
Contents

Overview	1
Weeks 6-10: Analysis	1
Weeks 11-15: Number systems	3
Chapter 1. The real numbers	5
§1.1. Intuitive picture of \mathbb{R} as points on the number line	6
§1.2. The field axioms	10
§1.3. Order axioms	11
§1.4. The Least Upper Bound Property of \mathbb{R}	12
§1.5. Intervals	23
§1.6. Absolute value $ \cdot $ and distance in \mathbb{R}	24
§1.7. (*) Remark on the construction of \mathbb{R}	27
Appendix: Binomial theorem	28
Chapter 2. Sequences and their convergence	31
§2.1. Limit of a convergent sequence	33
§2.2. Bounded and monotone sequences	41
§2.3. Algebra of limits	47
§2.4. Sandwich theorem	52
§2.5. Subsequences	56
Appendix (*)	63
Chapter 3. Continuity	65
§3.1. Definition of continuity	65
§3.2. Continuous functions preserve convergence	70
§3.3. Intermediate Value Theorem	76

§3.4. Extreme value theorem	79
Chapter 4. Number systems	85
§4.1. Equivalence relations	85
§4.2. Natural numbers	88
§4.3. Integers	93
§4.4. Rational numbers	101
§4.5. Real numbers	104
§4.6. Irrational numbers and the Rational Zeroes Theorem	112
§4.7. Cardinality	115
Appendix: Proof of the Least Upper Bound Property (*)	123
Chapter 5. The ring of integers	127
§5.1. The Division Algorithm	127
§5.2. Divisibility, gcd, and the Euclid Algorithm	128
§5.3. Prime numbers and the Fundamental Theorem of Arithmetic	134
§5.4. Modular arithmetic and the ring \mathbb{Z}_n	137
Solutions	143
Solutions to the exercises from Chapter 1	143
Solutions to the exercises from Chapter 2	156
Solutions to the exercises from Chapter 3	174
Solutions to the exercises from Chapter 4	187
Solutions to the exercises from Chapter 5	212
Bibliography	225
Index	227

Overview

Weeks 6-10: Analysis

In weeks 6-10, we will learn some of the basic concepts from the subject of ‘Mathematical Analysis’ or briefly ‘Analysis’. Roughly speaking, Analysis is Calculus made rigorous.

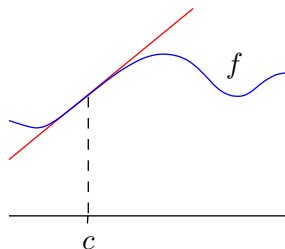
We are familiar with Calculus as a branch of mathematics in which the focus is on two main things: Given a real-valued function of a real variable,

- (Differentiation) What is the rate of change of the function at a point?
- (Integration) What is the area under the graph of the function over an interval?

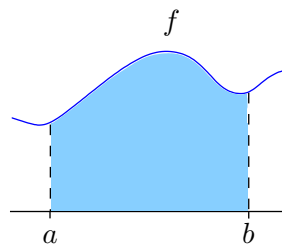
Differentiation

and

Integration



What is the slope of f
at the point c ?



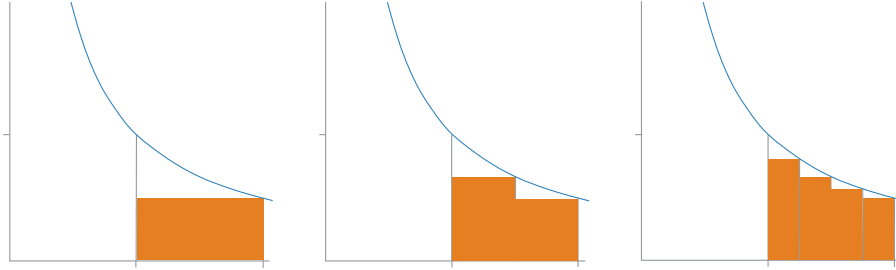
What is the area under the graph of f
over an interval from a to b ?

In Calculus, while there are these two quite different topics of study, the Fundamental Theorem of Calculus is a bridge between these different worlds, saying that the processes of differentiation and integration are inverses of each other:

$$\int_a^b f'(x)dx = f(b) - f(a) \quad \text{and} \quad \frac{d}{dx} \int_a^x f(\xi)d\xi = f(x).$$

This interaction between differentiation and integration provides a powerful body of understanding and calculational technique, called ‘Calculus’.

A thorough treatment of Calculus, i.e., of the subject of Analysis, must start with a careful study of the number system in which the action takes place, namely the set \mathbb{R} of real numbers. To see why, let's consider an example. Suppose we want to find the area under the graph of the function $f(x) = 1/x$ from $x = 1$ to $x = 2$.



We are trying to obtain the area by approximating it via the sum of rectangular areas, each time doubling the number of rectangles, hence ‘exhausting’ more and more of the required area. The area at the n^{th} stage is

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

The idea is then that if A is the area we seek, and a_n is the area at the n th step, then for large n , a_n approximates A . Clearly $a_1 \leq a_2 \leq a_3 \leq \cdots$, and they are all less than some big number¹. Since a_n misses A by smaller and smaller amounts as n increases, we expect that A should be the ‘smallest’ number exceeding the numbers a_1, a_2, a_3, \dots . Does such a number always exist? We seem to need the fact that

$$(F) \left\{ \begin{array}{l} \text{For an increasing sequence } a_1, a_2, a_3, \dots, \text{ of numbers} \\ \text{all of which are less than a certain number,} \\ \text{there is a smallest number which is bigger than each of } a_1, a_2, a_3, \dots \end{array} \right.$$

Note that each $a_n \in \mathbb{Q}$ (set of rationals). Does (F) hold for rational numbers?

This question might seem frivolous to a scientist who is just interested in ‘real world applications’. But such a sloppy attitude can lead to trouble. Indeed, results in Calculus during the 16th to the 18th century relying on a mixture of deductive reasoning and intuition, involving vaguely defined terms, were later shown to be *incorrect*. To give a quick example of how things might easily go wrong, one might naively, but incorrectly², guess that the answer to the question above is ‘yes’. This prompts the question of whether there is a bigger set of numbers than the rational numbers for which the property happens to be true? The answer is ‘yes’, and this is the **real number system** \mathbb{R} .

¹Take a square of height and width 1. Then each a_n is less than (the area of this square, which is) 1.

²In fact, for our sequence a_1, a_2, a_3, \dots above, there is no smallest rational number which is bigger than each of the a_n s. If we consider the sequence in \mathbb{R} , then $A = \log_e 2$, which can be shown to be irrational!

So the subject of Analysis must start with a careful study of the real number system \mathbb{R} , and this is where our journey begins. After learning about the key properties of the real numbers, we will discuss two useful notions in Analysis:

- the concept of convergence of a sequence of real numbers, and
- the concept of continuity.

These are the first fundamental notions with which one can embark on a more detailed study of Analysis (to be continued in later courses such as MA203).

Weeks 11-15: Number systems

In the weeks 6-10, we begin our study of the subject of Analysis by stipulating carefully the properties of the real number system, and proceed from there. But we do not address the issue of what exactly the set of real numbers is, i.e., how one constructs it as a mathematical object. Thus, in some sense, in the first part we hit the ground running, accepting on faith the properties we need and making quick progress from that starting point.

In the second part, i.e., in the weeks 11-15, we will spend some time learning about the foundations of the number systems, starting with the natural number system \mathbb{N} of the ‘counting numbers’, and progressively enlarging the number system set whenever we meet an arithmetic hurdle, until we meet an ‘analytical’ hurdle with the rationals, which is finally remedied by the real number system:

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

We will first learn about the manner in which these number systems are constructed. Subsequently, we shall also delve deeper into the ‘algebraic structure’ of the integers obtained from its arithmetic, since it plays a fundamental role in all of Mathematics. In particular, we will learn about

- the Division Algorithm,
- divisibility, the greatest common divisor, and Euclid’s Algorithm,
- prime numbers, and the Fundamental Theorem of Arithmetic, and
- modular arithmetic.

Challenging exercises within these notes are indicated with an asterisk symbol (*). Nonexaminable sections or remarks are labelled by (*).

Before beginning, we fix some notation and terminology. If X, Y are sets, and f a function from X to Y , then we write $f : X \rightarrow Y$. We refer to the set X as the *domain* of f , the set Y as the *codomain* of f , and the set $f(X) := \{f(x) : x \in X\}$ as the *image/range* of f . Sometimes we write $X \ni x \mapsto f(x) \in Y$, which is read as ‘element x belonging to X is mapped to the element $f(x)$ belonging to Y ’.

Chapter 1

The real numbers

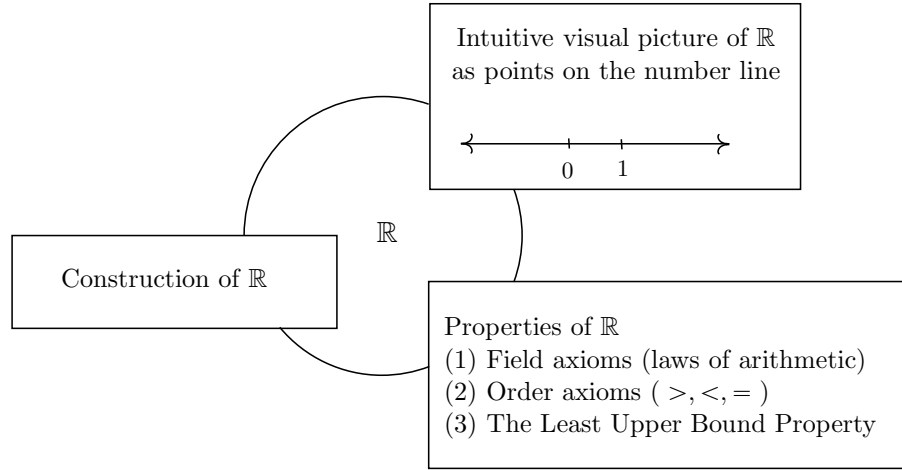
From the considerations in the overview, it is clear that one needs to begin the subject of Analysis by studying the real numbers carefully. The plan is as follows:

- (1) **An intuitive, visual picture of \mathbb{R} : the number line.** We will begin our understanding of \mathbb{R} intuitively as points on the ‘number line’. So we will have a mental picture of \mathbb{R} , in order to begin stating the precise properties of the real numbers that we will need later. It is a legitimate issue to worry about the construction of the set of real numbers, and we will say something about this in Section 1.7 (and do it more carefully in Chapter 4).
- (2) **Properties of \mathbb{R} .** Having a rough feeling for the real numbers as being points of the real line, we will proceed to state the precise properties of the real numbers we will need. So we will think of \mathbb{R} as a given (undefined) set for now, and just state rigorously what properties we need this set \mathbb{R} to have. These desirable properties¹ fall under three categories:
 - (a) the *field axioms*, which tell us about what laws the arithmetic of the real numbers should follow,
 - (b) the *order axiom*, telling us that comparison of real numbers is possible with an order $>$ and what properties this order relation has, and
 - (c) the *Least Upper Bound Property* of \mathbb{R} , which tells us roughly that unlike the set of rational numbers, the real number line has ‘no holes’. This property is the most important in Analysis. *Had* the set \mathbb{Q} of rationals possessed this property, then we wouldn’t have bothered studying \mathbb{R} , and instead we would have just used \mathbb{Q} for Calculus.
- (3) **The construction of \mathbb{R} .** Although we will think of real numbers intuitively as ‘numbers which can be depicted on the number line’, this is not acceptable as a rigorous mathematical definition. So we ask:

Is there a set \mathbb{R} that can be constructed with the properties (2)(a),(b),(c) above?

Answer: Yes. This will be outlined in Section 1.7 and detailed in Chapter 4.

¹These will be given in detail in Sections 1.2, 1.3, 1.4.



1.1. Intuitive picture of \mathbb{R} as points on the number line

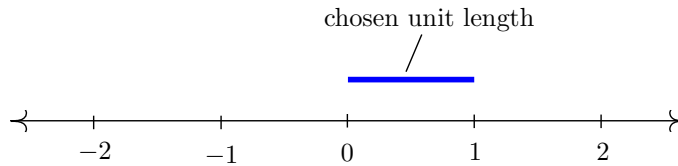
In elementary school, we learn about

the natural numbers $\mathbb{N} := \{1, 2, 3, \dots\}$

the integers $\mathbb{Z} := \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$, and

the rational numbers $\mathbb{Q} := \left\{ \frac{n}{d} : n, d \in \mathbb{Z}, d \neq 0 \right\}$,

and we are accustomed to visualizing these numbers on the ‘number line’. The *number line* is any line in the plane, on which we have chosen a point O as the ‘origin’, representing the number 0, and chosen a unit length by marking off a point on the right of O , where the number 1 is placed. In this way, we get all the positive integers, $1, 2, 3, 4, \dots$ by repeatedly marking off successively the unit length towards the right, and all the negative integers $-1, -2, -3, \dots$ by repeatedly marking off successively the unit length towards the left.



Just like the integers can be depicted on the number line, we can also depict all rational numbers on it as follows. Firstly, here is a procedure for dividing a unit length on the number line into d ($\in \mathbb{N}$) equal parts, allowing us to construct the rational number $1/d$ on the number line. See Figure 1. The steps are as follows.

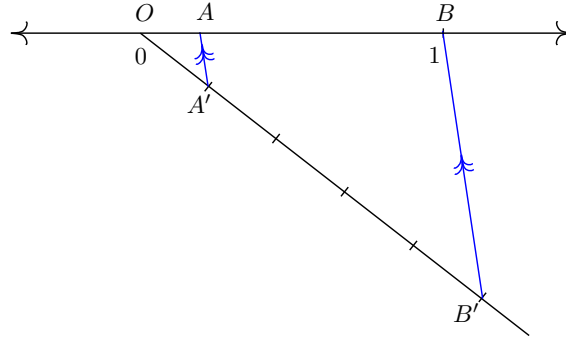


Figure 1. Construction of rationals: Given the length 1 ($= \ell(OB)$), we can construct the length $1/5$, and so A corresponds to the rational number $1/5$.

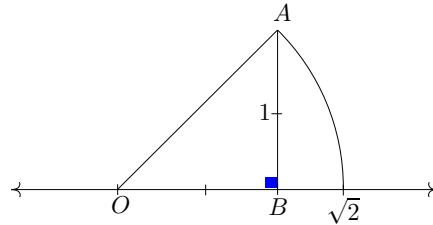
- (1) Take an arbitrary length $\ell(OA')$ along a ray starting at O in any direction other than that of the number line itself.
- (2) Let B' be a point on the ray such that $\ell(OB') = d \cdot \ell(OA')$.
- (3) Draw AA' parallel to BB' to meet the number line at A .

Conclusion: $\triangle OAA'$ is similar to $\triangle OBB'$, and so $\ell(OA) = 1/d$.

Having obtained $1/d$, we can now construct n/d on the number line for *any* $n \in \mathbb{Z}$, by repeating the length $1/d$ n times towards the right of 0 if $n > 0$, and towards the left $-n$ times from 0 if n is negative.

Hence we can depict all the rational numbers on the number line. Does this exhaust the number line? That is, suppose that we start with all the points on the number line being coloured black, and suppose that at a later time, we colour all the rational ones by red: are there any black points left over? The answer is ‘yes’, and we demonstrate this below. We will show that there does ‘exist’, based on geometric reasoning, a point on the number line, whose square is 2, but we will also argue that this number, denoted by $\sqrt{2}$, is not a rational number.

Firstly, the following picture shows that $\sqrt{2}$ exists as a point on the number line. Indeed, in the right angled triangle $\triangle OBA$, by the Pythagoras Theorem, we have $(\ell(OA))^2 = (\ell(OB))^2 + (\ell(AB))^2 = 1^2 + 1^2 = 2$, and so $\ell(OA)$ is a number, denoted say by $\sqrt{2}$, whose square is 2. Taking O as the center and radius $\ell(OA)$, we draw a circle intersecting the number line at a point C , corresponding to the number $\sqrt{2}$. Is $\sqrt{2}$ a rational number? We will now show that it isn’t!

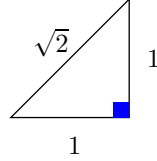


Exercise 1.1. Depict $-11/6$ and $\sqrt{3}$ on the number line.

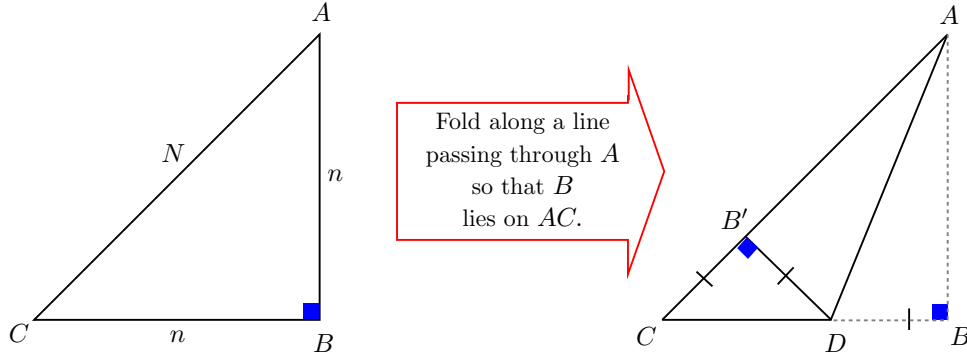
Theorem 1.1 (An ‘origami’ proof of the irrationality of $\sqrt{2}$).

There is no rational number $q \in \mathbb{Q}$ such that $q^2 = 2$.

Proof. Suppose that $\sqrt{2}$ is a rational number. Then some scaling of the triangle



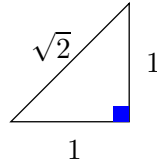
by an integer will produce a similar triangle, all of whose sides are integers. Choose the smallest such triangle, say $\triangle ABC$, with integer lengths $\ell(BC) = \ell(AB) = n$, and $\ell(AC) = N$, $n, N \in \mathbb{N}$. Now do the following origami: fold along a line passing through A so that B lies on AC , giving rise to the point B' on AC . The ‘crease’ in the paper is actually the angle bisector AD of the angle $\angle BAC$.



In $\triangle CB'D$, $\angle CB'D = 90^\circ$, $\angle B'CD = 45^\circ$. So $\triangle CB'D$ is an isosceles right triangle. We have $\ell(CB') = \ell(B'D) = \ell(AC) - \ell(AB') = N - n \in \mathbb{N}$, while

$$\ell(CD) = \ell(CB) - \ell(DB) = n - \ell(B'C) = n - (N - n) = 2n - N \in \mathbb{N}.$$

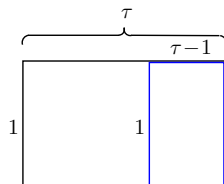
So $\triangle CB'D$ is similar to the triangle



has integer side lengths, and is smaller than $\triangle ABC$, contradicting the choice of $\triangle ABC$. So there is no rational number q such that $q^2 = 2$. \square

You had met a different proof earlier, and we will meet a different proof yet again in §4.6, based on the ‘Rational Zeroes Theorem’.

Exercise 1.2. (*) The number $\tau := \frac{1+\sqrt{5}}{2} \approx 1.618 \dots$ is called the *golden ratio*. This ratio is believed to create geometrical figures of particularly pleasing proportions, for example, the *golden rectangle*, with sides in the ratio $1 : \tau$. Using the smaller side, if a square is separated from the golden rectangle, we obtain yet another golden rectangle, thanks to the relation $\tau^2 - \tau - 1 = 0$, that is, $\tau - 1 = 1/\tau$.



Give a geometric proof that τ can't be rational. Conclude that $\sqrt{5}$ is not rational either.

Thus we have seen that the elements of \mathbb{Q} can be depicted on the number line, and that not all the points on the number line belong to \mathbb{Q} . We think of \mathbb{R} as *all* the points on the number line. As mentioned before, if we take out everything on the number line (the black points) except for the rational numbers \mathbb{Q} (the red points), then there will be holes amongst the rational numbers (for example there will be a missing black point where $\sqrt{2}$ lies on the number line). We can think of the real numbers as ‘filling in’ these holes between the rational numbers. We will say more about this when we make remarks about the construction of \mathbb{R} . Right now, we just have an intuitive picture of the set of real numbers as a bigger set than the rational numbers, and we think of the real numbers as points on the number line. Admittedly, this is certainly not a mathematical definition, and is extremely vague. In order to be precise, in Analysis we just can't rely on this vague intuitive picture of the real numbers. So we now turn to the precise properties of the real numbers which we are allowed to use. While stating these properties, we will think of the set \mathbb{R} as an (as yet) undefined set containing \mathbb{Q} which will satisfy the properties of

- (1) the field axioms (laws of arithmetic in \mathbb{R}),
- (2) the order axioms (allowing us to compare real numbers with $>$, $<$, $=$), and
- (3) the Least Upper Bound Property (making Calculus possible in \mathbb{R}),

stipulated below.

It is a pertinent question if one can construct (if there really exists) such a set \mathbb{R} satisfying the above properties (1), (2), (3). The answer to this question is ‘yes’, but it is tedious. So in this first part of the course, we will not worry ourselves too much with it². We will actually give some idea about the construction of the

²It is a bit like the process of learning a new language: If one starts painfully memorising systematically all the rules of grammar first then not much progress will be made. Instead, a more fruitful method is to start practicing simple phrases, listening to news, reading comics, and so on. Along the way grammar rules can be picked up, and a formal study can be done at leisure later, resulting in better comprehension.

real numbers in § 1.7, and return to it in §4.5. Right now, we just accept on faith that the construction of \mathbb{R} possessing the desired properties above can be done. To have a concrete object in mind, we rely on our familiarity with the number line to think of the real numbers when we study the properties (1), (2), (3) listed above.

We also remark that property (3) (the Least Upper Bound Property) of \mathbb{R} will turn out to be crucial in Analysis. The properties (1), (2) are also possessed by the rational number system \mathbb{Q} , but we will see that (3) fails for \mathbb{Q} .

1.2. The field axioms

This section's content³ can be summarised in one sentence: $(\mathbb{R}, +, \cdot)$ forms a field. What does this mean? It is a compact way of saying the following: \mathbb{R} is a set, equipped with two maps, namely

addition $+: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, sending a pair (x, y) of reals to their *sum* $x + y$, and

multiplication $\cdot: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, sending a pair of reals (x, y) to their *product* $x \cdot y$,

and these two operations $+$ and \cdot satisfy certain laws, called the ‘field axioms’⁴. The field axioms for \mathbb{R} are listed below:

$$\begin{aligned}
 + \left\{ \begin{array}{ll} \text{(F1) (Associativity)} & \text{For all } x, y, z \in \mathbb{R}, x + (y + z) = (x + y) + z. \\ \text{(F2) (Additive identity)} & \text{For all } x \in \mathbb{R}, x + 0 = x = 0 + x. \\ \text{(F3) (Inverses)} & \text{For all } x \in \mathbb{R}, \text{ there exists } -x \in \mathbb{R} \\ & \text{such that } x + (-x) = 0 = -x + x. \\ \text{(F4) (Commutativity)} & \text{For all } x, y \in \mathbb{R}, x + y = y + x. \end{array} \right. \\
 \cdot \left\{ \begin{array}{ll} \text{(F5) (Associativity)} & \text{For all } x, y, z \in \mathbb{R}, x \cdot (y \cdot z) = (x \cdot y) \cdot z. \\ \text{(F6) (Multiplicative identity)} & 1 \neq 0 \text{ and for all } x \in \mathbb{R}, x \cdot 1 = x = 1 \cdot x. \\ \text{(F7) (Inverses)} & \text{For all } x \in \mathbb{R} \setminus \{0\}, \text{ there exists } x^{-1} \in \mathbb{R} \\ & \text{such that } x \cdot x^{-1} = 1 = x^{-1} \cdot x. \\ \text{(F8) (Commutativity)} & \text{For all } x, y \in \mathbb{R}, x \cdot y = y \cdot x. \end{array} \right. \\
 +, \cdot \left\{ \begin{array}{ll} \text{(F9) (Distributivity)} & \text{For all } x, y, z \in \mathbb{R}, x \cdot (y + z) = x \cdot y + x \cdot z. \end{array} \right.
 \end{aligned}$$

With these axioms, it is possible to prove the usual arithmetic manipulations we are accustomed to. Here are a couple of examples.

Example 1.1. For every $a \in \mathbb{R}$, $a \cdot 0 = 0$.

Let $a \in \mathbb{R}$. Then we have $a \cdot 0 \stackrel{(F2)}{=} a \cdot (0 + 0) \stackrel{(F9)}{=} a \cdot 0 + a \cdot 0$. So with $x := a \cdot 0$, we have $x + x = x$. Adding $-x$ on both sides (F3), and using (F1) we obtain:

³This section has more detail that we need. The student may skip it, and begin reading from Section 1.4 onwards. For MA103 students, after learning about the construction of the real numbers in Chapter 4 (covered in the Lent Term), we will define real number addition and multiplication carefully and prove the properties (F1)-(F9).

⁴There are other number systems, for example the rational numbers \mathbb{Q} which also obey similar laws of arithmetic, and so $(\mathbb{Q}, +, \cdot)$ is also deemed to be a field. So the word ‘field’ is invented to describe the situation that one has a number system \mathbb{F} with corresponding operations $+: \mathbb{F} \times \mathbb{F} \rightarrow \mathbb{F}$ and $\cdot: \mathbb{F} \times \mathbb{F} \rightarrow \mathbb{F}$ which obey the usual laws of arithmetic, rather than listing all of these laws.

$$0 = x + (-x) = (x + x) + (-x) \stackrel{(F1)}{=} x + (x + (-x)) \stackrel{(F3)}{=} x + 0 \stackrel{(F2)}{=} x = a \cdot 0. \diamond$$

Example 1.2. If $a, b \in \mathbb{R}$, and $a \cdot b = 0$, then $a = 0$ or $b = 0$.

If $a = 0$, then we are done. Let $a \neq 0$. By (F7), there exists a real number a^{-1} such that $a \cdot a^{-1} = a^{-1} \cdot a = 1$. So $b = 1 \cdot b = (a^{-1} \cdot a) \cdot b = a^{-1} \cdot (a \cdot b) = a^{-1} \cdot 0 = 0$. So if $a \neq 0$, then $b = 0$. Thus $(a, b \in \mathbb{R} \text{ such that } a \cdot b = 0) \Rightarrow (a = 0 \text{ or } b = 0)$. \diamond

In this part of the course, we won't do such careful justifications every time we need to manipulate real numbers. We have listed the above laws to once and for all stipulate the laws of arithmetic for real numbers which justify the usual calculational rules we are familiar with, so that we know the *source* of it all. As examples, we consider the following exercises of giving a rigorous justification based on (F1) to (F9) of facts that are well-known to us.

Exercise 1.3. (*) Using the field axioms of \mathbb{R} , prove the following:

- (1) Additive inverses are unique.
- (2) For all $a \in \mathbb{R}$, $(-1) \cdot a = -a$.
- (3) Show that $(-1) \cdot (-1) = 1$.

1.3. Order axioms

We now turn to order axioms⁵ for the real numbers. This is the source of the inequality ' $>$ ' that we are used to, enabling one to compare two real numbers. The relation $>$ between real numbers arises from a special subset \mathbb{P} of the real numbers.

Order axiom. There exists a subset \mathbb{P} of \mathbb{R} such that

- (O1) If $x, y \in \mathbb{P}$, then $x + y \in \mathbb{P}$ and $x \cdot y \in \mathbb{P}$.
- (O2) For every $x \in \mathbb{R}$, *one and only one* of the following statements is true:

$$1^\circ \quad x = 0. \qquad 2^\circ \quad x \in \mathbb{P}. \qquad 3^\circ \quad -x \in \mathbb{P}.$$

Definition 1.1 (Positive numbers).

The elements of \mathbb{P} are called *positive numbers*. For real numbers x, y , we say that

$$\begin{aligned} x > y & \text{ if } x - y \in \mathbb{P}, \\ x < y & \text{ if } y - x \in \mathbb{P}, \\ x \geq y & \text{ if } x = y \text{ or } x > y, \\ x \leq y & \text{ if } x = y \text{ or } x < y. \end{aligned}$$

It is clear from (O2) that 0 is *not* a positive number. Also, from (O2) it follows that for real numbers x, y , *one and only one* of the following statements is true ('Trichotomy Law'):

$$1^\circ \quad x = y. \qquad 2^\circ \quad x > y. \qquad 3^\circ \quad x < y.$$

⁵This section has more detail that we need. The student may skip this section, and begin reading from Section 1.4 onwards. For MA103 students, we will define the order relation for the real numbers carefully in Chapter 4 (covered in the Lent Term), and also prove the Trichotomy Law.

Indeed, if $x \neq y$, then $x - y \neq 0$, and so by (O2), we have the mutually exclusive possibilities $x - y \in \mathbb{P}$ or $y - x = -(x - y) \in \mathbb{P}$, that is, either $x > y$ or $x < y$.

Example 1.3. $1 > 0$.

We have three possible, mutually exclusive cases:

$$1^\circ \quad 1 = 0. \qquad 2^\circ \quad 1 \in \mathbb{P}. \qquad 3^\circ \quad -1 \in \mathbb{P}.$$

As $1 \neq 0$, we know that 1° is not possible.

Suppose that 3° holds, that is, $-1 \in \mathbb{P}$. We had seen in Exercise 1.3.(3) that $(-1) \cdot (-1) = 1$. From (O1), and the fact that $-1 \in \mathbb{P}$, it then follows that $1 = (-1) \cdot (-1) \in \mathbb{P}$. So if we assume that 3° holds, then *both* 2° and 3° are true, which is impossible as it violates (O2).

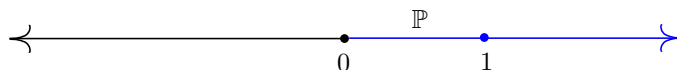
Thus by (O2), the only remaining case, namely 2° must hold, that is, $1 \in \mathbb{P}$. \diamond

Exercise 1.4. (*) Using the order axioms for \mathbb{R} , show the following:

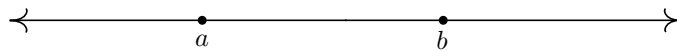
- (1) For all $a \in \mathbb{R}$, $a^2 \geq 0$.
- (2) There is no real number x such that $x^2 + 1 = 0$.

Again, just like we the field axioms, it is enough to know that if challenged, one can derive all the usual laws of manipulating inequalities among real numbers based on these order axioms, but we will not do this at every instance we meet an inequality.

From our intuitive picture of \mathbb{R} as points on the number line, what is the set \mathbb{P} ? \mathbb{P} is simply the set of all points/real numbers to the right of the origin O .



Also, geometrically on the number line, the inequality $a < b$ between real numbers a, b means that b lies to the right of a on the number line.



1.4. The Least Upper Bound Property of \mathbb{R}

This property is crucial in Analysis, and when we prove the key results (Bolzano-Weierstrass Theorem, Intermediate Value Theorem, Extreme Value Theorem, etc.), we will gradually learn to appreciate the key role played by it.

Definition 1.2 (Upper bound of a set).

Let S be a subset of \mathbb{R} . A real number u is said to be an *upper bound* of S if for all $x \in S$, $x \leq u$.

If we think of the set S as some blob on the number line, then u should be any point on the number line which lies to the right of the points of the blob.



Example 1.4.

- (1) If $S = \{0, 1, 9, 7, 6, 1976\}$, then 1976 is an upper bound of S . In fact, any real number $u \geq 1976$ is an upper bound of S . So S has lots of upper bounds.
- (2) Let $S := \{x \in \mathbb{R} : x < 1\}$. Then 1 is an upper bound of S . In fact, any real number $u \geq 1$ is an upper bound of S .
- (3) If $S = \mathbb{R}$, then S has no upper bound. Why? Suppose that $u \in \mathbb{R}$ is an upper bound of \mathbb{R} . Consider $u + 1 \in S = \mathbb{R}$. Then $S \ni u + 1 \leq u$ (upper bound of S). So $1 \leq 0$, a contradiction!
- (4) Let $S = \emptyset$ (the empty set, containing no elements). *Every* $u \in \mathbb{R}$ is an upper bound. For if $u \in \mathbb{R}$ is not an upper bound of S , then there must exist an element $x \in S$ which prevents u from being an upper bound of S , that is,

It is not the case that $x \leq u$.

But S has no elements at all, much less an element such that $\boxed{\dots}$ holds.

(This is an example of a ‘vacuous truth’. Consider the statement

Every man with 9 legs is intelligent.

This is considered a true statement in Mathematics. The argument is:

Can you show me a man with 9 legs for which the claimed property (namely of being intelligent) is not true? No! Because there are no men with 9 legs!

By the same argumentation, also the following statement is true:

Every man with 9 legs is not intelligent.

◇

Definition 1.3 (Set bounded above).

If $S \subset \mathbb{R}$ and S has an upper bound (that is, the set of upper bounds of S is not empty), then S is said to be *bounded above*.

Example 1.5. The set \mathbb{R} is not bounded above.

Each of the sets $\{0, 1, 9, 7, 6, 1976\}$, \emptyset , $\{x \in \mathbb{R} : x < 1\}$ is bounded above.

◇

The notions of ‘lower bound’, and ‘bounded below’ are defined analogously.

Definition 1.4 (Lower bound of a set; set bounded below).

Let S be a subset of \mathbb{R} . A real number ℓ is said to be a *lower bound of S* if for all $x \in S$, $\ell \leq x$.

If $S \subset \mathbb{R}$ and S has a lower bound (that is, the set of lower bounds of S is not empty), then S is said to be *bounded below*.

If we think of the set S as some blob on the number line, then ℓ should be any point on the number line which lies to the left of the points of the blob.



Example 1.6.

- (1) If $S = \{0, 1, 9, 7, 6, 1976\}$, then 0 is an lower bound of S . In fact any real number $\ell \leq 0$ serves as a upper bound of S . So S is bounded below.
- (2) Let $S := \{x \in \mathbb{R} : x < 1\}$. Then S is not bounded below. Let us show this. Suppose that, on the contrary, S does have a lower bound, say $\ell \in \mathbb{R}$. Let $x \in S$. Then $\ell \leq x < 1$. We have

$$\ell - 1 < \ell \leq x < 1,$$

and so $\ell - 1 < 1$. Thus $\ell - 1 \in S$, and as ℓ is a lower bound of S , we must have $\ell \leq \ell - 1$, that is, $1 < 0$, a contradiction! So our original assumption that S is bounded below must be false. So S is not bounded below. (This claim was intuitively obvious too, since the set of points in S on the number line is the entire ray of points on the left of 1, leaving no room for points on \mathbb{R} to be on the ‘left of S ’.)

- (3) If $S = \mathbb{R}$, then S has no lower bound: If $\ell \in \mathbb{R}$ is a lower bound of \mathbb{R} , then

$$\underbrace{\ell}_{\text{lower bound of } S} \leq \underbrace{\ell - 1}_{\in S},$$

and so $1 \leq 0$, a contradiction. Thus \mathbb{R} is not bounded below.

- (4) Let $S = \emptyset$ (the empty set, containing no elements). Every $\ell \in \mathbb{R}$ is a lower bound. If $\ell \in \mathbb{R}$ is not an lower bound of S , then there must exist an element $x \in S$ which prevents ℓ from being an lower bound of S , that is, it is not the case that $\ell \leq x$. As S is empty, this is impossible. So S is bounded below. \diamond

Definition 1.5 (Bounded set).

Let $S \subset \mathbb{R}$. S is called *bounded* if S is bounded below and bounded above.

Example 1.7.

S	An upper bound	Bounded above?	A lower bound	Bounded below?	Bounded?
$\{0, 1, 9, 7, 6, 1976\}$	1976 Any $u \geq 1976$	Yes	0 Any $\ell \leq 0$	Yes	Yes
$\{x \in \mathbb{R} : x < 1\}$	1 Any $u \geq 1$	Yes	Doesn't exist	No	No
\mathbb{R}	Doesn't exist	No	Doesn't exist	No	No
\emptyset	Every $u \in \mathbb{R}$	Yes	Every $\ell \in \mathbb{R}$	Yes	Yes

\diamond

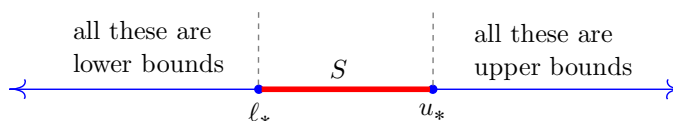
We now introduce the notions of a least upper bound (also called supremum) and a greatest lower bound (also called infimum) of a subset S of \mathbb{R} .

Definition 1.6 (Supremum and infimum).

Let S be a subset of \mathbb{R} .

- $u_* \in \mathbb{R}$ is called a *least upper bound of S* (or a *supremum of S*) if
 - (1) u_* is an upper bound of S , and
 - (2) if u is an upper bound of S , then $u_* \leq u$.
- $\ell_* \in \mathbb{R}$ is called a *greatest lower bound of S* (or an *infimum of S*) if
 - (1) ℓ_* is a lower bound of S , and
 - (2) if ℓ is a lower bound of S , then $\ell \leq \ell_*$.

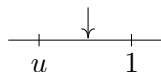
Pictorially, the supremum is the leftmost point amongst the upper bounds, and the infimum is the rightmost point amongst the lower bound of a set.

**Example 1.8.**

- (1) If $S = \{0, 1, 9, 7, 6, 1976\}$, then $u_* = 1976$ is a least upper bound of S because
- (a) 1976 is an upper bound of S , and
 - (b) if u is an upper bound of S , then $(S \ni) 1976 \leq u$, that is $u_* \leq u$.

Similarly, 0 is a greatest lower bound of S .

- (2) Let $S = \{x \in \mathbb{R} : x < 1\}$. Then $u_* = 1$ is a least upper bound of S . Indeed:
- (a) 1 is an upper bound of S : If $x \in S$, then $x < 1 = u_*$.
 - (b) Let u be an upper bound of S . We want to show that $u_* = 1 \leq u$. Suppose the contrary, that is, $1 > u$. Then there is a gap between u and 1.



