^n = x$.*

Proof. The proof proceeds along similar lines to Theorem 3.16. Let us consider the set

$$A = \{r \in \mathbb{R} \,:\, r^n < x\}.$$

(We could have equally said $A = \{q \in \mathbb{Q} \,:\, q^n < x\}$.) Clearly A is bounded: if $x \leq 1$ then 1 is a bound; if $x > 1$ then x is a bound. So, by Theorem 3.37, A has the least upper bound y, say.

We show that if $y^n < x$ then y is not an upper bound and if $y^n > x$ then y is not the least upper bound of A, to conclude that $y^n = x$.

Case 1: $y^n < x$. It is enough to find $\alpha > 0$ such that $(y+\alpha)^n < x$. But

$$\begin{aligned} x-(y+\alpha)^n &= (x-y^n)-\big((y+\alpha)^n-y^n\big) \\ &= (x-y^n)-\big((y+\alpha)-y\big)\big((y+\alpha)^{n-1}+(y+\alpha)^{n-2}y+\cdots+(y+\alpha)y^{n-2}+y^{n-1}\big) \\ &\geq (x-y^n)-\alpha n(y+\alpha)^{n-1} \\ &\geq (x-y^n)-\alpha n(y+1)^{n-1} \end{aligned}$$

provided $0 < \alpha < 1$. Hence, taking some $0 < \alpha < 1$ such that $\alpha < \frac{x-y^n}{n(y+1)^{n-1}}$ we get $x-(y+\alpha)^n > 0$, contradicting that y is an upper bound of A.

Case 2: $y^n > x$. We will find $0 < \alpha < y$ such that $(y-\alpha)^n > x$. But

$$\begin{aligned} (y-\alpha)^n - x &= (y^n-x)-\big(y^n-(y-\alpha)^n\big) \\ &= (y^n-x)-\big(y-(y-\alpha)\big)\big(y^{n-1}+y^{n-2}(y-\alpha)+\cdots+y(y-\alpha)^{n-2}+(y-\alpha)^{n-1}\big) \\ &\geq (y^n-x)-\alpha n y^{n-1} \end{aligned}$$

Hence, taking some $0 < \alpha < y$ such that $\alpha < \frac{y^n-x}{ny^{n-1}}$ we get $(y-\alpha)^n - x > 0$, contradicting that y is the least upper bound of A. □

3.5 Complex Numbers

From the definition the product of cuts we see that $x^2 \geq 0$ for all$x \in \mathbb{R}$, so it follows that negative numbers do not have square roots (or, indeed, any even roots). In particular, there is no number $x \in \mathbb{R}$ satisfying $x^2 = -1$. We can 'invent' such a number, and 'adjoin' it to $\mathbb{R}$. That is how we build complex numbers. But then something quite remarkable happens: it turns out that *all* real numbers have *all* roots. Even better, $\sqrt[n]{z}$ exists for every complex number z and every natural number n (i.e. all complex numbers have all roots). Even better than that, every polynomial equation $P(z) = 0$, where P is a polynomial with complex coefficients, has a solution in complex numbers. This last result is called the Fundamental Theorem of Algebra.

Compared to the construction of reals from rational, or even the construction of rationals from integers, the construction of complex numbers from reals is straightforward.

Definition 3.40. $\mathbb{C} = \mathbb{R} \times \mathbb{R} = \{(x,y) : x,y \in \mathbb{R}\}$.

In other words, complex numbers are ordered pairs of reals. Addition and multiplication are defined as follows:

$$(a,b)+(c,d) = (a+c, b+d), \quad (a,b)\cdot(c,d) = (ac-bd, ad+bc).$$

Theorem 3.41. *With theses definitions $\mathbb{C}$ is a field with zero $(0,0)$ and identity $(1,0)$. The negative of (a,b) is $(-a,-b)$ and the multiplicative inverse of (a,b) is* $(a/(a^2+b^2), -b/(a^2+b^2))$.

Proof. It is straightforward to check the field properties listed in Section 3.1. □

There is an obvious identification of real numbers with the first coordinate of these pairs.

Proposition 3.42. *The mapping $f : \mathbb{R} \to \mathbb{C}$ defined by $f(x) = (x,0)$ identifies $\mathbb{R}$ as a subset of $\mathbb{C}$, i.e. f is injective and*

$$f(x+y) = f(x) + f(y), \quad f(xy) = f(x)f(y)$$

for all $x, y \in \mathbb{R}$.

Proof. Obvious □

It will come as no surprise that we think of the pair (x,y) as $x+yi$ where 'i' is a label we think of as the 'square root of -1'. (Compare this with thinking of the pair of integers (a,b) as the rational number a/b.)

The following Theorem is a formal justification of this.

Theorem 3.43. *If we let*

$$i = (0,1)$$

and identify every $x \in \mathbb{R}$ with $(x,0) \in \mathbb{C}$ then

$$i^2 = -1$$

and every $z = (x,y) \in \mathbb{C}$ can be written as $z = x+yi$.

Proof. We have $i^2 = (0,1)(0,1) = (-1,0) = -1$
and $x+yi = (x,0)+(y,0)(0,1) = (x,0)+(0,y) = (x,y) = z$. □

In what follows we will need the notion of the conjugate and modulus of a complex number.

Definition 3.44. *For a complex number $z = a+bi$ its* conjugate *is $\bar{z} = a-bi$, and its* modulus *is*

$$|z| = \sqrt{a^2+b^2}$$

Note in particular that

$$|z|^2 = a^2+b^2 = z\bar{z}.$$

For almost any arguments involving complex numbers, the following two inequalities are essential.

Proposition 3.45. *For all $z, w \in \mathbb{C}$:*

(i) $|zu| = |z||w|$*;*

(ii) $|z+w| \leq |z| + |w|$*. (the triangle inequality);*

Proof. (i) Let $z = a+bi$, $w = c+di$. We have

$$|zu|^2 = |(ac-bd)+(ad+bc)i|^2 = (ac-bd)^2+(ad+bc)^2$$

$$= (ac)^2+(bd)^2+(ad)^2+(bc)^2 = (a^2+b^2)(c^2+d^2) = |z|^2|w|^2.$$

(ii) We have

$$\begin{aligned}
|z+w|^2 &= (z+w)\overline{(z+w)} \\
&= (z+w)(\overline{z}+\overline{w}) \\
&= z\overline{z}+(z\overline{w}+w\overline{z})+w\overline{w} \\
&= |z|^2+2\mathrm{Re}(z\overline{w})+|w|^2 \quad \text{(where } \mathrm{Re}(z)=a \text{ if } z=a+bi\text{)} \\
&\leq |z|^2+2|z\overline{w}|+|w|^2 \\
&\leq |z|^2+2|z||w|+|w|^2 \quad \text{(using (i))} \\
&= (|z|+|w|)^2.
\end{aligned}$$

□

The Fundamental Theorem of Algebra

We now prove the Fundamental Theorem of Algebra which is one of the main reasons why complex numbers are important.

Theorem 3.46. Fundamental Theorem of Algebra *Every non-constant polynomial with complex (or real) coefficients*

$$p(z) = a_nz^n + a_{n-1}z^{n-1} + \ldots a_1z + a_0 \quad (n \geq 1, a_i \in \mathbb{C}, a_n \neq 0)$$

has a zero in $\mathbb{C}$*, i.e. there exists* $z \in \mathbb{C}$ *such that* $p(z) = 0$*.*

There are many proofs of this theorem, but all of them depend at least to some extent on ideas from analysis, such as continuity. If you take a course on complex analysis you will see slick proofs based on properties of analytic functions (in particular Liouville's theorem).

We first state the analytic properties that we will use.

(1) For every $w \in \mathbb{C}$ and positive integer k there exists z such that $z^k = w$. [To see this, note that we can write any $w \in \mathbb{C}$ as $w = r(\cos\theta + i\sin\theta)$ where $r > 0, \theta \in \mathbb{R}$, and $\big(r^{1/k}((\cos\theta/k + i\sin\theta/k))\big)^k = w$.]

(2) Every complex polynomial is a continuous function $\mathbb{C} \to \mathbb{C}$.