(But then this gap between u and 1 contains elements of S which are to *right* of the supposed upper bound u , and this should give the contradiction we seek.) To this end, consider the number $(1 + u)/2$. We have

$$\frac{1 + u}{2} < \frac{1 + 1}{2} = 1$$

and so $(1 + u)/2$ belongs to S . As u is an upper bound of S , we must have

$$\frac{1 + u}{2} \leq u,$$

which upon rearranging gives $1 \leq u$, a contradiction.

S does not have a lower bound, and so certainly no greatest lower bound either (a greatest lower bound has to be first of all a lower bound!).

- (3) \mathbb{R} does not have a supremum, and no infimum either.
- (4) \emptyset has no supremum. (We intuitively expect this: indeed every real number serves as an upper bound, but there is no smallest one among these!) Indeed, suppose on the contrary that $u_* \in \mathbb{R}$ is a supremum. Then $u_* - 1 \in \mathbb{R}$ is an upper bound of \emptyset (since it is *some* real number, and we had seen that *all* real numbers are upper bounds of \emptyset). As u_* is the least upper bound, we must have $u_* \leq u_* - 1$, that is, $1 \leq 0$, a contradiction.

Similarly, \emptyset has no infimum either. \diamond

A set may have many upper bounds and many lower bounds, but it is intuitively clear, based on our visual number line picture, that the supremum and infimum of a set, assuming they exist, must be unique. Here is a formal proof.

Theorem 1.2. *If a subset S of \mathbb{R} has a supremum, then it is unique.*

Proof. Let u_*, u'_* be two supremums of S . Then as u'_* is, in particular, an upper bound, and since u_* is the least upper bound, we must have

$$u_* \leq u'_*. \quad (1.1)$$

Similarly, since u_* is, in particular, an upper bound, and since u'_* is the least upper bound, we must also have

$$u'_* \leq u_*. \quad (1.2)$$

From (1.1) and (1.2), it now follows that $u_* = u'_*$. \square

So when S has **a** supremum, then it is **the** supremum. So we can give it special notation (since we know what it means unambiguously):

$$\sup S.$$

Similarly, if a set S has an infimum, it is unique and is denoted by

$$\inf S.$$

Example 1.9. We have

$$\begin{aligned} \sup\{0, 1, 9, 7, 6, 1976\} &= 1976, \\ \sup\{x \in \mathbb{R} : x < 1\} &= 1, \\ \inf\{x \in \mathbb{R} : x \geq 1\} &= 1. \end{aligned}$$

To see the last equality, we note that 1 is certainly a lower bound of the set $S := \{x \in \mathbb{R} : x \geq 1\}$, and if ℓ is any lower bound, then as 1 is an element of the set S , we have $\ell \leq 1$. \diamond

Comparing the first two examples above, when $S := \{0, 1, 9, 7, 6, 1976\}$, we have

$$\sup S \in S,$$

while in the case of $S := \{x \in \mathbb{R} : x < 1\}$, we have

$$\sup S \notin S.$$

It will be convenient to keep track of when the supremum (or for that matter infimum) of a set *belongs to* the set. So we introduce the following definitions and corresponding notation.

Definition 1.7 (Maximum, minimum of a set).

- If $\sup S \in S$, then $\sup S$ is called a *maximum of S* , denoted by $\max S$.
- If $\inf S \in S$, then $\inf S$ is called a *minimum of S* , denoted by $\min S$.

Example 1.10.

S	Supremum	Maximum	Infimum	Minimum
$\{0, 1, 9, 7, 6, 1976\}$	1976	1976	0	0
$\{x \in \mathbb{R} : x < 1\}$	1	Doesn't exist	Doesn't exist	Doesn't exist
\mathbb{R}	Doesn't exist	Doesn't exist	Doesn't exist	Doesn't exist
\emptyset	Doesn't exist	Doesn't exist	Doesn't exist	Doesn't exist
$\{x \in \mathbb{R} : x \geq 1\}$	Doesn't exist	Doesn't exist	1	1

◇

Exercise 1.5. Provide the following information about the set S

An upper bound	A lower bound	Is S bounded?	$\sup S$	$\inf S$	$\max S$	$\min S$

where S is given by:

- (1) $(0, 1] := \{x \in \mathbb{R} : 0 < x \leq 1\}$
- (2) $[0, 1] := \{x \in \mathbb{R} : 0 \leq x \leq 1\}$
- (3) $(0, 1) := \{x \in \mathbb{R} : 0 < x < 1\}$.

In the above Example 1.10, we note that if S is nonempty and bounded above, then its supremum exists. In fact this is a fundamental property of the real numbers, called the *least upper bound property* of the real numbers, which we state below:

If $S \subset \mathbb{R}$ is such that $S \neq \emptyset$ and S has an upper bound, then $\sup S$ exists.

Example 1.11.

- (1) $S = \{0, 1, 9, 7, 6, 1976\}$ is a subset of \mathbb{R} , it is nonempty, and it has an upper bound. So the Least Upper Bound Property of \mathbb{R} tells us that this set should have a least upper bound. This is indeed true, as we had seen earlier that S has 1976 as the supremum.
- (2) $S = \{x \in \mathbb{R} : x < 1\}$ is a subset of \mathbb{R} , it is nonempty ($0 \in S$), and it has an upper bound (for example 2). So the Least Upper Bound Property of \mathbb{R} tells us that this set should have a least upper bound. This is indeed true, as we had seen earlier that 1 is the supremum of S .

- (3) $S = \mathbb{R}$ is a subset of \mathbb{R} , it is nonempty, and it has no supremum. So what went wrong? Well, S isn't bounded above.
- (4) $S = \emptyset$ is a subset of \mathbb{R} and it is bounded above. But S has no supremum. There is no contradiction to the Least Upper Bound Property, because S is empty! \diamond

Example 1.12. Let $S := \{x \in \mathbb{R} : x^2 \leq 2\}$. Clearly S is a subset of \mathbb{R} and it is nonempty since $1 \in S$: $1^2 = 1 \leq 2$. Let us show that S is bounded above. In fact, 2 serves as an upper bound of S . Since if $x > 2$, then $x^2 > 4 > 2$. Thus if $x \in S$, then $x^2 \leq 2$, and so $x \leq 2$.

By the Least Upper Bound Property of \mathbb{R} , $u_* := \sup S$ exists in \mathbb{R} . Moreover, one can show that this u_* satisfies $u_*^2 = 2$ by showing that the cases $u_*^2 < 2$ and $u_*^2 > 2$ are both impossible.

First of all, $u_* \geq 1$ (as u_* is in particular an upper bound of S and $1 \in S$). Define

$$r := u_* - \frac{u_*^2 - 2}{u_* + 2} = \frac{2(u_* + 1)}{u_* + 2} > 0. \quad (1.3)$$

Then we have

$$r^2 - 2 = \frac{2(u_*^2 - 2)}{(u_* + 2)^2}. \quad (1.4)$$

- 1° Suppose $u_*^2 < 2$. Then (1.4) implies that $r^2 - 2 < 0$, and so $r \in S$. But from (1.3), $r > u_*$, contradicting the fact that u_* is an upper bound of S .
- 2° Suppose that $u_*^2 > 2$. If $r' > r$ (> 0), then $r'^2 = r' \cdot r' > r \cdot r' > r \cdot r = r^2$. From (1.4), $r^2 > 2$, and so from the above, we know that $r'^2 > 2$ as well. Hence $r' \notin S$. So we have shown that if $r' \in S$, then $r' \leq r$. This means that r is an upper bound of S . But this is impossible, since (1.3) shows that $r < u_*$, and u_* is the *least* upper bound of S .

So it must be the case that $u_*^2 = 2$. Note also that u_* is nonnegative (as $u_* \geq 1 \in S$). (We will denote this nonnegative $u_* \in \mathbb{R}$ satisfying $u_*^2 = 2$ by $\sqrt{2}$.) \diamond

Example 1.13 (\mathbb{Q} does not possess the Least Upper Bound Property).

Consider the set $S := \{x \in \mathbb{Q} : x^2 \leq 2\}$. Clearly S is a subset of \mathbb{Q} and it is nonempty since $1 \in S$: $1^2 = 1 \leq 2$. Let us show that S is bounded above. In fact, 2 serves as an upper bound of S . Since if $x > 2$, then $x^2 > 4 > 2$. Thus if $x \in S$, then $x^2 \leq 2$, and so $x \leq 2$.

If \mathbb{Q} has the Least Upper Bound Property, then the above nonempty subset of \mathbb{Q} which is bounded above must possess a least upper bound $u_* := \sup S \in \mathbb{Q}$. Once again, just as in the previous example, we can show that this $u_* \in \mathbb{Q}$ must satisfy that $u_*^2 = 2$ (and we have given the details below). But we know that this is impossible as we had shown that there is no rational number whose square is 2.

Firstly, $u_* \geq 1$ (as u_* is in particular an upper bound of S and $1 \in S$). Now define

$$r := u_* - \frac{u_*^2 - 2}{u_* + 2} = \frac{2(u_* + 1)}{u_* + 2} > 0. \quad (1.5)$$

As $u_* \in \mathbb{Q}$, the rightmost expression for r shows that $r \in \mathbb{Q}$ as well. Then

$$r^2 - 2 = \frac{2(u_*^2 - 2)}{(u_* + 2)^2}. \quad (1.6)$$

- 1° Suppose $u_*^2 < 2$. Then (1.6) implies that $r^2 - 2 < 0$, and so $r \in S$. But from (1.5), $r > u_*$, contradicting the fact that u_* is an upper bound of S .
- 2° Suppose that $u_*^2 > 2$. If $r' > r$ (> 0), then $r'^2 = r' \cdot r' > r \cdot r' > r \cdot r = r^2$. From (1.6), $r^2 > 2$, and so from the above, we know that $r'^2 > 2$ as well. Hence $r' \notin S$. So we have shown that if $r' \in S$, then $r' \leq r$. This means that r is an upper bound of S . But this is impossible, since (1.5) shows that $r < u_*$, and u_* is the *least* upper bound of S .

So it must be the case that $u_*^2 = 2$. But as we mentioned earlier, this is impossible by Theorem 1.1. Hence \mathbb{Q} does not possess the Least Upper Bound Property. \diamond

In order to get the useful results in Analysis (for example the fact that for an increasing sequence of numbers bounded above, there must be a smallest number bigger than each of the terms of the sequence – a fact needed to calculate the area as described at the outset), it turns out to be the case that the Least Upper Bound Property is indispensable. So it makes sense that when we set up the definitions and results in Analysis, we don't work with the rational number system \mathbb{Q} (which regrettably does *not* possess the Least Upper Bound Property), but rather with the larger real number system \mathbb{R} , which *does* possess the Least Upper Bound Property.

Exercise 1.6. Let a_1, a_2, a_3, \dots be an infinite list (or sequence) of real numbers such that $a_n \leq a_{n+1}$ for all $n \in \mathbb{N}$, that is, the sequence is increasing. Also suppose that

$$S := \{a_n : n \in \mathbb{N}\}$$

is bounded above. Show that there is a smallest real number L which is bigger than each of the a_n , $n \in \mathbb{N}$.

Exercise 1.7.

- (1) Let S be a nonempty subset of real numbers which is bounded below. Let $-S$ denote the set of all real numbers $-x$, where x belongs to S . Prove that $\inf S$ exists and $\inf S = -\sup(-S)$.
- (2) Conclude from here that \mathbb{R} also has the 'Greatest Lower Bound Property':

If S is a nonempty subset of \mathbb{R} having a lower bound, then $\inf S$ exists.

Exercise 1.8. Let S be a nonempty subset of \mathbb{R} which is bounded above, and let $\alpha > 0$. Show that $\alpha \cdot S := \{\alpha x : x \in S\}$ is also bounded above and that $\sup(\alpha \cdot S) = \alpha \cdot \sup S$. Similarly, if S is a nonempty subset of \mathbb{R} which is bounded below and $\alpha > 0$, then show that $\alpha \cdot S$ is bounded below, and that $\inf(\alpha \cdot S) = \alpha \cdot \inf S$.

Exercise 1.9. Let A and B be nonempty subsets of \mathbb{R} that are bounded above and such that $A \subset B$. Prove that $\sup A \leq \sup B$.

Exercise 1.10. For any nonempty bounded set S , prove that $\inf S \leq \sup S$, and that the equality holds if and only if S is a singleton set (that is a set with cardinality 1).

Exercise 1.11. Let A and B be nonempty subsets of \mathbb{R} that are bounded above. Prove that $\sup(A \cup B)$ exists and that $\sup(A \cup B) = \max\{\sup A, \sup B\}$.

Exercise 1.12. Determine whether the following statements are true or false.

- (1) If u is an upper bound of S ($\subset \mathbb{R}$), and $u' < u$, then u' is not an upper bound of S .
- (2) If u_* is the supremum of S ($\subset \mathbb{R}$), and $\epsilon > 0$, then $u_* - \epsilon$ is not an upper bound of S .
- (3) Every subset of \mathbb{R} has a maximum.
- (4) Every subset of \mathbb{R} has a supremum.
- (5) Every bounded subset of \mathbb{R} has a maximum.
- (6) Every bounded subset of \mathbb{R} has a supremum.
- (7) Every bounded nonempty subset of \mathbb{R} has a supremum.
- (8) Every subset of \mathbb{R} that has a supremum is bounded above.
- (9) For every subset of \mathbb{R} that has a maximum, the maximum belongs to the set.
- (10) For every subset of \mathbb{R} that has a supremum, the supremum belongs to the set.
- (11) For every subset S of \mathbb{R} that is bounded above, $|S|$ defined by $\{|x| : x \in S\}$ is bounded.
- (12) For every subset S of \mathbb{R} that is bounded, $|S|$ defined by $\{|x| : x \in S\}$ is bounded.
- (13) For every bounded subset S of \mathbb{R} , if $\inf S < x < \sup S$, then $x \in S$.

Exercise 1.13. Let A and B be nonempty subsets of \mathbb{R} that are bounded above and define

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

Prove that $\sup(A + B)$ exists and that $\sup(A + B) = \sup A + \sup B$.

Exercise 1.14. Let S be a nonempty set of positive real numbers. Define the set

$$S^{-1} := \{x^{-1} : x \in S\}.$$

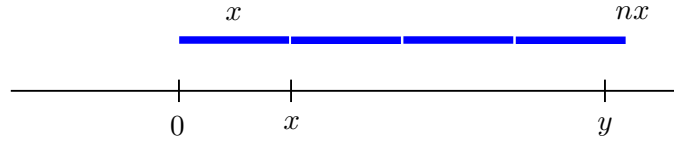
- (1) Show that S^{-1} is bounded above if and only if $\inf S > 0$.
- (2) Furthermore, if $\inf S > 0$, then show that $\sup S^{-1} = (\inf S)^{-1}$.

We now prove the following, called the *Archimedean property* of the real numbers.

Theorem 1.3 (Archimedean Property).

If $x, y \in \mathbb{R}$ and $x > 0$, then there exists an $n \in \mathbb{N}$ such that $y < nx$.

If $y \leq 0$ to begin with, then the above is just the trivial statement that $n \cdot x > 0 \geq y$, which works with every $n \in \mathbb{N}$. So the interesting content of the theorem is when $y > 0$. Then the above is telling us, that no matter how small x is, if we keep ‘tiling’ the real line with multiples of the length x , then eventually we will surpass y . Here is a picture to bear in mind.



Proof. Suppose that it is not the case that

‘there exists an $n \in \mathbb{N}$ such that $nx > y$ ’.

Then for *every* $n \in \mathbb{N}$, we must have $nx \leq y$. Let $S := \{nx : n \in \mathbb{N}\}$. Then S is a subset of \mathbb{R} , $S \neq \emptyset$ (indeed, $x = 1 \cdot x \in S$), and y is an upper bound of S . Thus by the Least Upper Bound Property of \mathbb{R} , $u_* := \sup S$ exists. As $x > 0$, the number $u_* - x$ is smaller than the least upper bound u_* of S . Hence $u_* - x$ can’t be an upper bound of S , which means that there is an element $mx \in S$, for some $m \in \mathbb{N}$, which prevents $u_* - x$ from being an upper bound: $mx > u_* - x$. Rearranging, we obtain $u_* < mx + x = (m + 1)x \in S$, contradicting the fact that u_* is an upper bound of S . Thus our original claim is false. In other words, there *does exist* an $n \in \mathbb{N}$ such that $nx > y$. \square

Example 1.14. Let $S = \left\{\frac{1}{n} : n \in \mathbb{N}\right\} = \left\{1, \frac{1}{2}, \frac{1}{3}, \dots\right\}$. We claim that $\inf S = 0$.

Clearly 0 is a lower bound of S since all the elements of S are positive.

Suppose that ℓ is a lower bound of S . We want to show that $\ell \leq 0$. Suppose on the contrary that $\ell > 0$. Then by the Archimedean property (with the real numbers x and y taken as $x = 1$ (> 0) and $y = 1/\ell$), there exists a $n \in \mathbb{N}$ such that

$$\frac{1}{\ell} = y < nx = n \cdot 1 = n,$$

and so

$$\frac{1}{n} < \ell,$$

contradicting the fact that ℓ is a lower bound of S . Thus any lower bound of S must be less than or equal to 0. Hence 0 is the infimum of S . \diamond

Exercise 1.15. Provide the following information about the set S

An upper bound	A lower bound	Is S bounded?	$\sup S$	$\inf S$	$\max S$	$\min S$

where S is given by:

- (1) $\left\{\frac{1}{n} : n \in \mathbb{Z} \setminus \{0\}\right\}$
- (2) $\left\{\frac{n}{n+1} : n \in \mathbb{N}\right\}$
- (3) $\left\{(-1)^n \left(1 + \frac{1}{n}\right) : n \in \mathbb{N}\right\}$

Exercise 1.16. Let $S := \{(xy - 1)^2 + x^2 : (x, y) \in \mathbb{R}^2\}$.

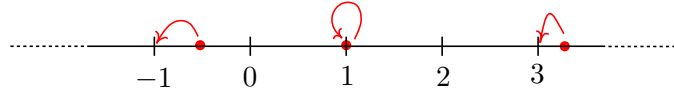
(a) Show that S is bounded below.

(b) What is $\inf S$? *Hint:* To justify your answer, consider $(x, y) = (1/n, n)$, $n \in \mathbb{N}$.

(c) Does $\min S$ exist?

Example 1.15 (The greatest integer part $\lfloor \cdot \rfloor$ of $x \in \mathbb{R}$).

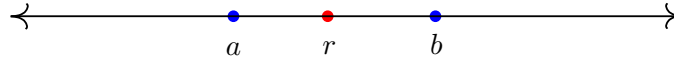
If we think of the real numbers as points of the line, then we see that along it, there are ‘milestones’ at each of the integers. So if we take any real number it is between two milestones. We take $\lfloor x \rfloor$ to be the milestone immediately to the left of x —in other words, it is the ‘greatest integer less than or equal to x ’. So for example $\lfloor 3.1 \rfloor = 3$, $\lfloor 0 \rfloor = 0$, $\lfloor n \rfloor = n$ for all integers n , $\lfloor -3.1 \rfloor = -4$, etc.



Using the Archimedean Property, one can give a rigorous justification of the fact that every real number *has to* belong to an interval $[n, n + 1)$ for some $n \in \mathbb{Z}$ (so that this $n = \lfloor x \rfloor$). By the Archimedean Property, there exists an $m_1 \in \mathbb{N}$ such that $m_1 \cdot 1 > x$. By the Archimedean Property, there exists an $m_2 \in \mathbb{N}$ such that $m_2 \cdot 1 > -x$. So there are integers m_1, m_2 such that $-m_2 < x < m_1$. Among the finitely many integers $k \in \mathbb{Z}$ such that $-m_2 \leq k \leq m_1$, we take as $\lfloor x \rfloor$ the largest one such that it is also $\leq x$. \diamond

Theorem 1.4 (Density of \mathbb{Q} in \mathbb{R}).

If $a, b \in \mathbb{R}$, and $a < b$, then there exists a $r \in \mathbb{Q}$ such that $a < r < b$.



This results says that ‘ \mathbb{Q} is dense in \mathbb{R} ’. In everyday language, we may say for example that ‘These woods have a dense growth of birch trees’, and the picture we then have in mind is that in any small area of the woods, we find a birch tree. A similar thing is conveyed by the above: no matter what ‘patch’ (described by the two numbers a and b) we take on the real line (thought of as the woods), we can find a rational number (analogous to birch trees) in that patch.

Proof. As $b - a > 0$ and since $1 \in \mathbb{R}$, by the Archimedean Property, there exists an $n \in \mathbb{N}$ such that $n(b - a) > 1$, that is, $na + 1 < nb$. Let $m := \lfloor na \rfloor + 1$. Then $\lfloor na \rfloor \leq na < \lfloor na \rfloor + 1$, that is, $m - 1 \leq na < m$. So

$$a < \frac{m}{n} \leq \frac{na + 1}{n} < \frac{nb}{n} = b.$$

With $r := \frac{m}{n} \in \mathbb{Q}$, the proof of the theorem is complete. \square

Exercise 1.17 (Density of irrationals in \mathbb{R}).

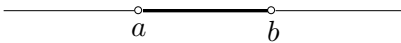
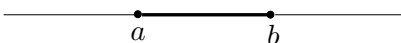
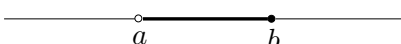
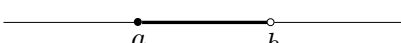
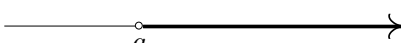




Show that if $a, b \in \mathbb{R}$ and $a < b$, then there exists an irrational number between a and b .

1.5. Intervals

In Analysis, we will consider real-valued functions of a real variable, and develop results about these. It will turn out that while doing so, we will keep meeting certain types of subsets of \mathbb{R} (for example subsets of this type will often be the ‘domains’ of our real-valued functions for which the results of Analysis hold). These special subsets of \mathbb{R} are called ‘intervals’, and we give the definition below. Roughly speaking, these are⁶ the ‘connected subsets’ of the real line, namely subsets of \mathbb{R} not having any ‘holes/gaps’.

Definition 1.8 (Interval).

An *interval* is a set consisting of all the real numbers between two given real numbers, or of all the real numbers on one side or the other of a given number. So an interval is a set of any of the following forms, where $a, b \in \mathbb{R}$ and $a < b$:

$(a, b) = \{x \in \mathbb{R} : a < x < b\}$	
$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$	
$(a, b] = \{x \in \mathbb{R} : a < x \leq b\}$	
$[a, b) = \{x \in \mathbb{R} : a \leq x < b\}$	
$(a, \infty) = \{x \in \mathbb{R} : a < x\}$	
$[a, \infty) = \{x \in \mathbb{R} : a \leq x\}$	
$(-\infty, b) = \{x \in \mathbb{R} : x < b\}$	
$(-\infty, b] = \{x \in \mathbb{R} : x \leq b\}$	
$(-\infty, \infty) = \mathbb{R}$	

In the above notation for intervals, a parenthesis ‘(’ or ‘)’ means that the respective endpoint is not included, and a square bracket ‘[’ or ‘]’ means that the endpoint is included. Thus $[0, 1)$ means the set of all real numbers x such that $0 \leq x < 1$. (Note that the use of the symbol ∞ in the notation for intervals is simply a matter of convenience and is not to be taken as suggesting that there is a number ∞ .)

Also, it will be convenient to give certain types of interval a special name.

Definition 1.9 (Open interval).

An interval of the form (a, b) , (a, ∞) , $(-\infty, b)$ or \mathbb{R} is called an *open interval*.

⁶That are nonempty and contain not just one point.

We note that if I is an open interval, then for every member $x \in I$, there exists a $\delta > 0$ such that $(x - \delta, x + \delta) \subset I$, that is, there is always some ‘room’ around x consisting only of elements of I .

Exercise 1.18. Show that if $a, b \in \mathbb{R}$, then the interval (a, b) has the following property:

For every $x \in (a, b)$, there exists a $\delta > 0$ such that $(x - \delta, x + \delta) \subset (a, b)$.

Show also that $[a, b]$ does not possess the above property.

Definition 1.10 (Compact interval).

If $a, b \in \mathbb{R}$ and $a < b$, then we call $[a, b]$ a *compact interval*.

Note that $\mathbb{R} \setminus [a, b]$ is the union of two open intervals, namely $(-\infty, a)$ and (b, ∞) and that $[a, b]$ is a bounded set.

Exercise 1.19. If $A_n, n \in \mathbb{N}$, is a collection of sets, then $\bigcap_{n \in \mathbb{N}} A_n$ denotes their intersection:

$$\bigcap_{n \in \mathbb{N}} A_n = \{x : \forall n \in \mathbb{N}, x \in A_n\},$$

and $\bigcup_{n \in \mathbb{N}} A_n$ denotes their union: $\bigcup_{n \in \mathbb{N}} A_n = \{x : \exists n \in \mathbb{N} \text{ such that } x \in A_n\}$. Prove that

$$(1) \quad \emptyset = \bigcap_{n \in \mathbb{N}} \left(0, \frac{1}{n}\right).$$

$$(2) \quad \{0\} = \bigcap_{n \in \mathbb{N}} \left[0, \frac{1}{n}\right].$$

$$(3) \quad (0, 1) = \bigcup_{n \in \mathbb{N}} \left[\frac{1}{n+2}, 1 - \frac{1}{n+2}\right].$$

$$(4) \quad [0, 1] = \bigcap_{n \in \mathbb{N}} \left(-\frac{1}{n}, 1 + \frac{1}{n}\right).$$

1.6. Absolute value $|\cdot|$ and distance in \mathbb{R}

In Analysis, in order to talk about notions such as continuity, convergence, etc., we will need a notion of ‘closeness/distance’ between real numbers. This is provided by the absolute value $|\cdot|$, and the distance between real numbers x and y is $|x - y|$. We give the definitions below.

Definition 1.11 (Absolute value and distance).

- The *absolute value* or *modulus* of a real number x is denoted by $|x|$, and it is defined as follows:

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

- The *distance* $d(x, y)$ between two real numbers x and y is the absolute value $|x - y|$ of their difference.

Thus $|1| = 1$, $|0| = 0$, $|-1| = 1$, and the distance between the real numbers -1 and 1 is equal to $d(-1, 1) = |-1 - 1| = |-2| = 2$. The distance gives a notion of closeness of two points, which is crucial in the formalization of the notions of

Analysis. We can now specify regions comprising points close to a certain point $c \in \mathbb{R}$ in terms of inequalities in absolute values, that is, by demanding that the distance of the points of the region, to the point c , is less than a certain positive number δ , say $\delta = 0.01$ or $\delta = 0.0000001$, and so on.

Theorem 1.5. *Let $c \in \mathbb{R}$ and $\delta > 0$. Then:*

$$\boxed{d(x, c) := |x - c| < \delta} \quad \Leftrightarrow \quad \boxed{c - \delta < x < c + \delta}.$$

Though the proof is trivial, it is worthwhile remembering Theorem 1.5, as such a manipulation will keep arising over and over again in our subsequent development of Analysis. See Figure 2.

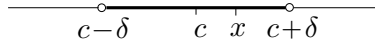


Figure 2. The interval $I = (c - \delta, c + \delta) = \{x \in \mathbb{R} : |x - c| < \delta\}$ is the set of all points in \mathbb{R} whose distance to the point c is strictly less than δ (> 0).

Proof.

- (\Rightarrow) Suppose that $|x - c| < \delta$. Then $x - c \leq |x - c| < \delta$, and $-(x - c) \leq |x - c| < \delta$. So $-\delta < x - c < \delta$, that is $c - \delta < x < c + \delta$.
- (\Leftarrow) If $c - \delta < x < c + \delta$, then $x - c < \delta$ and $-(x - c) = c - x < \delta$. Thus $|x - c| < \delta$, because $|x - c|$ is either $x - c$ or $-(x - c)$, and in both cases the numbers are less than δ . \square

If we think of the real numbers as points on the number line, and we think about the integers as milestones, then it is clear that the distance between, say -1 and 3 should be 4 miles, and we observe that $4 = |-1 - 3|$. So taking $|x - y|$ as the distance between $x, y \in \mathbb{R}$ is a sensible thing to do, based on our visual picture of \mathbb{R} as points on the number line.

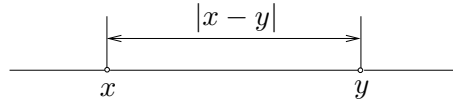


Figure 3. Distance between real numbers.

Exercise 1.20. Show that a subset S of \mathbb{R} is bounded if and only if there exists an $M \in \mathbb{R}$ such that for all $x \in S$, $|x| \leq M$.

The following properties of the absolute value will be useful in the sequel.

Theorem 1.6. *If x, y are real numbers, then*

$$|x \cdot y| = |x| \cdot |y|, \tag{1.7}$$

$$|x + y| \leq |x| + |y|. \tag{1.8}$$

(1.8) is called the *triangle inequality*.

Proof. We prove (1.7) by exhausting all possible cases:

1° $x = 0$ or $y = 0$. Then $|x| = 0$ or $|y| = 0$, and so $|x| |y| = 0$. On the other hand, as $x = 0$ or $y = 0$, it follows that $xy = 0$ and so $|xy| = 0$.

2° $x > 0$ and $y > 0$. Then $|x| = x$ and $|y| = y$, and so $|x| |y| = xy$. On the other hand, as $x > 0$ and $y > 0$, it follows that $xy > 0$ and so $|xy| = xy$.

3° $x > 0$ and $y < 0$. Then $|x| = x$ and $|y| = -y$, and so $|x| |y| = x(-y) = -xy$. On the other hand, as $x > 0$ and $y < 0$, it follows that $xy < 0$ and so $|xy| = -xy$.

4° $x < 0$ and $y > 0$. This follows from 3° above by interchanging x and y .

5° $x < 0$ and $y < 0$. Then $|x| = -x$ and $|y| = -y$, and so $|x| |y| = (-x)(-y) = xy$. On the other hand, as $x < 0$ and $y < 0$, it follows that $xy > 0$ and so $|xy| = xy$.

This proves (1.7).

Next we prove (1.8). First observe that from the definition of $|\cdot|$, it follows that for any real $x \in \mathbb{R}$, $|x| \geq x$: indeed if $x \geq 0$, then $|x| = x$, while if $x < 0$, then $-x > 0$, and so $|x| = -x > 0 > x$.

From (1.7), we also have $|-x| = |-1 \cdot x| = |-1||x| = 1|x| = |x|$, for all $x \in \mathbb{R}$, and so it follows that $|x| = |-x| \geq -x$ for all $x \in \mathbb{R}$.

We have the following cases:

1° $x + y \geq 0$. Then $|x + y| = x + y$. As $|x| \geq x$ and $|y| \geq y$, $|x| + |y| \geq x + y = |x + y|$.

2° $x + y < 0$. Then $|x + y| = -(x + y)$. Since $|x| \geq -x$ and $|y| \geq -y$, it follows that $|x| + |y| \geq -x + (-y) = -(x + y) = |x + y|$.

This proves (1.8). □

Using these, one can check that the ‘metric/distance function’ $d: \mathbb{R} \times \mathbb{R} \rightarrow [0, \infty)$ defined by $d(x, y) = |x - y|$ for all $x, y \in \mathbb{R}$, satisfies the following properties:

(D1) (Positive definiteness) For all $x, y \in \mathbb{R}$, $d(x, y) \geq 0$. If $d(x, y) = 0$ then $x = y$.

(D2) (Symmetry) For all $x, y \in \mathbb{R}$, $d(x, y) = d(y, x)$.

(D3) (Triangle inequality) For all $x, y, z \in \mathbb{R}$, $d(x, z) \leq d(x, y) + d(y, z)$.

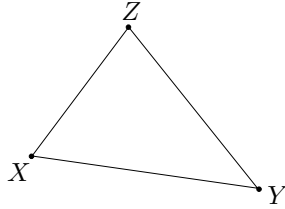


Figure 4. How the triangle inequality gets its name.

The reason (D3) is called the triangle inequality is that, for triangles in Euclidean geometry of the plane, we know that the sum of the lengths of two sides of a triangle is at least as much as the length of the third side: so for the points X, Y, Z in a plane forming the three vertices of a triangle: we know that $\ell(XZ) \leq \ell(XY) + \ell(YZ)$; see Figure 4. (D3) reminds us of this triangle inequality, and hence the name.

Exercise 1.21. Prove that if x, y are real numbers, then $||x| - |y|| \leq |x - y|$.

Exercise 1.22 (When does equality hold in the triangle inequality?).

(1) Show the generalized triangle inequality: if $n \in \mathbb{N}$ and a_1, \dots, a_n are real numbers, then $|a_1 + \dots + a_n| \leq |a_1| + \dots + |a_n|$.

(2) (*) We say that a_1, \dots, a_n *have the same sign* if either of the following cases is true:

$$1^\circ a_1 \geq 0, \dots, a_n \geq 0. \quad 2^\circ a_1 \leq 0, \dots, a_n \leq 0.$$

Thus the numbers have the same sign if on the number line either they all lie on the right of 0 including 0, or they all lie on the left of 0 including 0. Show that equality holds in the generalized triangle inequality if and only if the numbers have the same sign. *Hint:* Consider the $n = 2$ case first.

Exercise 1.23. For $a, b \in \mathbb{R}$, show that $\max\{a, b\} = \frac{a+b+|a-b|}{2}$ and $\min\{a, b\} = \frac{a+b-|a-b|}{2}$.

Exercise 1.24. (*) Let $\alpha > 0$ be an irrational number.

(1) Show that $\{n\alpha\} := n\alpha - \lfloor n\alpha \rfloor$, $n \in \mathbb{N}$, are all distinct, and belong to $(0, 1)$.

(2) Let $\epsilon > 0$. Show that there exists an $N \in \mathbb{N}$ such that

$$(0, 1) \subset [0, \epsilon) \cup [\epsilon, 2\epsilon) \cup \dots \cup [(N-1)\epsilon, N\epsilon).$$

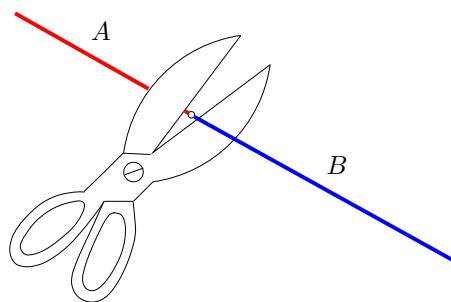
Using the Pigeonhole Principle, show that $|\{n\alpha\} - \{m\alpha\}| < \epsilon$ for some $m, n \in \mathbb{N}$.

(3) Prove that $\inf\{|n - m\alpha| : m \in \mathbb{N}, n \in \mathbb{N} \cup \{0\}\} = 0$.

1.7. (*) Remark on the construction of \mathbb{R}

We have treated the real number system \mathbb{R} as a given. But one might wonder if we can take the existence of real numbers on faith alone. A mathematical construction of \mathbb{R} can be given, and we will see this in the second part of the course.

There are several ways of doing this. One is via the ‘completion of \mathbb{Q} ’, where one considers ‘Cauchy sequences’ in \mathbb{Q} , and defines \mathbb{R} to be ‘equivalence classes of Cauchy sequences under a certain equivalence relation’. We will do this in §4.5.



Another way, which is more intuitive, is via ‘(Dedekind) cuts’, where we identify each real number by means of two sets A and B associated with it: A is the set of rationals less than the real number we are defining, and B is set of rational numbers at least as big as the real number we are trying to identify. In other words, if we

view the rational numbers lying on the number line, and think of the sets A and B (described above) corresponding to a real number, then this real number is the place along this rational number line where it can be cut, with A lying on the left side of this cut, and B lying on the right side of this cut. More precisely, a *cut* (A, B) in \mathbb{Q} is a pair of subsets A, B of \mathbb{Q} such that $A \cup B = \mathbb{Q}$, $A \neq \emptyset$, $B \neq \emptyset$, $A \cap B = \emptyset$, if $a \in A$ and $b \in B$ then $a < b$, and A contains no largest element. \mathbb{R} is then taken as the set of all cuts (A, B) . Here are two examples of cuts:

$$(A, B) = (\{r \in \mathbb{Q} : r < 0\}, \{r \in \mathbb{Q} : r \geq 0\}) \quad (\text{giving the real number '0'})$$

$$(A, B) = (\{r \in \mathbb{Q} : r \leq 0 \text{ or } r^2 < 2\}, \{r \in \mathbb{Q} : r > 0 \text{ and } r^2 \geq 2\}) \quad (\text{'}\sqrt{2}\text{'}).$$

It can then be shown that \mathbb{R} is a field containing \mathbb{Q} , and that it possesses the Least Upper Bound Property. The reader interested in this approach is referred to the Appendix to Chapter 1 in the classic textbook by Walter Rudin [R].

Appendix: Binomial theorem

Theorem 1.7. For any $x \in \mathbb{R}$ and any $n \in \mathbb{N}$, $(1 + x)^n = \sum_{k=0}^n \binom{n}{k} x^k$.

The *factorial* is defined by $0! = 1$ and for $n \in \mathbb{N}$, $n! = 1 \cdot 2 \cdots (n-1) \cdot n$. Also,

$$\binom{n}{k} := \frac{n!}{k!(n-k)!}.$$

Clearly, for all $n \in \mathbb{N}$, $\binom{n}{0} = 1 = \binom{n}{n}$. We will need:

Lemma 1.8. For all $n \in \mathbb{N}$ and all $0 \leq k < n$, $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$.

Proof. We have

$$\begin{aligned} \binom{n}{k} + \binom{n}{k+1} &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k+1)!(n-(k+1))!} \\ &= \frac{n!}{k!(n-(k+1))!} \left(\frac{1}{n-k} + \frac{1}{k+1} \right) \\ &= \frac{n!}{k!(n-(k+1))!} \frac{(n+1)}{(n-k)(k+1)} \\ &= \frac{(n+1)!}{(k+1)!(n-k)!} \\ &= \frac{(n+1)!}{(k+1)!((n+1)-(k+1))!} \\ &= \binom{n+1}{k+1}. \end{aligned}$$

□

Proof. (Of Theorem 1.7). We use induction on n . For $n = 1$, of course

$$(1+x)^1 = 1+x = \binom{1}{0} + \binom{1}{1}x.$$

If the result holds for some $n \in \mathbb{N}$, then

$$\begin{aligned} (1+x)^{n+1} &= (1+x)^n(1+x) = \left(\sum_{k=0}^n \binom{n}{k} x^k \right) (1+x) = \sum_{k=0}^n \binom{n}{k} x^k + \sum_{k=0}^n \binom{n}{k} x^{k+1} \\ &= \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{n-1}x^{n-1} + \binom{n}{n}x^n \\ &\quad + \binom{n}{0}x + \binom{n}{1}x^2 + \cdots + \binom{n}{n-2}x^{n-1} + \binom{n}{n-1}x^n + \binom{n}{n}x^n \\ &= \binom{n+1}{0} + \binom{n+1}{1}x + \binom{n+1}{2}x^2 + \cdots + \binom{n+1}{n-1}x^{n-1} + \binom{n+1}{n}x^n + \binom{n+1}{n+1}x^{n+1}. \end{aligned}$$

Here we used $\binom{n}{0} = 1 = \binom{n+1}{0}$, $\binom{n}{n} = 1 = \binom{n+1}{n+1}$, and Lemma 1.8. \square

Corollary 1.9. If $a, b \in \mathbb{R}$, and $n \in \mathbb{N}$, then $(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k$.

Proof. If $a = 0$, this is immediate. If $a \neq 0$, by the binomial theorem for $x := b/a$:

$$(a+b)^n = a^n \left(1 + \frac{b}{a}\right)^n = a^n \sum_{k=0}^n \binom{n}{k} \left(\frac{b}{a}\right)^k = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k. \quad \square$$

Remark 1.1 (The binomial coefficients are natural numbers).

Using Lemma 1.8, one can see that for all $n \in \mathbb{N}$ and $0 \leq k \leq n$, $\binom{n}{k} \in \mathbb{N}$.

One can use induction on n . First note that

$$\binom{1}{1} = 1 = \binom{1}{0}.$$

Let us suppose the claim has been shown for some $n \in \mathbb{N}$. We have

$$\binom{n+1}{0} = \frac{(n+1)!}{0!(n+1)!} = 1 = \frac{(n+1)!}{(n+1)!0!} = \binom{n+1}{n+1}.$$

For all $1 \leq k \leq n$, we have by Lemma 1.8 and the induction hypothesis that

$$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k} \in \mathbb{N}.$$

(Alternatively, $\binom{n}{k}$ is a natural number because it is the number of ways of choosing k objects from n distinct objects.) \ast

Chapter 2

Sequences and their convergence

The notion of a sequence occurs in ordinary conversation. For example when one says ‘an unfortunate sequence of events’, we imagine a *first* event, followed by a *second* event, followed by a *third* one, and so on.

Similarly, a sequence of real numbers is an infinite list

$$a_1, a_2, a_3, \dots$$

of real numbers, where

- a_1 is the first number/member/term of the sequence,
- a_2 is the second term of the sequence,
- a_3 is the third term of the sequence, and so on.

For example

$$1, \frac{1}{2}, \frac{1}{3}, \dots$$

is a sequence of real numbers, where 1 is the first term, $1/2$ is the second term, and in general, the n th term is $1/n$, $n \in \mathbb{N}$.

If in the sequence a_1, a_2, a_3, \dots , we think of a_1 as $f(1)$, a_2 as $f(2)$, a_3 as $f(3)$, and so on, then it becomes clear that a sequence is a special type of function, namely one with domain \mathbb{N} and co-domain \mathbb{R} .

Definition 2.1 (Sequence). A *sequence* is a function $f : \mathbb{N} \rightarrow \mathbb{R}$.

Only the notation is somewhat unusual. Instead of writing $f(n)$ for the value of f at a natural number n , we write a_n . The entire sequence is then referred to with the notation

$$(a_n)_{n \in \mathbb{N}}.$$

The n th term a_n of a sequence may be defined explicitly by a formula involving n , as in the example given above:

$$a_n = \frac{1}{n}, \quad n \in \mathbb{N}.$$

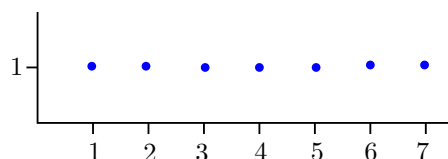
It might also sometimes be defined recursively. For example,

$$a_1 = 1, \quad a_{n+1} = \frac{n}{n+1}a_n \text{ for } n \in \mathbb{N}.$$

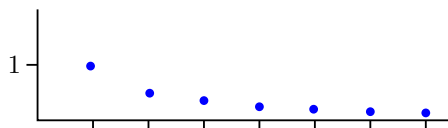
(Write down the first few terms of this sequence.)

Example 2.1. Here are a couple of examples of sequences. We have also listed and displayed the first few terms.

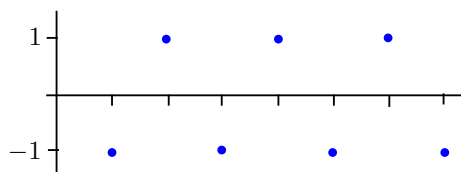
$$(1)_{n \in \mathbb{N}} \quad 1, 1, 1, \dots$$



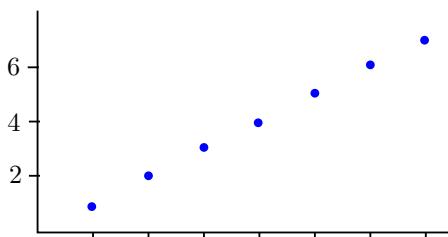
$$\left(\frac{1}{n}\right)_{n \in \mathbb{N}} \quad 1, \frac{1}{2}, \frac{1}{3}, \dots$$



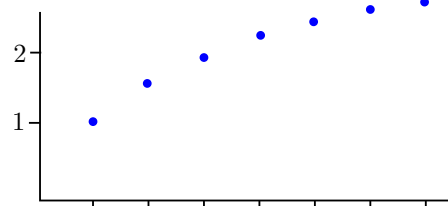
$$((-1)^n)_{n \in \mathbb{N}} \quad -1, 1, -1, 1, \dots$$



$$(n)_{n \in \mathbb{N}} \quad 1, 2, 3, \dots$$



$$\left(1 + \frac{1}{2} + \dots + \frac{1}{n}\right)_{n \in \mathbb{N}} \quad 1, 1 + \frac{1}{2}, 1 + \frac{1}{2} + \frac{1}{3}, \dots$$



◇

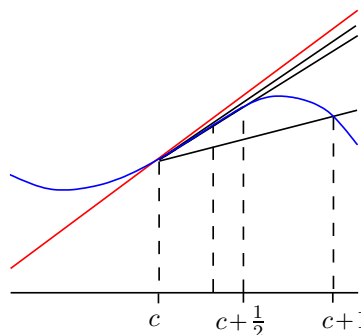
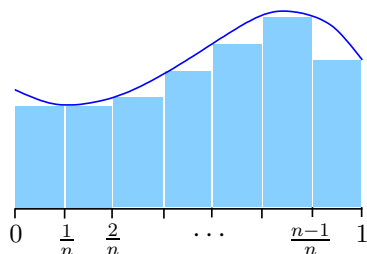
What do we want to know about sequences? In Analysis, we want to know ‘the limiting behaviour’ of the sequence, that is, what a_n behaves like for large n , and in particular, whether a_n gets closer and closer to some number L (called the *limit* of the sequence at hand).

What is the motivation for studying the limiting behaviour of sequences?

For example, the terms of the sequence might be the sum of the areas of the rectangles in the picture on the left below, or it might be the slopes of the chords in the picture on the right, and we might be interested in the limiting behaviour because we want to calculate the area under the graph (left picture) or the instantaneous rate of change of function at the point c (right picture). Thus we want to know what happens when n increases to the sequence $(a_n)_{n \in \mathbb{N}}$ where

$$\text{(Left picture)} \quad a_n = \sum_{k=1}^{n-1} m_k \cdot \frac{k}{n}, \text{ here } m_k := \text{height of } k\text{th shaded rectangle,}$$

$$\text{(Right picture)} \quad a_n = \frac{f(c + \frac{1}{n}) - f(c)}{\frac{1}{n}}.$$



2.1. Limit of a convergent sequence

We want to give a precise definition for

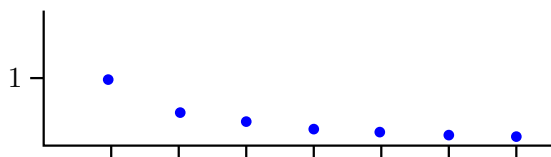
‘The sequence $(a_n)_{n \in \mathbb{N}}$ is convergent with limit L ’ or ‘ $\lim_{n \rightarrow \infty} a_n = L$ ’.

Intuitively, by the above, we mean that there is a number L such that the terms of the sequence are getting ‘closer and closer’ or are ‘settling down’ to L for larger and larger values of n . If there is no such finite number L to which the terms of the sequence get arbitrarily close, then the sequence is said to diverge.

For example, the sequence $\left(\frac{1}{n}\right)_{n \in \mathbb{N}}$ seems to be convergent with limit 0, that is,

$$\lim_{n \rightarrow \infty} \frac{1}{n} = 0.$$

This is consistent with the idea of convergence that we have in mind: a sequence $(a_n)_{n \in \mathbb{N}}$ converges to some real number L , if the terms a_n get ‘closer and closer’ to L as n ‘increases without bound’.

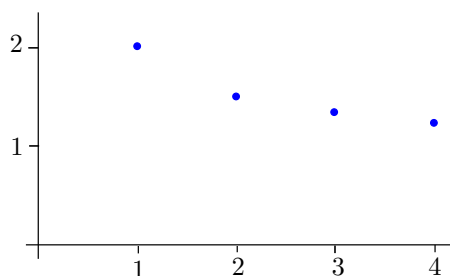


The problem with such a characterization is its imprecision. Exactly what does it mean when we say that the terms of a sequence get ‘closer and closer’ or ‘as close as we like’ or ‘arbitrarily close’ to some number L ? Even if we accept this ambiguity, how would we use the definition to prove theorems that involve sequences?

The terms of the sequence $\left(1 + \frac{1}{n}\right)_{n \in \mathbb{N}}$ are

$$2, \frac{3}{2}, \frac{4}{3}, \frac{5}{4}, \dots,$$

and the first few are plotted below.



The terms of this sequence get ‘closer and closer’ to 0 (indeed the distance to 0 keeps decreasing), but

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right) \neq 0,$$

rather

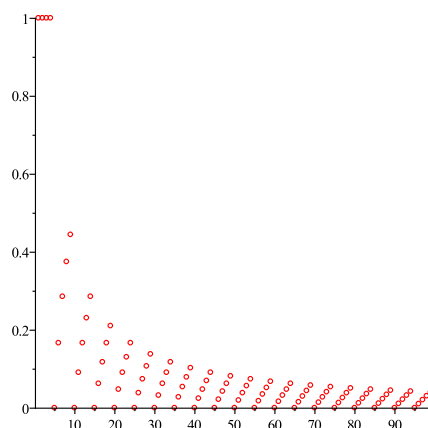
$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right) = 1.$$

One might say ‘but clearly the terms don’t get *arbitrarily* close to 0, but they *do* get arbitrarily close to 1!’

Also, we would also like to say that a sequence is convergent with limit L even if the *adjacent* terms of the sequence do not always *reduce* their distance to L , but it is nevertheless true that the distance to the limit can be made arbitrarily small provided we go far enough in the sequence: An example is the sequence

$$\left(\frac{n \bmod 5}{n}\right)_{n \in \mathbb{N}}.$$

Here $n \bmod 5$ denotes the remainder obtained when n is divided by 5. The graph of the sequence is shown below.



We notice that the limit of this sequence turns out to be 0, despite the fact that any two successive terms may not always reduce the distance to 0. However, given any small distance $\epsilon > 0$, there is some index N beyond which all the terms of the sequence *do* lie within a distance of ϵ from 0. In other words, the sequence *is* settling down to the value 0.

Based on the above examples, we would like to say that a sequence is deemed to be convergent with limit L if

‘No matter what distance ϵ is specified, there is an index N beyond which all the terms $a_{N+1}, a_{N+2}, a_{N+3}, \dots$ all have a distance smaller than ϵ to L .’

In other words

$\forall \epsilon > 0$	$\exists N \in \mathbb{N}$	such that $\forall n > N,$	$ a_n - L < \epsilon$
for every	there is	such that all terms	have distance to L
specified distance ϵ	an index	beyond that index	less than ϵ

(Here the symbol \forall means ‘for every’, and \exists means ‘there exists a/an’.)

With these introductory remarks, we now have the following concrete, precise mathematical definition for the convergence/divergence of a sequence.

Definition 2.2 (Convergent/Divergent sequence; limit).

A sequence $(a_n)_{n \in \mathbb{N}}$ is said to be *convergent with limit* L ($\in \mathbb{R}$) if for every $\epsilon > 0$, there exists¹ an $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$ with $n > N$, we have $|a_n - L| < \epsilon$. Then we write

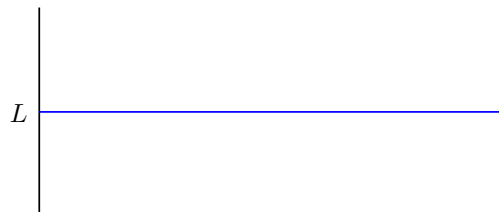
$$\lim_{n \rightarrow \infty} a_n = L.$$

If there is no $L \in \mathbb{R}$ such that $\lim_{n \rightarrow \infty} a_n = L$, then $(a_n)_{n \in \mathbb{N}}$ is called *divergent*.

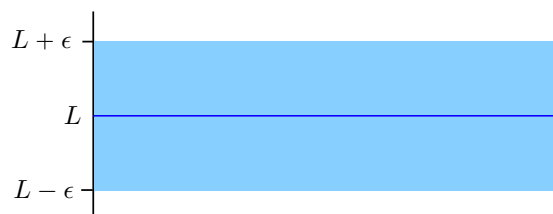
¹depending on ϵ

The picture below gives the geometric meaning of the definition of a sequence being convergent with limit L .

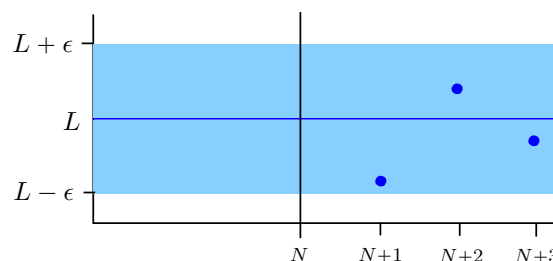
There exists an L



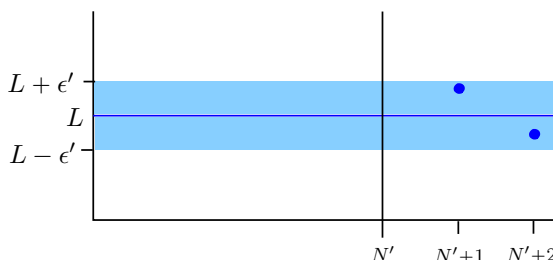
such that no matter what $\epsilon > 0$ we pick and consider a shaded strip of width ϵ around the horizontal line passing through L ,



there exists an index N such that all terms with indices $n > N$ lie in that strip.



Had we chosen a smaller ϵ , then perhaps a larger N' would work.

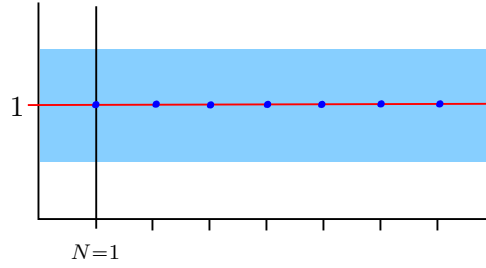


Let us consider some simple examples in order to illustrate the definition.

Example 2.2. $(1)_{n \in \mathbb{N}}$ is convergent with limit 1. We want to check if:

$$\forall \epsilon > 0, \exists N \in \mathbb{N} \text{ such that } \forall n > N, |a_n - L| < \epsilon. \quad (2.1)$$

Well, given $\epsilon > 0$, we have that $|a_n - L| = |1 - 1| = |0| = 0 < \epsilon$ *always*, that is for *all* $n \in \mathbb{N}$! So any $N \in \mathbb{N}$ works. Pictorially, no matter what the width of the shaded region is, *all* the terms of the sequence lie in that shaded strip. So for example, $N = 1$ works.



Here is a rigorous proof of ‘ $\lim_{n \rightarrow \infty} a_n = 1$ ’:

Let $\epsilon > 0$.

Let N be any natural number, say $N = 1$.

Then for all $n > N = 1$, we have $|a_n - L| = |1 - 1| = |0| = 0 < \epsilon$.

So we have checked that the statement in (2.1) holds. \diamond

Example 2.3. $\left(\frac{1}{n}\right)_{n \in \mathbb{N}}$ is a convergent sequence with limit 0.

Before one proceeds to give rigorous proof, we often need to do some rough work. Recall that in order to check the claim, we need to verify

$$\forall \epsilon > 0, \exists N \in \mathbb{N} \text{ such that } \forall n > N, |a_n - L| < \epsilon. \quad (2.2)$$

Thus given $\epsilon > 0$, the task is to find a special index N such that the inequality $|a_n - L| < \epsilon$ is satisfied for all $n > N$. So in order to find this N , we will work backwards, by first starting with the inequality $|a_n - L| < \epsilon$, and making an educated guess about what N is likely to work. Then we will give a formal proof.

(*Rough work:* Let $\epsilon > 0$. We want an N such that for all $n > N$, $|a_n - L| < \epsilon$, i.e.,

$$\left|\frac{1}{n} - 0\right| = \frac{1}{n} < \epsilon,$$

that is, $n > 1/\epsilon$. So we guess that we can take any $N \in \mathbb{N}$ such that $N > 1/\epsilon$, because then for $n > N$, $n > N > 1/\epsilon$, and we may retrace the steps above.)

Rigorous argument:

Let $\epsilon > 0$.

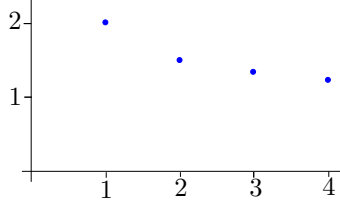
Let $N \in \mathbb{N}$ be such that $N > 1/\epsilon$.

(We use the Archimedean Property here with $y = 1/\epsilon$, $x = 1$: by Theorem 1.3, there exists an $N \in \mathbb{N}$ such that $Nx > y$, that is, $N > 1/\epsilon$.)

Then for all $n \in \mathbb{N}$ with $n > N$, we have $|a_n - L| = \left|\frac{1}{n} - 0\right| = \frac{1}{n} < \frac{1}{N} < \epsilon$.

So $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$. \diamond

Example 2.4. $\left(1 + \frac{1}{n}\right)_{n \in \mathbb{N}}$ is a convergent sequence with limit 1.



(*Rough work:* $|a_n - L| = \left|1 + \frac{1}{n} - 1\right| = \left|\frac{1}{n}\right| = \frac{1}{n} < \epsilon$ for $n > N > \frac{1}{\epsilon}$.)

Rigorous argument:

Let $\epsilon > 0$.

Let $N \in \mathbb{N}$ be such that $N > 1/\epsilon$.

Then for all $n \in \mathbb{N}$ with $n > N$, we have $|a_n - L| = \left|1 + \frac{1}{n} - 1\right| = \left|\frac{1}{n}\right| = \frac{1}{n} < \frac{1}{N} < \epsilon$.

So $\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right) = 1$.

We note that it is **not** the case that $\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right) = 0$. For, if on the contrary,

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right) = 0,$$

then the following statement holds:

$$\forall \epsilon > 0, \exists N \in \mathbb{N} \text{ such that } \forall n > N, |a_n - 0| = \left|1 + \frac{1}{n} - 0\right| = 1 + \frac{1}{n} < \epsilon.$$

But if we take $\epsilon = 1 > 0$, then the above gives the existence of an $N \in \mathbb{N}$ such that

$$\forall n > N, 1 + \frac{1}{n} < \epsilon = 1.$$

If we take $n = N + 1$, then this last inequality gives the contradiction that

$$\frac{1}{N+1} < 0.$$

(We will soon learn (in Theorem 2.1) that in fact if a sequence is convergent with a certain limit L , then it cannot converge to any *other* limit L' . So in light of this result, the last paragraph above is superfluous: indeed, since we proved that

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right) = 1,$$

we immediately know that for any $L' \neq 1$, it cannot be the case that

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right) = L',$$

and in particular, with $L' := 0 \neq 1$, we surely know that $\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right) \neq 0$. \diamond

Here is an example of a divergent sequence.

Example 2.5. $((-1)^n)_{n \in \mathbb{N}}$ is divergent.

We will prove this by contradiction. Let $((-1)^n)_{n \in \mathbb{N}}$ be convergent with limit L . Then

$$\forall \epsilon > 0, \exists N \in \mathbb{N} \text{ such that } \forall n > N, |a_n - L| = |(-1)^n - L| < \epsilon.$$

Take $\epsilon = 1/2$. (This choice is motivated by hindsight – we want to arrive at a contradiction, and it will turn out that this choice of ϵ delivers the contradiction. In order to make this transparent, let us keep working with a general ϵ in our argument below, and at a crucial last step, we will see the rationale behind our choice of $\epsilon = 1/2$!)

Then there exists an $N \in \mathbb{N}$ such that for all $n > N$, $|(-1)^n - L| < \epsilon$. But if we take any *even* $n > N$ (for example $2N, 4N, 6N, 8N, \dots$), then we obtain

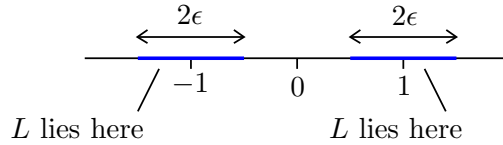
$$|(-1)^n - L| = |1 - L| < \epsilon. \quad (2.3)$$

(This inequality says that the distance of L to 1 is less than ϵ .) On the other hand, if we take any *odd* $n > N$ (for example $2N + 1, 4N + 1, 6N + 1, 8N + 1, \dots$), then

$$|(-1)^n - L| = |-1 - L| < \epsilon. \quad (2.4)$$

(This inequality says that the distance of L to -1 is less than ϵ .)

So pictorially our L is supposed to lie in an interval about 1 with width 2ϵ , and in an interval about -1 with width 2ϵ . But such intervals won't overlap if $\epsilon = 1/2$ (in fact any positive $\epsilon \leq 1$ will do the job!), and this will give us the contradiction.



Indeed, we have, using (2.3) and (2.4) that

$$2 = |-1 - L + L - 1| \leq |-1 - L| + |L - 1| < \epsilon + \epsilon = 2\epsilon = 2 \cdot \frac{1}{2} = 1,$$

a contradiction. Consequently, the sequence $((-1)^n)_{n \in \mathbb{N}}$ is divergent. \diamond

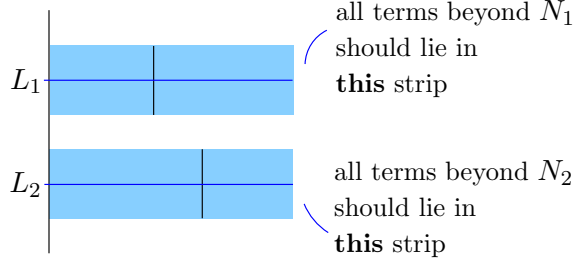
The notation

$$\lim_{n \rightarrow \infty} a_n$$

suggests that the limit of a convergent sequence is unique. Indeed this is the case, and we prove this now.

Theorem 2.1. *A convergent sequence has a unique limit.*

Proof. Let $(a_n)_{n \in \mathbb{N}}$ be a convergent sequence with limits L_1 and L_2 , with $L_1 \neq L_2$.



Let

$$\epsilon := \frac{|L_1 - L_2|}{3} > 0,$$

where the positivity of the ϵ defined above follows from the fact that $L_1 \neq L_2$. Since L_1 is a limit of the sequence $(a_n)_{n \in \mathbb{N}}$, $\exists N_1 \in \mathbb{N}$ such that

$$\text{for all } n > N_1, |a_n - L_1| < \epsilon.$$

Since L_2 is a limit of the sequence $(a_n)_{n \in \mathbb{N}}$, $\exists N_2 \in \mathbb{N}$ such that

$$\text{for all } n > N_2, |a_n - L_2| < \epsilon.$$

Consequently for $n > N_1 + N_2$, we have $n > N_1$ and $n > N_2$, and so

$$|L_1 - L_2| = |L_1 - a_n + a_n - L_2| \leq |L_1 - a_n| + |a_n - L_2| < \epsilon + \epsilon = 2\epsilon = \frac{2}{3}|L_1 - L_2|.$$

So we arrive at the contradiction that $1 < \frac{2}{3}$. Hence our original assumption was incorrect, and so a convergent sequence must have a unique limit. \square

Checking whether a sequence is convergent or not by using the definition is cumbersome. In the rest of the chapter, we will learn ways of deducing the convergence without having to do this hard work. Instead, we will establish results which allow us to deduce the convergence based on certain properties possessed by the sequence. One example of such a result is:

Bounded and monotone sequences are convergent.

So in the next section, among other things, we will study what is meant by a bounded sequence, a monotone sequence, and also see a proof of the result stated above.

Exercise 2.1. (*)

- (1) Can the limit of a convergent sequence be one of the terms of the sequence?
- (2) If none of the terms of a convergent sequence equal its limit, then prove that the terms of the sequence cannot consist of a finite number of distinct values.
- (3) Prove that the sequence $((-1)^n)_{n \in \mathbb{N}}$ is divergent using the above.

Exercise 2.2. In each of the cases listed below, give an example of a divergent sequence $(a_n)_{n \in \mathbb{N}}$ that satisfies the given conditions. Suppose that $L = 1$.

- (1) For all $\epsilon > 0$, there exists an N such that for infinitely many $n > N$, $|a_n - L| < \epsilon$.
- (2) There exists an $\epsilon > 0$ and a $N \in \mathbb{N}$ such that for all $n > N$, $|a_n - L| < \epsilon$.

Exercise 2.3. Suppose that S is a nonempty subset of \mathbb{R} such that S is bounded above. Show that there exists a sequence $(a_n)_{n \in \mathbb{N}}$ contained in S (that is, $a_n \in S$ for all $n \in \mathbb{N}$) and which is convergent with limit equal to $\sup S$.

Exercise 2.4. Suppose that $(a_n)_{n \in \mathbb{N}}$ is a sequence such that for all $n \in \mathbb{N}$, we have $a_n \geq 0$. Prove that if $(a_n)_{n \in \mathbb{N}}$ is convergent with limit L , then $L \geq 0$.

Exercise 2.5. Which of the following listed statements have the same meaning as ‘It is not the case that the sequence $(a_n)_{n \in \mathbb{N}}$ is convergent to L ’?

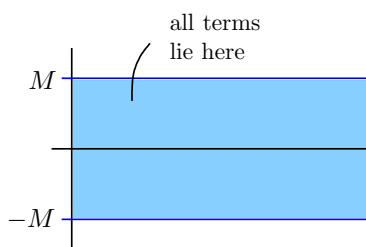
- (A) $\forall \epsilon > 0, \exists N \in \mathbb{N}$ such that $\forall n \in \mathbb{N}$ such that $n > N$, $|a_n - L| \geq \epsilon$.
- (B) $\forall \epsilon > 0, \exists N \in \mathbb{N}$ such that $\forall n \in \mathbb{N}$ such that $n \leq N$, $|a_n - L| \geq \epsilon$.
- (C) $\exists \epsilon > 0, \forall N \in \mathbb{N}, \exists n \in \mathbb{N}$ such that $n > N$ but $|a_n - L| \geq \epsilon$.
- (D) $\exists \epsilon > 0, \exists N \in \mathbb{N}, \forall n \in \mathbb{N}$ such that $n > N$ but $|a_n - L| \geq \epsilon$.

2.2. Bounded and monotone sequences

Bounded sequences.

Definition 2.3 (Bounded sequence).

A sequence $(a_n)_{n \in \mathbb{N}}$ is said to be *bounded* if there exists a $M > 0$ such that for all $n \in \mathbb{N}$, $|a_n| \leq M$.



Note that a sequence is bounded if and only if the set $S = \{a_n : n \in \mathbb{N}\}$ is bounded. (See Exercise 1.20 on page 25).

Example 2.6.

- (1) $(1)_{n \in \mathbb{N}}$ is bounded, since $|1| = 1 \leq 1$ for all $n \in \mathbb{N}$.
- (2) $\left(\frac{1}{n}\right)_{n \in \mathbb{N}}$ is bounded, since $\left|\frac{1}{n}\right| = \frac{1}{n} \leq 1$ for all $n \in \mathbb{N}$.
- (3) $((-1)^n)_{n \in \mathbb{N}}$ is bounded, since $|(-1)^n| = 1 \leq 1$ for all $n \in \mathbb{N}$.

(4) $(n)_{n \in \mathbb{N}}$ is not bounded.

(If there exists an $M > 0$ such that for all $n \in \mathbb{N}$, $|a_n| = |n| = n \leq M$, then this contradicts the Archimedean Property: we know there exists an $N \in \mathbb{N}$ such that with $x = 1$, $N = N \cdot x > M = y$.)

(5) The sequence $(a_n)_{n \in \mathbb{N}}$ is bounded, where

$$a_n = \frac{1}{1^1} + \frac{1}{2^2} + \frac{1}{3^3} + \cdots + \frac{1}{n^n}, \quad n \in \mathbb{N}.$$

Indeed this can be seen as follows:

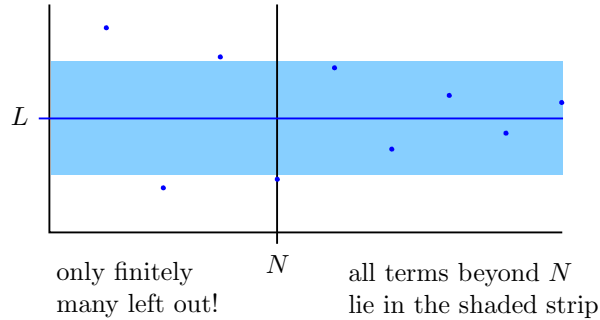
$$\begin{aligned} |a_n| &= \left| \frac{1}{1^1} + \frac{1}{2^2} + \frac{1}{3^3} + \cdots + \frac{1}{n^n} \right| = \frac{1}{1^1} + \frac{1}{2^2} + \frac{1}{3^3} + \cdots + \frac{1}{n^n} \\ &\leq \frac{1}{1^1} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots + \frac{1}{2^n} \\ &= \frac{1}{1^1} + \frac{1}{2} \left(1 - \frac{1}{2} \right) + \frac{1}{2^2} \left(1 - \frac{1}{2} \right) + \cdots + \frac{1}{2^{n-1}} \left(1 - \frac{1}{2} \right) \\ &= 1 + \frac{1}{2} - \frac{1}{2^2} + \frac{1}{2^2} - \frac{1}{2^3} + \cdots + \frac{1}{2^{n-1}} - \frac{1}{2^n} \\ &= 1 + \frac{1}{2} - \frac{1}{2^n} < \frac{3}{2}. \end{aligned}$$

Thus the sequence is bounded. \diamond

The sequences $(1)_{n \in \mathbb{N}}$, $(1/n)_{n \in \mathbb{N}}$ are convergent, and we have shown above that they are also bounded. This is not a coincidence, and in the next theorem we show that the set of all convergent sequences is contained in the set of all bounded sequences.

Theorem 2.2. *If a sequence is convergent, then it is bounded.*

Proof. Let $(a_n)_{n \in \mathbb{N}}$ be a convergent sequence with limit L . Let $\epsilon := 1 > 0$. Then there exists an $N \in \mathbb{N}$ such that for all $n > N$, $|a_n - L| < \epsilon = 1$. Hence for $n > N$, $|a_n| = |a_n - L + L| \leq |a_n - L| + |L| < 1 + |L|$. So all the terms with index beyond N lie in the shaded strip below.



But only finitely many are left out, and surely for $n = 1, \dots, N$,

$$|a_n| \leq \max\{|a_1|, \dots, |a_N|\}.$$

So if we set $M := \max\{|a_1|, \dots, |a_N|, 1 + |L|\}$, then for all $n \in \mathbb{N}$ $|a_n| \leq M$, and so $(a_n)_{n \in \mathbb{N}}$ is bounded. \square

Thus:

$$\text{convergent} \Rightarrow \text{bounded}.$$

But the reverse implication is not true, since for example $((-1)^n)_{n \in \mathbb{N}}$ is bounded, but not convergent. So:

$$\text{convergent} \not\Leftarrow \text{bounded}.$$

But we will see soon enough that if we add the property of being ‘monotone’ to boundedness, then we do get convergence:

$$\text{bounded and ‘monotone’} \Rightarrow \text{convergent}.$$

We will now study what we mean by a monotone sequence before proving this last implication.

Exercise 2.6.

- (1) Let $(b_n)_{n \in \mathbb{N}}$ be a bounded sequence. Prove that $(b_n/n)_{n \in \mathbb{N}}$ is convergent with limit 0.
- (2) Is the sequence $((\sin n)/n)_{n \in \mathbb{N}}$ convergent?

Exercise 2.7. (*)

- (1) If $(a_n)_{n \in \mathbb{N}}$ is a convergent sequence with limit L , then prove that the sequence $(s_n)_{n \in \mathbb{N}}$, where $s_n = \frac{a_1 + \dots + a_n}{n}$ for all $n \in \mathbb{N}$, is also convergent with limit L .
- (2) Give an example such that $(s_n)_{n \in \mathbb{N}}$ is convergent but $(a_n)_{n \in \mathbb{N}}$ is divergent.

Exercise 2.8. Let ℓ^∞ denote² the set of all bounded sequences. Define the set ℓ^2 of all ‘square summable’ sequences $\ell^2 = \{(a_n)_{n \in \mathbb{N}} : (a_1^2 + \dots + a_n^2)_{n \in \mathbb{N}} \text{ is convergent}\}$. Also, let c_{00} be the set of all sequences that are ‘eventually zero’, that is,

$$c_{00} = \{(a_n)_{n \in \mathbb{N}} : \exists N \in \mathbb{N} \text{ such that } \forall n > N, a_n = 0\}.$$

Prove that $c_{00} \subset \ell^2 \subset \ell^\infty$ *Hint:* For the last inclusion, use $a_n^2 \leq a_1^2 + \dots + a_n^2$.

Monotone sequences.

Definition 2.4 (Increasing, decreasing and monotone sequences).

- A sequence $(a_n)_{n \in \mathbb{N}}$ is said to be *increasing* if for all $n \in \mathbb{N}$, $a_n \leq a_{n+1}$, that is, if $a_1 \leq a_2 \leq a_3 \leq \dots$.
- A sequence $(a_n)_{n \in \mathbb{N}}$ is said to be *decreasing* if for all $n \in \mathbb{N}$, $a_n \geq a_{n+1}$, that is, if $a_1 \geq a_2 \geq a_3 \geq \dots$.
- A sequence is said to be *monotone* if it is increasing or decreasing.

²There is some rationale behind this notation, but we will not elaborate on this will now; in later courses in Analysis the motivation for this notation will become clearer.

Example 2.7.

Sequence	Is it increasing?	Is it decreasing?	Is it monotone?
$(n)_{n \in \mathbb{N}}$	Yes	No	Yes
$\left(\frac{1}{1^1} + \frac{1}{2^2} + \frac{1}{3^3} + \cdots + \frac{1}{n^n}\right)_{n \in \mathbb{N}}$	Yes	No	Yes
$(1)_{n \in \mathbb{N}}$	Yes	Yes	Yes
$((-1)^n)_{n \in \mathbb{N}}$	No	No	No
$\left(\frac{1}{n}\right)_{n \in \mathbb{N}}$	No	Yes	Yes

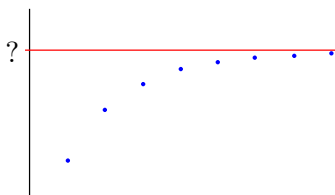
◇

The following theorem can be useful in showing that sequences converge when one does not know the limit beforehand. This is the central result of this section on bounded and monotone sequences.

Theorem 2.3. *If a sequence is monotone and bounded, then it is convergent.*

Proof.

1° We will first consider the case of *increasing* sequences which are bounded. Let $(a_n)_{n \in \mathbb{N}}$ be an increasing and bounded sequence. We want to show that $(a_n)_{n \in \mathbb{N}}$ is convergent. But with what limit?



The picture above suggests that the limit should be the smallest number bigger than each of the terms of this sequence, and if we recall Exercise 1.6, we know that this is the supremum of the set $S := \{a_n : n \in \mathbb{N}\}$. Since $(a_n)_{n \in \mathbb{N}}$ is bounded, it follows that the set S has an upper bound and so $\sup S$ exists. We show that in fact $(a_n)_{n \in \mathbb{N}}$ converges to $\sup S$. Let $\epsilon > 0$. Since $\sup S - \epsilon < \sup S$, we conclude that $\sup S - \epsilon$ is *not* an upper bound for S , and so there exists an $a_N \in S$ such that $\sup S - \epsilon < a_N$, that is $\sup S - a_N < \epsilon$. As $(a_n)_{n \in \mathbb{N}}$ is an increasing sequence, for $n > N$, we have $a_N \leq a_n$. Because $\sup S$ is an upper bound for S , $a_n \leq \sup S$, and so $|a_n - \sup S| = \sup S - a_n$. Thus for $n > N$ we obtain $|a_n - \sup S| = \sup S - a_n \leq \sup S - a_N < \epsilon$.