(3) The modulus of a complex polynomial $p(x)$ attains a minimum, i.e. there exists $c \in \mathbb{C}$ such that $|p(c)| = \min\{|p(z)| : z \in \mathbb{C}\}$. [This follows since $|p(z)|$ is continuous and gets very large for all large $|z|$.]

Proof of the Fundamental Theorem of Algebra

By Fact (3) let $|p(z)|$ take its minimum at $c \in \mathbb{C}$. We assume that $|p(c)| > 0$ and derive a contradiction.

We make the substitution $z = c + w$ and regard p as a polynomial in w with c as the origin; thus

$$\begin{aligned} p(c+w) &= a_n(c+w)^n + a_{n-1}(c+w)^{n-1} + \cdots + a_1(c+w) + a_0 \\ &= b_n w^n + b_{n-1}w^{n-1} + \cdots + b_1 w + b_0 \qquad \text{(for some } b_0, \ldots, b_n \in \mathbb{C}) \\ &= b_n w^n + b_{n-1}w^{n-1} + \cdots + b_k w^k + p(c) \end{aligned}$$

where k is the least integer such that $b_k \neq 0$, and noting that $b_0 = p(c) \neq 0$ (by setting $w = 0$).

Our strategy is to use that the dominant part of this polynomial when w is small is $b_k w^k + p(c)$; thus we write

$$p(c+w) = p(c) + b_k w^k \big(1 + wq(w)\big)$$

where $q(w)$ is some polynomial.

Using Fact (1) choose $a \in \mathbb{C}$ such that $a^k = -p(c)/b_k$, so that $b_k = -p(c)/a^k$. Then

$$p(c+w) = p(c)\left[1 - \frac{w^k}{a^k} - \frac{w^{k+1}}{a^k}q(w)\right].$$

By Fact (2), the polynomial $q(w)$ is continuous, so there is a number M such that $|aq(w)| \leq M$ if $|w| \leq |a|$. Let t be a real number with $0 < t \leq 1$. Then setting $w = ta$

$$p(c+ta) = p(c)[1 - t^k - t^{k+1}aq(ta)].$$

so, using the inequalities of Proposition 3.45,

$$\begin{aligned} |p(c+ta)| &\leq |p(c)|\big(|1-t^k| + t^{k+1}|aq(ta)|\big) \\ &\leq |p(c)|\big(1 - t^k + t^{k+1}M\big). \end{aligned}$$

It follows that if $0 < t < 1/M$ then $|p(c+ta)| < |p(c)|$ contradicting the minimality of $|p(c)|$.

We conclude that $|p(c)| = 0$, so that $p(c) = 0$, as desired. □

We have shown that a polynomial of degree n has at least one zero. We now show that it has n zeros. For this we need the Remainder Theorem.

Theorem 3.47. *Let $p(z)$ be a polynomial of degree $n \geq 1$ and let $\alpha \in \mathbb{C}$. Then we may write*

$$p(z) = q(z)(z - \alpha) + r \tag{3.2}$$

where $q(z)$ is a polynomial of degree $n-1$ and $r \in \mathbb{C}$.

In particular, $r = 0$ if and only if α is a zero of $p(z)$.

Proof. The identity (3.2) follows by long division of $p(z)$ by $(z-\alpha)$. □

Corollary 3.48. *Every polynomial of degree $n \geq 1$ has exactly n (not necessarily distinct) zeros, $\alpha_1, \ldots, \alpha_n$, i.e.*

$$p(z) = a(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_n),$$

where $a \neq 0$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$.

Proof. We prove this by induction on n.

If $n = 1$ then $p(z) = a_1 z + a_0 = a_1(z - (-a_0/a_1))$, where $a_1 \neq 0$, so the conclusion holds for $n = 1$.

So assume inductively that every polynomial of degree n has n zeros for some $n \geq 1$. Let $p(z)$ be a polynomial of degree $n+1$. By the Fundamental Theorem of Algebra, $p(z)$ has some zero, call it α_{n+1}, so by Theorem 3.47

$$p(z) = q(z)(z-\alpha_{n+1})$$

where q is a polynomial of degree n. By the inductive assumption there are $\alpha_1, \ldots, \alpha_n$ such that

$$q(z) = a(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_n),$$

so

$$p(z) = a(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_n)(z-\alpha_{n+1}).$$

□

Finally, we remark that there are explicit formulae for the zeros of polynomials of degrees 1,2,3 and 4, but not for degrees ≥ 5.

4 Some Number Theory

In this chapter we investigate properties of numbers in their own right, especially properties of prime numbers, factorisation, and whether certain numbers are rational or not.

4.1 Prime Numbers

We consider properties of the natural numbers $\mathbb{N} = \{1,2,3,\ldots\}$ and their factorisation. The division algorithm is almost obvious.

Proposition 4.1 (Division Algorithm)**.** *Let $a,b \in \mathbb{Z}$ with $b > 0$. Then $a = bq + r$ where $q,r \in \mathbb{Z}$ with $0 \leq r < b$.*

Proof. Consider $X = \{a - bt : t \in \mathbb{Z}\}$. This contains non-negative numbers, so let r be the least non-negative number in X, with $r = a - bt$ for some $t \in \mathbb{Z}$. If $r \geq b$ then $0 \leq r - b = a - b(t+1) \in X$ which contradicts the minimality of r. Thus $0 \leq r < b$. □

Definition 4.2. *If $a.b \in \mathbb{N}$ and $a = bq$ for some $q \in \mathbb{N}$ we say that a* divides *b, written $a|b$, or a* is a factor of *b.*

An integer $p \geq 2$ with no factors other than 1 and p is called a prime number.

If $a,b \in \mathbb{Z}$, the highest common factor *of a and b, written $\mathrm{HCF}(a,b)$ is the largest positive integer that divides both a and b.*

Integers a,b are called coprime *if $\mathrm{HCF}(a,b) = 1$.*

Examples: $5|15$; $\mathrm{HCF}(24,30) = 6$; the primes are $2,3,5,7,11,13,17,\ldots$.

Repeated application of the division algorithm leads to Euclid's Algorithm which states an important property of HCFs.

Proposition 4.3 (Euclid's Algorithm)**.** *Let $a_0, a_1 \in \mathbb{N}$. Then there exist integers x,y such that*

$$a_0x + a_1y = d \qquad \text{where } d = \mathrm{HCF}(a_0,a_1). \tag{4.3}$$

Proof. Assume that $a_0 > a_1$. Repeatedly applying the division algorithm we get

$$\begin{array}{rcllc}
a_0 & = & a_1q_1 + a_2 & (0 < a_2 < a_1) & (1) \\
a_1 & = & a_2q_2 + a_3 & (0 < a_3 < a_2) & (2) \\
\vdots & & \vdots & & \\
a_{n-3} & = & a_{n-2}q_{n-2} + a_{n-1} & (0 < a_{n-1} < a_{n-2}) & (n-2) \\
a_{n-2} & = & a_{n-1}q_{n-1} + a_n & (0 < a_n < a_{n-1}) & (n-1) \\
a_{n-1} & = & a_nq_n & & (n)
\end{array}$$

(since the a_i are decreasing the remainder will eventually be 0). From (n), $a_n|a_{n-1}$, so working upwards we get $a_n|a_{n-2}, a_n|a_{n-3}, \ldots, a_n|a_2, a_n|a_1, a_n|a_0$, so a_n is a common factor of a_0 and a_1. Moreover, by repeated substitution of the previous equation (starting with $(n-1)$ and working upwards), we get that $a_n = a_0x + a_1y$ for some integers x,y. Since any common factor of a_0 and a_1 must divide a_n, we get that $d \equiv a_n = \mathrm{HCF}(a_0,a_1)$. □

Primes are defined to be integers that cannot be factorised in a non-trivial way, and the property that a prime divides an individual term of any product that it divides is not obvious from this definition. However, this follows from Euclid's algorithm.

Proposition 4.4. *Let p be prime and suppose that $p|ab$. Then either $p|a$ or $p|b$.*
More generally if p is prime and $p|a_1a_2\ldots a_n$ then $p|a_i$ for some i.