2° Let $(a_n)_{n \in \mathbb{N}}$ be a *decreasing* and bounded sequence. Then the sequence $(-a_n)_{n \in \mathbb{N}}$ is increasing. Furthermore if $(a_n)_{n \in \mathbb{N}}$ is bounded, then $(-a_n)_{n \in \mathbb{N}}$ is bounded as well ($|-a_n| = |a_n| \leq M$). Hence by the case considered above, it follows that $(-a_n)_{n \in \mathbb{N}}$ is a convergent sequence with limit

$$\sup\{-a_n : n \in \mathbb{N}\} = -\inf\{a_n : n \in \mathbb{N}\} = -\inf S,$$

where $S = \{a_n : n \in \mathbb{N}\}$ (see Exercise 1.7 on page 19). So given $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that for all $n > N$, $|-a_n - (-\inf S)| < \epsilon$, that is, $|a_n - \inf S| < \epsilon$. Thus $(a_n)_{n \in \mathbb{N}}$ is convergent with limit $\inf S$. \square

Exercise 2.9. Fill in the blanks in the following proof of the fact that *every bounded decreasing sequence of real numbers converges*.

Let $(a_n)_{n \in \mathbb{N}}$ be a bounded decreasing sequence of real numbers. Let ℓ_* be the _____ lower bound of $\{a_n : n \in \mathbb{N}\}$. The existence of ℓ_* is guaranteed by the _____ of the set of real numbers. We show that ℓ_* is the _____ of $(a_n)_{n \in \mathbb{N}}$. Taking $\epsilon > 0$, we must show that there exists a positive integer N such that _____ for all $n > N$. Since $\ell_* + \epsilon > \ell_*$, $\ell_* + \epsilon$ is not _____ of $\{a_n : n \in \mathbb{N}\}$. Therefore there exists N with _____ $\leq a_N < \ell_* + \epsilon$. Since $(a_n)_{n \in \mathbb{N}}$ is _____, we have for all $n \geq N$ that $\ell_* - \epsilon < \ell_* \leq \text{_____} \leq a_N < \ell_* + \epsilon$, and so $|a_n - \ell_*| < \epsilon$. \square

The result in Theorem 2.3 gives a sufficient condition for convergence: namely by knowing the *properties* of monotonicity and boundedness (which can be checked by just looking at the terms a_n of the sequence), we can deduce convergence. We do not need to make a guess about what the limit of the sequence is, and we do not need to check the cumbersome Definition 2.2. Here is an example.

Example 2.8. We had seen earlier that the sequence $(a_n)_{n \in \mathbb{N}}$ given by

$$a_n = \frac{1}{1^1} + \frac{1}{2^2} + \frac{1}{3^3} + \cdots + \frac{1}{n^n}, \quad n \in \mathbb{N}$$

is monotone (indeed, it is increasing since

$$a_{n+1} - a_n = \frac{1}{(n+1)^{n+1}} > 0$$

for all $n \in \mathbb{N}$) and bounded (see Example 2.6.(5) on page 42). Thus it follows from Theorem 2.3 that this sequence is convergent. (Although it is known that this sequence is convergent to some limit $L \in \mathbb{R}$, which is the supremum of the terms of the sequence,

$$L = \sup_{n \in \mathbb{N}} \left(\frac{1}{1^1} + \frac{1}{2^2} + \frac{1}{3^3} + \cdots + \frac{1}{n^n} \right),$$

it is so far not even known if the limit³ L is rational or irrational, and this is still an open problem in mathematics!) \diamond

³Also associated with this sequence is the interesting identity $\sum_{n=1}^{\infty} \frac{1}{n^n} = \int_0^1 \frac{1}{x^x} dx$.

We remark that although [boundedness and monotonicity] is a sufficient condition for convergence, it is *not necessary*, as illustrated in the following example.

Example 2.9 (Convergence \nRightarrow [Monotone and bounded]).

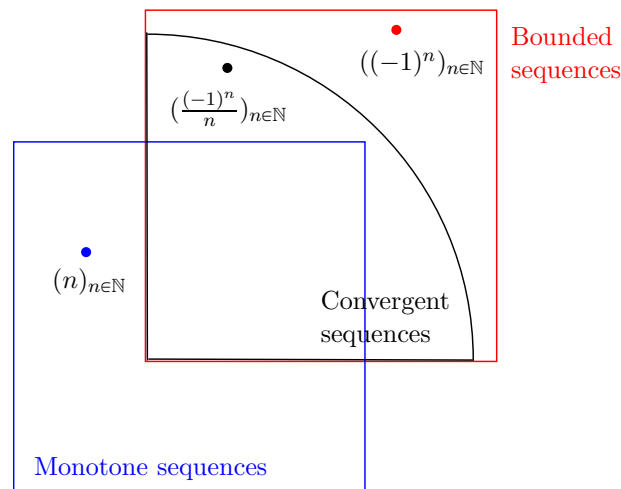
The sequence $(a_n)_{n \in \mathbb{N}}$, where $a_n = \frac{(-1)^n}{n}$ for $n \in \mathbb{N}$, is convergent with limit 0:

Suppose $\epsilon > 0$. Let $N \in \mathbb{N}$ be such that $N > 1/\epsilon$. Then for all $n > N$, we have that $|a_n - 0| = |\frac{(-1)^n}{n} - 0| = \frac{1}{n} < \frac{1}{N} < \epsilon$.

We note that although the sequence is bounded (all convergent sequences are!), it is not monotone: $a_1 = -1 < a_2 = \frac{1}{2} > a_3 = -\frac{1}{3}$. So the sequence is neither increasing (second inequality above), nor decreasing (first inequality above). \diamond

The table below gives a summary of the valid implications, and counterexamples to implications which are not true. See also the Venn diagram after the table.

Question	Answer	Reason/Counterexample
Is every convergent sequence bounded?	Yes	Theorem 2.2
Is every bounded sequence convergent?	No	$((-1)^n)_{n \in \mathbb{N}}$ is bounded, but not convergent.
Is every convergent sequence monotone?	No	$(\frac{(-1)^n}{n})_{n \in \mathbb{N}}$ is convergent, but not monotone.
Is every monotone sequence convergent?	No	$(n)_{n \in \mathbb{N}}$ is not convergent.
Is every bounded and monotone sequence convergent?	Yes	Theorem 2.3



Exercise 2.10. Let $(a_n)_{n \in \mathbb{N}}$ be defined by $a_1 = 1$, and $a_n = \frac{2n+1}{3n}a_{n-1}$ for $n \geq 2$.

- (1) Show that $(a_n)_{n \in \mathbb{N}}$ is bounded.
- (2) Show that $(a_n)_{n \in \mathbb{N}}$ is decreasing.
- (3) Conclude that $(a_n)_{n \in \mathbb{N}}$ is convergent.

Exercise 2.11. Given a bounded sequence $(a_n)_{n \in \mathbb{N}}$, define

$$\ell_k = \inf\{a_n : n \geq k\} \text{ and } u_k = \sup\{a_n : n \geq k\}, \quad k \in \mathbb{N}.$$

Show that the sequences $(\ell_n)_{n \in \mathbb{N}}$, $(u_n)_{n \in \mathbb{N}}$ are bounded and monotone, and conclude that they are convergent. Their respective limits are called the *limit superior* and *limit inferior*, respectively, and are denoted by $\liminf_{n \rightarrow \infty} a_n$ and $\limsup_{n \rightarrow \infty} a_n$.

Exercise 2.12 (Precursor to Euler's number, e).

Consider the sequence $(a_n)_{n \in \mathbb{N}}$, where $a_n := 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!}$, $n \in \mathbb{N}$.

- (1) Show that $(a_n)_{n \in \mathbb{N}}$ is increasing.
- (2) Show that $(a_n)_{n \in \mathbb{N}}$ is bounded.

Hint: $a_n = 1 + 1 + \frac{1}{2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{2 \cdot 3 \cdots n} \leq 1 + 1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^{n-1}} = 1 + \frac{1 - \frac{1}{2^n}}{1 - \frac{1}{2}} < 3$.

- (3) Conclude that $(a_n)_{n \in \mathbb{N}}$ is convergent. We set $a := \lim_{n \rightarrow \infty} a_n$.

Consider the sequence $(b_n)_{n \in \mathbb{N}}$, where $b_n := \left(1 + \frac{1}{n}\right)^n$, $n \in \mathbb{N}$. Using the binomial theorem,

$$\begin{aligned} b_n &= 1 + n \frac{1}{n} + \frac{n(n-1)}{2!} \frac{1}{n^2} + \cdots + \frac{n(n-1) \cdots 2 \cdot 1}{n!} \frac{1}{n^n} \\ &= 1 + 1 + \frac{1}{2!} \left(1 - \frac{1}{n}\right) + \cdots + \frac{1}{n!} \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{n-1}{n}\right). \end{aligned}$$

- (4) Show by replacing n by $n+1$ in factors of the type $(1 - \frac{k}{n})$ that $b_n \leq b_{n+1}$, $n \in \mathbb{N}$.
- (5) Show that $b_n \leq a_n < 3$.
- (6) Conclude that $(b_n)_{n \in \mathbb{N}}$ is convergent. We set $b := \lim_{n \rightarrow \infty} b_n$.

In Exercise 2.20, we will show that in fact $a = b$, and this common value is denoted by e .

Exercise 2.13. In continuation to Exercise 2.8, we also define the set ℓ^1 of 'summable sequences' as $\ell^1 = \{(a_n)_{n \in \mathbb{N}} : (|a_1| + \cdots + |a_n|)_{n \in \mathbb{N}} \text{ is a convergent sequence}\}$.

- (1) Show that $\ell^1 \subset \ell^\infty$. *Hint:* $|a_n| \leq |a_1| + \cdots + |a_n|$.
- (2) Show that $\ell^1 \subset \ell^2$. *Hint:* If $|a_1| + \cdots + |a_n| \leq M$, then $a_1^2 + \cdots + a_n^2 \leq M(|a_1| + \cdots + |a_n|)$.
- (3) It will be show in Example 2.17 that $(1 + \frac{1}{2} + \cdots + \frac{1}{n})_{n \in \mathbb{N}}$ diverges. On the other hand, show that $(1 + \frac{1}{2} + \cdots + \frac{1}{n^2})_{n \in \mathbb{N}}$ converges. *Hint:* $\frac{1}{n^2} \leq \frac{1}{n(n-1)} = \frac{1}{2} \left(\frac{1}{n-1} - \frac{1}{n}\right)$ for $n \geq 2$.

Remark: Thus $(\frac{1}{n})_{n \in \mathbb{N}} \in \ell^2 \setminus \ell^1$.

2.3. Algebra of limits

In this section we will learn that if we 'algebraically' combine the terms of convergent sequences, then the new sequence which is obtained, is again convergent, and moreover the limit of this sequence is the same algebraic combination of the limits.

In this manner we can sometimes prove the convergence of complicated sequences by breaking them down and writing them as an algebraic combination of simple sequences. Thus, we conveniently apply arithmetic rules to compute the limits of sequences if the terms are the sum, product, quotient of terms of simpler sequences with a known limit. For instance, using the formal definition of a limit, one can show that the sequence $(a_n)_{n \in \mathbb{N}}$ defined by

$$a_n = \frac{4n^2 + 9}{3n^2 + 7n + 11}$$

converges to $\frac{4}{3}$. However, it is simpler to observe that

$$a_n = \frac{n^2 \left(4 + \frac{9}{n^2}\right)}{n^2 \left(3 + \frac{7}{n} + \frac{11}{n^2}\right)} = \frac{4 + \frac{9}{n^2}}{3 + \frac{7}{n} + \frac{11}{n^2}},$$

and by a repeated application of Theorem 2.4 given below, we obtain

$$\lim_{n \rightarrow \infty} a_n = \frac{\lim_{n \rightarrow \infty} \left(4 + \frac{9}{n^2}\right)}{\lim_{n \rightarrow \infty} \left(3 + \frac{7}{n} + \frac{11}{n^2}\right)} = \frac{\lim_{n \rightarrow \infty} 4 + \lim_{n \rightarrow \infty} \frac{9}{n^2}}{\lim_{n \rightarrow \infty} 3 + \lim_{n \rightarrow \infty} \frac{7}{n} + \lim_{n \rightarrow \infty} \frac{11}{n^2}} = \frac{4 + 0}{3 + 0 + 0} = \frac{4}{3}.$$

Theorem 2.4. *If $(a_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}}$ are convergent sequences, then:*

- (1) *For all $\alpha \in \mathbb{R}$, $(\alpha a_n)_{n \in \mathbb{N}}$ is a convergent sequence and $\lim_{n \rightarrow \infty} \alpha a_n = \alpha \lim_{n \rightarrow \infty} a_n$.*
- (2) *$(|a_n|)_{n \in \mathbb{N}}$ is a convergent sequence and $\lim_{n \rightarrow \infty} |a_n| = \left| \lim_{n \rightarrow \infty} a_n \right|$.*
- (3) *$(a_n + b_n)_{n \in \mathbb{N}}$ is convergent and $\lim_{n \rightarrow \infty} (a_n + b_n) = \lim_{n \rightarrow \infty} a_n + \lim_{n \rightarrow \infty} b_n$.*
- (4) *$(a_n b_n)_{n \in \mathbb{N}}$ is a convergent sequence and $\lim_{n \rightarrow \infty} a_n b_n = \left(\lim_{n \rightarrow \infty} a_n \right) \left(\lim_{n \rightarrow \infty} b_n \right)$.*
- (5) *For all $k \in \mathbb{N}$, $(a_n^k)_{n \in \mathbb{N}}$ is a convergent sequence and $\lim_{n \rightarrow \infty} a_n^k = \left(\lim_{n \rightarrow \infty} a_n \right)^k$.*
- (6) *If for all $n \in \mathbb{N}$, $b_n \neq 0$ and $\lim_{n \rightarrow \infty} b_n \neq 0$, then $\left(\frac{1}{b_n} \right)_{n \in \mathbb{N}}$ is convergent, and*

$$\lim_{n \rightarrow \infty} \frac{1}{b_n} = \frac{1}{\lim_{n \rightarrow \infty} b_n}.$$

Proof. Let $(a_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}}$ converge to L_a and L_b , respectively.

- (1) If $\alpha = 0$, then $\alpha a_n = 0$ for all $n \in \mathbb{N}$ and clearly $(0)_{n \in \mathbb{N}}$ is a convergent sequence with limit 0. Thus

$$\lim_{n \rightarrow \infty} \alpha a_n = 0 = 0L_a = \alpha \lim_{n \rightarrow \infty} a_n.$$

If $\alpha \neq 0$, then given $\epsilon > 0$, let $N \in \mathbb{N}$ be such that for all $n > N$,

$$|a_n - L_a| < \frac{\epsilon}{|\alpha|},$$

that is,

$$|\alpha a_n - \alpha L_a| = |\alpha| |a_n - L_a| < |\alpha| \frac{\epsilon}{|\alpha|} = \epsilon.$$

So $(\alpha a_n)_{n \in \mathbb{N}}$ is convergent with limit αL_a , i.e., $\lim_{n \rightarrow \infty} \alpha a_n = \alpha L_a = \alpha \lim_{n \rightarrow \infty} a_n$.

- (2) Given $\epsilon > 0$, let $N \in \mathbb{N}$ be such that for all $n > N$, $|a_n - L_a| < \epsilon$. Then we have for all $n > N$: $||a_n| - |L_a|| \leq |a_n - L_a| < \epsilon$. Hence $(|a_n|)_{n \in \mathbb{N}}$ is convergent with limit $|L_a|$, that is,

$$\lim_{n \rightarrow \infty} |a_n| = |L_a| = \left| \lim_{n \rightarrow \infty} a_n \right|.$$

- (3) Given $\epsilon > 0$, let $N_a \in \mathbb{N}$ be such that for all $n > N_a$,

$$|a_n - L_a| < \frac{\epsilon}{2}.$$

Let $N_b \in \mathbb{N}$ be such that for all $n > N_b$,

$$|b_n - L_b| < \frac{\epsilon}{2}.$$

Then for all $n > N := \max\{N_a, N_b\}$, we have

$$|a_n + b_n - (L_a + L_b)| = |a_n - L_a + b_n - L_b| \leq |a_n - L_a| + |b_n - L_b| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Hence $(a_n + b_n)_{n \in \mathbb{N}}$ is convergent with limit $L_a + L_b$, that is,

$$\lim_{n \rightarrow \infty} (a_n + b_n) = L_a + L_b = \lim_{n \rightarrow \infty} a_n + \lim_{n \rightarrow \infty} b_n.$$

- (4) Note that

$$\begin{aligned} |a_n b_n - L_a L_b| &= |a_n b_n - L_a b_n + L_a b_n - L_a L_b| \leq |a_n b_n - L_a b_n| + |L_a b_n - L_a L_b| \\ &= |a_n - L_a| |b_n| + |L_a| |b_n - L_b|. \end{aligned} \tag{2.5}$$

Given $\epsilon > 0$, we need to find a N such that for all $n > N$,

$$|a_n b_n - L_a L_b| < \epsilon.$$

This can be achieved by finding a N such that each of the summands in (2.5) is less than $\epsilon/2$ for $n > N$. This can be done as follows.

Step 1. Since $(b_n)_{n \in \mathbb{N}}$ is convergent, by Theorem 2.2 it follows that it is bounded: $\exists M > 0$ such that for all $n \in \mathbb{N}$, $|b_n| \leq M$. Let $N_a \in \mathbb{N}$ be such that for $n > N_a$,

$$|a_n - L_a| < \frac{\epsilon}{2M}.$$

Step 2. Let $N_b \in \mathbb{N}$ be such that for all $n > N_b$,

$$|b_n - L_b| < \frac{\epsilon}{2(|L_a| + 1)}.$$

(We add $+1$ in the denominator to take care of the case when $L_a = 0$.) Thus for $n > N := \max\{N_a, N_b\}$, we have

$$|a_n b_n - L_a L_b| \leq |a_n - L_a| |b_n| + |L_a| |b_n - L_b| < \frac{\epsilon}{2M} M + |L_a| \frac{\epsilon}{2(|L_a|+1)} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

So $(a_n b_n)_{n \in \mathbb{N}}$ is a convergent sequence with limit $L_a L_b$, that is,

$$\lim_{n \rightarrow \infty} a_n b_n = L_a L_b = \left(\lim_{n \rightarrow \infty} a_n \right) \left(\lim_{n \rightarrow \infty} b_n \right).$$

- (5) This can be shown by using induction on k and part 2.4 above. It is trivially true with $k = 1$. Suppose it holds for some k : then $(a_n^k)_{n \in \mathbb{N}}$ is convergent and

$$\lim_{n \rightarrow \infty} a_n^k = \left(\lim_{n \rightarrow \infty} a_n \right)^k.$$

Hence by part 2.4 above applied to the sequences $(a_n)_{n \in \mathbb{N}}$ and $(a_n^k)_{n \in \mathbb{N}}$, we obtain that the sequence $(a_n \cdot a_n^k)_{n \in \mathbb{N}}$ is convergent and

$$\lim_{n \rightarrow \infty} a_n a_n^k = \left(\lim_{n \rightarrow \infty} a_n \right) \left(\lim_{n \rightarrow \infty} a_n^k \right) = \left(\lim_{n \rightarrow \infty} a_n \right) \left(\lim_{n \rightarrow \infty} a_n \right)^k = \left(\lim_{n \rightarrow \infty} a_n \right)^{k+1}.$$

Thus $(a_n^{k+1})_{n \in \mathbb{N}}$ is convergent and $\lim_{n \rightarrow \infty} a_n^{k+1} = \left(\lim_{n \rightarrow \infty} a_n \right)^{k+1}$.

- (6) Let $N_1 \in \mathbb{N}$ be such that for all $n > N_1$, $|b_n - L_b| < \frac{|L_b|}{2}$. Thus for all $n > N_1$,

$$|L_b| - |b_n| \leq ||L_b| - |b_n|| \leq |b_n - L_b| < \frac{|L_b|}{2},$$

and so $|b_n| \geq \frac{|L_b|}{2}$. Let $\epsilon > 0$, and let $N_2 \in \mathbb{N}$ be such that for all $n > N_2$,

$$|b_n - L_b| < \frac{\epsilon |L_b|^2}{2}.$$

Hence for $n > N := \max\{N_1, N_2\}$, we have

$$\left| \frac{1}{b_n} - \frac{1}{L_b} \right| = \frac{|b_n - L_b|}{|b_n| |L_b|} < \frac{\epsilon |L_b|^2}{2} \frac{2}{|L_b|} \frac{1}{|L_b|} = \epsilon.$$

So $\left(\frac{1}{b_n} \right)_{n \in \mathbb{N}}$ is convergent and $\lim_{n \rightarrow \infty} \frac{1}{b_n} = \frac{1}{L_b} = \frac{1}{\lim_{n \rightarrow \infty} b_n}$. □

Example 2.10. Consider the sequence $(a_n)_{n \in \mathbb{N}}$, where

$$a_n := \frac{1}{n^3} + \frac{2^2}{n^3} + \frac{3^2}{n^3} + \cdots + \frac{n^2}{n^3}, \quad n \in \mathbb{N}.$$

A novice observes that

$$\lim_{n \rightarrow \infty} \frac{1}{n^3} = 0, \quad \lim_{n \rightarrow \infty} \frac{2^2}{n^3} = 0, \quad \lim_{n \rightarrow \infty} \frac{3^2}{n^3} = 0, \quad \cdots, \quad \lim_{n \rightarrow \infty} \frac{n^2}{n^3} = 0,$$

and hastily concludes that

$$\text{'by the Algebra of Limits, } \lim_{n \rightarrow \infty} a_n = 0 + 0 + 0 + \cdots + 0 = 0.'$$

Where does the error in this argument lie?

Note that by Theorem 2.4.(3), we do have that the termwise sum of a *finite fixed* number of sequences is convergent with the limit of the sum being the sum of the limits. In other words, if

$$\begin{aligned} a_{n,1} &\xrightarrow{n \rightarrow \infty} L_1, \\ a_{n,2} &\xrightarrow{n \rightarrow \infty} L_2, \\ a_{n,3} &\xrightarrow{n \rightarrow \infty} L_3, \\ &\dots \\ a_{n,k} &\xrightarrow{n \rightarrow \infty} L_k, \end{aligned}$$

then we do have that $a_{n,1} + a_{n,2} + a_{n,3} + \dots + a_{n,k} \xrightarrow{n \rightarrow \infty} L_1 + L_2 + L_3 + \dots + L_k$.

However, in the application above, the number of sequences wasn't fixed. In fact, knowing the following formula for the sum of squares (which can easily be shown by induction)

$$1^2 + 2^2 + 3^2 + \dots + n^2 = \frac{n(n+1)(2n+1)}{6}, \quad n \in \mathbb{N},$$

we have

$$\begin{aligned} a_n &= \frac{1}{n^3} + \frac{2^2}{n^3} + \frac{3^2}{n^3} + \dots + \frac{n^2}{n^3} = \frac{1^2 + 2^2 + 3^2 + \dots + n^2}{n^3} \\ &= \frac{n(n+1)(2n+1)/6}{n^3} = \frac{1}{6} \left(1 + \frac{1}{n}\right) \left(2 + \frac{1}{n}\right), \end{aligned}$$

and so by the Algebra of Limits,

$$\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \frac{1}{6} \left(1 + \frac{1}{n}\right) \left(2 + \frac{1}{n}\right) = \frac{1}{6} \cdot (1+0) \cdot (2+0) = \frac{1}{3}. \quad \diamond$$

Exercise 2.14. Is the following manipulation justified based on Theorem 2.4?

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = \left(\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)\right)^n = \left(1 + \lim_{n \rightarrow \infty} \frac{1}{n}\right)^n = (1+0)^n = 1^n = 1.$$

Exercise 2.15. Suppose that the sequence $(a_n)_{n \in \mathbb{N}}$ is convergent, and assume that the sequence $(b_n)_{n \in \mathbb{N}}$ is bounded. Prove that the sequence $(c_n)_{n \in \mathbb{N}}$ defined by

$$c_n = \frac{a_n b_n + 5n}{a_n^2 + n}, \quad n \in \mathbb{N},$$

is convergent, and find its limit.

Exercise 2.16. Let $(a_n)_{n \in \mathbb{N}}$ be a convergent sequence with limit L and suppose that $a_n \geq 0$ for all $n \in \mathbb{N}$. Prove that the sequence $(\sqrt{a_n})_{n \in \mathbb{N}}$ is also convergent, with limit \sqrt{L} .

Hint: First show that $L \geq 0$. Let $\epsilon > 0$. If $L = 0$, then choose $N \in \mathbb{N}$ large enough so that for $n > N$, $|a_n - L| = a_n < \epsilon^2$. If $L > 0$, then choose $N \in \mathbb{N}$ large enough so that for $n > N$, $|\sqrt{a_n} - \sqrt{L}| |\sqrt{a_n} + \sqrt{L}| = |a_n - L| < \epsilon \sqrt{L}$.

Exercise 2.17. Show that $(\sqrt{n^2 + n} - n)_{n \in \mathbb{N}}$ is a convergent sequence and find its limit.

Hint: 'Rationalize the numerator' by using $\sqrt{n^2 + n} + n$.

Exercise 2.18.

- (1) Let $(a_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}}$ be convergent sequences such that for all $n \in \mathbb{N}$, $a_n \leq b_n$. Show that $\lim_{n \rightarrow \infty} a_n \leq \lim_{n \rightarrow \infty} b_n$. *Hint:* Use Exercise 2.4 on page 41.
- (2) With the same notation as in Exercise 2.11, show that for a bounded sequence $(a_n)_{n \in \mathbb{N}}$, $\liminf_{n \rightarrow \infty} a_n \leq \limsup_{n \rightarrow \infty} a_n$. Given an example to show that there can be a strict inequality.

Exercise 2.19. The *Fibonacci sequence* $(F_n)_{n \in \mathbb{N}}$ is defined recursively by $F_1 = F_2 = 1$ and for $n \geq 2$, $F_{n+1} = F_n + F_{n-1}$. Thus the sequence has the terms 1, 1, 2, 3, 5, 8, 13, \dots . Define the new sequence $(x_n)_{n \in \mathbb{N}}$ by $x_n = \frac{F_{2n-1}}{F_{2n}}$, $n \geq 1$.

- (1) Show that $x_{n+1} = \frac{1+x_n}{2+x_n}$, $n \geq 1$.
- (2) Prove that $x_{n+1} - x_n = \frac{x_n - x_{n-1}}{(2+x_n)(2+x_{n-1})}$, $n \geq 2$.
- (3) Show that $(x_n)_{n \in \mathbb{N}}$ is convergent.
- (4) Determine the limit of $(x_n)_{n \in \mathbb{N}}$.
Hint: If $(x_n)_{n \in \mathbb{N}}$ converges with limit L , then $(x_{n+1})_{n \in \mathbb{N}}$ converges to L too.
- (5) A French mathematician travelling by car in the UK notices the following curiosity:
In order to approximately convert a Fibonacci number of miles into kilometers, all one needs to do is to take the next Fibonacci number! Can you explain why?

Exercise 2.20 (Euler's constant, e).

Let us revisit Exercise 2.12.

- (1) Fix $m \in \mathbb{N}$. Show that for $n \geq m$, we have

$$b_n \geq 1 + 1 + \frac{1}{2!} \left(1 - \frac{1}{n}\right) + \dots + \frac{1}{m!} \left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{m-1}{n}\right).$$

Use Exercise 2.18(1) to conclude first that $b \geq a_m$, and next that $b \geq a$.

- (2) Use the inequality from Exercise 2.12(5) to conclude that also $b \leq a$.
- (3) Conclude that $b = a$.

We call this number *Euler's number*, denoted by $e \in \mathbb{R}$:

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \dots + \frac{1}{n!}\right).$$

We will see later in Theorem 4.27 that $e \notin \mathbb{Q}$.

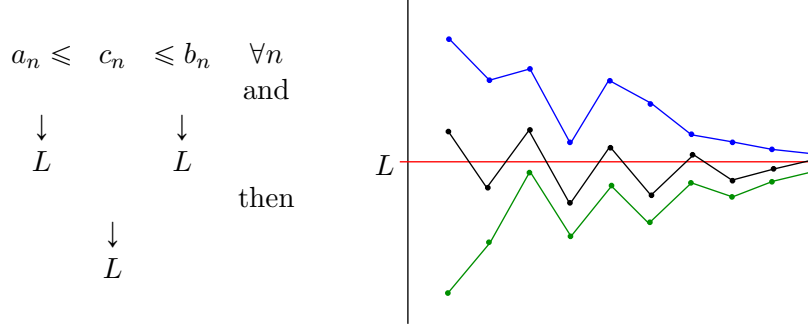
Exercise 2.21. Let the sequence $(x_n)_{n \in \mathbb{N}}$ be defined by $x_1 = 0$ and for $n \geq 1$,

$$x_{n+1} = x_n + \left(\frac{1}{2019} + x_n^2\right)^{2020}.$$

Prove that $(x_n)_{n \in \mathbb{N}}$ diverges. *Hint:* If $(x_n)_{n \in \mathbb{N}}$ converges with limit L , then $(x_{n+1})_{n \in \mathbb{N}}$ converges to L too.

2.4. Sandwich theorem

Another useful theorem for proving that sequences are convergent and in determining their limits is the so-called Sandwich Theorem. Roughly speaking, it says that if a sequence is 'sandwiched' between two convergent sequences with the *same* limit, then the sandwiched sequence is also convergent with the same limit.



Theorem 2.5 (Sandwich theorem).

Let $(a_n)_{n \in \mathbb{N}}$, $(b_n)_{n \in \mathbb{N}}$ be convergent sequences with the same limit, that is,

$$\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n.$$

If $(c_n)_{n \in \mathbb{N}}$ is a third sequence such that

$$\text{for all } n \in \mathbb{N}, \quad a_n \leq c_n \leq b_n,$$

then $(c_n)_{n \in \mathbb{N}}$ is also convergent with the same limit, that is,

$$\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} c_n = \lim_{n \rightarrow \infty} b_n.$$

Proof. Let L denote the common limit of $(a_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}}$:

$$\lim_{n \rightarrow \infty} a_n = L = \lim_{n \rightarrow \infty} b_n.$$

Given $\epsilon > 0$, let $N_a \in \mathbb{N}$ be such that for all $n > N_a$, $|a_n - L| < \epsilon$. For $n > N_a$,

$$L - a_n \leq |L - a_n| = |a_n - L| < \epsilon,$$

and so $L - a_n < \epsilon$, that is,

$$L - \epsilon < a_n.$$

Let $N_b \in \mathbb{N}$ be such that for all $n > N_b$, $|b_n - L| < \epsilon$. For $n > N_b$, $b_n - L < \epsilon$, i.e.,

$$b_n < L + \epsilon.$$

Thus for $n > N := \max\{N_a, N_b\}$, we have

$$L - \epsilon < a_n \leq c_n \leq b_n < L + \epsilon,$$

and so $L - \epsilon < c_n < L + \epsilon$. Consequently, $c_n - L < \epsilon$ and $-(c_n - L) < \epsilon$, and so

$$|c_n - L| < \epsilon.$$

This proves that $(c_n)_{n \in \mathbb{N}}$ is convergent with limit L . □

Example 2.11 (The geometric progression).

The aim of this example is to show that if $|r| < 1$, then $\lim_{n \rightarrow \infty} r^n = 0$.

First let us consider the case when $r \in (0, 1)$. Then $h := \frac{1}{r} - 1 > 0$. For $n \in \mathbb{N}$,

$$\frac{1}{r^n} = \underbrace{(1+h)^n \geq 1+nh \geq nh}_{(*)}. \quad (2.6)$$

One can show the inequality $(*)$ using induction as follows. Clearly when $n = 1$,

$$(1+h)^1 = 1+h = 1+1 \cdot h.$$

If $(1+h)^n \geq 1+nh$ for some n , then

$$(1+h)^{n+1} = (1+h)^n(1+h) \geq (1+nh)(1+h) = 1+(n+1)h+nh^2 \geq 1+(n+1)h,$$

and so the inequality is true for all n .

Hence we obtain $0 \leq r^n \leq \frac{1}{nh}$ for all $n \in \mathbb{N}$. Since

$$\lim_{n \rightarrow \infty} 0 = 0 = \lim_{n \rightarrow \infty} \frac{1}{nh},$$

it follows by the Sandwich Theorem that $\lim_{n \rightarrow \infty} r^n = 0$ too.

When $r = 0$, $r^n = 0$ for all $n \in \mathbb{N}$, and so clearly $\lim_{n \rightarrow \infty} r^n = 0$.

Now suppose that $|r| < 1$. Then $|r| \in [0, 1)$, and so by the above,

$$\lim_{n \rightarrow \infty} |r|^n = 0.$$

By the Algebra of Limits, $\lim_{n \rightarrow \infty} -|r|^n = 0$ as well. Since

$$-|r|^n \leq r^n \leq |r|^n \text{ for all } n \in \mathbb{N},$$

it follows again by the Sandwich Theorem that $\lim_{n \rightarrow \infty} r^n = 0$.

As a consequence of the above, we can show that if $r \in (-1, 1)$, then the ‘sequence of partial sums’ $(1 + r + r^2 + \cdots + r^n)_{n \in \mathbb{N}}$ converges because

$$\begin{aligned} 1 + r + r^2 + \cdots + r^n &= \frac{(1-r)(1+r+r^2+\cdots+r^n)}{1-r} \\ &= \frac{1+r+\cdots+r^n - (r+r^2+\cdots+r^{n+1})}{1-r} \\ &= \frac{1-r^{n+1}}{1-r} = \frac{1-r \cdot r^n}{1-r}, \end{aligned}$$

and so $\lim_{n \rightarrow \infty} (1 + r + r^2 + \cdots + r^n) = \lim_{n \rightarrow \infty} \frac{1-r \cdot r^n}{1-r} = \frac{1-r \cdot 0}{1-r} = \frac{1}{1-r}$. ◇

Exercise 2.22 (Decimal expansions). By a decimal expansion

$$N.d_1d_2d_3\cdots$$

where $N \in \mathbb{N}$, and $d_n \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, we mean the real number

$$L := N + \lim_{n \rightarrow \infty} \left(\frac{d_1}{10} + \frac{d_2}{10^2} + \frac{d_3}{10^3} + \cdots + \frac{d_n}{10^n} \right).$$

(1) Show that $(\frac{d_1}{10} + \frac{d_2}{10^2} + \frac{d_3}{10^3} + \cdots + \frac{d_n}{10^n})_{n \in \mathbb{N}}$ is convergent for any sequence $(d_n)_{n \in \mathbb{N}}$.

(2) Prove that $0.999\cdots = 1.000\cdots$.

(3) Suppose the decimal expansion is nonterminating and repeating, that is, of the form

$$r = N.d_1 \cdots d_n \boxed{d_{n+1} \cdots d_{n+m}} \boxed{d_{n+1} \cdots d_{n+m}} \boxed{d_{n+1} \cdots d_{n+m}} \cdots$$

where a block of digits $\boxed{d_{n+1} \cdots d_{n+m}}$ keeps repeating. Show that $r \in \mathbb{Q}$.

(Conversely, every nonnegative rational number has either a terminating or a repeating decimal expansion; see the Appendix to this chapter.)

(4) Is $0.123456878910111213\cdots$ a rational number?

Example 2.12. $\lim_{n \rightarrow \infty} a^{1/n} = 1$ for $a > 1$.

For concreteness, let us take $a = 2$, but the proof is the same, *mutatis mutandis*⁴, for any $a > 1$. As $2 > 1$, we have $2^{1/n} > 1$ for all $n \in \mathbb{N}$. So we can write $2^{1/n} = 1 + h$, where $h := 2^{1/n} - 1 > 0$. Thus

$$2 = (1 + h)^n = 1 + nh + \underbrace{\binom{n}{2}h^2 + \cdots + h^n}_{\geq 0} \geq 1 + nh,$$

(where the inequality above can also be shown as the justification of $(*)$ in (2.6)), and so $1 \geq nh$. This gives

$$\frac{1}{n} \geq h = 2^{1/n} - 1 > 0 \text{ for all } n \in \mathbb{N}.$$

By the Sandwich Theorem, $\lim_{n \rightarrow \infty} (2^{1/n} - 1) = 0$, that is, $\lim_{n \rightarrow \infty} 2^{1/n} = 1$. \diamond

Exercise 2.23. Show that for all $a > 0$, $\lim_{n \rightarrow \infty} a^{1/n} = 1$.

Example 2.13. For any $a, b \in \mathbb{R}$, $\lim_{n \rightarrow \infty} (|a|^n + |b|^n)^{\frac{1}{n}} = \max\{|a|, |b|\}$.

Let $M := \max\{|a|, |b|\}$. Then $|a| \leq M$ gives $|a|^n \leq M^n$, and similarly $|b|^n \leq M^n$. Thus $|a|^n + |b|^n \leq 2M^n$, and so $(|a|^n + |b|^n)^{1/n} \leq 2^{1/n}M$. Also, $|a|^n + |b|^n \geq M^n$ gives $(|a|^n + |b|^n)^{1/n} \geq M$. So we have

$$M \leq (|a|^n + |b|^n)^{1/n} \leq 2^{1/n}M \text{ for all } n \in \mathbb{N}.$$

Since $\lim_{n \rightarrow \infty} 2^{1/n} = 1$, we have

$$\lim_{n \rightarrow \infty} M = M = \lim_{n \rightarrow \infty} (2^{1/n}M).$$

⁴Latin phrase meaning “changing only those things which need to be changed”

It follows from the Sandwich Theorem that $\lim_{n \rightarrow \infty} (|a|^n + |b|^n)^{1/n} = M = \max\{|a|, |b|\}$. In particular, with $a = 27$ and $b = 2014$, we have that

$$\lim_{n \rightarrow \infty} (27^n + 2014^n)^{\frac{1}{n}} = 2014.$$

The first few terms of the sequence are given (upon rounding to three decimal places) by 2041, 2014.181, 2014.002, 2014.000, \dots . \diamond

Exercise 2.24. Prove that the sequence $\left(\frac{n!}{n^n}\right)_{n \in \mathbb{N}}$ is convergent and that $\lim_{n \rightarrow \infty} \frac{n!}{n^n} = 0$.

Hint: Observe that $0 \leq \frac{n!}{n^n} = \frac{1}{n} \cdot \frac{2}{n} \cdot \dots \cdot \frac{n}{n} \leq \frac{1}{n} \cdot 1 \cdot \dots \cdot 1 \leq \frac{1}{n}$.

Exercise 2.25. Prove that for all $k \in \mathbb{N}$, $\lim_{n \rightarrow \infty} \frac{1^k + 2^k + 3^k + \dots + n^k}{n^{k+2}} = 0$.

Exercise 2.26 ($\lim_{n \rightarrow \infty} n^{\frac{1}{n}} = 1$).

(1) Using induction, prove that if $x \geq -1$ and $n \in \mathbb{N}$, then $(1+x)^n \geq 1+nx$.

(2) Show that for all $n \in \mathbb{N}$, $1 \leq n^{\frac{1}{n}} < (1+\sqrt{n})^{\frac{2}{n}} \leq (1+\frac{1}{\sqrt{n}})^2$.

Hint: Take $x = \frac{1}{\sqrt{n}}$ in the inequality above.

(3) Prove that $(n^{\frac{1}{n}})_{n \in \mathbb{N}}$ is convergent and find its limit.

Exercise 2.27. Suppose $(a_n)_{n \in \mathbb{N}}$ is contained in the interval (a, b) (i.e., $\forall n \in \mathbb{N}$, $a < a_n < b$). If $(a_n)_{n \in \mathbb{N}}$ is convergent with limit L , then show that $L \in [a, b]$. *Hint:* Exercise 2.4. Give an example to show that L need not belong to (a, b) .

Exercise 2.28. Let $(a_n)_{n \in \mathbb{N}}$ be a convergent sequence, and let $(b_n)_{n \in \mathbb{N}}$ satisfy $|b_n - a_n| < \frac{1}{n}$ for all $n \in \mathbb{N}$. Show that $(b_n)_{n \in \mathbb{N}}$ is also convergent. What is its limit?

Hint: Observe that $-\frac{1}{n} + a_n < b_n < a_n + \frac{1}{n}$ for all $n \in \mathbb{N}$.

Exercise 2.29. (*) See Exercises 2.11 and 2.18. Prove that a bounded sequence $(a_n)_{n \in \mathbb{N}}$ is convergent if and only if

$$\liminf_{n \rightarrow \infty} a_n = \limsup_{n \rightarrow \infty} a_n.$$

Moreover, then $\lim_{n \rightarrow \infty} a_n = \liminf_{n \rightarrow \infty} a_n = \limsup_{n \rightarrow \infty} a_n$.

2.5. Subsequences

In this section we prove an important result in Analysis, called the Bolzano-Weierstrass Theorem, which says that

Every bounded sequence has a convergent subsequence.

We begin this section by defining what we mean by a ‘subsequence’ of a sequence.

Definition 2.5 (Subsequence of a sequence).

Let $(a_n)_{n \in \mathbb{N}}$ be a sequence and let $n_1 < n_2 < n_3 < \dots$ be a strictly increasing sequence of natural numbers. Then $(a_{n_k})_{k \in \mathbb{N}}$ is called a *subsequence* of $(a_n)_{n \in \mathbb{N}}$. Thus the terms of the subsequence are $a_{n_1}, a_{n_2}, a_{n_3}, \dots$.

Example 2.14. For example, the sequence $(a_{n^2})_{n \in \mathbb{N}} = (\frac{1}{n^2})_{n \in \mathbb{N}}$

$$1, \frac{1}{4}, \frac{1}{9}, \frac{1}{16}, \frac{1}{25}, \dots$$

is a subsequence of the sequence $(a_n)_{n \in \mathbb{N}} = (\frac{1}{n})_{n \in \mathbb{N}}$. However, the sequence

$$\frac{1}{9}, \frac{1}{4}, \frac{1}{16}, \frac{1}{25}, \dots$$

is *not* a subsequence of $(\frac{1}{n^2})_{n \in \mathbb{N}}$, since terms of subsequence are not in the same order as the original sequence:

$$a_3 = \frac{1}{9}, \quad a_2 = \frac{1}{4},$$

and $3 > 2$. But

$$\frac{1}{9}, \frac{1}{4}, \frac{1}{16}, \frac{1}{25}, \dots$$

is a subsequence of

$$1, \frac{1}{4}, \frac{1}{9}, \frac{1}{4}, \frac{1}{25}, \dots$$

The sequences

$$\begin{aligned} ((-1)^{2n})_{n \in \mathbb{N}} & \quad (\text{that is, the constant sequence } 1, 1, 1, \dots) \text{ and} \\ ((-1)^{2n-1})_{n \in \mathbb{N}} & \quad (\text{that is the constant sequence } -1, -1, -1, \dots) \end{aligned}$$

are both subsequences of $((-1)^n)_{n \in \mathbb{N}}$. Here are some more examples:

$n_1 < n_2 < n_3 < \dots$	Subsequence of $(a_n)_{n \in \mathbb{N}}$	Subsequence of $(\frac{1}{n})_{n \in \mathbb{N}}$
$1 < 2 < 3 < \dots$	$(a_n)_{n \in \mathbb{N}} \quad a_1, a_2, a_3, \dots$	$1, \frac{1}{2}, \frac{1}{3}, \dots$
$2 < 3 < 4 < \dots$	$(a_{n+1})_{n \in \mathbb{N}} \quad a_2, a_3, a_4, \dots$	$\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$
$2 < 4 < 6 < 8 < \dots$	$(a_{2n})_{n \in \mathbb{N}} \quad a_2, a_4, a_6, a_8, \dots$	$\frac{1}{2}, \frac{1}{4}, \frac{1}{6}, \frac{1}{8}, \dots$
$2 < 4 < 8 < 16 < \dots$	$(a_{2^n})_{n \in \mathbb{N}} \quad a_2, a_4, a_8, a_{16}, \dots$	$\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \dots$
$1 < 4 < 27 < 64 < \dots$	$(a_{n^n})_{n \in \mathbb{N}} \quad a_1, a_4, a_{27}, a_{64}, \dots$	$1, \frac{1}{4}, \frac{1}{27}, \frac{1}{64}, \dots$
$2 < 3 < 5 < 7 < \dots$	$(a_{p_n})_{n \in \mathbb{N}} \quad a_2, a_3, a_5, a_7, \dots$ (p_n denotes the n th prime)	$\frac{1}{2}, \frac{1}{3}, \frac{1}{5}, \frac{1}{7}, \dots$
$1 < 2 < 6 < 24 < \dots$	$(a_{n!})_{n \in \mathbb{N}} \quad a_1, a_2, a_6, a_{24}, \dots$	$1, \frac{1}{2}, \frac{1}{6}, \frac{1}{24}, \dots$

◇

Exercise 2.30. Is $(\frac{1}{n^4})$ a subsequence of $(\frac{1}{n^2})_{n \in \mathbb{N}}$? Is $(\frac{1}{n^3})$ a subsequence of $(\frac{1}{n^2})_{n \in \mathbb{N}}$?

Exercise 2.31. (*) Beginning with 2 and 7, the sequence 2, 7, 1, 4, 7, 4, 2, 8, 2, 8, ... is constructed by multiplying successive pairs of its terms and adjoining the result as the next one or two members of the sequence depending on whether the product is a one- or two-digit number. Thus we start with 2 and 7, giving the product 14, and so the next two terms are 1, 4. Proceeding in this manner, we get subsequent terms as follows:

$\frac{2, 7}{2, 7, 1, 4}$
 $\frac{2, 7, 1, 4}{2, 7, 1, 4, 7}$
 $\frac{2, 7, 1, 4, 7}{2, 7, 1, 4, 7, 4}$
 $\frac{2, 7, 1, 4, 7, 4}{2, 7, 1, 4, 7, 4, 2, 8}$
 $\frac{2, 7, 1, 4, 7, 4, 2, 8}{2, 7, 1, 4, 7, 4, 2, 8, 2, 8}$
 ...

Prove that this sequence has the constant subsequence 6, 6, 6, ...

Hint: Show that 6 appears an infinite number of times as follows. Since the terms 2, 8, 2, 8 are adjacent, they give rise to the adjacent terms 1, 6, 1, 6 at some point, which in turn give rise to the adjacent terms 6, 6, 6 eventually, and so on. Proceeding in this way, find out if you get a loop containing the term 6.

If $(n_k)_{k \in \mathbb{N}}$ is a strictly increasing sequence in \mathbb{N} , then $n_k \geq k$. (This follows by induction on k : $n_1 \geq 1$, and if $n_k \geq k$, then $n_{k+1} > n_k \geq k$ gives $n_{k+1} \geq k+1$.)

Theorem 2.6.

Any subsequence of a convergent sequence is convergent with the same limit.

Proof. Let $(a_{n_k})_{k \in \mathbb{N}}$ be a subsequence of a convergent sequence $(a_n)_{n \in \mathbb{N}}$ with limit L . Given $\epsilon > 0$, let $N \in \mathbb{N}$ be such that for all $n > N$, $|a_n - L| < \epsilon$. Since the sequence $n_1 < n_2 < n_3 < \dots$ of natural numbers is increasing, it follows that $n_N \geq N$. Then for all $k > N$, $n_k > n_N \geq N$. Hence for $k > N$, $|a_{n_k} - L| < \epsilon$, and so $(a_{n_k})_{k \in \mathbb{N}}$ is convergent with limit L . \square

Example 2.15. From Example 2.14 and the fact that $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$, it follows that

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} = \lim_{n \rightarrow \infty} \frac{1}{2n} = \lim_{n \rightarrow \infty} \frac{1}{n^2} = \lim_{n \rightarrow \infty} \frac{1}{2^n} = \lim_{n \rightarrow \infty} \frac{1}{n^n} = \lim_{n \rightarrow \infty} \frac{1}{p_n} = \lim_{n \rightarrow \infty} \frac{1}{n!} = 0.$$

In the above p_n denotes the n th prime number. \diamond

Example 2.16. Let us give a proof of the fact that $((-1)^n)_{n \in \mathbb{N}}$ is divergent based on Theorem 2.6. Suppose on the contrary, that $((-1)^n)_{n \in \mathbb{N}}$ is convergent with limit L . Then the terms with odd indices give the subsequence $-1, -1, -1, \dots$, which is convergent with limit -1 , and so (by uniqueness of limits!) $L = -1$. On the other hand, the terms with even indices give the subsequence $1, 1, 1, \dots$, which is convergent with limit 1 , and so $L = 1$. So we have arrived at the contradiction that $-1 = L = 1$. Hence $((-1)^n)_{n \in \mathbb{N}}$ is divergent. \diamond

Example 2.17 ('The harmonic series diverges.'). Consider $(s_n)_{n \in \mathbb{N}}$, where

$$s_n := 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}, \quad n \in \mathbb{N}.$$

Suppose that $(s_n)_{n \in \mathbb{N}}$ is convergent with limit L . Then its subsequence $(s_{2n})_{n \in \mathbb{N}}$ would also be convergent with limit L , and so by the Algebra of Limits, the sequence $(s_{2n} - s_n)_{n \in \mathbb{N}}$ must converge to $L - L = 0$. But

$$\begin{aligned} s_{2n} - s_n &= \cancel{1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}} + \frac{1}{n+1} + \cdots + \frac{1}{2n} - \left(\cancel{1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}} \right) \\ &= \frac{1}{n+1} + \cdots + \frac{1}{2n} > \underbrace{\frac{1}{2n} + \cdots + \frac{1}{2n}}_{n \text{ times}} = n \cdot \frac{1}{2n} = \frac{1}{2}. \end{aligned}$$

Hence $|(s_{2n} - s_n) - 0| = s_{2n} - s_n > \frac{1}{2}$, showing that it is *not* the case that

$$\lim_{n \rightarrow \infty} (s_{2n} - s_n) = 0,$$

a contradiction. So $\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right)_{n \in \mathbb{N}}$ diverges. \diamond

Exercise 2.32. Recall the convergent sequence $(a_n)_{n \in \mathbb{N}}$ from Exercise 2.10 on page 47:

$$a_1 = 1 \text{ and } a_n = \frac{2n+1}{3n} a_{n-1} \text{ for } n \geq 2.$$

What is its limit?

Exercise 2.33. Determine if the following statements are true or false.

- (1) Every subsequence of a convergent real sequence is convergent.
- (2) Every subsequence of a divergent real sequence is divergent.
- (3) Every subsequence of a bounded real sequence is bounded.
- (4) Every subsequence of an unbounded real sequence is unbounded.
- (5) Every subsequence of a monotone real sequence is monotone.
- (6) Every subsequence of a nonmonotone real sequence is nonmonotone.
- (7) If every subsequence of a real sequence converges, the sequence itself converges.
- (8) If $(a_{2n})_{n \in \mathbb{N}}$ and $(a_{2n+1})_{n \in \mathbb{N}}$ both converge, then $(a_n)_{n \in \mathbb{N}}$ converges.
- (9) If $(a_{2n})_{n \in \mathbb{N}}$ and $(a_{2n+1})_{n \in \mathbb{N}}$ both converge to the same limit, then $(a_n)_{n \in \mathbb{N}}$ converges.

Exercise 2.34. (*) Show that if $(a_n)_{n \in \mathbb{N}}$ is a sequence that does not converge to L , then there exists an $\epsilon > 0$ and there exists a subsequence $(a_{n_k})_{k \in \mathbb{N}}$ of $(a_n)_{n \in \mathbb{N}}$ such that for all $k \in \mathbb{N}$, $|a_{n_k} - L| \geq \epsilon$.

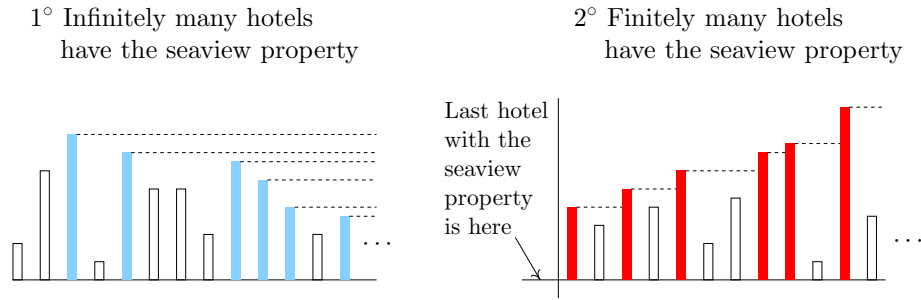
Exercise 2.35. Let $(a_n)_{n \in \mathbb{N}}$ be given by $a_1 = \sqrt{2}$ and $a_{n+1} = \sqrt{2 + a_n}$ for all $n \in \mathbb{N}$.

(The first few terms are $\sqrt{2}$, $\sqrt{2 + \sqrt{2}}$, $\sqrt{2 + \sqrt{2 + \sqrt{2}}}$, \cdots .)

- (a) Show that for all $n \in \mathbb{N}$, $a_n \leq 2$. *Hint:* Use induction on n .
- (b) Show that $(a_n)_{n \in \mathbb{N}}$ is increasing. *Hint:* Consider $a_{n+1}^2 - a_n^2$.
- (c) Is $(a_n)_{n \in \mathbb{N}}$ convergent? If so, find its limit.

Theorem 2.7. *Every sequence has a monotone subsequence.*

Before giving the formal proof, we give an illustration of the idea behind this proof⁵. If $(a_n)_{n \in \mathbb{N}}$ is the given sequence, then imagine that there is an infinite chain of hotels along a line, where the n th hotel has height a_n , and at the horizon, there is a sea. A hotel is said to have the *seaview property* if it is higher than all hotels following it (so that from the roof of the hotel, one can view the sea). Now there are only two possibilities, as illustrated below.



1° There are infinitely many hotels with the seaview property.	2° There are finitely many hotels with the seaview property.
Then by taking successively the heights of the hotels the seaview property we get a <i>decreasing</i> subsequence.	Then after the last hotel with the seaview property, one can start with any hotel and then always find one that is at least as high, which is taken as the next hotel, and then finding yet another that is at least as high as that one, and so on. The heights of these hotels form an <i>increasing</i> subsequence.

Proof. Let $(a_n)_{n \in \mathbb{N}}$ be a sequence, and let

$$S = \{m \in \mathbb{N} : \text{for all } n > m, a_n < a_m\}.$$

(This is the collection of indices of hotels with the seaview property.)

Then we have the following two cases.

1° S is infinite.

Arrange the elements of S in increasing order: $n_1 < n_2 < n_3 < \dots$.

Then $(a_{n_k})_{k \in \mathbb{N}}$ is a *decreasing* subsequence of $(a_n)_{n \in \mathbb{N}}$.

⁵This illustrative analogy stems from [B]. The proof seems to go back to [?]. See also [?].

2° S is finite.

We will define inductively an increasing subsequence $(a_{n_k})_{k \in \mathbb{N}}$ of $(a_n)_{n \in \mathbb{N}}$.

If S empty, then define $n_1 = 1$, and otherwise let $n_1 = \max S + 1$.

Suppose that for some $k \in \mathbb{N}$, $n_1 < \dots < n_k$ have been constructed such that $a_{n_1} \leq \dots \leq a_{n_k}$. Define $n_{k+1} = \min\{m \in \mathbb{N} : m > n_k \text{ and } a_m \geq a_{n_k}\}$.

(Thus n_{k+1} is the first hotel blocking the view from the top of the n_k^{th} hotel.)

The minimum exists as $S' := \{m \in \mathbb{N} : m > n_k \text{ and } a_m \geq a_{n_k}\}$ is a nonempty subset of \mathbb{N} . (Indeed, otherwise if $S' = \emptyset$, then for all $m > n_k$, we have $a_n < a_{n_k}$, so that $n_k \in S$. But this is impossible if S was empty, and also impossible if S was not empty, since we know $n_k \geq n_1 > \max S$.) By the definition of S' , we have $a_{n_{k+1}} \geq a_{n_k}$. Then $(a_{n_k})_{k \in \mathbb{N}}$ is an *increasing* subsequence of $(a_n)_{n \in \mathbb{N}}$.

Thus every sequence $(a_n)_{n \in \mathbb{N}}$ has a monotone subsequence. \square

An important consequence of the above theorem is the following result.

Theorem 2.8 (Bolzano-Weierstrass Theorem).

Every bounded sequence has a convergent subsequence.

Proof. Let $(a_n)_{n \in \mathbb{N}}$ be a bounded sequence. Then there exists an $M > 0$ such that for all $n \in \mathbb{N}$, $|a_n| \leq M$. From Theorem 2.7 above, it follows that the sequence $(a_n)_{n \in \mathbb{N}}$ has a monotone subsequence, say $(a_{n_k})_{k \in \mathbb{N}}$. Then clearly for all $k \in \mathbb{N}$, $|a_{n_k}| \leq M$ and so the sequence $(a_{n_k})_{k \in \mathbb{N}}$ is also bounded. Since $(a_{n_k})_{k \in \mathbb{N}}$ is monotone and bounded, it follows from Theorem 2.3 that it is convergent. \square

Example 2.18 ('Compactness' of $[a, b]$).

Consider any sequence $(a_n)_{n \in \mathbb{N}}$ in $[a, b]$, i.e., for all $n \in \mathbb{N}$, $a_n \in [a, b]$, or equivalently, $a \leq a_n \leq b$. Then $(a_n)_{n \in \mathbb{N}}$ is bounded, and so it has a convergent subsequence, say $(a_{n_k})_{k \in \mathbb{N}}$. Then for all $k \in \mathbb{N}$, $a \leq a_{n_k} \leq b$. By Exercise 2.18, $a \leq \lim_{k \rightarrow \infty} a_{n_k} \leq b$. So:

Every sequence in $[a, b]$ has a convergent subsequence,
and the limit of this subsequence belongs to $[a, b]$. \diamond

Example 2.19. Let $(a_n)_{n \in \mathbb{N}}$ be the sequence of fractional parts of integral multiples of $\sqrt{2}$, i.e., $a_n := \{n\sqrt{2}\} := n\sqrt{2} - \lfloor n\sqrt{2} \rfloor$, for $n \in \mathbb{N}$. The first few terms are:

$$\begin{array}{ll} \sqrt{2} = 1.414213 \dots & a_1 = 0.414213 \dots \\ 2\sqrt{2} = 2.828427 \dots & a_2 = 0.828427 \dots \\ 3\sqrt{2} = 4.242640 \dots & a_3 = 0.242640 \dots \\ 4\sqrt{2} = 5.656854 \dots & a_4 = 0.656854 \dots \\ 5\sqrt{2} = 7.071067 \dots & a_5 = 0.071067 \dots \end{array}$$

The sequence $(a_n)_{n \in \mathbb{N}}$ is bounded: Indeed, $0 \leq a_n < 1$. By the Bolzano-Weierstrass Theorem it has a convergent subsequence⁶. \diamond

⁶In fact, it can be shown that these fractional parts a_n are dense in $(0, 1)$. Thus given any number $L \in (0, 1)$, there exists a subsequence of the sequence $(a_n)_{n \in \mathbb{N}}$ above that converges to L . See [N].

Example 2.20 ('Nested interval property of \mathbb{R} '). We will show the following:

Let $I_m := \{x \in \mathbb{R} : a_m \leq x \leq b_m\}$, $m \in \mathbb{N}$, be such that

$$I_1 \supset I_2 \supset I_3 \supset \cdots.$$

Then

$$\bigcap_{m \in \mathbb{N}} I_m \neq \emptyset.$$

For $n \in \mathbb{N}$, $a_n \in I_n \subset I_1$, and so $a_1 \leq a_n \leq b_1$. Thus $(a_n)_{n \in \mathbb{N}}$ is bounded, and by the Bolzano-Weierstrass Theorem, possesses a convergent subsequence, say $(a_{n_k})_{k \in \mathbb{N}}$. Let a denote the limit of $(a_{n_k})_{k \in \mathbb{N}}$.

Claim A. For all $m \in \mathbb{N}$, $a_m \leq a$.

Suppose there exists an $m \in \mathbb{N}$ such that $a_m > a$, that is, $\epsilon := a_m - a > 0$. As $(a_{n_k})_{k \in \mathbb{N}}$ is convergent with limit a , there exists a $K \in \mathbb{N}$ such that for all $k > K$, $|a_{n_k} - a| < \epsilon = a_m - a$. Then for $k > K$,

$$a_{n_k} - a \leq |a_{n_k} - a| < a_m - a,$$

and so

$$a_{n_k} < a_m. \quad (\star)$$

For $k > m$, $n_k \geq k > m$, and so $I_{n_k} \subset I_m$. Thus for $k > m$,

$$a_{n_k} \geq a_m. \quad (\star\star)$$

Now if $k > \max\{m, K\}$, then both (\star) and $(\star\star)$ hold, which is impossible. This proves Claim A.

Claim B. For all $m \in \mathbb{N}$, $a \leq b_m$.

For $k \geq m$, $n_k \geq k \geq m$, and so $a_{n_k} \in I_{n_k} \subset I_m$. Thus for all $k \geq m$, $a_{n_k} \leq b_m$. Passing to the limit as $k \rightarrow \infty$, we obtain $a \leq b_m$, proving Claim B.

From Claims A and B, for all $m \in \mathbb{N}$, $a_m \leq a \leq b_m$, that is, $a \in I_m$. Consequently,

$$a \in \bigcap_{m \in \mathbb{N}} I_m,$$

and so $\bigcap_{m \in \mathbb{N}} I_m \neq \emptyset$. ◇

Exercise 2.36. Does the sequence $(\sin n)_{n \in \mathbb{N}}$ have a convergent subsequence? What about the sequence $(n)_{n \in \mathbb{N}}$?

Exercise 2.37. (*) Consider the bounded divergent sequence $((-1)^n)_{n \in \mathbb{N}}$. Note that there exist two subsequences $(-1, -1, -1, \dots)$ and $(1, 1, 1, \dots)$ which have distinct limits $(-1 \neq 1)$. In this exercise we show that this is a general phenomenon. Show that if $(a_n)_{n \in \mathbb{N}}$ is bounded and divergent, then it has two subsequences which converge to distinct limits. *Hint:* Use the Bolzano-Weierstrass theorem twice, and also Exercise 2.34.

Appendix (*)

Let us show that every nonnegative rational number has either a terminating or a repeating decimal expansion, as was mentioned in Exercise 2.22. This section is not examinable, and may be skipped.

Let

$$x = \frac{p}{q}$$

where p is a nonnegative integer and $q \in \mathbb{N}$. We can factorize $q = 2^i 5^j q'$, where i, j are nonnegative integers, and $2, 5$ do not divide $q' \in \mathbb{N}$. Choose a natural number $n > i, j$. Then $10^n / (2^i 5^j)$ is a natural number, and so

$$10^n x = 10^n \frac{p}{q} = \frac{p'}{q'},$$

where $p' = 2^{n-i} 5^{n-j} p$ is a nonnegative integer, and q' is coprime to 10. But if we look at the remainders we get when we divide $10, 10^2, 10^3, \dots$ by q' , then for some integers $K > k \geq 1$, $10^K, 10^k$ leave the same remainder when divided by q' . So

$$10^K - 10^k = 10^k(10^{K-k} - 1)$$

is divisible by q' . But q' is coprime to 10, and hence also to 10^k . Thus q' must divide $10^{K-k} - 1$. So we have with $m := K - k$ that

$$10^n(10^m - 1)x = (10^{K-k} - 1)\frac{p'}{q'}$$

is an integer. Now let $x = N.d_1 d_2 d_3 \dots$. Then⁷

$$\begin{aligned} 10^{n+m}x - 10^n x &= Nd_1 \dots d_{n+m} - Nd_1 \dots d_n \\ &+ \left(\frac{d_{n+m+1}}{10} + \frac{d_{n+m+2}}{10^2} + \frac{d_{n+m+3}}{10^3} + \dots - \frac{d_{n+1}}{10} - \frac{d_{n+2}}{10^2} - \frac{d_{n+3}}{10^3} - \dots \right), \end{aligned}$$

is an integer. This implies that the part in the brackets above, henceforth denoted by Δ , is also an integer:

$$\Delta = \frac{d_{n+m+1} - d_{n+1}}{10} + \frac{d_{n+m+2} - d_{n+2}}{10^2} + \frac{d_{n+m+3} - d_{n+3}}{10^3} + \dots \in \mathbb{Z}.$$

We claim that this gives $d_{n+m+k} = d_{n+k}$ for all $k \in \mathbb{N}$, giving the desired conclusion.

⁷Here by $Nd_1 \dots d_n$ we mean the number which in decimal notation has the digits of N followed by the digits d_1, \dots, d_n . The number $Nd_1 \dots d_{n+m}$ has a similar connotation.

1° It cannot be the case that the sequence $(d_{n+m+k} - d_k)_{n \in \mathbb{N}}$ is eventually the constant sequence $9, 9, 9, \dots$. Indeed, then there exists some K such that $d_{n+m+k} - d_{n+k} = 9$ for all $k > K$. This means that $d_{n+m+k} = 9$ and $d_{n+k} = 0$ for all $k > K$, which is impossible.

2° Similar to 1°, it can't be the case that the sequence $(d_{n+m+k} - d_k)_{n \in \mathbb{N}}$ is eventually the constant sequence $-9, -9, -9, \dots$ either.

3° Suppose that $k_* \in \mathbb{N}$ is the smallest number such that $d_{n+m+k_*} \neq d_{n+k_*}$.

If $d_{n+m+k_*} > d_{n+k_*}$, then by 2°, there exists a number $K \in \mathbb{N}$ such that $d_{n+m+k_*+K} - d_{n+k_*+K} \neq -9$, and so

$$\Delta \geq \frac{1}{10^{k_*}} - \frac{9}{10^{k_*+1}} - \dots - \frac{9}{10^{k_*+K-1}} - \frac{8}{10^{k_*+K}} - \frac{9}{10^{k_*+K+1}} \dots = \frac{1}{10^{k_*+K}}.$$

We also have in light of case 1° that

$$1 \geq \frac{1}{10^{k_*-1}} = \frac{9}{10^{k_*}} + \frac{9}{10^{k_*+1}} + \frac{9}{10^{k_*+2}} + \dots > \Delta,$$

a contradiction to the fact that Δ is an integer.

Now we consider the other possibility, namely, $d_{n+m+k_*} < d_{n+k_*}$, then by 1°, there exists a K such that $d_{n+m+k_*+K} - d_{n+k_*+K} \neq 9$, and so

$$\Delta \leq \frac{-1}{10^{k_*}} + \frac{9}{10^{k_*+1}} + \dots + \frac{9}{10^{k_*+K-1}} + \frac{8}{10^{k_*+K}} + \frac{9}{10^{k_*+K+1}} \dots = -\frac{1}{10^{k_*+K}}.$$

We also have in light of case 2° that

$$-1 \leq -\frac{1}{10^{k_*-1}} = -\frac{9}{10^{k_*}} - \frac{9}{10^{k_*+1}} - \frac{9}{10^{k_*+2}} + \dots < \Delta.$$

Hence, again, we have a contradiction to the fact that Δ is an integer.

Consequently $d_{n+m+k} = d_{n+k}$ for all $k \in \mathbb{N}$, which means that the block of digits $d_{n+1} \dots d_{n+m}$ keeps repeating.

Chapter 3

Continuity

Let I be an interval in \mathbb{R} . So I is a set of the form (a, b) or $[a, b]$ or $(-\infty, b)$, etc. Among all possible functions $f : I \rightarrow \mathbb{R}$, there is a ‘nice’ class of functions, namely ones which are *continuous on I* .

What’s so nice about continuous functions? Continuous functions have properties that make them easy to work with in Analysis. For example, we will see that continuous functions possess two important properties, given by

- the Intermediate Value Theorem, and
- the Extreme Value Theorem.

We will learn the statements and proofs of these in the course of this chapter. Functions which aren’t continuous may fail to possess these properties.

Many bizarre functions make appearances in Analysis, and in order to avoid falling into pitfalls with simplistic thinking, we need definitions and assumptions of theorems to be stated carefully and clearly.

3.1. Definition of continuity

In everyday speech, a ‘continuous’ process is one that proceeds without gaps of interruptions or sudden changes.

What does it mean for a function $f : \mathbb{R} \rightarrow \mathbb{R}$ to be continuous? Roughly, f is said to be continuous on I if f has ‘no breaks’ at any point of I . If a break does occur in f , then this break will occur at some point of I . So we realize that in order to define continuity, we need to define what is meant by the notion of ‘ f being continuous at a point $c \in I$ ’.

So (based on this visual view of continuity), we first give the formal definition of the continuity of a function *at a point*. Next, if a function is continuous at *each* point, then it is called continuous. If a function has a break at a point c , then even if points x are close to c , the points $f(x)$ do not get close to $f(c)$. See Figure 1.

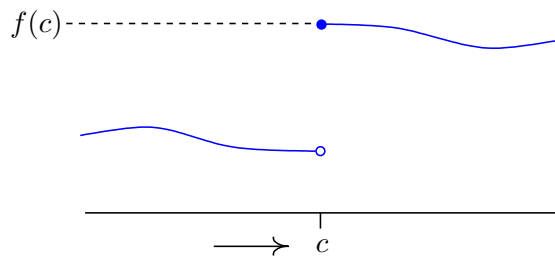


Figure 1. A function with a break at c . If x lies to the left of c , then $f(x)$ is not close to $f(c)$, no matter how close x comes to c .

So ‘no break in f at c ’ should mean that $f(x)$ stays close to $f(c)$ whenever x is close to c . This motivates the following definition of continuity, which guarantees that if a function is continuous at a point c , then we can make $f(x)$ as close as we like to $f(c)$, by choosing x sufficiently close to c . See Figure 2.

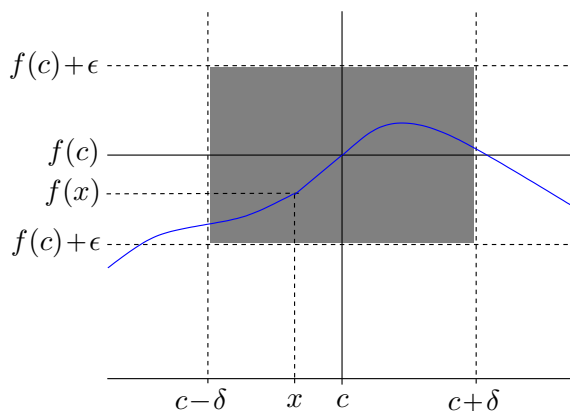


Figure 2. The definition of continuity of a function at point c . If the function is continuous at c , then given any $\epsilon > 0$ (which determines a strip around the line $y = f(c)$ of width 2ϵ), there exists a $\delta > 0$ (which determines an interval $(c - \delta, c + \delta)$ of width 2δ around the point c) such that whenever x lies in this interval (so that x satisfies $c - \delta < x < c + \delta$, that is, $|x - c| < \delta$), then $f(x)$ satisfies $f(c) - \epsilon < f(x) < f(c) + \epsilon$, that is, $|f(x) - f(c)| < \epsilon$.

Definition 3.1 (Continuity at a point; Continuous function).

Let I be an interval in \mathbb{R} , $c \in I$ and $f : I \rightarrow \mathbb{R}$.

The function f is *continuous at c* if for every $\epsilon > 0$, there exists a $\delta > 0$ such that for all $x \in I$ satisfying $|x - c| < \delta$, $|f(x) - f(c)| < \epsilon$.

The function f is *continuous (on I)* if for every $x \in I$, f is continuous at x .

Remark 3.1.

- (1) **Continuity is a ‘local’ concept.** That is, we can decide the continuity of f on an interval by looking at each point of the domain f and checking if f is continuous at that point, and moreover, what matters for continuity of f at a point, roughly speaking, is what the function is doing ‘locally’ in arbitrarily small neighbourhoods of the point, that is, ‘near the point’, and what happens away from the point is irrelevant.
- (2) **History of the notion of continuity.** In the early development of Analysis, there was no rigorous definition of continuity offered. Only in the 18th century mathematicians started examining this notion, in connection with Fourier’s work on the theory of heat, where discontinuous functions arose naturally in various kinds of physical problems. A satisfactory mathematical definition of continuity was first formulated by Cauchy in 1821. *

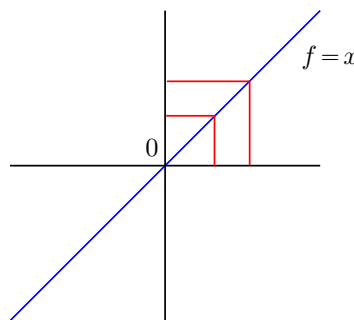
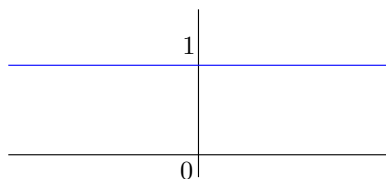
Example 3.1 (The constant function).

Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be given by $f(x) = 1$ for all $x \in \mathbb{R}$. Then f is continuous.

Let $c \in \mathbb{R} = (-\infty, \infty)$. Let $\epsilon > 0$. For $x \in \mathbb{R}$, $|f(x) - f(c)| = |1 - 1| = 0 < \epsilon$. So any $\delta > 0$ will do! For example, set $\delta = 1$. Then if $x \in \mathbb{R}$ and $|x - c| < \delta = 1$, we have:

$$|f(x) - f(c)| = |1 - 1| = |0| = 0 < \epsilon.$$

So f is continuous at c . Since the choice of $c \in \mathbb{R}$ was arbitrary, it follows that f is continuous on \mathbb{R} . See the picture on the left below. ◇

**Example 3.2** (The identity function).

$f : \mathbb{R} \rightarrow \mathbb{R}$ given by $f(x) = x$ for all $x \in \mathbb{R}$ is continuous.

Let $c \in \mathbb{R}$. Let $\epsilon > 0$.

(*Rough work:* $|f(x) - f(c)| = |x - c| < \delta \leq \epsilon$, if for example $\delta := \epsilon$.)

Let $\delta = \epsilon$. Then if $x \in \mathbb{R}$ and $|x - c| < \delta$, we have:

$$|f(x) - f(c)| = |x - c| < \delta = \epsilon.$$

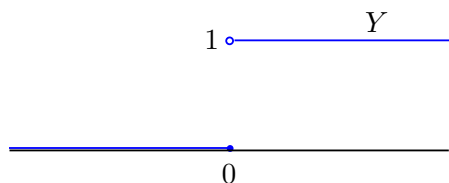
So f is continuous at c . Since the choice of $c \in \mathbb{R}$ was arbitrary, it follows that f is continuous on \mathbb{R} . See the picture on the right above. ◇

Example 3.3 (The Heaviside¹ function).

Let $Y : \mathbb{R} \rightarrow \mathbb{R}$ be given by

$$Y(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

From the graph of Y displayed below, we see clearly that there is a ‘break’ or ‘jump’ at $x = 0$, and so we guess that Y is not continuous at 0. Let us show this using the definition of continuity at a point.



Suppose that Y is continuous at 0. Let $\epsilon = \frac{1}{2} > 0$. Suppose that there exists a $\delta > 0$ such that whenever $|x - 0| < \delta$, we have $|Y(x) - Y(0)| = |Y(x) - 0| < \epsilon = \frac{1}{2}$. Take $x = \frac{\delta}{2}$. Then $|x - 0| = |\frac{\delta}{2} - 0| = \frac{\delta}{2} < \delta$, and so we must have

$$|Y(x) - Y(0)| = \left| Y\left(\frac{\delta}{2}\right) - 0 \right| = |1 - 0| = 1 < \epsilon = \frac{1}{2},$$

a contradiction. So Y is not continuous at 0. ◇

Example 3.4 (The reciprocal function).