Proof. Suppose $p|ab$. If $p \not| a$ then $\mathrm{HCF}(a,p) = 1$, so by Euclid's algorithm $ax+py = 1$, for some integers x,y. Thus $abx + pby = b$, so $p|b$ (as p divides the LHS of this equation).
The second statement follows inductively. □

We recall the Fundamental Theorem of Arithmetic, Theorem 1.5.

Theorem 4.5 (Fundamental Theorem of Arithmetic)**.** *Every natural number $n \geq 2$ is a product of prime numbers, i.e.*

$$n = p_1^{k_1}p_2^{k_2}\cdots p_m^{k_m},$$

for some distinct primes $p_1,\ldots,p_m$ and natural numbers $k_1,\ldots,k_m$. Moreover, up to the order of the p_i, this decomposition is unique.

A natural question to ask is how many prime numbers there are, and how big they are. We start with Euclid's famous theorem and proof, that there are infinitely many primes.

Theorem 4.6. *There are infinitely many prime numbers. In fact, writing p_k for the kth prime,*

$$p_k \leq 2^{2^k}.$$

Proof. Suppose there are finitely many primes, $p_1, p_2, \ldots, p_k$, say. Consider the number

$$m = p_1p_2,\cdots p_k - 1. \tag{4.4}$$

Then m must have some prime factor p (which may be m if m itself is prime). But m is not divisible by any of $p_1, p_2, \ldots, p_k$, so p is a prime distinct from $p_1, p_2, \ldots, p_k$, a contradiction.

Moreover, it follows from (4.4) that $p_{k+1} \leq p_1p_2,\cdots p_k$. Assume inductively that $p_k \leq 2^{2^k}$. This is true for $n = 1$ since $p_1 = 2 \leq 2^{2^1} = 4$. If the result holds for the first k primes, then

$$p_{k+1} \leq p_1p_2,\cdots p_k \leq 2^{2^1}2^{2^2}\cdots 2^{2^k} < 2^{2^{k+1}}$$

(since $2^1 + 2^2 + \cdots + 2^k = 2^{k+1} - 1$). The conclusion follows by induction. □

In fact there are infinitely many primes of various specific types. The following result concerns primes in arithmetic progression.

Theorem 4.7 (Dirichlet's Theorem)**.** *Let $a,b \in \mathbb{N}$ be coprime integers. Then there are infinitely many primes of the form $ak+b$ $(k \in \mathbb{N})$. In other words the arithmetic progression*

$$a+b, 2a+b, 3a+b, \ldots$$

contains infinitely many primes, i.e. there are infinitely many primes of the form $ak+b$.

Proof. This is hard to prove in general. However, some special cases just require a minor variant of the proof of Theorem 4.6.

For example, to show that there are infinitely many primes of the form $4k+3$, first note that if $n = 4k+1$ and $n' = 4k'+1$ then $nn' = 4k''+1$ for some k''. Now if $p_1, p_2, \ldots, p_j$ are the only primes of this form consider the number

$$m = 4p_1 p_2 \cdots p_j - 1.$$

Either m itself is prime or is a product of prime factors which must all be distinct from $p_1, p_2, \ldots, p_j$. These prime factors cannot all be of the form $4k+1$ otherwise m would be of the form $4k+1$. Thus at least one of these new primes must be of the form $4k+3$, to give the required contradiction. □

Although there are infinitely many primes, it is important to know roughly how large the kth prime is, or equivalently how many primes there are $\leq n$. It is customary to write $\pi(n)$ to denote the number of primes $\leq n$. Thus we have shown that $\pi(2^{2^k}) \geq k$. Letting $n = 2^{2^k} < e^{e^k}$, so that $\log\log n < k$, we get that $\pi(n) \geq \log\log n$. This is a very poor estimate for $\pi(n)$. The Prime Number Theorem gives a much better estimate.

Theorem 4.8 (Prime Number Theorem)**.**

$$\pi(n) \sim \frac{n}{\log n} \qquad \textit{as } n \to \infty,$$

that is

$$\frac{\pi(n)}{n/\log n} \to 1 \qquad \textit{as } n \to \infty.$$

The various proofs of the Prime Number Theorem are hard, so we give a weaker result due to Chebyshev that nevertheless gives a fairly accurate estimate on the size of primes, namely that $\pi(n)$ increases roughly at the rate of $n/\log n$. We first require some properties of binary coefficients. Recall that $\binom{m}{n} = m!/n!(m-n)!$, which is always an integer.

Lemma 4.9.

(1) *For $n \geq 1$ we have*

$$2^n \leq \binom{2n}{n} \leq 2^{2n}$$

(2) *For $n \geq 1$*

$$\prod_{n<p\leq 2n} p \,\bigg|\, \binom{2n}{n},$$

where the product is over all primes between n and $2n$.

(3) *For each prime p let $r(p)$ be the integer such that $p^{r(p)} \leq 2n < p^{r(p)+1}$. Then*

$$\binom{2n}{n} \,\bigg|\, \prod_{p\leq 2n} p^{r(p)}.$$

Proof. (1)

$$\binom{2n}{n} = \frac{(2n)!}{n!n!} = \frac{2n}{n}\,\frac{2n-1}{n-1}\,\frac{2n-2}{n-2}\cdots\frac{n+1}{1} \geq 2^n$$

(since all the fractions are ≥ 2). On the other hand, $2^{2n} = (1+1)^{2n} \geq \binom{2n}{n}$ (as one of the binomial coefficients).

(2) For every prime p with $n < p \leq 2n$ we have $p \not| n!$ but $p|(2n)! = n!n!\frac{(2n)!}{n!n!}$, so $p|\frac{(2n)!}{n!n!} = \binom{2n}{n}$.

(3) For each prime $2 \leq p \leq 2n$ and each positive integer k, the number p^k divides at most one more of the numbers $\{n+1, n+2, \ldots, 2n\}$ than of the numbers $\{1, 2, \ldots, n\}$. Also, if $p^k > 2n$ then p^k does not divide any or these numbers. Hence, considering prime factorisation,

$$\binom{2n}{n} = \frac{2n(2n-1)\ldots(n+1)}{n(n-1)\ldots 1} \left| \prod_{p \leq 2n} p^{r(p)} \right. .$$

□

Theorem 4.10 (Chebyshev's Prime Number Theorem)**.**

$$0.347\frac{n}{\log n} \simeq \tfrac{1}{2}\log 2\,\frac{n}{\log n} \leq \pi(n) \leq 8\log 2\,\frac{n}{\log n} \simeq 5.55\frac{n}{\log n}.$$

Proof. LH Inequality. By (1) and (3) of Lemma 4.9,

$$2^n \leq \binom{2n}{n} \leq \prod_{p \leq 2n} p^{r(p)} \leq (2n)^{\pi(2n)}$$

(recall that $r(p)$ is the integer such that $p^{r(p)} \leq 2n < p^{r(p)+1}$). Taking logs

$$n\log 2 \leq \pi(2n)\log(2n).$$

For $m = 2n$ even this becomes

$$\tfrac{1}{2}m\log 2 \leq \pi(m)\log(m) \quad \text{that is} \quad \pi(m) \geq \tfrac{1}{2}\log 2\frac{m}{\log m}.$$

If m is odd, $m+1$ is even, so by the even case

$$\pi(m) = \pi(m+1) \geq \tfrac{1}{2}\log 2\frac{m+1}{\log(m+1)} \geq \tfrac{1}{2}\log 2\frac{m}{\log m},$$

since $x/\log x$ increases as x increases.