Let $h : (0, \infty) \rightarrow \mathbb{R}$ be the function given by $h(x) = \frac{1}{x}$ for all $x \in (0, \infty)$. Then h is continuous (on $(0, \infty)$). See Figure 3.

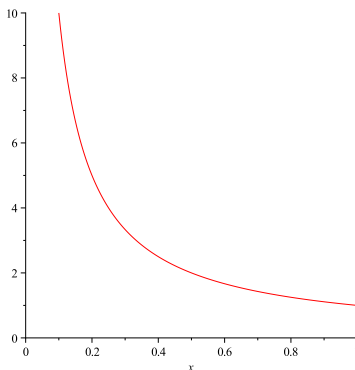


Figure 3. $x \mapsto \frac{1}{x} : (0, \infty) \rightarrow \mathbb{R}$ is continuous on $(0, \infty)$.

¹Named after the mathematical physicist Oliver Heaviside (1850-1925).

Let $c \in (0, \infty)$. Let $\epsilon > 0$.

(*Rough work:* We want $\delta > 0$ such that for $|x - c| < \delta$, $|h(x) - h(c)| < \epsilon$. We have

$$|h(x) - h(c)| = \left| \frac{1}{x} - \frac{1}{c} \right| = \frac{|x - c|}{|x||c|}.$$

We know that if x is close to c , then the numerator $|x - c|$ can be made small. But what about the denominator $|x||c|$. Well, $|c| > 0$ is just a constant, and so it is harmless really. What about $|x|$? If it gets small, then it has the effect of making $|h(x) - h(c)|$ big, something which we want to avoid. But we note that when x is close to c , $|x|$ will be close to $|c|$, and so $|x|$ can be bounded below by some positive constant. Indeed, by the triangle inequality,

$$|c| - |x| \leq ||c| - |x|| \leq |c - x| = |x - c|,$$

and so if we choose the $\delta \leq |c|/2$, then for x satisfying $|x - c| < \delta$ we will obtain from the above that $|x| \geq |c| - |c - x| \geq |c| - \delta \geq |c| - |c|/2 = |c|/2$. So for such x ,

$$|h(x) - h(c)| = \frac{|x - c|}{|x||c|} < \frac{\delta}{(|c|/2) \cdot |c|},$$

and the last quantity can be made smaller than ϵ by further ensuring that the δ also satisfies that $\delta < \epsilon \frac{|c|^2}{2}$. Hence $\delta := \min\{\frac{|c|}{2}, \epsilon \frac{|c|^2}{2}\}$ should do the job! We remark that this is just one choice among many other equally good δ s which will also work. End of *Rough Work*.)

Set $\delta = \min\left\{\frac{c}{2}, \frac{\epsilon c^2}{2}\right\} (> 0)$. Then if $x \in (0, \infty)$ and $|x - c| < \delta$, we have

$$|c| - |x| \leq |x - c| < \delta \leq \frac{|c|}{2}$$

and so $\frac{|c|}{2} < |x|$, that is, $\frac{1}{|x|} < \frac{2}{|c|}$. Thus if $x \in (0, \infty)$ and $|x - c| < \delta$, then

$$\left| \frac{1}{x} - \frac{1}{c} \right| = \frac{|c - x|}{|x||c|} = \frac{|x - c|}{|x||c|} < \delta \cdot \frac{2}{|c|} \cdot \frac{1}{|c|} = \frac{2\delta}{c^2} \leq \epsilon.$$

So f is continuous at c . As $c \in (0, \infty)$ was arbitrary, f is continuous on $(0, \infty)$. \diamond

Exercise 3.1. Let the function $f : \mathbb{R} \rightarrow \mathbb{R}$ be given by $f(x) = x^2$.

(1) Prove that f is continuous at 0.

(2)(*) Suppose that c is a nonzero real number. Prove that f is continuous at c .

In Exercise 3.7, we will give a slick proof of the fact that f is continuous on \mathbb{R} .

Exercise 3.2. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ satisfy $f(x + y) = f(x) + f(y)$ for all $x, y \in \mathbb{R}$.

(1) Suppose that f is continuous at some real number c . Prove that f is continuous on \mathbb{R} .

Hint: Since f is continuous at c , given $\epsilon > 0$, $\exists \delta > 0$ such that for all $x \in \mathbb{R}$ satisfying $|x - c| < \delta$, $|f(x) - f(c)| < \epsilon$. Show that given any other point $c' \in \mathbb{R}$, the function f is continuous at c' by showing that the same δ works (for this ϵ).

(2) Give an example of such a continuous, additive function.

Exercise 3.3. Suppose that $f : \mathbb{R} \rightarrow \mathbb{R}$ and there exists an $M > 0$ such that for all $x \in \mathbb{R}$, $|f(x)| \leq M|x|$. Prove that f is continuous at 0. *Hint:* Find $f(0)$.

Exercise 3.4. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be defined by

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is rational,} \\ 1 & \text{if } x \text{ is irrational.} \end{cases}$$

Prove that for every $c \in \mathbb{R}$, f is not continuous at c .

Hint: Use the fact that there are irrational numbers arbitrarily close to any rational number and rational numbers arbitrarily close to any irrational number.

Exercise 3.5. Let $f : (a, b) \rightarrow \mathbb{R}$ be a continuous function. Prove that if for some $c \in (a, b)$, $f(c) > 0$, then there exists a $\delta > 0$ such that for all $x \in (c - \delta, c + \delta)$, $f(x) > 0$.

Exercise 3.6. Show that in the definition of continuity of a function at a point, we may replace the symbol $<$ with \leq , that is, the following statements are equivalent for $f : I \rightarrow \mathbb{R}$, and c belonging to the interval I :

- (1) $\forall \epsilon > 0, \exists \delta > 0$ such that whenever $x \in I$ satisfies $|x - c| < \delta$, $|f(x) - f(c)| < \epsilon$.
- (2) $\forall \epsilon > 0, \exists \delta > 0$ such that whenever $x \in I$ satisfies $|x - c| < \delta$, $|f(x) - f(c)| \leq \epsilon$.
- (3) $\forall \epsilon > 0, \exists \delta > 0$ such that whenever $x \in I$ satisfies $|x - c| \leq \delta$, $|f(x) - f(c)| \leq \epsilon$.
- (4) $\forall \epsilon > 0, \exists \delta > 0$ such that whenever $x \in I$ satisfies $|x - c| \leq \delta$, $|f(x) - f(c)| < \epsilon$.

3.2. Continuous functions preserve convergence

In Example 3.4, we had to work hard in order to prove the continuity of the reciprocal function. We will now learn about a result which will make life considerably simpler. Roughly speaking, this result says that a function is continuous at a point if and only if it preserves convergence of sequences with limit c .

Theorem 3.1. Let I be an interval in \mathbb{R} , $c \in I$ and $f : I \rightarrow \mathbb{R}$. Then

$$\boxed{f \text{ is continuous at } c}$$

if and only if

$$\boxed{\text{for every convergent sequence } (x_n)_{n \in \mathbb{N}} \text{ contained in } I \text{ with limit } c, \\ (f(x_n))_{n \in \mathbb{N}} \text{ is convergent with limit } f(c).} \quad (3.1)$$

Proof.

Only if: Suppose that f is continuous at $c \in I$.

Let $(x_n)_{n \in \mathbb{N}}$ be a convergent sequence contained in I with limit c .

Since f is continuous at $c \in I$, given $\epsilon > 0$, there exists a $\delta > 0$ such that for all $x \in I$ satisfying $|x - c| < \delta$, $|f(x) - f(c)| < \epsilon$.

As $(x_n)_{n \in \mathbb{N}}$ is convergent with limit c , there exists an $N \in \mathbb{N}$ such that for all $n > N$, $|x_n - c| < \delta$.

Consequently, for $n > N$, $|f(x_n) - f(c)| < \epsilon$. So $(f(x_n))_{n \in \mathbb{N}}$ is convergent with limit $f(c)$. This completes the proof of the ‘Only if’ part.

If: Now suppose that (3.1) holds. Then we need to show that f is continuous at c and we prove this by contradiction. Assume that f is not continuous at c , that is,

$$\neg [\forall \epsilon > 0 \exists \delta > 0 \text{ such that } \forall x \in I \text{ such that } |x - c| < \delta, |f(x) - f(c)| < \epsilon]$$

that is, $\exists \epsilon > 0$ such that $\forall \delta > 0 \exists x \in I$ such that $|x - c| < \delta$ but $|f(x) - f(c)| \geq \epsilon$. Hence if $\delta = \frac{1}{n}$, then we can find an $x_n \in I$ such that we have $|x_n - c| < \delta = \frac{1}{n}$, but $|f(x_n) - f(c)| \geq \epsilon$.

Claim 1: The sequence $(x_n)_{n \in \mathbb{N}}$ is contained in I and is convergent with limit c .

We have for all $n \in \mathbb{N}$ that $|x_n - c| < 1/n$, that is, $c - \frac{1}{n} < x_n < c + \frac{1}{n}$.

As $\lim_{n \rightarrow \infty} c - \frac{1}{n} = c = \lim_{n \rightarrow \infty} c + \frac{1}{n}$, the Sandwich Theorem gives $\lim_{n \rightarrow \infty} x_n = c$ too.

Claim 2: The sequence $(f(x_n))_{n \in \mathbb{N}}$ does not converge to $f(c)$.

Indeed for all $n \in \mathbb{N}$, we have $|f(x_n) - f(c)| \geq \epsilon$. Thus for instance $\frac{\epsilon}{2} > 0$, but it is not possible to find a large enough $N \in \mathbb{N}$ such that for all $n > N$, we have $|f(x_n) - f(c)| < \frac{\epsilon}{2}$ (for if this were possible, then we would arrive at the contradiction $\epsilon \leq |f(x_n) - f(c)| < \frac{\epsilon}{2}$).

Claims 1 and 2 show that (3.1) does not hold, a contradiction. Hence f is continuous at c . \square

Let us revisit some of our examples from the previous section in light of this result.

Example 3.5 (The reciprocal function). Let us revisit the function h considered in Example 3.4. Let $c \in (0, \infty)$ and $(x_n)_{n \in \mathbb{N}}$ be any convergent sequence in $(0, \infty)$ with limit c . Then by the Algebra of Limits, $(h(x_n))_{n \in \mathbb{N}} = (1/x_n)_{n \in \mathbb{N}}$ is convergent with limit $1/c = h(c)$. By Theorem 3.1, it follows that h is continuous at c . As the choice of $c \in (0, \infty)$ was arbitrary, h is continuous on $(0, \infty)$. Done! \diamond

Example 3.6 (The Heaviside function). Let us revisit the function Y considered in Example 3.3. Consider the convergent sequence $(1/n)_{n \in \mathbb{N}}$ with limit 0. Then $(Y(1/n))_{n \in \mathbb{N}} = (1)_{n \in \mathbb{N}}$ is convergent with limit $1 \neq 0 = Y(0)$.

But if Y was continuous at 0, then by Theorem 3.1, $(Y(1/n))_{n \in \mathbb{N}}$ should have been convergent with limit $Y(0) = 0$. Thus Y is not continuous at 0. \diamond

Exercise 3.7. Recall Exercise 3.1: $f : \mathbb{R} \rightarrow \mathbb{R}$ is given by $f(x) = x^2$ for $x \in \mathbb{R}$. Using the characterization of continuity provided in Theorem 3.1, prove that f is continuous on \mathbb{R} .

Exercise 3.8. Let $c \in \mathbb{R}$, $\delta > 0$ and $f : (c - \delta, c] \rightarrow \mathbb{R}$ be continuous and strictly increasing on $(c - \delta, c]$. Show that f is strictly increasing on $(c - \delta, c]$.

Exercise 3.9. Prove that if $f : \mathbb{R} \rightarrow \mathbb{R}$ is continuous and $f(x) = 0$ if x is rational, then $f(x) = 0$ for all $x \in \mathbb{R}$. Revisit Exercise 3.4.

Hint: Given $c \in \mathbb{R}$, there exists a sequence $(r_n)_{n \in \mathbb{N}}$ of rationals that converges to c .

Exercise 3.10. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ ‘preserve divergent sequences’, that is, for every divergent sequence $(x_n)_{n \in \mathbb{N}}$, $(f(x_n))_{n \in \mathbb{N}}$ is divergent as well. Prove that f is one-to-one.

Hint: Let x_1, x_2 be distinct real numbers, and consider the sequence $x_1, x_2, x_1, x_2, \dots$.

Exercise 3.11. Let I be an interval, $c \in I$, $f : I \rightarrow \mathbb{R}$. Show the following are equivalent:

- (1) f is continuous at c .
- (2) For every $(x_n)_{n \in \mathbb{N}}$ contained in I such that $(x_n)_{n \in \mathbb{N}}$ converges to c , $(f(x_n))_{n \in \mathbb{N}}$ converges.

Exercise 3.12. Consider the function $f : \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$f(x) = \begin{cases} x & \text{if } x \text{ is rational,} \\ -x & \text{if } x \text{ is irrational.} \end{cases}$$

Prove that f is continuous only at 0.

Hint: For every rational, there is a sequence of irrational numbers that converges to it, and for every irrational number, there is a sequence of rational numbers that converges to it.

Exercise 3.13. (*) Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a continuous function such that for all $x, y \in \mathbb{R}$,

$$f(x + y) = f(x) + f(y).$$

Show that there exists a real number a such that for all $x \in \mathbb{R}$, $f(x) = ax$.

Hint: Show first that for natural numbers n , $f(n) = nf(1)$. Extend this to integers n , and then to rational numbers n/d . Finally use the density of \mathbb{Q} in \mathbb{R} to prove the claim.

Exercise 3.14. Determine if the following are always true for two continuous $f, g : \mathbb{R} \rightarrow \mathbb{R}$.

- (1) If $f\left(\frac{1}{2n+7}\right) = g\left(\frac{n}{n^2+1}\right)$ for all $n \in \mathbb{N}$, then $f(0) = g(0)$.
- (2) If $f(n) = g(n^2)$ for all n , and $\lim_{n \rightarrow \infty} g(n) = L$, then $\lim_{n \rightarrow \infty} f(n)$ also exists, and equals L .

Using Theorem 3.1, we obtain the following useful result which says that the point-wise sum, product, etc. of continuous functions is continuous. But before we state this result, we introduce some convenient notation.

Let I be an interval in \mathbb{R} . Given functions $f : I \rightarrow \mathbb{R}$ and $g : I \rightarrow \mathbb{R}$, we define:

- (1) If $\alpha \in \mathbb{R}$, then we define the function $\alpha f : I \rightarrow \mathbb{R}$ by

$$(\alpha f)(x) = \alpha f(x), \quad x \in I.$$

- (2) We define the *absolute value of f* , $|f| : I \rightarrow \mathbb{R}$ by

$$|f|(x) = |f(x)|, \quad x \in I.$$

- (3) The *sum of f and g* , $f + g : I \rightarrow \mathbb{R}$ is defined by

$$(f + g)(x) = f(x) + g(x), \quad x \in I.$$

- (4) The *product of f and g* , $fg : I \rightarrow \mathbb{R}$ is defined by

$$(fg)(x) = f(x)g(x), \quad x \in I.$$

- (5) If $k \in \mathbb{N}$, then we define the *k th power of f* , $f^k : I \rightarrow \mathbb{R}$ by

$$f^k(x) = (f(x))^k, \quad x \in I.$$

- (6) If for all $x \in I$, $g(x) \neq 0$, then we define $\frac{1}{g} : I \rightarrow \mathbb{R}$ by

$$\left(\frac{1}{g}\right)(x) = \frac{1}{g(x)}, \quad x \in I.$$

Theorem 3.2. Let I be an interval in \mathbb{R} and let $c \in I$. Suppose that $f : I \rightarrow \mathbb{R}$ and $g : I \rightarrow \mathbb{R}$ are continuous at c . Then:

- (1) For all $\alpha \in \mathbb{R}$, αf is continuous at c .
- (2) $|f|$ is continuous at c .
- (3) $f + g$ is continuous at c .
- (4) fg is continuous at c .
- (5) For all $k \in \mathbb{N}$, f^k is continuous at c .
- (6) If for all $x \in I$, $g(x) \neq 0$, then $\frac{1}{g}$ is continuous at c .

Proof. Suppose $(x_n)_{n \in \mathbb{N}}$ is a convergent sequence contained in I , with limit c . Since f and g are continuous at c , from Theorem 3.1, it follows that $(f(x_n))_{n \in \mathbb{N}}$ and $(g(x_n))_{n \in \mathbb{N}}$ are convergent with limits $f(c)$ and $g(c)$, respectively. Hence from Theorem 2.4, it follows that:

- (1) $(\alpha \cdot f(x_n))_{n \in \mathbb{N}}$ is convergent with limit $\alpha \cdot f(c)$, i.e., $((\alpha f)(x_n))_{n \in \mathbb{N}}$ is convergent with limit $(\alpha f)(c)$. So from Theorem 3.1, it follows that αf is continuous at c .
- (2) $(|f(x_n)|)_{n \in \mathbb{N}}$ is convergent with limit $|f(c)|$, that is, $(|f|(x_n))_{n \in \mathbb{N}}$ is convergent with limit $|f|(c)$. So from Theorem 3.1, it follows that $|f|$ is continuous at c .
- (3) $(f(x_n) + g(x_n))_{n \in \mathbb{N}}$ is convergent with limit $f(c) + g(c)$, i.e., $((f + g)(x_n))_{n \in \mathbb{N}}$ is convergent with limit $(f + g)(c)$. By Theorem 3.1, $f + g$ is continuous at c .
- (4) $(f(x_n)g(x_n))_{n \in \mathbb{N}}$ is convergent with limit $f(c)g(c)$, i.e., the sequence $((fg)(x_n))_{n \in \mathbb{N}}$ is convergent with limit $(fg)(c)$. By Theorem 3.1, fg is continuous at c .
- (5) $((f(x_n))^k)_{n \in \mathbb{N}}$ is convergent with limit $(f(c))^k$, that is, $(f^k(x_n))_{n \in \mathbb{N}}$ is convergent with limit $f^k(c)$. So from Theorem 3.1, it follows that f^k is continuous at c .
- (6) $(\frac{1}{g(x_n)})_{n \in \mathbb{N}}$ is convergent with limit $\frac{1}{g(c)}$ (since for all $x \in I$, $g(x) \neq 0$, in particular $g(x_n) \neq 0$ and $g(c) \neq 0$), that is, $(\frac{1}{g}(x_n))_{n \in \mathbb{N}}$ is convergent with limit $(\frac{1}{g})(c)$. So from Theorem 3.1, it follows that $\frac{1}{g}$ is continuous at c . \square

Example 3.7 (Polynomials are continuous). Since $f : \mathbb{R} \rightarrow \mathbb{R}$ given by $f(x) = x$ for $x \in \mathbb{R}$ is continuous (see Example 3.2 on page 67), it follows that for all $k \in \mathbb{N}$, x^k is continuous. Thus given arbitrary scalars c_0, c_1, \dots, c_d in \mathbb{R} , it follows that the functions $c_0 \cdot 1, c_1 \cdot x, \dots, c_d \cdot x^d$ are continuous. Consequently the *polynomial function* $p : \mathbb{R} \rightarrow \mathbb{R}$ defined by $p(x) = c_0 + c_1x + \dots + c_dx^d$, $x \in \mathbb{R}$, is continuous. \diamond

Example 3.8 (The reciprocal function). Let us revisit the function h considered in Example 3.4. As $x \mapsto x : (0, \infty) \rightarrow \mathbb{R}$ is continuous, and since $g(x) = x \neq 0$ for all $x \in (0, \infty)$, it follows that $h = \frac{1}{g} : (0, \infty) \rightarrow \mathbb{R}$, given by $h(x) = \frac{1}{x}$ for $x > 0$, is continuous too. \diamond

Exercise 3.15. Show that the rational function $f : \mathbb{R} \rightarrow \mathbb{R}$, given by $f(x) = \frac{x^2}{1+x^2}$ for $x \in \mathbb{R}$, is continuous on \mathbb{R} .

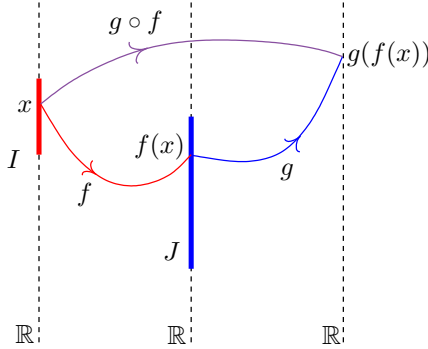
The composition of continuous functions is continuous.

Let I, J be intervals in \mathbb{R} , and $f : I \rightarrow \mathbb{R}$, $g : J \rightarrow \mathbb{R}$ be functions such that

$$f(I) := \{f(x) : x \in I\} \subset J,$$

that is, the range of f is contained in the domain of g . Then the *composition of g with f* , denoted by $g \circ f$ is the function $g \circ f : I \rightarrow \mathbb{R}$ defined by

$$(g \circ f)(x) = g(f(x)), \quad x \in I.$$



Theorem 3.3. Let I, J be intervals in \mathbb{R} , and $f : I \rightarrow \mathbb{R}$, $g : J \rightarrow \mathbb{R}$ be two functions such that

- $f(I) := \{f(x) : x \in I\} \subset J$,
- f is continuous at c , and
- g is continuous at $f(c)$ ($\in f(I) \subset J$).

Then their composition $g \circ f : I \rightarrow \mathbb{R}$ is continuous at c .

Proof. Let $(x_n)_{n \in \mathbb{N}}$ be any sequence in I with limit c . As f is continuous at c , $(f(x_n))_{n \in \mathbb{N}}$ converges to $f(c)$. But for all $n \in \mathbb{N}$, $f(x_n) \in f(I) \subset J$, and $f(c) \in f(I) \subset J$. As g is continuous at $f(c)$, $(g(f(x_n)))_{n \in \mathbb{N}}$ converges to $g(f(c))$, that is, $((g \circ f)(x_n))_{n \in \mathbb{N}}$ converges to $(g \circ f)(c)$. Hence $g \circ f$ is continuous at c . \square

Example 3.9. We know the polynomial function $x \mapsto 1+x^2 : \mathbb{R} \rightarrow \mathbb{R}$ is continuous, and the reciprocal function $x \mapsto 1/x : (0, \infty) \rightarrow \mathbb{R}$ is continuous. Also, the range of p , $p(\mathbb{R}) = \{1+x^2 : x \in \mathbb{R}\} \subset (0, \infty) = \text{domain of } h$. So their composition, namely

$$x \mapsto \frac{1}{1+x^2} : \mathbb{R} \rightarrow \mathbb{R}$$

is continuous too.

More generally, suppose q is a polynomial such that $q(x) > 0$ for all x in an interval I . Then for any polynomial p , the rational function $r : I \rightarrow \mathbb{R}$ given by $r(x) = \frac{p(x)}{q(x)}$ for $x \in I$, is continuous. \diamond

Exercise 3.16. Define $f : \mathbb{R} \rightarrow \mathbb{R}$ by $f(x) = |x+1| - |x|$, $x \in \mathbb{R}$. Find $\lim_{n \rightarrow \infty} (f \circ f)\left(\frac{(-1)^n}{n}\right)$.

Exercise 3.17. Determine if the following are always true for $f, g : \mathbb{R} \rightarrow \mathbb{R}$ and $a \in \mathbb{R}$.

- (1) If $g \circ f$ is continuous at a , then f is continuous at a and g is continuous at $f(a)$.
- (2) If $g \circ f$ is continuous at a , then f is continuous at a or g is continuous at $f(a)$.
- (3) If $g \circ f$ isn't continuous at a , then f isn't continuous at a and g isn't continuous at $f(a)$.
- (4) If $g \circ f$ isn't continuous at a , then f isn't continuous at a or g isn't continuous at $f(a)$.

Exercise 3.18. Show that $f : \mathbb{R} \rightarrow \mathbb{R}$ given by $f(x) = \begin{cases} x \sin \frac{1}{x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$ is continuous. Use Maple to plot the graph of f .

Exercise 3.19. Suppose I is an interval, and $f, g : I \rightarrow \mathbb{R}$ are continuous functions on I . Define the function $\max\{f, g\} : I \rightarrow \mathbb{R}$ by $(\max\{f, g\})(x) = \max\{f(x), g(x)\}$ for all $x \in I$. Is $\max\{f, g\}$ continuous on I ? *Hint:* Exercise 1.23.

In the next two sections, we will learn two fundamental results concerning continuous functions $f : [a, b] \rightarrow \mathbb{R}$ on a compact interval $[a, b]$, namely:

- (1) The Intermediate Value Theorem, saying that f assumes all the values between $f(a)$ and $f(b)$. Geometrically, this means the following. Consider the graph of f in the Cartesian plane. If we choose any number y lying between $f(a)$ and $f(b)$ and draw a horizontal line through the point y on the y -axis, then this horizontal line must meet the graph of f at some point. This is 'clear' since f , being continuous, should have a graph having 'no breaks'.

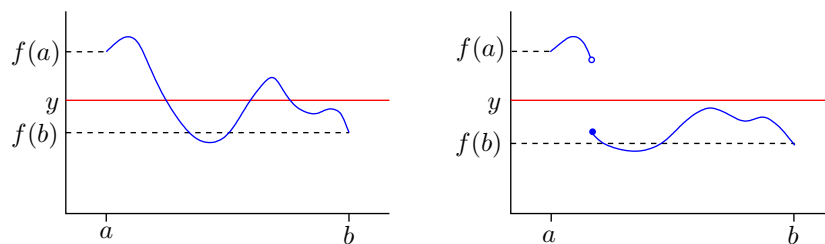
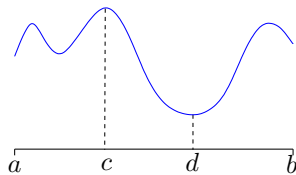


Figure 4. The right picture shows that the continuity condition can't be dropped.

- (2) The Extreme Value Theorem, saying that f has a maximiser and a minimiser on $[a, b]$ (i.e., the function f assumes the *extreme* values of the range $f([a, b])$). Geometrically, this means that if we consider the graph of f , then there must be a point in $c \in [a, b]$, where the graph y coordinate is highest, and a point $d \in [a, b]$ where the graph y coordinate is lowest.



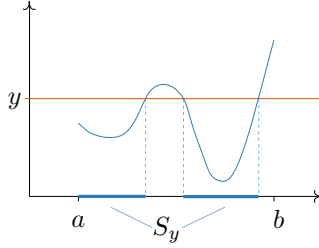
Although these two properties might seem 'obvious' when interpreted geometrically, they require proofs. We will see that the Least Upper Bound Property of \mathbb{R} will be used crucially in the proofs. We will begin with the Intermediate Value Property.

3.3. Intermediate Value Theorem

Roughly speaking, the Intermediate Value Theorem says that a continuous function on a compact interval cannot ‘hop over’ intermediate values. For instance, if the height of a mountain is 1976 meters above sea level, then given any number between 0 and 1976, say 399, there must exist a point on the mountain that is exactly 399 meters above sea level. The picture shown in Figure 4 shows that the continuity of the function is an essential requirement.

Theorem 3.4 (Intermediate Value Theorem). *If $f : [a, b] \rightarrow \mathbb{R}$ is continuous and $y \in \mathbb{R}$ lies between $f(a)$ and $f(b)$, (that is, $f(a) \leq y \leq f(b)$ or $f(b) \leq y \leq f(a)$), then there exists a $c \in [a, b]$ such that $f(c) = y$.*

Proof. Consider first the case $f(a) \leq y \leq f(b)$. Define $S_y = \{x \in [a, b] : f(x) \leq y\}$. (Pictorially, this set can be visualized like this: imagine again the horizontal line through y , and look at the portion of the graph of f that lies below y . S_y is the shadow on the x axis of this portion with a light source very high up above.)



S_y is a subset of \mathbb{R} , it is nonempty (as $a \in S_y$) and S_y is bounded above (by b). By the Least Upper Bound Property of \mathbb{R} , $c := \sup S_y$ exists. As b is an upper bound of S_y , and c is the *least* upper bound of S_y , clearly $c \leq b$. As $a \in S_y$, we also have $a \leq c$. Summarizing, $c \in [a, b]$. We now claim that this c does the job.

Claim: $f(c) = y$.

We will show that $f(c) \leq y$ as well as $f(c) \geq y$, and this will prove the claim.

$f(c) \leq y$: For every $n \in \mathbb{N}$, $c - \frac{1}{n}$ is not an upper bound of S_y . So there must be an $x_n \in S_y$ such that $x_n > c - \frac{1}{n}$. Hence for all n , $c - \frac{1}{n} < x_n \leq c$. By the Sandwich Theorem, $\lim_{n \rightarrow \infty} x_n = c$. As f is continuous, $\lim_{n \rightarrow \infty} f(x_n) = f(c)$. As $f(x_n) \leq y$, $n \in \mathbb{N}$, (since $x_n \in S_y$), $f(c) = \lim_{n \rightarrow \infty} f(x_n) \leq y$.

$f(c) \geq y$: If $c = b$, then we are done, since $y \leq f(b) = f(c)$. So we suppose that $c < b$. Define for $n \in \mathbb{N}$, $x_n := c + \frac{b-c}{n}$ ($\leq c + \frac{b-c}{1} = b$). Then $x_n \in [a, b]$, and $(x_n)_{n \in \mathbb{N}}$ is convergent with limit c . Because f is continuous, $(f(x_n))_{n \in \mathbb{N}}$ converges to $f(c)$. But $x_n > c$ for each $n \in \mathbb{N}$, and so $x_n \notin S_y$ for each n . Hence for all n , $f(x_n) > y$. Thus $f(c) \geq y$.

Consequently, $f(c) = y$, proving the claim.

Thus the proof of the theorem is complete when $f(a) \leq y \leq f(b)$.

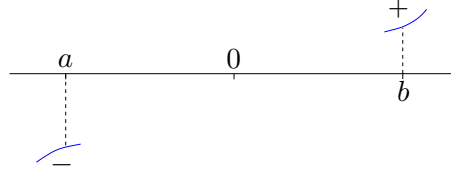
Now suppose that $f(b) \leq y \leq f(a)$. Then $(-f)(a) \leq -y \leq (-f)(b)$. By the continuity of f , $-f$ is continuous too. So applying the previous result (with $-f$ instead of f , and $-y$ instead of y), it follows that there is a $c \in [a, b]$ such that $(-f)(c) = -y$, that is, $-f(c) = -y$ or $f(c) = y$. This completes the proof. \square

Example 3.10. Consider the polynomial $p : \mathbb{R} \rightarrow \mathbb{R}$ given by

$$p(x) = x^{2014} + x^{1976} - \frac{1}{399}, \quad x \in \mathbb{R}.$$

Then p is continuous, and $p(0) = 0 + 0 - \frac{1}{399} = -\frac{1}{399} < 0$, $p(1) = 1 + 1 - \frac{1}{399} > 0$. As $p(0) \leq y := 0 \leq p(1)$, and since $p : [0, 1] \rightarrow \mathbb{R}$ is continuous, it follows by the Intermediate Value Theorem, that there exists a $c \in [0, 1]$ such that $p(c) = 0$. In other words p has a real root in $[0, 1]$.

More generally, one can show that *any* odd degree polynomial p with real coefficients must have at least one real root. The reason is that for large positive values of x , $p(x)$ will have the same sign as the leading coefficient c_d , while for large² negative values of x , $p(x)$ will have the opposite sign as that of c_d (since d is *odd*). Consequently, p must vanish somewhere in between these two extremes of large positive and negative x s.



We give a proof below. Suppose for some $m \in \mathbb{N}$, $p(x) = c_{2m-1}x^{2m-1} + \cdots + c_1x + c_0$ ($x \in \mathbb{R}$), where c_0, \dots, c_{2m-1} are real numbers and $c_{2m-1} \neq 0$. Then $p(x) = c_{2m-1}x^{2m-1}r(x)$, where

$$r(x) = 1 + \frac{c_{2m-2}}{c_{2m-1}x} + \cdots + \frac{c_1}{c_{2m-1}x^{2m-2}} + \frac{c_0}{c_{2m-1}x^{2m-1}}.$$

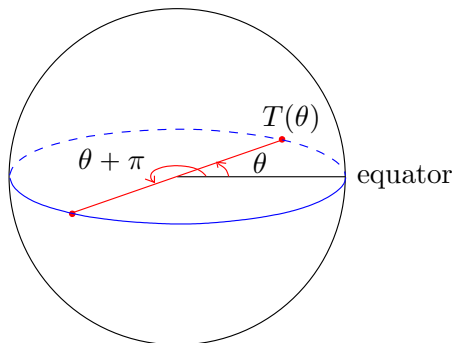
By the Algebra of Limits, $\lim_{n \rightarrow \infty} r(n) = 1 = \lim_{n \rightarrow \infty} r(-n)$. So there exists an $N \in \mathbb{N}$ such that for all $n > N$, $|r(n) - 1| < \frac{1}{2}$ and $|r(-n) - 1| < \frac{1}{2}$. Thus $1 - r(n) \leq |r(n) - 1| < \frac{1}{2}$, so that $r(n) > \frac{1}{2} > 0$ for $n > N$. Similarly, $r(-n) > \frac{1}{2} > 0$ for $n > N$. Using $p(n) = c_{2m-1}n^{2m-1}r(n)$ and $p(-n) = -c_{2m-1}n^{2m-1}r(-n)$, we obtain the following table of signs for all $n > N$:

	$c_{2m-1} > 0$	$c_{2m-1} < 0$
$p(n)$	> 0	< 0
$p(-n)$	< 0	> 0

By the Intermediate Value Theorem for $p|_{[-(N+1), N+1]} : [-(N+1), N+1] \rightarrow \mathbb{R}$, we conclude that since the values at the end points have opposite signs, p must vanish somewhere on $[-(N+1), N+1]$. \diamond

²that is, $x < 0$ and $|x|$ large

Example 3.11. At any given instant of time, there exists a pair of diametrically opposite points on the equator of the earth which have the same temperature.



Let $T(\theta)$ denote the surface temperature at the point on the equator with longitude θ . Then $\theta \mapsto T(\theta)$ is continuous on the interval $[0, 2\pi]$ (with longitude measured in radians³). Note that $T(0) = T(2\pi)$. Let $S : [0, \pi] \rightarrow \mathbb{R}$ be given by

$$S(\theta) = T(\theta) - T(\theta + \pi), \quad \theta \in [0, \pi].$$

Then S is continuous, and

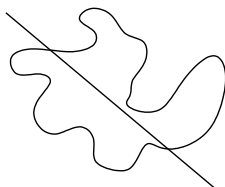
$$S(\pi) = T(\pi) - T(2\pi) = T(\pi) - T(0) = -(T(0) - T(\pi)) = -S(0).$$

So 0 lies between $S(\pi)$ and $S(0) = -S(\pi)$. By the Intermediate Value Theorem, there exists a $\theta_* \in [0, \pi]$ such that $S(\theta_*) = 0$, that is, $T(\theta_*) = T(\theta_* + \pi)$. \diamond

Exercise 3.20. Suppose that $f : [0, 1] \rightarrow \mathbb{R}$ is a continuous function such that for all $x \in [0, 1]$, $0 \leq f(x) \leq 1$. Prove that there exists at least one $c \in [0, 1]$ such that $f(c) = c$. *Hint:* Consider the function $g(x) = f(x) - x$, and use the Intermediate Value Theorem.

Exercise 3.21. Let $f : [0, 1] \rightarrow \mathbb{R}$ be continuous. Show that there exists a $c \in [0, 1]$ such that $f(c) - f(1) = (f(0) - f(1))c$. *Hint:* Consider $f(x) - f(1) - (f(0) - f(1))x$.

Exercise 3.22. Consider a flat pancake of arbitrary shape. Show that there is a straight line cut that divides the pancake into two parts having equal areas. Can the direction of the straight line cut be chosen arbitrarily?



Exercise 3.23. True or false? There is real number x such that $x^{399} + \frac{1976}{1 + x^2(\cos x)^2} = 28$.

³If the reader is not familiar with the radian angle measure, one may just think of T as a function on the interval $[0, 360]$, with the angle θ measured in degrees.

Exercise 3.24. At 8:00 a.m. on Saturday, a hiker begins walking up the side of a mountain to his weekend campsite. On Sunday morning at 8:00 a.m., he walks back down the mountain along the same trail. It takes him one hour to walk up, but only half an hour to walk down. At some point on his way down, he realizes that he was at the same spot at exactly the same time on Saturday. Prove that he is right.

Hint: Let $u(t)$ and $d(t)$ be the position functions for the walks up and down, and apply the Intermediate Value Theorem to $f(t) = u(t) - d(t)$.

Exercise 3.25. Show that p , where $p(x) := 2x^3 - 5x^2 - 10x + 5$, has a real root in $[-1, 2]$.

Exercise 3.26. Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous and such that for all $x \in [a, b]$, $f(x) \neq 0$. Show that f assumes only positive values or f assumes only negative values.

Exercise 3.27. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be continuous. If $S := \{f(x) : x \in \mathbb{R}\}$ is neither bounded above nor bounded below, prove that $S = \mathbb{R}$.

Hint: If $y \in \mathbb{R}$, then since S is neither bounded above nor bounded below, there exist $x_0, x_1 \in \mathbb{R}$ such that $f(x_0) < y < f(x_1)$.

Exercise 3.28. (*) Show that given any continuous function $f : \mathbb{R} \rightarrow \mathbb{R}$, there exists an $x_0 \in [0, 1]$ and an $m \in \mathbb{Z} \setminus \{0\}$ such that $f(x_0) = mx_0$. In other words, the graph of f intersects some nonhorizontal line $y = mx$ at some point x_0 in $[0, 1]$.

Hint: If $f(0) = 0$, take $x_0 = 0$ and any $m \in \mathbb{Z} \setminus \{0\}$. If $f(0) > 0$, then choose $N \in \mathbb{N}$ satisfying $N > f(1)$, and apply the intermediate value theorem to the continuous function $g(x) = f(x) - Nx$ on the interval $[0, 1]$. If $f(0) < 0$, then first choose a $N \in \mathbb{N}$ such that $N > -f(1)$, and consider the function $g(x) = f(x) + Nx$, and proceed in a similar manner.

Exercise 3.29. (*) Prove that there does not exist a continuous function $f : \mathbb{R} \rightarrow \mathbb{R}$ which assumes rational values at irrational numbers, and irrational values at rational numbers, that is, $f(\mathbb{Q}) \subset \mathbb{R} \setminus \mathbb{Q}$ and $f(\mathbb{R} \setminus \mathbb{Q}) \subset \mathbb{Q}$.

Hint: Note that for each $m \in \mathbb{Z} \setminus \{0\}$, there is no $x_0 \in \mathbb{R}$ such that $f(x_0) = mx_0$.

Exercise 3.30. In each of the following cases, give an example of a continuous function $f : S \rightarrow \mathbb{R}$ such that $f(S) = T$, or explain why such an f can't exist.

- (1) $S = (0, 1)$, $T = (0, 1]$.
- (2) $S = (0, 1)$, $T = \{0, 1\}$.

3.4. Extreme value theorem

Theorem 3.5 (Extreme Value Theorem).

If $f : [a, b] \rightarrow \mathbb{R}$ is continuous, then

- (1) $S := \{f(x) : x \in [a, b]\} =: f([a, b])$ is bounded.
- (2) $\sup S$ and $\inf S$ exist.
- (3) $\sup S$ and $\inf S$ are attained, that is, there exist $c, d \in [a, b]$ such that $f(c) = \sup S = \max S$ and $f(d) = \inf S = \min S$.

Thus, in the above conclusion, we have $f(c) \geq f(x)$ for all $x \in [a, b]$ (so that c is a *maximiser* of f), and $f(d) \leq f(x)$ for all $x \in [a, b]$ (so that d is a *minimiser* of f).

The continuity of f says something locally about f at each point of its domain. However, the conclusion says something globally about f . This miracle happens because $[a, b]$ is ‘compact’. We will later see examples that show that maximisers/minimisers may fail to exist if either the domain of f is not compact or if f is not continuous. First, let us prove the Extreme Value Theorem. We will use the following observation:

Claim: A subset $S \subset \mathbb{R}$ is bounded if and only if $|S| := \{|x| : x \in \mathbb{R}\}$ is bounded.

Indeed, if $|S|$ is bounded, then in particular, $|S|$ is bounded above, and so there exists a $u \in \mathbb{R}$ such that for all $x \in S$, $|x| \leq u$, giving $-u \leq x \leq u$. Thus S is bounded. Vice versa, if S is bounded, then there exist $u, \ell \in \mathbb{R}$ such that for all $x \in S$, $\ell \leq x \leq u$, giving $x \leq u \leq \max\{u, -\ell\} =: M$ and $-x \leq -\ell \leq \max\{u, -\ell\} = M$. Hence for all $x \in S$, $0 \leq |x| \leq M$, showing that $|S|$ is bounded.

Proof. (Of the Extreme Value Theorem).

- (1) We first show that f is bounded, that is, $S := \{f(x) : x \in [a, b]\}$ is bounded. Suppose S is not bounded. Then $|S|$ is not bounded. But $|S|$ is bounded below (by 0). So $|S|$ is not bounded above. Let $n \in \mathbb{N}$. Then n is not an upper bound of $|S|$. So there exists some $x_n \in [a, b]$ such that $|f(x_n)| > n$. In this way, we get a sequence $(x_n)_{n \in \mathbb{N}}$. Since $a \leq x_n \leq b$ for all $n \in \mathbb{N}$, $(x_n)_{n \in \mathbb{N}}$ is bounded. By the Bolzano-Weierstrass Theorem (Theorem 2.8), $(x_n)_{n \in \mathbb{N}}$ has a convergent subsequence, say $(x_{n_k})_{k \in \mathbb{N}}$, that converges to some limit L . For all $k \in \mathbb{N}$, we have $a \leq x_{n_k} \leq b$. It follows that $a \leq L \leq b$, i.e., $L \in [a, b]$. As f is continuous in particular at L , $(f(x_{n_k}))_{k \in \mathbb{N}}$ is convergent, and in particular bounded. So there exists an $M > 0$ such that for all $k \in \mathbb{N}$, $|f(x_{n_k})| \leq M$. So for all $k \in \mathbb{N}$, we have $k \leq n_k < |f(x_{n_k})| \leq M$, a contradiction. Thus S is bounded.
- (2) $S \neq \emptyset$ (as $f(a) \in S$). S is bounded. So by the Least Upper Bound Property of \mathbb{R} , $\sup S$ exists, and by the Greatest Lower Bound Property of \mathbb{R} , $\inf S$ exists.
- (3) We claim that there exists a $c \in [a, b]$ such that $f(c) = M := \sup S$. Let $n \in \mathbb{N}$. Then $M - \frac{1}{n}$ is not an upper bound of S . So there exists a $y_n \in S$ such that $M - \frac{1}{n} < y_n \leq M$. As this y_n belongs to the range S of f , $y_n = f(x_n)$ for some $x_n \in [a, b]$. By Bolzano-Weierstrass Theorem, there is a subsequence, say $(x_{n_k})_{k \in \mathbb{N}}$, of $(x_n)_{n \in \mathbb{N}}$ which converges with limit, say c . As $a \leq x_{n_k} \leq b$ for all k , it follows that $c \in [a, b]$. We have $M - \frac{1}{n_k} < f(x_{n_k}) \leq M$ for all $k \in \mathbb{N}$. By the Sandwich Theorem, we conclude that $(f(x_{n_k}))_{k \in \mathbb{N}}$ is convergent with limit M . But since f is continuous at c , and since $(x_{n_k})_{k \in \mathbb{N}}$ converges to c , we must have $(f(x_{n_k}))_{k \in \mathbb{N}}$ is convergent with limit $f(c)$. By the uniqueness of limits, $f(c) = M = \sup S$. But $f(c) \in S$. So $\max S$ exists.

Finally, consider $-f : [a, b] \rightarrow \mathbb{R}$. As f is continuous, $-f$ is continuous too. By the above, there exists a $d \in [a, b]$ such that

$$(-f)(d) = \sup\{(-f)(x) : x \in [a, b]\} = \sup\{-f(x) : x \in [a, b]\} = \sup(-S) = -\inf S,$$

and so $f(d) = \inf S$. Since $f(d) \in S$, it follows that $\min S$ exists. \square

Example 3.12. There is no continuous function $f : [0, 1] \rightarrow \mathbb{R}$ onto \mathbb{R} . Indeed, by the Extreme Value Theorem, there exist m, M such that for all $x \in [0, 1]$, $m \leq f(x) \leq M$, that is the range $f([0, 1])$ of f is a bounded set, and so it can't equal the unbounded set \mathbb{R} . \diamond

Example 3.13.

(1) Let $f_1 : (0, 1) \rightarrow \mathbb{R}$ be defined by $f_1(x) := \frac{1}{x}$ for $x \in (0, 1)$.

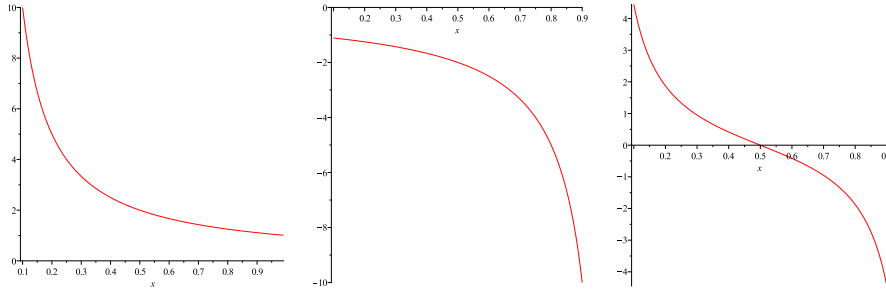
Then f_1 is continuous, but $(0, 1)$ is not a compact interval. We have

$$f_1((0, 1)) = \{1/x : x \in (0, 1)\} = \{y : y > 1\} = (1, \infty),$$

and so $\sup f_1((0, 1)) = \sup(1, \infty)$ does not exist. Also,

$$\inf f_1((0, 1)) = \inf(1, \infty) = 1,$$

but it is not attained: There does not exist a $d \in (0, 1)$ such that $f(d) = 1$. (Indeed, for all $d \in (0, 1)$, $f(d) = 1/d > 1$.)



Graphs of $x \mapsto \frac{1}{x}, \frac{1}{x-1}, \frac{x - \frac{1}{2}}{x(x-1)} : (0, 1) \rightarrow \mathbb{R}$.

(2) Let $f_2 : (0, 1) \rightarrow \mathbb{R}$ given by $f_2(x) = \frac{1}{x-1}$, $x \in (0, 1)$.

Then it can be shown that $f_2((0, 1)) = (-\infty, -1)$, and so $\sup f_2((0, 1)) = -1$, but it is not attained, and $\inf f_2((0, 1))$ does not exist.

(3) Similarly, consider $f_3 : (0, 1) \rightarrow \mathbb{R}$ given by

$$f_3(x) = \frac{x - \frac{1}{2}}{x(x-1)}, \quad x \in (0, 1).$$

It can be shown that $f_3((0, 1)) = \mathbb{R}$. So neither $\sup f_3((0, 1))$ nor $\inf f_3((0, 1))$ exist.

(4) Let $f_4 : (0, 1) \rightarrow \mathbb{R}$ be given by $f_4(x) = x$, $x \in (0, 1)$. Then f_4 is continuous, but $(0, 1)$ is not compact, and

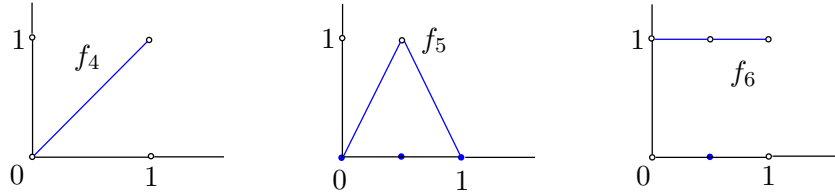
$$f_4((0, 1)) = \{f(x) : 0 < x < 1\} = \{x : 0 < x < 1\} = (0, 1).$$

$f_4((0, 1))$ is bounded, $\sup(0, 1) = 1$, but there is no $c \in (0, 1)$ such that $f_4(c) = 1$, and $\inf(0, 1) = 0$, but there is no $d \in (0, 1)$ such that $f_4(d) = 0$.

(5) Let $f_5 : [0, 1] \rightarrow \mathbb{R}$ be given by

$$f_5(x) = \begin{cases} 2x & \text{if } 0 \leq x < \frac{1}{2}, \\ 0 & \text{if } x = \frac{1}{2}, \\ 2 - 2x & \text{if } \frac{1}{2} < x \leq 1. \end{cases}$$

Then $[0, 1]$ is compact, but f_5 is *not* continuous. We have $f_5([0, 1]) = [0, 1]$, and there is no $c \in [0, 1]$ such that $f_5(c) = \sup f_5([0, 1]) = 1$.



(6) (Continuity or compactness is not necessary for existence of maximisers and minimisers.) Let $f_6 : (0, 1) \rightarrow \mathbb{R}$ be given by

$$f_6(x) = \begin{cases} 1 & \text{if } 0 < x < \frac{1}{2}, \\ 0 & \text{if } x = \frac{1}{2}, \\ 1 & \text{if } \frac{1}{2} < x < 1. \end{cases}$$

Then $(0, 1)$ is not compact, and f is not continuous. But $f_6([0, 1]) = \{0, 1\}$, and there do exist maximizers and a minimizer:

$$f(1/2) = 0 = \inf f((0, 1)), \quad \text{and} \quad f(3/4) = 1 = \sup f((0, 1)).$$

We summarize the above examples in the following table. ◇

Function $f : I \rightarrow \mathbb{R}$	I compact?	f continuous?	$\sup f(I)$ exists?	$\inf f(I)$ exists?	$\sup f(I)$ attained?	$\inf f(I)$ attained?
f_1	No	Yes	No	Yes	-	No
f_2	No	Yes	Yes	No	No	-
f_3	No	Yes	No	No	-	-
f_4	No	Yes	Yes	Yes	No	No
f_5	Yes	No	Yes	Yes	No	Yes
f_6	No	No	Yes	Yes	Yes	Yes

The utility of the Extreme Value Theorem in Optimisation.

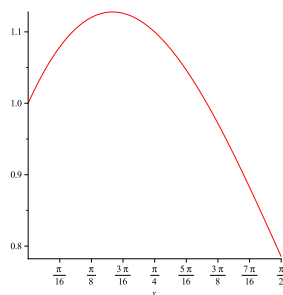
The Extreme Value Theorem (and its multivariable generalisation saying that a real-valued continuous function on a ‘compact set’ $K \subset \mathbb{R}^n$ has a maximiser and a minimiser) is useful in Optimisation Theory. In Optimisation Theory, one often meets *necessary* conditions for maximisers, that is, results of the following form:

If x_* is a maximiser of $f : S \rightarrow \mathbb{R}$, then x_* satisfies * * *.

(Where $\boxed{***}$ are certain mathematical conditions, such as the Lagrange multiplier equations.) Now such a result has limited use, since even if we find all $x_*(s)$ which satisfy $\boxed{***}$, we cannot conclude that there is one among these is actually a maximiser. But if we had an existence result (like the Extreme Value Theorem), then we know that a maximiser exists, and so we know that it must be among the (few) $x_*(s)$ in S that satisfy $\boxed{***}$. For example, we will later on learn that:

If x_* is a maximiser of $f : (a, b) \rightarrow \mathbb{R}$, then $f'(x_*) = 0$.

As an example, consider $f : [0, \pi/2] \rightarrow \mathbb{R}$, where $f(x) = \cos x + (x/2)$ for $x \in [0, \pi/2]$. Then $f(0) = 1$, $f(\pi/2) = \pi/4 < 1$, and $f(\pi/3) = (1/2) + (\pi/6) > (1/2) + (3/6) = 1 = f(0)$. By the Extreme Value Theorem, f has a maximiser $x_* \in [0, \pi/2]$. But the above calculation shows that $x_* \neq 0$ and $x_* \neq \pi/2$. Thus $x_* \in (0, \pi/2)$. Hence $f'(x_*) = 0$, that is, $-\sin x_* + \frac{1}{2} = 0$, and so $x_* = \pi/6$.



Exercise 3.31. A function $f : \mathbb{R} \rightarrow \mathbb{R}$ is called *periodic* if there exists a $T > 0$ such that for all $x \in \mathbb{R}$, $f(x + T) = f(x)$. If $f : \mathbb{R} \rightarrow \mathbb{R}$ is continuous and periodic, then prove that f is bounded, that is, the set $S = \{f(x) : x \in \mathbb{R}\}$ is bounded.

Exercise 3.32. True or false? If $f : [a, b] \rightarrow \mathbb{R}$ is continuous and $f(x) > 0$ for all $x \in [a, b]$, then f is in fact ‘bounded away from 0’, that is, there exists a $\delta > 0$ such that $f(x) \geq \delta$ for all $x \in [a, b]$.

Exercise 3.33. Let $f : [0, 3] \rightarrow [3, 9]$ be a continuous function such that $f(0) = 3$ and $f(3) = 6$. Which of the following statements is/are always true?

- (A) There exists a unique $c \in [0, 3]$ such that $f(c) = 4$.
- (B) The range of f contains the interval $[3, 6]$.
- (C) $f([0, 3]) = [3, 6]$.
- (D) There cannot exist a $c \in [0, 3]$ such that $f(c) = 9$.

Exercise 3.34. Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous on $[a, b]$, and define f_* as follows:

$$f_*(x) = \begin{cases} f(a) & \text{if } x = a, \\ \max\{f(y) : y \in [a, x]\} & \text{if } x \in (a, b]. \end{cases}$$

- (1) Show that f_* is a well-defined function.
- (2) If $f : [0, 1] \rightarrow \mathbb{R}$ is given by $f(x) = x - x^2$, then find f_* .

Exercise 3.35. Let the function $f : [a, b] \rightarrow \mathbb{R}$ be continuous.

Show that for any $c_1, \dots, c_n \in [a, b]$, there is a $c \in [a, b]$ such that $f(c) = \frac{f(c_1) + \dots + f(c_n)}{n}$.

Exercise 3.36. Let $f : [a, b] \rightarrow \mathbb{R}$ be a continuous function, having the property that for every $x \in [a, b]$, there exists a $y_x \in [a, b]$ such that $|f(y_x)| \leq |f(x)|/2$. Show that there exists a $c \in [a, b]$ such that $f(c) = 0$. *Hint:* Consider a minimiser of $|f|$.

Chapter 4

Number systems

In this chapter, beginning with an axiomatic framework, we will construct the natural numbers, the integers, the rational numbers, and the real numbers. We begin by recalling the important notion of equivalence relations again, because it will play an important role in the rest of the course. For example, to construct the integers from natural numbers, we will need to identify pairs of natural numbers using an equivalence relation.

4.1. Equivalence relations

Definition 4.1 (Relation).

A *relation* R on a set S is a subset of the $S \times S := \{(a, b) : a, b \in S\}$. If $(a, b) \in R$, then we write aRb .

For example, if we take S to be the set of all human beings, then

$$R_{\text{sibling}} := \{(a, b) \in S \times S : a, b \text{ have the same biological parents}\}$$

is a relation. As another example, we can take the set $S = \mathbb{Z}$, the set of all integers, and $R_{\text{mod } 2} = \{(m, n) \in \mathbb{Z} \times \mathbb{Z} : m - n \text{ is divisible by } 2\}$. Sometimes we use the symbol \sim to denote a relation, and then instead of aRb , we will write $a \sim b$.

Definition 4.2 (Equivalence relation).

A relation R on a set S is called an *equivalence relation* if it satisfies the following:

- (ER1) R is *reflexive*, that is, for all $a \in S$, aRa .
- (ER2) R is *symmetric*, that is, if aRb , then bRa .
- (ER3) R is *transitive*, that is, if aRb and bRc , then aRc .

In our example above, where $S = \{\text{all human beings}\}$, R_{sibling} can easily be checked to be an equivalence relation¹. Similarly $R_{\text{mod } 2}$ is an equivalence relation on \mathbb{Z} .

Why are equivalence relations useful? They help ‘partition’ the set into ‘equivalence classes’, and help to break down the big set into smaller subsets, such that

¹Here we accept that a person is one’s own sibling.

all the elements in each subset are related to each other, and hence ‘equivalent’ in some way. For example, R_{sibling} enables one to partition the set of human beings into equivalence classes consisting of groups of brothers/sisters. On the other hand, $R_{\text{mod } 2}$ partitions \mathbb{Z} into the sets {even integers} and {odd integers}.

Definition 4.3 (Equivalence class).

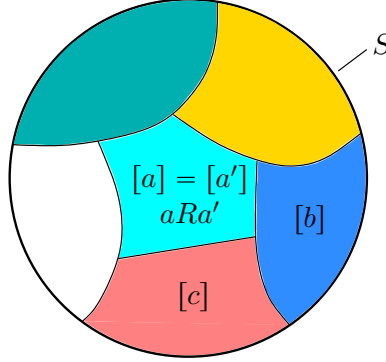
If R is an equivalence relation of a set S , then the *equivalence class of a* , denoted by $[a]$, is defined to be the set $[a] = \{b \in S : aRb\}$.

Given any $a, b \in S$, either $[a] = [b]$ or $[a] \cap [b] = \emptyset$. Indeed, let $[a] \cap [b] \neq \emptyset$. Suppose $c \in [a] \cap [b]$, that is, aRc and bRc . By symmetry, cRb . As aRc and cRb , by transitivity, we obtain aRb , and again by symmetry, bRa . If $d \in [a]$, then aRd . As bRa and aRd , by transitivity, bRd . So $d \in [b]$ too. So we have shown that $[a] \subset [b]$. In the same way, one can show $[b] \subset [a]$ as well. So $[a] = [b]$.

Clearly, $\bigcup_{a \in S} [a] \subset S$. For $a \in S$, aRa (reflexivity), and so $a \in [a]$. Thus $S \subset \bigcup_{a \in S} [a]$. So

$$S = \bigcup_{a \in S} [a].$$

As any two distinct equivalence classes do not overlap at all, it follows that S is partitioned into equivalence classes by R , as shown in the schematic picture below.



So the idea is that an equivalence relation is really an ‘attention focusing device’, where we have chosen to ignore other distinguishing features of objects which are related, and have put them together in an equivalence class. So an equivalence relation gives one a ‘pair of glasses’ through which we ‘clump together’ things which are ‘essentially the same’ (equivalent under the relation) and see them as one object! For example, if our set is the collection of children in a school bus and we consider the equivalence relation R_1 of ‘having the same sex’, then through these glasses, we see only two equivalence classes: boys and girls. On the other hand, if we consider the equivalence relation R_2 of ‘having the same age’, then through these glasses, we see groups of children sorted by age.

Exercise 4.1. Find out which of the properties of reflexivity, symmetry or transitivity are valid for each of the following relations R on the given set S , and hence determine which amongst them are equivalence relations.

- (1) $S = \mathbb{R}$, xRy if $x < y + 1$.
- (2) $S = \mathbb{Z}$, mRn if $m|n$, that is, if n is divisible by m .
- (3) $S = \mathbb{R} \setminus \{0\}$, xRy if $x/y \in \mathbb{Q}$.
- (4) $S = \mathbb{N}$, mRn if there exists an $k \in \mathbb{Z}$ such that $m/n = 2^k$.

Exercise 4.2. Let R be a relation on the set \mathbb{R} of real numbers. Viewing R as a subset of the (x, y) -plane, explain the geometric meaning of the reflexive and symmetric properties.

Exercise 4.3. Which of the following $R \subset \mathbb{R} \times \mathbb{R}$ defines an equivalence relation on \mathbb{R} ?

- (1) $R = \{(x, x) \in \mathbb{R}^2 : x \in \mathbb{R}\}$.
- (2) $R = \emptyset$.
- (3) $R = \{(x, y) \in \mathbb{R}^2 : y = 0\}$.
- (4) $R = \{(x, y) \in \mathbb{R}^2 : xy + 1 = 0\}$.
- (5) $R = \{(x, y) \in \mathbb{R}^2 : x^2y - xy^2 - x + y = 0\}$.

Exercise 4.4 (Equivalence relation induced by a map $f : X \rightarrow Y$).

Let $f : X \rightarrow Y$ be a map. Define the relation \sim on X by $x \sim x'$ if $f(x) = f(x')$.

- (1) Show that \sim is an equivalence relation.
- (2) For $y \in \text{ran } f = \{f(x) : x \in X\}$, consider the inverse image $f^{-1}\{y\} = \{x \in X : f(x) = y\}$. Show that for all $x \in X$, $[x] = f^{-1}\{f(x)\}$.
- (3) Denote the set of all equivalence classes of X by \overline{X} . Define $\overline{f} : \overline{X} \rightarrow \text{ran } f$ as follows. For any equivalence class $S \in \overline{X}$, take any $x \in S$, and set $\overline{f}(S) = f(x)$. Show that \overline{f} is *well-defined*, i.e., it does not depend on the choice of the representative selected from the equivalence class S .
- (4) Prove that $\overline{f} : \overline{X} \rightarrow \text{ran } f$ is bijective.
- (5) For the following maps, consider the corresponding induced equivalence relation on their respective domains. Determine the equivalence classes. Sketch these in the domain.
 - (a) The complex absolute value $|\cdot| : \mathbb{C} \rightarrow \mathbb{R}$.
 - (b) $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ given by $f(x, y) = xy$.

Exercise 4.5. (*) A *Cauchy sequence in \mathbb{Q}* is a sequence $(a_n)_{n \in \mathbb{N}}$ of rational numbers such that for every rational $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that whenever $m, n > N$, we have $|a_n - a_m| < \epsilon$.

- (1) Show that $(1/n)_{n \in \mathbb{N}}$ is a Cauchy sequence in \mathbb{Q} .
- (2) Show that every sequence $(a_n)_{n \in \mathbb{N}}$ of rational numbers that is convergent² is a Cauchy sequence in \mathbb{Q} .
- (3) Show that a sequence that is a Cauchy sequence in \mathbb{Q} need not be convergent with its limit belonging to \mathbb{Q} , by considering the sequence

$$0.1, 0.101, 0.101001, 0.1010010001, 0.101001000100001, \dots$$

²Right now we just use the notion of convergence in \mathbb{R} of a sequence of real numbers. We will later revisit this exercise to *construct* the real numbers using rational numbers – then we will be more careful, and explain what we mean by convergence in \mathbb{Q} of a sequence of rational numbers.

- (4) Let \mathcal{C} denote the set of all Cauchy sequences in \mathbb{Q} . Define the relation \sim on \mathcal{C} by $(a_n)_{n \in \mathbb{N}} \sim (b_n)_{n \in \mathbb{N}}$ if the sequence $(a_n - b_n)_{n \in \mathbb{N}}$ converges to 0. Show that \sim is an equivalence relation. The set of all equivalence classes of \mathcal{C} under \sim is denoted by \mathcal{C}/\sim .
- (5) Given a $q \in \mathbb{Q}$, the constant sequence $(q)_{n \in \mathbb{N}}$ is clearly a Cauchy sequence in \mathbb{Q} . Show that the map $\mathbb{Q} \ni q \mapsto [(q)_{n \in \mathbb{N}}] \in \mathcal{C}/\sim$ is injective.

(We will return to this equivalence relation when we construct \mathbb{R} from \mathbb{Q} : The set of real numbers will be \mathcal{C}/\sim . The last part of this exercise shows that \mathbb{Q} can be considered to be contained in \mathbb{R} . We will clarify later how the arithmetic operations are defined in \mathcal{C}/\sim .)

4.2. Natural numbers

We are familiar with natural numbers from an early age when we learn to count. The basic idea is that for each natural number, there is ‘next one’ or a ‘successor’ (which is used to label the next object in the set we are counting). Thus the natural numbers are built, intuitively speaking, as a sequence of objects, starting with 1 and then taking ‘successive successors’. The Italian mathematician Giuseppe Peano (1858-1932) formulated axioms for the natural numbers in 1889 which reflect the above intuition. The familiar properties of the natural numbers can then be proved as theorems.

The *Peano axioms* for a set \mathbb{N} and a (successor) function $\sigma : \mathbb{N} \rightarrow \mathbb{N}$ state that

- (N1) The set \mathbb{N} contains an element 1 such that for every $n \in \mathbb{N}$, $\sigma(n) \neq 1$.
- (N2) The map σ is injective.
- (N3) (*Induction axiom*) Suppose that a subset $S \subset \mathbb{N}$ has the properties that
 - $1 \in S$, and
 - if $n \in S$, then $\sigma(n) \in S$.

Then $S = \mathbb{N}$.

The element $\sigma(n)$ is called the *successor* of $n \in \mathbb{N}$. The property (N3) is the induction property, and forms the basis for why induction works.

Remark 4.1. (*) The existence/construction of such a set \mathbb{N} and a successor function σ that satisfies the Peano axioms relies on the axioms of set theory. For example, a possible definition is

$$\begin{aligned} 1 &:= \{\emptyset\}, \\ 2 &:= \{\emptyset, \{\emptyset\}\}, \\ 3 &:= \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \end{aligned}$$

and so on. The successor of $n \in \mathbb{N}$ is $\sigma(n) = n \cup \{n\}$. To talk about the set of *all* natural numbers, one requires the so-called ‘axiom of infinity’ from set theory. We will not get into these matters here. Instead, we simply take for granted the existence of a set \mathbb{N} and a successor function σ that satisfies the Peano axioms. From this starting point, in the rest of the section, we will prove the familiar arithmetic properties of the natural numbers, which will in turn be used in order to construct the integers, and also to derive similar properties for the integers. *

Let us show that every natural number not equal to 1 has a ‘predecessor’.

Theorem 4.1. *If $n \in \mathbb{N} \setminus \{1\}$, then there exists an $m \in \mathbb{N}$ such that $\sigma(m) = n$.*

Proof. Consider the set $S = \{1\} \cup \{k \in \mathbb{N} : \exists m \in \mathbb{N} \text{ such that } \sigma(m) = k\}$. Since 1 is not a successor of any natural number, $1 \notin \{k \in \mathbb{N} : \exists m \in \mathbb{N} \text{ such that } \sigma(m) = k\}$. Clearly $1 \in S$. If $n \in S \subset \mathbb{N}$, then $\sigma(n) \in \{k \in \mathbb{N} : \exists m \in \mathbb{N} \text{ such that } \sigma(m) = k\} \subset S$. By the induction axiom (N3), $S = \mathbb{N}$. Hence if $n \in \mathbb{N} \setminus \{1\} = S \setminus \{1\}$, then there exists an $m \in \mathbb{N}$ such that $\sigma(m) = n$. \square

Exercise 4.6. Consider the set $S = \{n \in \mathbb{N} : \sigma(n) \neq n\}$. Using Peano’s axioms, show that $S = \mathbb{N}$. Conclude that for all $n \in \mathbb{N}$, $\sigma(n) \neq n$.

Remark 4.2 (Recursive definitions).

The Peano axioms allow us to make ‘recursive/inductive definitions’. This means that we can define a sequence of objects C_n indexed by the natural numbers, via

$$(*) \left\{ \begin{array}{l} \bullet \text{ specifying the initial value } C_1, \\ \bullet \text{ giving a rule for determining } C_{\sigma(n)} \text{ from } C_n \\ \quad \text{(so that each object is defined using the preceding one).} \end{array} \right.$$

It is intuitively clear that $(*)$ defines the sequence C_1, C_2, C_3, \dots uniquely, though proving this using the Peano axioms is cumbersome. A natural approach to show this would be to define S to be the set of n such that $(*)$ determines C_k for $k \leq n$. Then $1 \in S$. Also whenever $n \in S$, also $\sigma(n) \in S$. So (N3) would yield $S = \mathbb{N}$. The problem is the usage of \leq , which has not been defined yet, and so is inadmissible in the above argument. As mentioned earlier, we will not see the proof of the recursive definition theorem here, but can be found e.g. in [C]. \ast

Addition and multiplication. Using Peano’s axioms, we can define addition and multiplication in \mathbb{N} . For each $n \in \mathbb{N}$, we define

$$n + 1 = \sigma(n).$$

Instead of the notation $\sigma(n)$, we often simply use $n + 1$ instead. If we assume that $n + m$ has been defined, we define $n + (m + 1)$ by setting

$$n + (m + 1) = (n + m) + 1.$$

By the induction axiom (N3), this defines $n + m$ for all $m \in \mathbb{N}$.

Example 4.1 (What is $2 + 2$?).

We have $2 + 2 = 2 + (1 + 1) = (2 + 1) + 1 = 3 + 1 = 4$. Thus, $2 + 2$ is by definition

the successor of $2 + 1$,

that is, the successor of the successor of 2,

that is, the successor of 3, which we call 4. \diamond

Remark 4.3 (Principle of Induction). Let $P(1), P(2), P(3), \dots$ be a sequence of statements, one for each $n \in \mathbb{N}$. Suppose $P(1)$ is true, and whenever $P(n)$ is true, also $P(n + 1)$ is true. Let $S := \{n \in \mathbb{N} : P(n) \text{ is true}\}$. Thus $1 \in S$, and whenever $n \in S$, also $\sigma(n) \in S$. By (N3), $S = \mathbb{N}$. So $P(n)$ is true for all $n \in \mathbb{N}$. \ast

Similarly, if $n \in \mathbb{N}$, then we define $n \times 1$ by

$$n \times 1 = n.$$

If we assume that $n \times m$ has been defined, then we define $n \times (m + 1)$ by

$$n \times (m + 1) = (n \times m) + n.$$

By the induction axiom, this defines $n \times m$ for all $m \in \mathbb{N}$.

Exercise 4.7. Determine 2×2 using the above definition.

We often skip writing \times , and so we use the notation $m n$ instead of $m \times n$ for natural numbers $m, n \in \mathbb{N}$.

Peano playing. Using the Peano axioms and the above definitions of addition and multiplication, the usual rules of arithmetic, namely the commutativity and associativity of addition and of multiplication, and the distributive law can be proved. As examples, we give two verifications.

Theorem 4.2 (Associativity of $+$). *For all $n, m, \ell \in \mathbb{N}$, $(n + m) + \ell = n + (m + \ell)$.*

Proof. Let $S = \{\ell \in \mathbb{N} : (n + m) + \ell = n + (m + \ell) \text{ for all } n, m \in \mathbb{N}\}$. Then $1 \in S$ because $(n + m) + 1 = n + (m + 1)$ by the definition of $n + (m + 1)$. If $\ell \in S$, then

$$\begin{aligned} (n + m) + (\ell + 1) &= ((n + m) + \ell) + 1 && \text{(definition of addition)} \\ &= (n + (m + \ell)) + 1 && (\ell \in S) \\ &= n + ((m + \ell) + 1) && \text{(definition of addition)} \\ &= n + (m + (\ell + 1)) && \text{(definition of addition)} \end{aligned}$$

for all $m, n \in \mathbb{N}$. Thus $\sigma(\ell) \in S$. By the induction axiom (N3), $S = \mathbb{N}$, that is, for all $\ell \in \mathbb{N}$, $(n + m) + \ell = n + (m + \ell)$ for all $n, m \in \mathbb{N}$. \square

Exercise 4.8. $(*)$ (Commutativity of $+$).

- (1) By considering the set $S = \{n \in \mathbb{N} : n + 1 = 1 + n\}$ and using (N3), show that $S = \mathbb{N}$.
- (2) Consider $S' = \{m \in \mathbb{N} : \text{for all } n \in \mathbb{N}, n + m = m + n\}$. Show that $S' = \mathbb{N}$.

Exercise 4.9. Show that for all $n \in \mathbb{N}$, $1 \times n = n$.

Theorem 4.3 (Commutativity of \times). *For all $n, m \in \mathbb{N}$, $n \times m = m \times n$.*

Proof.

1° We first show $(m + 1) \times n = (m \times n) + n$ for all $m, n \in \mathbb{N}$. To this end, we define $S = \{n \in \mathbb{N} : (m + 1) \times n = (m \times n) + n \text{ for all } m \in \mathbb{N}\}$. Then $1 \in S$ because $(m + 1) \times 1 = m + 1 = (m \times 1) + 1$. If $n \in S$, then for $m \in \mathbb{N}$

$$\begin{aligned} (m + 1) \times (n + 1) &= ((m + 1) \times n) + (m + 1) && \text{(definition of multiplication)} \\ &= ((m \times n) + n) + (m + 1) && (n \in S) \\ &= ((m \times n) + m) + (n + 1) && \text{(various laws for addition)} \\ &= (m \times (n + 1)) + (n + 1) && \text{(definition of multiplication)} \end{aligned}$$

By (N3), $S = \mathbb{N}$, i.e., for all $n \in \mathbb{N}$, $(m + 1) \times n = (m \times n) + n$ for all $m \in \mathbb{N}$.

2° By Exercise 4.9, $1 \times n = n$ for all $n \in \mathbb{N}$. So $1 \times n = n = n \times 1$ for all $n \in \mathbb{N}$.

3° Let $S' = \{m \in \mathbb{N} : m \times n = n \times m \text{ for all } n \in \mathbb{N}\}$. By 2°, $1 \in S'$. If $m \in S'$, then

$$\begin{aligned} n \times (m + 1) &= (n \times m) + n \quad (\text{definition of multiplication}) \\ &= (m \times n) + n \quad (m \in S') \\ &= (m + 1) \times n \quad (\text{by } 1^\circ) \end{aligned}$$

for all $n \in \mathbb{N}$. By (N3), $S' = \mathbb{N}$. So for all $m, n \in \mathbb{N}$, $m \times n = n \times m$. \square

Exercise 4.10. (*) (Distributive law). Show that $m \times (n + \ell) = (m \times n) + (m \times \ell)$ for all $\ell, m, n \in \mathbb{N}$. *Hint:* Consider $S = \{\ell \in \mathbb{N} : m \times (n + \ell) = (m \times n) + (m \times \ell) \text{ for all } m, n \in \mathbb{N}\}$. Show that $1 \in S$, and that if $\ell \in S$, then also $\sigma(\ell) \in S$.

(By the commutativity of \times , also $(n + \ell) \times m = (n \times m) + (\ell \times m)$ for all $\ell, m, n \in \mathbb{N}$.)

Exercise 4.11. (*) (Associativity of \times). Show that $(m \times n) \times \ell = m \times (n \times \ell)$ for all $\ell, m, n \in \mathbb{N}$. *Hint:* Consider $S = \{\ell \in \mathbb{N} : (m \times n) \times \ell = m \times (n \times \ell) \text{ for all } m, n \in \mathbb{N}\}$.

Example 4.2 (No ‘additive identity’ in \mathbb{N} , that is, no ‘zero’ in \mathbb{N}).

We show that there does not exist an $m \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, $n = n + m$. Suppose, on the contrary, that there exists such an m . If we take $n = 1$, then $1 = n + 1 = \sigma(n)$, a contradiction to (N1). \diamond

Example 4.3 (Order relation $<$ on \mathbb{N}).

For $k, K \in \mathbb{N}$, we say that $k < K$ if there exists an $m \in \mathbb{N}$ such that $K = k + m$. The notation $K > k$ means $k < K$.

(a) For instance $1 < 2$, because we can write $2 = 1 + m$, with $m = 1 \in \mathbb{N}$.

In fact, for all $n \in \mathbb{N}$, $n < \sigma(n)$ since $\sigma(n) = n + 1$.

(b) The equation $1 = 2 + m$ does not have a solution in the unknown $m \in \mathbb{N}$.

Otherwise $1 = (1 + 1) + m = 1 + (m + 1) = \sigma(m + 1)$, a contradiction to (N1). \diamond

Examples 4.2 and 4.3(b), show arithmetic ‘flaws’ with the natural numbers, and this will be remedied by the integers.