RH Inequality. By (1) and (2) of Lemma 4.9,

$$n^{\pi(2n)-\pi(n)} \leq \prod_{n<p\leq 2n} p \leq \binom{2n}{n} \leq 2^{2n}.$$

Taking logs,

$$\big(\pi(2n)-\pi(n)\big)\log n \le 2n\log 2.$$

Thus

$$\begin{aligned}\pi(2n)\log 2n-\pi(n)\log n &= \pi(2n)\big(\log 2n-\log n\big)+\big(\pi(2n)-\pi(n)\big)\log n\\ &\le 2n\log 2+2n\log 2=(4\log 2)n.\end{aligned}$$

Now taking $n=1,2,2^2,\ldots,2^{k-1}$,

$$\begin{aligned}\pi(2)\log 2-\pi(1)\log 1 &\le (4\log 2)1\\ \pi(2^2)\log 2^2-\pi(2)\log 2 &\le (4\log 2)2\\ \pi(2^3)\log 2^3-\pi(2^2)\log 2^2 &\le (4\log 2)2^2\\ \vdots\qquad & \qquad\vdots\\ \pi(2^k)\log 2^k-\pi(2^{k-1})\log 2^{k-1} &\le (4\log 2)2^{k-1}.\end{aligned}$$

Adding, most terms cancel, to give

$$\pi(2^k)\log 2^k-\pi(1)\log 1\le(4\log 2)(1+2+2^2+\cdots+2^{k-1})\le(4\log 2)2^k.$$

Thus

$$\pi(2^k)\le(4\log 2)\frac{2^k}{\log 2^k}.$$

Finally, given n, let k be the integer with $2^{k-1}<n\le 2^k$. Then

$$\pi(n)\le\pi(2^k)\le(4\log 2)\frac{2n}{\log n}=8\log 2\frac{n}{\log n}.$$

□

Corollary 4.11. *For all $n\in\mathbb{N}$ there exists at least one prime p with $n<p\le 40n$.*

Proof. Exercise - Use the bounds of Theorem 4.10 to show that $\pi(40n)-\pi(n)>0$ if $n\ge 40$. □

In fact a stronger property is true. 'Bertrand's postulate' states that for all $n\in\mathbb{N}$ there is a prime p with $n\le p\le 2n$. It is conjectured, but not known, that there is always a prime betwteen n^2 and $(n+1)^2$.

Formulae for Primes

Is there a formula that gives prime numbers? Euler noted that n^2-n+41 is prime for $0\le n\le 40$ (why not for $n=41$?), and $n^2-79n+1601$ is prime for $0\le n\le 79$.

We can artificially create a formula for primes.

Proposition 4.12. *Let* $\alpha = \sum_{n=1}^{\infty} p_n 10^{-2^n} = 0.0203000500000070\ldots$ *Then*

$$p_n = \lfloor 10^{2^n}\alpha \rfloor - 10^{2^{n-1}} \lfloor 10^{2^{n-1}}\alpha \rfloor$$

where $\lfloor x \rfloor$ *denotes the largest integer* $\leq x$.

In fact no polynomial can just give prime values.

Proposition 4.13. *There is no polynomial* $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ *(with integer coefficients) such that* $f(n)$ *is prime for all* $n \in \mathbb{N}$.

Proof. We may assume $a_n > 0$. Suppose $f(1) = p$, a (positive) prime. Then for each $k \in \mathbb{N}$

$$\begin{aligned} f(1+kp) &= a_n(1+kp)^n + a_{n-1}(1+kp)^{n-1} + \cdots + a_1(1+kp)^n + a_0 \\ &= a_n + a_{n-1} + \cdots + a_1 + a_0 + pg(p) \\ &= f(1) + pg(p) = p(1+g(p)) \end{aligned}$$

where g is some polynomial with integer coefficients. If k is large enough $p < f(1+kp)$ and this is divisible by p so is not a prime. □

Unsolved problems on the distribution of primes

1. Find some very large primes: largest known to date is $2^{43112609} - 1$ which has 12978189 digits.
2. Twin primes: Are their infinitely many prime pairs, e.g. (3,5), (17,19), (29,31) ...?
3. Goldbach's conjecture: Every $n \geq 2$ is the sum of two primes. E.g. 12= 5+7.
4. Are there infinitely many primes of the form $k^2 + 1$?
5. Is there always a prime betwteen n^2 and $(n+1)^2$?

The Riemann Hypothesis

The Riemann Hypothesis is undoubtedly the most important unsolved question in pure mathematics. (There is a prize of \$1m for a valid solution.) If it is true then immediately a great deal more would be known about the distribution of primes, for example it would imply that

$$\left| \pi(n) - \int_0^n \frac{dt}{\log t} \right| < \frac{1}{8\pi}\sqrt{n}\log n$$

(for $n \geq 2657$) which is a much more precise result than the prime number theorem.

The Riemann zeta function is defined by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{3^s} + \cdots \quad (s \in \mathbb{C});$$

this series converges if $\mathrm{Re}(s) > 1$ and the function can be extended to the rest of the complex plane in a natural way (by 'analytic continuation').

It is not hard to show that $\zeta(s) = 0$ if s is a negative even integer, i.e. $s = -2, -4, -6, \ldots$. The Riemann Hypothesis states that *all other zeros of* $\zeta(s)$ *have real part* $\mathrm{Re}(s) = \frac{1}{2}$. There is enormous computational and, indeed mathematical, evidence to support this, but a proof still seems a long way off.

The relationship between $\zeta(s)$ and the prime numbers is given by the 'Euler product'

$$\begin{aligned}
\prod_{p \text{ prime}} \frac{1}{1-p^{-s}} &= \frac{1}{1-2^{-s}} \frac{1}{1-3^{-s}} \frac{1}{1-5^{-s}} \frac{1}{1-7^{-s}} \cdots \\
&= (1+2^{-s}+2^{-2s}+\cdots)(1+3^{-s}+3^{-2s}+\cdots)(1+5^{-s}+5^{-2s}+\cdots)\cdots \\
&= \frac{1}{1^s}+\frac{1}{2^s}+\frac{1}{3^s}+\frac{1}{4^s}+\frac{1}{5^s}\cdots = \zeta(s).
\end{aligned}$$

4.2 Irrational Numbers

Definition 4.14. *A real number is* irrational *if it is not rational.*

Example 4.15. $\sqrt{2}$ *is irrational (See Proposition 3.13).*

Example 4.16. *Prove that* $\sqrt{3}$ *and* $\sqrt{6}$ *are both irrational.*

More generally, we have:

Theorem 4.17. *For any natural numbers* m,n, *either* $\sqrt[m]{n}$ *is a natural number or else it is irrational.*

Proof. Suppose $\sqrt[m]{n}$ is rational, then we can write $\sqrt[m]{n} = a/b$ with $\mathrm{HCF}(a,b) = 1$. If $p \geq 2$ is any prime factor of b then $p|b^m n = a^m$ giving that $p|a$ by Proposition 4.4. Hence every prime factor of b must also be a prime factor of a, so since a and b are coprime, $b = 1$ making $\sqrt[m]{n} = a/b$ an integer. □

Example 4.18. $\sqrt{2}+\sqrt{3}$ *is irrational. For otherwise* $\sqrt{2}+\sqrt{3} = r \in \mathbb{Q}$, *so*

$$(\sqrt{2}+\sqrt{3})^2 = r^2 \quad \Rightarrow \quad 2+2\sqrt{6}+3 = r^2 \quad \Rightarrow \quad \sqrt{6} = \frac{r^2-5}{2},$$

implying that $\sqrt{6}$ *is rational, a contradiction.*

More generally again, we may consider roots of polynomials with integer coefficients.

Theorem 4.19. *Let* x *satisfy the equation*

$$x^n + c_{n-1}x^{n-1} + \ldots + c_1 x + c_0 = 0, \tag{4.5}$$

where $c_0, \ldots, c_{n-1} \in \mathbb{Z}$. *Then either* x *is irrational, or* x *is an integer in which case* $x \mid c_0$.

Proof. If x is not irrational, suppose that $x = a/b$, where $\mathrm{HCF}(a,b) = 1$, substitute into (4.5) and multiply by b^n:

$$a^n + c_{n-1}a^{n-1}b + \ldots + c_1ab^{n-1} + c_0b^n = 0, \tag{4.6}$$

so

$$a^n = b(-c_{n-1}a^{n-1} - \ldots - c_1ab^{n-2} - c_0b^{n-1}).$$

Hence if p is a prime such that $p|b$ then $p|a$. But since a and b are co-prime, this can happen only if b has no prime divisors, i.e. if $b = 1$, in which case $x = a$ is an integer.

Moreover, (4.6) can now be written as

$$c_0 = a(-a^{n-1} - c_{n-1}a^{n-2} - \ldots - c_1),$$

implying that a divides c_0. □

4.3 e and π

In this section we will define the two famous real numbers e and π, and prove that they are both irrational.

Definition 4.20. $e = \exp(1) = \sum_{n=0}^{\infty} \frac{1}{n!} = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \ldots$.
(Note that this series converges.)

Theorem 4.21. *e is irrational.*

Proof. For a contradiction, suppose that $e = a/b \in \mathbb{Q}$, $a,b \in \mathbb{Z}^+$. Then

$$\begin{aligned}
0 \ < \ e - \left(1 + \frac{1}{1!} + \frac{1}{2!} + \ldots + \frac{1}{b!}\right) & \\
&= \frac{1}{(b+1)!} + \frac{1}{(b+2)!} + \frac{1}{(b+3)!} + \ldots \\
&= \frac{1}{b!}\left(\frac{1}{b+1} + \frac{1}{(b+1)(b+2)} + \frac{1}{(b+1)(b+2)(b+3)} + \ldots\right) \\
&< \frac{1}{b!}\left(\frac{1}{b+1} + \frac{1}{(b+1)^2} + \frac{1}{(b+1)^3} + \ldots\right) \\
&= \frac{1}{b!}\left(\frac{1}{b+1} \cdot \frac{1}{1 - \frac{1}{b+1}}\right) \ = \ \frac{1}{b!}\frac{1}{b} \ < \ \frac{1}{b!}.
\end{aligned}$$

But

$$0 \ < \ e \ - \left(1 + \frac{1}{1!} + \frac{1}{2!} + \ldots + \frac{1}{b!}\right) = \frac{a}{b} \ - \left(1 + \frac{1}{1!} + \frac{1}{2!} + \ldots + \frac{1}{b!}\right) = \frac{m}{b!} < \frac{1}{b!},$$

for some integer m which satisfies $0 < m < 1$, a contradiction. □

Irrationality of π is rather harder. We first note the analytic definition of π.

Definition 4.22. *π is the smallest positive zero of* $\sin(x)$.

This definition assumes quite a lot from analysis, including the definition of 'sin' (formally in terms of the sine series) and its properties such as continuity and that it does indeed have a positive zero. Further work is needed to relate this definition of π to the the geometry of the circle, i.e. that the area and circumference of a unit circle are π and 2π respectively. In the proof that follows we will also assume various calculus properties of sin, cos, etc.

Theorem 4.23. *π^2, and hence π itself, is irrational.*

Proof. For a given $n \in \mathbb{N}$ (to be specified later) we define a polynomial

$$p(x) = \frac{x^n(1-x)^n}{n!}. \tag{4.7}$$

Because of the factor x^n, the polynomial $p(x)$ has no terms of degrees $0, 1, \dots, n-1$. Hence we can write

$$p(x) = \frac{1}{n!}\sum_{k=n}^{2n} a_k x^k = \frac{1}{n!}\left(a_n x^n + a_{n+1}x^{n+1} + \cdots + a_{2n}x^{2n}\right), \tag{4.8}$$

where all the a_k are integers.

The following SubLemma summarises the properties of the polynomial p that we will need. We write $p^{(k)}(x)$ for the k-th derivative of p evaluated at x.

Lemma 4.24.

(1) *For* $0 < x < 1$ *we have* $0 < p(x) < 1/n!$.

(2) *For all $k \geq 0$, the numbers $p^{(k)}(0)$ and $p^{(k)}(1)$ are integers.*

Proof. (1) Follows immediately from (4.7).

(2) It follows from (4.8) that $p^{(k)}(0) = 0$ for $0 \leq k < n$ and for $k > 2n$. For $n \leq k \leq 2n$ we have

$$p^{(k)}(0) = \frac{k!}{n!}a_k \in \mathbb{Z}.$$

Now note that $p(x) = p(1-x)$. It follows on differentiating k times that

$$p^{(k)}(x) = (-1)^k p^{(k)}(1-x) \quad (k = 0, 1, 2, 3\dots).$$

Hence

$$p^{(k)}(1) = (-1)^k p^{(k)}(0) \in \mathbb{Z}.$$

□

Continuation of main proof. Let us now suppose that

$$\pi^2 = \frac{a}{b} \quad (a, b \in \mathbb{Z}^+) \tag{4.9}$$

is a rational. Define a function

$$\begin{aligned} G(x) &= b^n \left(\pi^{2n} p(x) - \pi^{2n-2} p^{(2)}(x) + \ldots + (-1)^n p^{(2n)}(x)\right) & (4.10)\\ &= a^n p(x) - b a^{n-1} p^{(2)}(x) + \ldots + (-1)^n b^n p^{(2n)}(x). & (4.11)\end{aligned}$$

Differentiating (4.10) twice,

$$G''(x) = b^n \left(\pi^{2n} p^{(2)}(x) - \pi^{2n-2} p^{(4)}(x) + \ldots\ldots + (-1)^{n-1}\pi^2 p^{(2n)}(x) + (-1)^n p^{(2n+2)}(x)\right).$$

Note that the right hand term in the above sum equals 0. It follows that

$$\begin{aligned} &\frac{d}{dx}\left(G'(x)\sin\pi x - \pi G(x)\cos\pi x\right)\\ &\quad = G''(x)\sin\pi x + \pi G'(x)\cos\pi x - \pi G'(x)\cos\pi x + \pi^2 G(x)\sin\pi x\\ &\quad = \left(G''(x) + \pi^2 G(x)\right)\sin\pi x\\ &\quad = b^n \pi^{2n+2} p(x)\sin\pi x\\ &\quad = \pi^2 a^n p(x)\sin\pi x,\end{aligned}$$

using (4.9). Dividing by π and integrating we obtain:

$$\pi\int_0^1 a^n p(x)\sin(\pi x)dx = \left[\frac{G'(x)\sin\pi x}{\pi} - G(x)\cos\pi x\right]_0^1 = G(0) + G(1), \tag{4.12}$$

which is an integer using (4.11) and Lemma 4.24 (2).

On the other hand, from Lemma 4.24 (1)

$$0 < \pi\int_0^1 a^n p(x)\sin(\pi x)dx < \pi\int_0^1 \frac{a^n}{n!}dx = \pi\frac{a^n}{n!}$$

But we may choose n large enough to ensure

$$0 < \pi\frac{a^n}{n!} < 1$$

which contradicts that (4.12) is an integer. □

Transcendental Numbers

In fact, for e and π a much stronger assertion than irrationality is true. A number $x \in \mathbb{R}$ (or $\mathbb{C}$) is called *algebraic* if it satisfies a polynomial equation with integer coefficients, i.e.

$$a_n x^n + \cdots + a_1 x + a_0 = 0 \quad (a_i \in \mathbb{Z}).$$

A number $x \in \mathbb{R}$ (or $\mathbb{C}$) that is not algebraic is called *transcendental*. If $x = a/b$ is rational then it satisfies $bx = a$ so is certainly algebraic. Thus transcendental numbers are irrational. It may be shown that e and π are both transcendental. In fact there is the following very general (hard!) theorem.

Theorem 4.25 (Hermite-Lindemann Transcendence Theorem)**.** *Let $a_1,\dots,a_n$ be non-zero algebraic numbers and let $\alpha_1,\dots,\alpha_n$ be distinct algebraic numbers. Then*

$$a_1e^{\alpha_1}+a_2e^{\alpha_2}+\cdots+a_ne^{\alpha_n}\neq 0.$$

Corollary 4.26. *e and π are transcendental.*

Proof. Since by the H-L theorem $a_ne^n+a_{n-1}e^{n-1}+\cdots+a_1e+a_0\neq 0$ for all integers a_i, e is transcendental.

By Euler's identity $e^{i\pi}+1=0$, so the H-L Theorem implies $i\pi$ is not algebraic, so π is not algebraic. □

Corollary 4.27. *There is no point (x,y) on the curve $y=e^x$ other than $(0,1)$ such that both x and y are algebraic. In particular the curve passes through no point with both coordinates rational except for $(0,1)$.*

Proof. Note that if $x\neq 0$ and y are both algebraic the H-L theorem implies that $ye^0-e^x\neq 0$. □

Finally, it is not known whether $\pi+e$, πe, π^e, 2^e, 2^π,... are irrational or not.

5 Countability and Cardinality

5.1 Size and Similarity of Sets

When do two sets have the same size? When is one set larger than another?

For finite sets, this is easy: two sets X and Y have the same size if they contain the same number of elements. The set X is larger than the set Y if X contains a greater number of elements. But, if you think about it, what is the 'number'? We determine the number of elements of a finite set by establishing a one-one correspondence between the set and another set for which we 'know' its number of elements (e.g. the 'set' of fingers on our hand). For example $\{1,2,3\}$ and $\{3,4,5\}$ have the same size because there is a one-one correspondence between them: $1 \leftrightarrow 3$, $2 \leftrightarrow 4$, $3 \leftrightarrow 5$. The set $\{1,2,3\}$ is smaller than the set $\{5,6,7,8\}$ because no such correspondence can be established, but there is an injection $1 \mapsto 5$, $2 \mapsto 6$, $3 \mapsto 7$.

What about infinite sets? For example, which set is larger: $\mathbb{N} = \{1,2,3,\dots\}$ or $\mathbb{N} \setminus \{1\} = \{2,3,4,\dots\}$? One may be tempted to say that $\mathbb{N}$ is larger as it contains all the elements of $\mathbb{N} \setminus \{1\}$ plus an extra element. But there is nevertheless a bijection or correspondance between the two sets of numbers:

$$\begin{array}{ccccccccc} \mathbb{N} & 1 & 2 & 3 & 4 & 5 & 6 & \dots \\ & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ \mathbb{N} \setminus \{1\} & 2 & 3 & 4 & 5 & 6 & 7 & \dots \end{array}$$

– removing an element from an infinite set does not change its size from this point of view! But how about $\mathbb{N}$ and $2\mathbb{N} = \{2,4,6,\dots\}$? This time we have removed 'lots' of elements: 1, 3, 5,.... Nonetheless, they still 'look the same'. More precisely, there is a bijection between them:

$$\begin{array}{cccccccc} \mathbb{N} & 1 & 2 & 3 & 4 & 5 & 6 & \dots \\ & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ 2\mathbb{N} & 2 & 4 & 6 & 8 & 10 & 12 & \dots \end{array}$$

This leads us to the definition of 'similarity' of sets.

Definition 5.1. *Two sets A and B are* similar *or have the* same cardinality *if there is a bijection $f : A \to B$, in which case we write $A \simeq B$.*

Proposition 5.2. *Similarity $\simeq$ is an equivalence relation.*

Proof. (Reflexive) The identity mapping $i_A : A \to A$, $i_A(x) = x$ is a bijection.

(Symmetric) If $f : A \to B$ is a bijection then so is $f^{-1} : B \to A$; see Proposition 2.19.

(Transitive) If $f : A \to B$ and $g : B \to C$ are bijections then so is their composition $g \circ f : A \to C$ see Proposition 2.21. □

Example 5.3. $\{1,2,3\} \simeq \{4,5,6\}$ *since $x \mapsto x+3$ is a bijection between the sets.*

$\{1,2,3,\dots\} \simeq \{4,5,6,\dots\}$ *since $x \mapsto x+3$ is a bijection between the sets.*

The ‘size’ of finite sets does not pose a problem.

Definition 5.4. *We say that a finite set A has* cardinality $n \in \mathbb{N}$ *if it is similar to* $\{1,2,\ldots,n\}$*, that is if there is a bijection from* $\{1,2,\ldots,n\}$ *to A, in which case we write* $|A| = n$.

Proposition 5.5. *Two finite sets A, B have the same cardinality if and only if $A \simeq B$.*

Proof. If A and B both have cardinality n then $\{1,2,\ldots,n\} \simeq A$ and $\{1,2,\ldots,n\} \simeq B$ so it follows from the properties of equivalence relations that $A \simeq B$.

If $A \simeq B$ and $|A| = n$ then $\{1,2,\ldots,n\} \simeq A$ so by transitivity $\{1,2,\ldots,n\} \simeq B$ so $|B| = n$. □

The situation with infinite sets gets rather more complicated as some examples indicate.

Example 5.6.

1. $\mathbb{N} \simeq \mathbb{N}\setminus\{1\} \equiv \{2,3,4,\ldots\}$ since the mapping $f : \mathbb{N} \to \mathbb{N}\setminus\{1\}$ given by $f(x) = x+1$ is a bijection.

2. $\mathbb{N} \simeq 5\mathbb{N} \equiv \{5,10,15,\ldots\}$ since $f : \mathbb{N} \to 5\mathbb{N}$ given by $f(n) = 5n$ is a bijection; thus we have the following 1-1 correspondance:

$$\begin{array}{cccccccc}
\mathbb{N} & 1 & 2 & 3 & 4 & 5 & 6 & \ldots \\
 & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\
2\mathbb{N} & 5 & 10 & 15 & 20 & 25 & 30 & \ldots
\end{array}$$

3. $\mathbb{N} \simeq \mathbb{Z}$ since $f : \mathbb{N} \to \mathbb{Z}$ given by

$$f(n) = \begin{cases} n/2 & \text{if } n \text{ is even,} \\ -((n-1)/2) & \text{if } n \text{ is odd.} \end{cases}$$

is a bijection. Thus we have the pairing:

$$\begin{array}{cccccccc}
\mathbb{N} & 1 & 2 & 3 & 4 & 5 & 6 & \ldots \\
 & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\
2\mathbb{N} & 0 & 1 & -1 & 2 & -2 & 3 & \ldots
\end{array}$$

4. $\mathbb{R} \simeq \mathbb{R}^+$ since $\exp : \mathbb{R} \to \mathbb{R}^+$ is a bijection.

5. $(-\pi/2, \pi/2) \simeq \mathbb{R}$ since $\tan : (-\pi/2, \pi/2) \to \mathbb{R}$ is a bijection.

Whilst these examples might seem reasonable, we will shortly encounter rather more surprising pairs of similar and non-similar sets. In particular, we will show that $\mathbb{N} \simeq \mathbb{Q}$ but $\mathbb{N} \not\simeq \mathbb{R}$. Thus the cardinality of $\mathbb{R}$ is ‘strictly larger’ than the cardinality of $\mathbb{N}$ and $\mathbb{Q}$, in other words the infinite sets $\mathbb{Q}$ and $\mathbb{R}$ are of different sizes.

5.2 Countable Sets

Definition 5.7. *An infinite set A is* countable *if $\mathbb{N} \simeq A$, that is if there exists a bijection $f : \mathbb{N} \to A$. We can think of such a bijection as a* list *or* enumeration *of the elements of A:*

$$\begin{array}{ccccccccc} \mathbb{N} & 1 & 2 & 3 & 4 & 5 & 6 & \ldots \\ & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ A & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & \ldots \end{array}$$

where each element of A appears exactly once as an a_i.

Thus to show that a set A is countable it is enough to show that we can *list*, *count*, or *enumerate* its elements as a sequence $(a_1, a_2, a_3, a_4, \ldots)$ so that each element of A occurs somewhere in the sequence. (Bear in mind that such a list is just a way of specifying a bijection $\mathbb{N} \to A$).

Alternatively, since $\simeq$ is an equivalence relation, if we can show that $A \simeq B$ for a set B that is known to be countable, then A must be countable.

Example 5.8.

1. $\mathbb{Z}$ is countable since we may enumerate $\mathbb{Z}$ as $0, 1, -1, 2, -2, 3, -3, 4, \ldots$.

2. $\mathbb{Q} \cap (0,1)$, i.e. the set of rational numbers between 0 and 1, is countable since it may be enumerated in the logical sequence:

$$\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \cancel{\frac{2}{4}}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \frac{1}{6}, \cancel{\frac{2}{6}}, \ldots$$

where we delete any fraction that has already appeared in another form. (We will see later that $\mathbb{Q}$ itself is countable.)

3. The set $\mathbb{N} \times \mathbb{N}$ of all pairs of natural numbers is countable. To see this write $\mathbb{N} \times \mathbb{N}$ in an array as below (as coordinates of points in the plane with both coordinates natural numbers) and enumerate in the manner indicated by superscripts:

$$\begin{array}{ccccccccc} \vdots & \nwarrow & \vdots & & \vdots & & \vdots & & \\ {}^{10}(4,1) & & {}^{14}(4,2) & & {}^{19}(4,3) & & {}^{25}(4,4) & & \ldots \\ & \nwarrow & & \nwarrow & & \nwarrow & & & \\ {}^{6}(3,1) & & {}^{9}(3,2) & & {}^{13}(3,3) & & {}^{18}(3,4) & & \ldots \\ & \nwarrow & & \nwarrow & & \nwarrow & & & \\ {}^{3}(2,1) & & {}^{5}(2,2) & & {}^{8}(2,3) & & {}^{12}(2,4) & & \ldots \\ & \nwarrow & & \nwarrow & & \nwarrow & & \nwarrow & \\ {}^{1}(1,1) & & {}^{2}(1,2) & & {}^{4}(1,3) & & {}^{7}(1,4) & & \ldots \end{array}$$

Thus we may enumerate $\mathbb{N} \times \mathbb{N}$ as

$$(1,1), (1,2), (2,1), (1,3), (2,2), (3,1), (1,4), (2,3), (3,2), (4,1), (1,5), \ldots.$$

More formally, the mapping $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$

$$f(a,b) = \sum_{k=1}^{a+b-2} k + a$$

is a bijection that gives the position of (a,b) in the list.

4. By writing pairs in an array In the same way as (3) we may show that if A and B are countable then the product $A \times B = \{(a,b) : a \in A, b \in B\}$ is countable. It follows that $\mathbb{Z}$, $\mathbb{Z} \times \mathbb{Z}$, $\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}, \ldots$ are all countable.

Sometimes it is easier to set up an injection or surjection than to find a bijection between $\mathbb{N}$ and a given set.

Proposition 5.9. *(i) Every subset of a countable set is countable or finite.*
(ii) If $f : A \to B$ is a surjection and A is countable then B is countable or finite.
(iii) If $f : A \to B$ is an injection and B is countable then A is countable or finite.

Proof. (i) If A is countable we may list its elements as $(a_1, a_2, a_3, \ldots)$. Any subset may be enumerated by deleting elements not in the subset, i.e. as $(a_{i_1}, a_{i_2}, a_{i_3}, \ldots)$ where $1 \leq i_1 < i_2 < i_3 < \cdots$ and this will either terminate or give an enumeration of a countable set.

(ii) List A as $(a_1, a_2, a_3, \ldots)$. Then define a subset $A' \subseteq A$ by $A' = \{a_i \in A$ such that $f(a_i) \neq f(a_j)$ for all $1 \leq j < i\}$. Then $A' \simeq B$ and by (i) A' is countable or finite, so B is countable or finite.

(iii) We have $A \simeq f(A) \subseteq B$, so using (i) $f(A)$ and thus A is countable or finite. □

Example 5.10.
The map $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ given by $f(a,b) = 2^a 3^b$ is an injection, since if $2^a 3^b = 2^{a'} 2^{b'}$ then $(a,b) = (a', b')$ by unique factorisation. By Proposition 5.9 (iii) $\mathbb{N} \times \mathbb{N}$ is finite. (This is an alternative to the direct proof above).

5.3 Some uncountable sets

An infinite set that is not countable is called *uncountable*. We start with Cantor's classical 'diagonal argument' that demonstrates that the real numbers are uncountable.

Proposition 5.11. $\mathbb{R} \cap (0,1)$ *is uncountable and so* $\mathbb{R}$ *is uncountable.*

Proof. Suppose, for a contradiction, that $\mathbb{R} \cap (0,1)$ is countable, so that we may enumerate its elements as a list $(a_1, a_2, a_3, a_4, \ldots)$ which must contain every number in $(0,1)$. We may express these numbers in decimal form

$$\begin{array}{rcl} a_1 & = & 0.\underline{a_{11}}\,a_{12}\,a_{13}\,a_{14}\ldots \\ a_2 & = & 0.a_{21}\,\underline{a_{22}}\,a_{23}\,a_{24}\ldots \\ a_3 & = & 0.a_{31}\,a_{32}\,\underline{a_{33}}\,a_{34}\ldots \\ a_4 & = & 0.a_{41}\,a_{42}\,a_{43}\,\underline{a_{44}}\ldots \\ \vdots & & \quad\vdots \end{array}$$

so that a_{ij} is the jth decimal digit of a_i. (If a_i is a number with two decimal expansions, take the one ending in a string of 0s rather than that ending in a string of 9s.)

Define

$$b = 0.b_1\,b_2\,b_3\,b_4\ldots \text{ where } b_j = \begin{cases} 5 & \text{if} \quad a_{ii} \neq 5 \\ 7 & \text{if} \quad a_{ii} = 5 \end{cases}.$$

Then $b \neq a_i$ for all i, since $b_i \neq a_{ii}$, that is b differs from a_i in the ith decimal place. Thus b is not in the list, which contradicts the assumption that the list contains all real numbers in $(0,1)$. □

Recall that $\mathcal{P}(A)$ denotes the power set of A, that is the set of all subsets of A. The following result, which is really just a variant of the previous one, intuitively says that the power set of a set A has cardinality strictly larger than that of A itself.

Theorem 5.12. *Let A be a non-empty set. Then $\mathcal{P}(A) \not\approx A$.*

Proof. Suppose, for a contradiction, that there exists a bijection $f : A \to \mathcal{P}(A)$. Let $B = \{x \in A \text{ such that } x \notin f(x)\}$. Since f is a bijection, $B = f(a)$ for some $a \in A$.

From the definition of B, $a \in B$ iff $a \notin f(a) = B$, a contradiction. Thus there that there can be no bijection from A to B. □

It follows, at least intuitively, that, given any infinite set there is a strictly larger one, so there are infinitely many 'different sizes' of infinity.

The following property, often stated as 'A countable union of countable sets is countable', is useful in showing certain sets are countable.

Theorem 5.13 (Cantor's Theorem)**.** *Let $\{A_i\}_{i\in I}$ be a countable family of countable (or finite) sets. Then $\bigcup_{i\in I} A_i$ is countable.*

Proof. For each $i = 1,2,3\ldots$ let $A_i = \{a_{i1}, a_{i2}, a_{i3}, \ldots\}$. We could now enumerate $\bigcup_{i\in I} A_i$ in a similar manner to Example 5.8 (3). Alternatively, we may define an injection $f : \bigcup_{i\in I} A_i \to \mathbb{N}$ by $f(a_{ij}) = 2^i 3^j$ so that countability follows from Propostion 5.9 (iii). □

Example 5.14. *(i) The set $\mathbb{Q}$ of all rationals is countable.*

(ii) The set of all algebraic numbers is countable.

Proof. (i) For $n = 1,2,3,\ldots$ let $A_n = \{r/n : r \in \mathbb{Z}\}$. Then A_n is countable, so by Theorem 5.13 $\mathbb{Q} = \bigcup_{n=1}^{\infty} A_n$ is countable.

(ii) For $n = 1,2,3,\ldots$ let A_n be the set of all zeros of polynomials of degree $\leq n$. There are countably many such polynomials (since the list of coefficients are in bijective correspondence with the $(n+1)$-fold product $\mathbb{Z} \times \cdots \times \mathbb{Z}$), and each such polynomial has at most n distinct roots, so each A_n is countable. But the set of algebraic numbers is $\bigcup_{n=1}^{\infty} A_n$ which is countable by Theorem 5.13. □

Example 5.15. *(i) The set of all subsets of* $\mathbb{N}$ *is uncountable.*

(ii) However, the set of all finite *subsets of* $\mathbb{N}$ *is countable.*

Proof. (i) This follows from Theorem 5.12.

(i) For each n let A_n be the set of all subsets of $\{1,2,\ldots,n\}$. Then A_n is finite; indeed $|A_n| = 2^n$. The set of all finite subsets of $\mathbb{N}$ is just $\bigcup_{n=1}^{\infty} A_n$ so is countable by Theorem 5.13. □

5.4 Cardinality

We return to thinking of cardinality as representing the size of sets – can the idea of $|A|$ denoting the number of elements in a finite set A be extended in a meaningful manner to infinite sets? We recall and extend the Definition 5.1 to allow comparison as well as equality of cardinalities – we start to think of $|A|$ as the size of A even if A is infinite.

Definition 5.16. *Two sets A and B have the* same cardinality *or are* similar *if there is a bijection* $f : A \to B$*, in which case we now write* $|A| = |B|$.

We say that the set A has cardinality less than or equal to *B if there is an injection* $f : A \to B$*, and we write this as* $|A| \leq |B|$*. We say that the set A has* cardinality strictly less than *B if* $|A| \leq |B|$ *and* $|A| \neq |B|$ *in which case we write* $|A| < |B|$

Example 5.17. $|\mathbb{Z}| < |\mathbb{R}|$*, since* $f : \mathbb{Z} \to \mathbb{R}$ *given by* $f(n) = n$ *is an injection and* $|\mathbb{Z}| \neq |R|$.

Theorem 5.18. *Let A, B, C be sets. Then*

(i) $|A| = |A|$,

(ii) If $|A| = |B|$ *then* $|B| = |A|$,

(iii) If $|A| = |B|$ *and* $|B| = |C|$ *then* $|A| = |C|$,

(iv) $|A| \leq |A|$,

(v) If $|A| \leq |B|$ *and* $|B| \leq |C|$ *then* $|A| \leq |C|$.

Proof. Parts (i)-(iii) are just a restatement of Proposition 5.1 concerning properties of bijections.

Part (iv) follows since the identity map is an injection, and (v) follows since if $f : A \to B$ and $g : B \to C$ are injections then $g \circ f : A \to C$ is an injection. □

Despite the notation ‘$=$’ and ‘$\leq$’ with its intuitive connotations, one thing is missing from the above list of properties, namely that if $|A| \leq |B|$ and $|B| \leq |A|$ then $|A| = |B|$. Without this ‘antisymmetry’ property one might have both $|A| < |B|$ and $|B| < |A|$ and cardinality would be a rather limited notion. Of course, this can be proved, but it is a serious theorem known as the Schroeder–Bernstein Theorem.

Theorem 5.19 (Schroeder–Bernstein)**.** *Given two sets A and B, if there exist injections $f : A \to B$ and $g : B \to A$ then there exists a bijection $h : A \to B$. More succinctly:*

$$|A| \leq |B| \text{ and } |B| \leq |A| \implies |A| = |B|.$$

Proof. If $a \in A$ is such that $f(a) = b$, call a the *parent* of b. Similarly, $b \in B$ is the parent of $c \in A$ if $g(b) = c$. (Notice that the injectivity of f and g means that an element can have at most one parent.)

Let $z \in A \cup B$. An *ancestral chain* for z is a sequence $z_0, z_1, \ldots$ such that $z_0 = z$ and z_{i+1} is the parent of z_i for each i. (An ancestral chain may be of finite or infinite length.) If there is no infinite ancestral chain for z, then the *depth* of z is the index of the last element in the unique longest ancestral chain for z; otherwise z has *infinite depth*. (Observe that z may have depth 0.)

Let A_e, B_e be the subsets of A, B consisting of even-depth elements; A_o, B_o be their subsets consisting of odd-depth elements; and A_∞, B_∞ be their subsets consisting of infinite-depth elements.

Notice that f maps A_e to B_o, A_o to B_e, and A_∞ to B_∞. Similarly, g maps B_e to A_o, B_o to A_e, and B_∞ to A_∞.

Observe that elements of $A_o \cup B_o \cup A_\infty \cup B_\infty$ always have parents; this may not be true for elements of $A_e \cup B_e$, since an element of this set may have depth 0.

Define $h : A \to B$ by

$$h(a) = \begin{cases} f(a) & \text{if } a \in A_e \cup A_\infty, \\ g^{-1}(a) & \text{if } a \in A_o. \end{cases}$$

This mapping is defined everywhere since $g^{-1}(a)$ exists for all $a \in A_o$ and is unique by the injectivity of g.

Suppose $a_1, a_2 \in A$ are such that $h(a_1) = h(a_2)$. If $h(a_1) = h(a_2)$ lies in B_e, then $a_1, a_2 \in A_o$. So $g^{-1}(a_1) = h(a_1) = h(a_2) = g^{-1}(a_2)$. So $a_1 = g(g^{-1}(a_1)) = g(g^{-1}(a_2)) = a_2$. If $h(a_1) = h(a_2)$ lies in $B_o \cup B_\infty$, then $a_1, a_2 \in A_e \cup A_\infty$. So $f(a_1) = h(a_1) = h(a_2) = f(a_2)$. But f is injective, so $a_1 = a_2$. Thus h is injective.

Choose $b \in B$. If $b \in B_e$, let $a = g(b)$. Then $h(a) = g^{-1}(g(b)) = b$. If $b \in B_o \cup B_\infty$, then b has a parent $a \in A$ with $f(a) = b$. In fact, a must lie in $A_e \cup A_\infty$, so $h(a) = f(a) = b$. Thus h is surjective.

Therefore h is a bijection from A to B and thus $|A| = |B|$. □

This fact that $\leq$ is a total order allows us to extend some of our earlier results.

Corollary 5.20. *(i) Let A be a non-empty set. Then* $|A| < |\mathcal{P}(A)|$.

(ii) There are infinitely many different infinite cardinalities.

(iii) There is no largest cardinality.

(iv) There is no set containing all sets.

Proof. (i) We showed in Theorem 5.12 that $|A| \neq |\mathcal{P}(A)|$, but clearly $|A| \leq |\mathcal{P}(A)|$ since $a \mapsto \{a\}$ is an injection. Then $|A| < |\mathcal{P}(A)|$.

(ii) We may define a sequence by $A_1 = \mathbb{N}$ and $A_n = \mathcal{P}(A_{n-1})$ for $n \geq 2$. By (i) $|A_1| < |A_2| < |A_3| < \cdots$.

(iii) This follows from (i).

(iv) If there was such a set it would have to have cardinality strictly greater than its own, by (i). □

So far we have not given a meaning to $|A|$ outside the context of '=' and '$\leq$'. Motivated by finite sets, we can think of $|A|$ as an 'infinite number' or *cardinal* representing the cardinality of A, and so of any set B such that $|A| \simeq |B|$.

Example 5.21. *We write* $\aleph_0$ *('aleph nought') for the cardinality of* $\mathbb{N}$ *so a set is countable if* $|A| = \aleph_0$. *Thus* $|\mathbb{N}| = |\mathbb{Q}| = \aleph_0$.

We write $\mathfrak{c}$ *for the cardinality of* $\mathbb{R}$ *or 'cardinality of the continuum' . Thus* $|\mathbb{R}| = |\mathbb{C}| = \mathfrak{c}$ *and* $\aleph_0 < \mathfrak{c}$.

It is possible to define an arithmetic on cardinals:

Definition 5.22. *Let A and B be disjoint infinite sets. The sum and product of cardinals* $|A|$ *and* $|B|$ *is defined by*

$$|A| + |B| = |A \cup B|, \; |A| \cdot |B| = |A \times B|.$$

It may be checked that these operations are well-defined, i.e. independent of which sets of given cardinality are chosen. The basic arithmetic for infinite cardinals is very simple:

Theorem 5.23. *For any two infinite sets A and B we have*

$$|A| + |B| = |A| \cdot |B| = \max(|A|, |B|).$$

Typically, if you are given specific A and B this is reasonably easy to prove. However, proving the statement in general is quite difficult, and is beyond our scope here.

Example 5.24. $|\mathbb{C}| = |\mathbb{R} \times \mathbb{R}| = \mathfrak{c} \times \mathfrak{c} = \mathfrak{c}$.

The Continuum Hypothesis

Does there exist a cardinal $\mathfrak{a}$ such that $\aleph_0 < \mathfrak{a} < \mathfrak{c}$, or to put it another way, is there a set X such that $|\mathbb{N}| < |X| < |\mathbb{R}|$? That is, is there an uncountable set that is 'smaller' than the reals?

The conjecture, originally made by Cantor, that no such set exists is called the *Continuum Hypothesis*.

Gödel showed that one cannot *disprove* the Continuum Hypothesis using standard mathematical logic, and later Cohen showed that one cannot *prove* the Continuum Hypothesis using standard mathematical logic.

Therefore the Continuum Hypothesis is independent of the usual axioms of mathematics and one can choose whether to assume that the Continuum Hypothesis is true or that its negation is true.