Exercise 4.12 (Transitivity of $<$).

Suppose that $m, n, k \in \mathbb{N}$ are such that $m < n$ and $n < k$. Prove that $m < k$.

Exercise 4.13. Let $m, n \in \mathbb{N}$ be such that $m < n$. Show that for all $k \in \mathbb{N}$, $m + k < n + k$ and $m \times k < n \times k$.

Theorem 4.4 (Trichotomy Law).

For all $m, n \in \mathbb{N}$, one and only one of the following three statements holds:

- 1° $m = n$.
- 2° $m < n$.
- 3° $m > n$.

To show this we will use the following generalisation of Exercise 4.6.

Lemma 4.5. For $m, n \in \mathbb{N}$, $m + n \neq n$.

Proof. Fix $m \in \mathbb{N}$, and define $S = \{n \in \mathbb{N} : m+n \neq n\}$. Then $1 \in S$, because $m+1 = \sigma(m) \neq 1$ by (N1). Let $n \in \mathbb{N}$, i.e., $m+n \neq n$. By (N2), $\sigma(m+n) \neq \sigma(n)$. Then $m + \sigma(n) = \sigma(m+n) \neq \sigma(n)$. Hence whenever $n \in S$, we have $\sigma(n) \in S$ as well. By (N3) it follows that $S = \mathbb{N}$. \square

Proof. (of the Trichotomy Law). By Lemma 4.5, the cases 1° and 2° are mutually exclusive. Also, 1° and 3° are mutually exclusive. Finally, 2° and 3° are also mutually exclusive, since otherwise we have $n = m+k$ (by $m < n$) and $m = n+\ell$ (by $m > n$), giving $n = m+k = (n+\ell)+k = n+(\ell+k)$, contradicting Lemma 4.5. Now fix $m \in \mathbb{N}$. Define $S = \{n \in \mathbb{N} : \text{one (and only one) of the cases } 1^\circ, 2^\circ, 3^\circ \text{ holds}\}$. We claim that $1 \in S$. Indeed, if $m = 1$, then this is true since 1° holds. If $m \neq 1$, then it has a predecessor $\ell \in \mathbb{N}$ by Theorem 4.1: $m = \sigma(\ell) = \ell + 1 = 1 + \ell$, and so the statement 3° holds for $n = 1$. Thus $1 \in S$.

Now suppose $n \in S$, that is, one and only one of the cases $1^\circ, 2^\circ, 3^\circ$ holds.

We want to show that $\sigma(n) \in S$. We consider the three cases separately.

1° $m = n$. Then $\sigma(n) = \sigma(m) = m + 1$, showing that 2° holds for $\sigma(n)$.

2° $m < n$. Then $n = m + k$ for a $k \in \mathbb{N}$. So $\sigma(n) = \sigma(m+k) = (m+k)+1 = m+(k+1)$, showing 2° holds for $\sigma(n)$.

3° $m > n$. Then $m = n + k$.

If $k = 1$, then $\sigma(n) = n + 1 = m$, and so we have case 1° for $\sigma(n)$.

If $k \neq 1$, then it has a predecessor: $k = \sigma(\ell)$ for an $\ell \in \mathbb{N}$. Then we have that $m = n + k = n + \sigma(\ell) = n + (\ell + 1) = (n + 1) + \ell = \sigma(n) + \ell$, and so we have case 3° for $\sigma(n)$.

Hence whenever $n \in S$, $\sigma(n) \in S$ too. By (N3), $S = \mathbb{N}$. As $m \in \mathbb{N}$ was arbitrary, we have shown the trichotomy law for all $m, n \in \mathbb{N}$. \square

The following exercise justifies calling $\sigma(n)$ the successor of $n \in \mathbb{N}$.

Exercise 4.14. For any $n \in \mathbb{N}$, there does not exist an $m \in \mathbb{N}$ such that $n < m < n + 1$. *Hint:* Argue by contradiction. Write $m = n + k$. Consider the cases $k = 1$ and $k \neq 1$.

Exercise 4.15 (Additive and multiplicative cancellation rules). Let $m, n, k \in \mathbb{N}$.

(1) Show that if $m + k = n + k$, then $m = n$.

(2) Show that if $m \times k = n \times k$, then $m = n$.

Hints: If $m \neq n$, what does the trichotomy law allow? Use Exercise 4.13.

If $m, n \in \mathbb{N}$, and either $m = n$ or $m < n$, then we write $m \leq n$. The symbol \geq is defined similarly: $m \geq n$ if $m = n$ or $m > n$.

Exercise 4.16. Show that for all $n \in \mathbb{N}$, $1 \leq n$.

Exercise 4.17. Let $m, n \in \mathbb{N}$ be such that $m < n$. Prove that $m + 1 \leq n$.

Definition 4.4. Let S be a nonempty subset of \mathbb{N} . An element $\ell \in S$ is called a *least element* of S if for all $n \in S$, $\ell \leq n$. An element $u \in S$ is called a *greatest element* of S if for all $n \in \mathbb{N}$, $n \leq u$.

Exercise 4.18. Show that \mathbb{N} has no greatest element.

Hint: If u is a greatest element, consider $\sigma(u)$ and use Exercise 4.6.

Theorem 4.6 (Well-Ordering principle).

Any nonempty subset of \mathbb{N} has a least element.

Proof. Let S be a nonempty subset of \mathbb{N} . Define

$$\Lambda = \{\ell \in \mathbb{N} : \text{for all } n \in S, \ell \leq n\}.$$

Then $1 \in \Lambda$ by Exercise 4.16. Also, $\Lambda \neq \mathbb{N}$: Indeed if $n \in S$, then $\sigma(n) > n$, and so by the trichotomy law, $\neg(\sigma(n) \leq n)$, showing that $\sigma(n) \notin \Lambda$.

It follows that there exists a $\ell_* \in \Lambda$ with $\sigma(\ell_*) \notin \Lambda$ (otherwise by (N3), $\Lambda = \mathbb{N}$). We claim that $\ell_* \in S$. Suppose $\ell_* \notin S$. We know that $\ell_* \leq n$ for all $n \in S$. But as ℓ_* can't equal any $n \in S$, the trichotomy law shows $\ell_* < n$ for all $n \in S$. By Exercise 4.17, $\ell_* + 1 \leq n$ for all $n \in S$. But then $\sigma(\ell_*) = \ell_* + 1 \in \Lambda$, a contradiction. Now $\ell_* \in \Lambda$ and $\ell_* \in S$ means that ℓ_* is a least element of S . \square

Exercise 4.19. Let S be a nonempty subset of \mathbb{N} . Show that the least element of S (which exists, by the Well-Ordering Principle) is unique. The least element of S is called its *minimum*, denoted by $\min S$. (Similarly a greatest element, if it exists, is unique, and it is called the *maximum* of S , and is denoted by $\max S$.)

4.3. Integers

By the definition of $<$, given $m, n \in \mathbb{N}$, the equation $m + x = n$ is solvable in the unknown $x \in \mathbb{N}$ if and only if $m < n$. By the Trichotomy Law, it follows that that if $m \geq n$, then the equation $m + x = n$ is *not* solvable for $x \in \mathbb{N}$. Now the aim is to 'extend' the natural number system to a 'bigger' system, the integers \mathbb{Z} , with an extension of the operation $+$, which allows a unique solution to such an equation for *arbitrary* $m, n \in \mathbb{N}$. Intuitively, we just think of the symbol $n - m$ as being the integer solution x . As the ordered pair (m, n) determines this x , we could think of x as a pair (m, n) , where $m, n \in \mathbb{N}$. But then we realise that the equation $m + x = n$ is equivalent to $(m + 1) + x = n + 1$, and so the integer x ought to be also the pair $(m + 1, n + 1)$. So we must identify the pairs (m, n) and $(m + 1, n + 1)$. In fact, we need to identify all pairs $(m, n), (m + 1, n + 1), (m + 2, n + 2), \dots$, and think of it as being the integer x . To formalise this, we introduce an equivalence relation \sim on $\mathbb{N} \times \mathbb{N}$, the equivalence classes of which will then be integers. We should deem the pairs $(k, \ell), (m, n)$ as equivalent, that is, $(k, \ell) \sim (m, n)$, if they yield the same solution x , and so we should have $\ell - k = n - m$, or rearranging, $k + n = \ell + m$. This prompts defining the relation \sim by $(k, \ell) \sim (m, n)$ if $k + n = \ell + m$. This is a relation, and we check that it is an *equivalence* relation below.

Theorem 4.7. Let \sim be the relation on $\mathbb{N} \times \mathbb{N}$ defined as follows: For $k, \ell, m, n \in \mathbb{N}$, $(k, \ell) \sim (m, n)$ if $k + n = \ell + m$. Then \sim is an equivalence relation.

Proof. We check that \sim is reflexive, symmetric, and transitive.

Reflexivity: Let $(m, n) \in \mathbb{N} \times \mathbb{N}$. Then $(m, n) \sim (m, n)$ because $m + n = n + m$.

Symmetry: Suppose $(k, \ell), (m, n) \in \mathbb{N} \times \mathbb{N}$, and $(k, \ell) \sim (m, n)$. Then $k + n = \ell + m$. Thus also $m + \ell = n + k$, showing that $(m, n) \sim (k, \ell)$.

Transitivity: Let $(k, \ell), (m, n), (p, q) \in \mathbb{N} \times \mathbb{N}$, $(k, \ell) \sim (m, n)$, and $(m, n) \sim (p, q)$. Then $k + n = \ell + m$ and $m + q = n + p$. Hence $(k + n) + (m + q) = (\ell + m) + (n + p)$, and so $(k + q) + (m + n) = (\ell + p) + (m + n)$. By the additive cancellation rule (Exercise 4.15), we conclude that $k + q = \ell + p$, and so $(k, \ell) \sim (p, q)$. \square

The equivalence relation partitions $\mathbb{N} \times \mathbb{N}$ into equivalence classes. The equivalence class of $(m, n) \in \mathbb{N} \times \mathbb{N}$ is

$$[(m, n)] = \{(k, \ell) \in \mathbb{N} \times \mathbb{N} : (k, \ell) \sim (m, n)\} = \{(k, \ell) \in \mathbb{N} \times \mathbb{N} : k + n = \ell + m\}.$$

Definition 4.5 (\mathbb{Z} , the set of integers).

The *set \mathbb{Z} of integers* is the set of equivalence classes of $\mathbb{N} \times \mathbb{N}$ under the equivalence relation described in Theorem 4.7.

The symbol \mathbb{Z} comes from ‘Zahlen’, meaning ‘numbers’ in German.

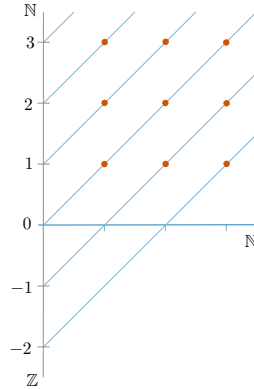
Example 4.4. Intuitively $[(1, 1)]$ represents the integer ‘ $(1 - 1 =) 0$ ’. We have

$$\begin{aligned} [(1, 1)] &= \{(m, n) \in \mathbb{N} \times \mathbb{N} : m + 1 = n + 1\} = \{(m, n) \in \mathbb{N} \times \mathbb{N} : m = n\} \\ &= \{(m, m) : m \in \mathbb{N}\} = \{(1, 1), (2, 2), (3, 3), \dots\}. \end{aligned}$$

Similarly, we think of $[(2, 1)]$ as representing the integer ‘ $(1 - 2 =) -1$ ’. We have

$$\begin{aligned} [(2, 1)] &= \{(m, n) \in \mathbb{N} \times \mathbb{N} : m + 1 = n + 2\} = \{(m, n) \in \mathbb{N} \times \mathbb{N} : m = n + 1\} \\ &= \{(n + 1, n) : n \in \mathbb{N}\} = \{(2, 1), (3, 2), (4, 3), \dots\}. \end{aligned}$$

The following picture depicts the equivalence classes visually. \diamond



Exercise 4.20 ($\mathbb{N} \subset \mathbb{Z}$). Show that the map $\mathbb{N} \ni n \mapsto [(1, \sigma(n))] \in \mathbb{Z}$ is injective.

Addition of integers. We know that we can think of the integer $[(m, n)]$ as $n - m$ in the notation from elementary school. So to add $[(k, \ell)]$ and $[(m, n)]$, we ought to get $(\ell - k) + (n - m) = (\ell + n) - (k + m)$, which is the integer $[(k + m, \ell + n)]$. This motivates the following definition.

Definition 4.6 (Addition in \mathbb{Z}).

The *sum* $\mathbf{a} + \mathbf{b}$ of $\mathbf{a} = [(k, \ell)]$ and $\mathbf{b} = [(m, n)] \in \mathbb{Z}$, is defined to be the integer $\mathbf{a} + \mathbf{b} = [(k, \ell)] + [(m, n)] := [(k + m, \ell + n)]$.

We need to check that the above notion is well-defined. What does this mean? Firstly, an integer $\mathbf{a} \in \mathbb{Z}$ is an equivalence class, and so it is a set with many members, and when we write $\mathbf{a} = [(k, \ell)]$, we have just picked *one* member (k, ℓ) belonging to the equivalence class \mathbf{a} . Then we know that the equivalence class $[(k, \ell)]$ corresponding to this (k, ℓ) is equal to \mathbf{a} . Secondly, after picking such representatives (k, ℓ) and (m, n) for \mathbf{a} and \mathbf{b} , we are defining the sum $\mathbf{a} + \mathbf{b}$ by taking the equivalence class of $(k + m, \ell + n)$. But if we had chosen different representatives, say $(k', \ell') \in \mathbf{a}$, and $(m', n') \in \mathbf{b}$, do we get the same integer? Thus we ask: Is $[(k' + m', \ell' + n')] = [(k + m, \ell + n)]$? This is the question of well-defined-ness. We now check that the answer is ‘yes’.

Theorem 4.8 (Addition is well-defined).

If $k, \ell, m, n, k', \ell', m', n' \in \mathbb{N}$ are such that $(k, \ell) \sim (k', \ell')$ and $(m, n) \sim (m', n')$, then $(k + m, \ell + n) \sim (k' + m', \ell' + n')$. (And so $[(k + m, \ell + n)] = [(k' + m', \ell' + n')]$.)

Proof. As $(k, \ell) \sim (k', \ell')$ and $(m, n) \sim (m', n')$, we know that

$$\begin{aligned} k + \ell' &= \ell + k', \\ m + n' &= n + m'. \end{aligned}$$

Adding these we get $(k + \ell') + (m + n') = (\ell + k') + (n + m')$. Using associativity and commutativity of natural number addition, $(k + m) + (\ell' + n') = (\ell + n) + (k' + m')$. Hence $(k + m, \ell + n) \sim (k' + m', \ell' + n')$. \square

Ideally, we should use a different symbol for addition, since we already used $+$ for the addition of natural numbers, but the following result shows that the new addition is in fact an extension of the old addition. Recall that by putting on our ‘integer glasses’, the natural number $n \in \mathbb{N}$ is the integer $[(1, \sigma(n))] \in \mathbb{Z}$.

Theorem 4.9. If $m, n \in \mathbb{N}$, then $[(1, \sigma(m))] + [(1, \sigma(n))] = [(1, \sigma(m + n))]$.

$$\begin{array}{ccc} \mathbb{N} \times \mathbb{N} & \xrightarrow{+} & \mathbb{N} \\ \downarrow \text{glasses} & & \downarrow \text{glasses} \\ \mathbb{Z} \times \mathbb{Z} & \xrightarrow{+} & \mathbb{Z} \end{array} \qquad \begin{array}{ccc} (m, n) & \xrightarrow{+} & m + n \\ \downarrow \text{glasses} & & \downarrow \text{glasses} \\ ([1, \sigma(m)], [1, \sigma(n)]) & \mapsto & [(1 + 1, \sigma(m) + \sigma(n))] \\ & & = [(1, \sigma(m + n))] \end{array}$$

Proof. To show that $[(1 + 1, \sigma(m) + \sigma(n))] = [(1, \sigma(m + n))]$, we must show $(1 + 1, \sigma(m) + \sigma(n)) \sim (1, \sigma(m + n))$. But this is clear using the commutativity and associativity of natural number addition:

$$\begin{aligned} (1 + 1) + \sigma(m + n) &= (1 + 1) + ((m + n) + 1) = ((m + 1) + (n + 1)) + 1 \\ &= (\sigma(m) + \sigma(n)) + 1. \end{aligned} \quad \square$$

Example 4.5 (**0** and additive inverses).

We define *zero* to be the integer $\mathbf{0} := [(1, 1)] \in \mathbb{Z}$. For any integer $\mathbf{a} = [(m, n)] \in \mathbb{Z}$,

$$\mathbf{a} + \mathbf{0} = [(m, n)] + [(1, 1)] = [(m + 1, n + 1)] \stackrel{(*)}{=} [(m, n)] = \mathbf{a}.$$

Justification of $(*)$: $(m + 1, n + 1) \sim (m, n)$ because $(m + 1) + n = (n + 1) + m$. In Exercise 4.22, we'll show that addition in \mathbb{Z} is commutative, using which it follows that $\mathbf{0} + \mathbf{a} = \mathbf{a} + \mathbf{0} = \mathbf{a}$.

For an integer $\mathbf{a} = [(m, n)] \in \mathbb{Z}$, we define its *additive inverse* to be the³ integer $-\mathbf{a} := [(n, m)]$. Then we have

$$\mathbf{a} + (-\mathbf{a}) = [(m, n)] + [(n, m)] = [(m + n, n + m)] = [(1, 1)] = \mathbf{0}.$$

In particular, if $n \in \mathbb{N}$, and we view this as the integer $[(1, \sigma(n))]$, then its additive inverse in \mathbb{Z} is the integer $[(\sigma(n), 1)]$. \diamond

Exercise 4.21. Prove that addition in \mathbb{Z} is associative.

Exercise 4.22. Prove that addition in \mathbb{Z} is commutative.

We now show that result that prompted the construction of \mathbb{Z} .

Theorem 4.10. *Let $\mathbf{a}, \mathbf{b} \in \mathbb{Z}$. Then there exists a unique solution $\mathbf{x} \in \mathbb{Z}$ to $\mathbf{a} + \mathbf{x} = \mathbf{b}$.*

Proof. Set $\mathbf{x} := \mathbf{b} + (-\mathbf{a}) \in \mathbb{Z}$. Then using the commutativity and associativity of addition in \mathbb{Z} , we obtain $\mathbf{a} + \mathbf{x} = \mathbf{a} + (\mathbf{b} + (-\mathbf{a})) = \mathbf{b} + (\mathbf{a} + (-\mathbf{a})) = \mathbf{b} + \mathbf{0} = \mathbf{b}$.

To show uniqueness, suppose that $\mathbf{x}, \mathbf{x}' \in \mathbb{Z}$ are such that $\mathbf{a} + \mathbf{x} = \mathbf{b} = \mathbf{a} + \mathbf{x}'$. Adding $-\mathbf{a}$ to both sides yields $-\mathbf{a} + (\mathbf{a} + \mathbf{x}) = -\mathbf{a} + (\mathbf{a} + \mathbf{x}')$, and thanks to associativity, $(-\mathbf{a} + \mathbf{a}) + \mathbf{x} = (-\mathbf{a} + \mathbf{a}) + \mathbf{x}'$, i.e., $\mathbf{0} + \mathbf{x} = \mathbf{0} + \mathbf{x}'$. Thus $\mathbf{x} = \mathbf{x}'$. \square

Theorem 4.11. *For any integer $\mathbf{a} \in \mathbb{Z}$, one and exactly one of the following hold:*

1° $\mathbf{a} = [(1, \sigma(k))]$ for a $k \in \mathbb{N}$. (Integer corresponding to the natural number $k \in \mathbb{N}$.)

2° $\mathbf{a} = \mathbf{0}$.

3° $\mathbf{a} = [(\sigma(k), 1)] = -[(1, \sigma(k))]$ for a $k \in \mathbb{N}$.

(Additive inverse of the integer corresponding to the natural number $k \in \mathbb{N}$.)

³This is well-defined, since if $(m, n) \sim (m', n')$, then $(n, m) \sim (n', m')$.

Proof. 1° and 2° are mutually exclusive, since otherwise $(1, \sigma(k)) \sim (1, 1)$, giving $1 + 1 = \sigma(k) + 1$, so that $\sigma(k) = 1$, contradicting (N1). Similarly 2° and 3° are mutually exclusive. Finally 1° and 3° are mutually exclusive, because otherwise $(\sigma(k'), 1) \sim (1, \sigma(k))$ for some $k, k' \in \mathbb{N}$, i.e., $1 + k' + 1 + k = 1 + 1$, so that $\sigma(k + k') = 1$, which is again a contradiction to (N1).

Let $\mathbf{a} = [(m, n)] \in \mathbb{Z}$. By the Trichotomy Law for \mathbb{N} , one of the following hold:

1° $m < n$. Then $n = m + k$ for some $k \in \mathbb{N}$. If $m = 1$, then we have $\mathbf{a} = [(1, \sigma(k))]$.

If $m \neq 1$, then m has a predecessor, say $\ell \in \mathbb{N}$, that is, $m = \sigma(\ell) = \ell + 1$, and so $\mathbf{a} = [(\ell + 1, \ell + 1 + k)] = [(1, \sigma(k))]$.

2° $m = n$. Then $\mathbf{a} = \mathbf{0}$.

3° $m > n$. Just as in 1° , if $m = n + k$, then we have $\mathbf{a} = [(\sigma(k), 1)]$. □

So the set \mathbb{Z} is partitioned into three mutually disjoint subsets:

- the ‘positive integers’ (integers corresponding to the natural numbers),
- zero, and
- the ‘negative integers’ (additive inverses of the natural numbers).

Review the picture on page 94.

Multiplication of integers. To define multiplication formally, again we note that we expect the product of $[(m, n)]$ and $[(k, \ell)]$ to be the integer (in elementary school notation) $(\ell - k)(n - m) = \ell n + km - kn - \ell m$, that is, $[(kn + \ell m, km + \ell n)]$. This motivates the following:

Definition 4.7 (Multiplication in \mathbb{Z}).

The *product* $\mathbf{a} \cdot \mathbf{b}$ of $\mathbf{a} = [(k, \ell)]$ and $\mathbf{b} = [(m, n)] \in \mathbb{Z}$ is defined to be the integer $\mathbf{a} \cdot \mathbf{b} = [(k, \ell)] \cdot [(m, n)] := [(kn + \ell m, km + \ell n)]$.

Again we need to check well-definedness.

Theorem 4.12 (Multiplication is well-defined).

If $k, \ell, m, n, k', \ell', m', n' \in \mathbb{N}$ are such that $(k, \ell) \sim (k', \ell')$ and $(m, n) \sim (m', n')$, then

$$(kn + \ell m, km + \ell n) \sim (k'n' + \ell'm', k'm' + \ell'n').$$

(Thus $[(k'n' + \ell'm', k'm' + \ell'n')] = [(kn + \ell m, km + \ell n)]$.)

Proof. As $(k, \ell) \sim (k', \ell')$ and $(m, n) \sim (m', n')$,

$$k + \ell' = \ell + k' \quad (\star)$$

$$m + n' = n + m'. \quad (\star\star)$$

We want to show $(kn + \ell m, km + \ell n) \sim (k'n' + \ell'm', k'm' + \ell'n')$, that is,

$$\textcolor{teal}{kn} + \textcolor{teal}{\ell m} + \textcolor{teal}{k'm'} + \textcolor{teal}{\ell'n'} = \textcolor{brown}{km} + \textcolor{brown}{\ell n} + \textcolor{brown}{k'n'} + \textcolor{brown}{\ell'm'}. \quad (*)$$

The terms on the left-hand side suggest considering $n \cdot (\star)$, $m \cdot (\star)$, $k' \cdot (\star\star)$, $\ell' \cdot (\star\star)$, which are, respectively:

$$\begin{aligned} kn + \ell'n &= \ell n + k'n \\ \ell m + k'm &= km + \ell'm \\ k'm' + k'n &= k'n' + k'm \\ \ell'n' + \ell'm &= \ell'm' + \ell'n. \end{aligned}$$

Adding these, we obtain

$$\left. \begin{aligned} &kn + \ell m + \ell'n' + k'm' \\ &+ (k'n + \ell'm + k'm + \ell'n) \end{aligned} \right\} = \left\{ \begin{aligned} &\ell n + km + k'n' + \ell'm' \\ &+ (k'n + \ell'm + k'm + \ell'n) \end{aligned} \right.$$

and so by the additive cancellation rule in \mathbb{N} , we obtain $(*)$. \square

Next we show that multiplication in \mathbb{Z} is an extension of the multiplication in \mathbb{N} .

Theorem 4.13. *If $m, n \in \mathbb{N}$, then $[(1, \sigma(m))] \cdot [(1, \sigma(n))] = [(1, \sigma(mn))]$.*

In the proof below, for simplicity, we have denoted multiplication in \mathbb{N} also with \cdot instead of the symbol \times used in the previous section.

Proof. We must show $[(1 \cdot \sigma(n) + \sigma(m) \cdot 1, 1 \cdot 1 + \sigma(m)\sigma(n))] = [(1, \sigma(mn))]$, that is, $(\sigma(m) + \sigma(n), 1 + \sigma(m)\sigma(n)) \sim (1, \sigma(mn))$, that is,

$$(\sigma(m) + \sigma(n)) + \sigma(mn) = (1 + \sigma(m)\sigma(n)) + 1.$$

But this is readily verified using the arithmetic rules in \mathbb{N} :

$$\begin{aligned} (\sigma(m) + \sigma(n)) + \sigma(mn) &= ((m+1) + (n+1)) + (mn+1) \\ &= mn + m \cdot 1 + 1 \cdot n + 1 \cdot 1 + 1 + 1 \\ &= m(n+1) + 1(n+1) + 1 + 1 = (m+1)(n+1) + 1 + 1 \\ &= \sigma(m)\sigma(n) + 1 + 1 = (1 + \sigma(m)\sigma(n)) + 1. \end{aligned} \quad \square$$

In accordance with what we are used to, we often skip writing \cdot to denote integer multiplication.

Exercise 4.23. For all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{Z}$, show that

- (Associativity) $\mathbf{a}(\mathbf{bc}) = (\mathbf{ab})\mathbf{c}$.
- (Commutativity) $\mathbf{ab} = \mathbf{ba}$.
- (Distributivity) $\mathbf{a}(\mathbf{b} + \mathbf{c}) = \mathbf{ab} + \mathbf{ac}$.

Example 4.6 (Multiplicative identity).

We know that for any natural number $n \in \mathbb{N}$, $1 \times n = n = n \times 1$. We now show that if we view $1 \in \mathbb{N}$ as an integer, that is, $\mathbf{1} := [(1, \sigma(1))] = [(1, 2)]$, then it continues to serve as a multiplicative identity, but now in the bigger set \mathbb{Z} . Thus we want to check that for all $\mathbf{a} = [(m, n)] \in \mathbb{Z}$, we have

$$\mathbf{a} \cdot \mathbf{1} = \mathbf{a} = \mathbf{1} \cdot \mathbf{a}.$$

We have $m \cdot 2 = m \cdot (1 + 1) = m \cdot 1 + m \cdot 1 = m + m$ and $n \cdot 2 = n + n$. Thus

$$\begin{aligned} \mathbf{a} \cdot \mathbf{1} &= [(m, n)][(1, 2)] = [(m \cdot 2 + n \cdot 1, m \cdot 1 + n \cdot 2)] \\ &= [(m + m + n, m + n + n)] = [(m, n)], \end{aligned}$$

where the last equality follows thanks to $(m + m + n, m + n + n) \sim (m, n)$. \diamond

Exercise 4.24. Let $\mathbf{a}, \mathbf{b} \in \mathbb{Z}$, and suppose that $\mathbf{a} = \mathbf{0}$ or $\mathbf{b} = \mathbf{0}$. Prove that $\mathbf{a} \cdot \mathbf{b} = \mathbf{0}$.

The converse of the result from Exercise 4.24 holds.

Theorem 4.14. Let $\mathbf{a}, \mathbf{b} \in \mathbb{Z}$ and $\mathbf{a} \cdot \mathbf{b} = \mathbf{0}$. Then $\mathbf{a} = \mathbf{0}$ or $\mathbf{b} = \mathbf{0}$.

Proof. Let $\mathbf{a} = [(k, \ell)]$ and $\mathbf{b} = [(m, n)]$. Then

$$[(1, 1)] = \mathbf{0} = \mathbf{ab} = [(k, \ell)][(m, n)] = [(kn + \ell m, km + \ell n)],$$

and so $1 + km + \ell n = 1 + kn + \ell m$, giving $km + \ell n = kn + \ell m$. By the trichotomy law, we have the following three mutually exclusive cases:

- 1° $k = \ell$. Then $\mathbf{a} = [(k, k)] = [(1, 1)] = \mathbf{0}$. (Here we used $(k, k) \sim (1, 1)$.)
- 2° $k > \ell$. Then $k = \ell + p$ for some $p \in \mathbb{N}$. Hence $km + \ell n = kn + \ell m$ gives $\ell m + pm + \ell n = \ell n + pn + \ell m$, i.e., $pm = pn$ by the additive cancellation rule. The multiplicative cancellation rule gives $m = n$. So $\mathbf{b} = [(m, m)] = [(1, 1)] = \mathbf{0}$.
- 3° $k < \ell$. Then $\ell = k + q$ for some $q \in \mathbb{N}$. Hence $km + \ell n = kn + \ell m$ gives $km + kn + qn = kn + km + qm$, i.e., $qn = qm$ by the additive cancellation rule. The multiplicative cancellation rule gives $m = n$. So $\mathbf{b} = [(m, m)] = [(1, 1)] = \mathbf{0}$. \square

Exercise 4.25. Consider the set $C[0, 1]$ of all continuous functions on $[0, 1]$. For $f, g \in C[0, 1]$, define $f \cdot g \in C[0, 1]$ by $(f \cdot g)(x) = f(x)g(x)$ for all $x \in [0, 1]$. Let $\mathbf{0} \in C[0, 1]$ be the function that is identically 0. Give an example to show that $f \cdot g = \mathbf{0}$, but neither f nor g equals $\mathbf{0}$.

Exercise 4.26 (Multiplicative cancellation rule).

Let $\mathbf{a} \in \mathbb{Z} \setminus \{\mathbf{0}\}$, $\mathbf{b}, \mathbf{c} \in \mathbb{Z}$ be such that $\mathbf{ab} = \mathbf{ac}$. Show that $\mathbf{b} = \mathbf{c}$. *Hint:* Bring to one side.

Order. We can extend the order relation from \mathbb{N} to \mathbb{Z} as follows. Given $\mathbf{a}, \mathbf{b} \in \mathbb{Z}$, we say $\mathbf{a} < \mathbf{b}$ if for some $n \in \mathbb{N}$, $\mathbf{b} = \mathbf{a} + [(1, \sigma(n))]$. We write equivalently $\mathbf{b} > \mathbf{a}$.

Exercise 4.27. Let $m, n \in \mathbb{N}$. Show that $m < n$ if and only if $[(1, \sigma(m))] < [(1, \sigma(n))]$. *Hint:* E.g. use Theorem 4.9.

Exercise 4.28 (Transitivity of $<$).

Let $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{Z}$ be such that $\mathbf{a} < \mathbf{b}$ and $\mathbf{b} < \mathbf{c}$. Show that $\mathbf{a} < \mathbf{c}$. *Hint:* Use Theorem 4.9.

Exercise 4.29. Let $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{Z}$ be such that $\mathbf{a} < \mathbf{b}$ and $\mathbf{c} > \mathbf{0}$. Show that $\mathbf{a} \cdot \mathbf{c} < \mathbf{b} \cdot \mathbf{c}$.

Exercise 4.30.

Let $\mathbf{a} \in \mathbb{Z}$ and $\mathbf{1} = [(1, 2)]$. Show that there is no $\mathbf{b} \in \mathbb{Z}$ such that $\mathbf{a} < \mathbf{b} < \mathbf{a} + \mathbf{1}$.

Theorem 4.15 (Trichotomy Law).

Let $\mathbf{a}, \mathbf{b} \in \mathbb{Z}$. Then one and exactly one of the following holds:

1° $\mathbf{a} < \mathbf{b}$.

2° $\mathbf{a} = \mathbf{b}$.

3° $\mathbf{a} > \mathbf{b}$.

Proof. By Theorem 4.10, there exists a unique $\mathbf{x} \in \mathbb{Z}$ such that $\mathbf{b} = \mathbf{a} + \mathbf{x}$. By Theorem 4.11, for the integer \mathbf{x} , there holds one and exactly one of the following three mutually exclusive options:

1°° $\mathbf{x} = [(1, \sigma(k))]$ for some $k \in \mathbb{N}$. Then $\mathbf{a} < \mathbf{b}$, that is, 1° holds.

2°° $\mathbf{x} = \mathbf{0}$. Then $\mathbf{b} = \mathbf{a}$, that is, 2° holds.

3°° $\mathbf{a} = [(\sigma(k), 1)]$ for some $k \in \mathbb{N}$. Then case 3° holds, namely $\mathbf{a} > \mathbf{b}$, because

$$\begin{aligned} \mathbf{b} + [(1, \sigma(k))] &= (\mathbf{a} + \mathbf{x}) + [(1, \sigma(k))] = \mathbf{a} + (\mathbf{x} + [(1, \sigma(k))]) \\ &= \mathbf{a} + [(\sigma(k) + 1, 1 + \sigma(k))] = \mathbf{a} + [(1, 1)] = \mathbf{a} + \mathbf{0} = \mathbf{a}. \quad \square \end{aligned}$$

Exercise 4.31. Let $\mathbf{a}, \mathbf{b} \in \mathbb{Z}$. Show that the following are equivalent:

(1) $\mathbf{a} \cdot \mathbf{b} > \mathbf{0}$.

(2) $[\mathbf{a} > \mathbf{0} \text{ and } \mathbf{b} > \mathbf{0}]$ or $[\mathbf{a} < \mathbf{0} \text{ and } \mathbf{b} < \mathbf{0}]$.

Exercise 4.32. Show that $(-1)(-1) = 1$.

Exercise 4.33. Let $\mathbf{a} \in \mathbb{Z}$. Prove that $(-1)\mathbf{a} = -\mathbf{a}$.

Exercise 4.34 (Failure of the Well-Ordering Principle for \mathbb{Z}).

If S is a nonempty subset of \mathbb{Z} , then an element $\ell \in S$ is called a *least element* of S if for all $\mathbf{n} \in S$, $\ell \leq \mathbf{n}$. Show that $S := \mathbb{Z}$ does not have a least element.

Exercise 4.35.*(Validity of the Well-Ordering Principle for subsets of \mathbb{Z} *bounded below*).

Let S be a nonempty subset of \mathbb{Z} , which has a *lower bound*, namely an element $\ell \in \mathbb{Z}$ such that $\ell \leq \mathbf{m}$ for all $\mathbf{m} \in S$. Show that S has a least element. *Hint:* Consider $\{\mathbf{m} - \ell : \mathbf{m} \in S\}$.

Example 4.7. There is no integer $\mathbf{x} \in \mathbb{Z}$ such that $\mathbf{2} \cdot \mathbf{x} = \mathbf{1}$, where $\mathbf{2} = [(1, \sigma(2))]$ and $\mathbf{1} = [(1, \sigma(1))]$. Indeed, as $\mathbf{1} > \mathbf{0}$ and $\mathbf{2} > \mathbf{0}$, it follows from Exercise 4.31 that $\mathbf{x} > \mathbf{0}$. So $\mathbf{x} = [(1, \sigma(n))]$ for some $n \in \mathbb{N}$. Clearly $n \neq 1$ (otherwise $\mathbf{2} \cdot \mathbf{x} = \mathbf{2} \neq \mathbf{1}$). So n has a predecessor, say $\ell \in \mathbb{N}$. Then $\mathbf{x} = [(1, 2)] + [(1, \sigma(\ell))]$, showing $\mathbf{x} > \mathbf{1}$. By Exercise 4.29, $\mathbf{2} \cdot \mathbf{x} > \mathbf{2} \cdot \mathbf{1} = \mathbf{2} > \mathbf{1}$. By the Trichotomy Law, $\mathbf{2} \cdot \mathbf{x} \neq \mathbf{1}$. \diamond

The previous example shows that given $\mathbf{a}, \mathbf{b} \in \mathbb{Z}$, with $\mathbf{a} \neq \mathbf{0}$, the equation $\mathbf{a} \cdot \mathbf{x} = \mathbf{b}$ in the unknown $\mathbf{x} \in \mathbb{Z}$ is not always solvable. The set \mathbb{Q} of rational number remedies this ‘flaw’ with integers. In the next section, we will learn about the construction of \mathbb{Q} and its arithmetic. From now on, for $n \in \mathbb{N}$, we will often denote the integer

$$\begin{aligned} [(1, \sigma(n))] &\text{ by } n, \\ [(1, 1)] &\text{ by } 0, \\ [(\sigma(n), 1)] &\text{ by } -n. \end{aligned}$$

By Theorem 4.11, $\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$. The elements of $\{1, 2, 3, \dots\}$ are called the *positive integers*, and those of $\{-1, -2, -3, \dots\}$ the *negative integers*.

4.4. Rational numbers

From elementary school, we think of a fraction as an expression $\frac{n}{d}$, where $n \in \mathbb{Z}$, $d \in \mathbb{Z} \setminus \{0\}$. Two fractions $\frac{n}{d}, \frac{n'}{d'}$ are deemed to be equivalent/same if $nd' = n'd$. We now formalise this by starting with the ‘equivalence relation’ that produces equivalence classes of fractions that are equivalent, and then formally define \mathbb{Q} as the set of these equivalence classes. Then we formally define addition and multiplication in \mathbb{Q} .

Definition 4.8 (Equivalence relation \sim on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$).

The relation \sim on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ is defined as follows: $(n, d) \sim (n', d')$ if $nd' = n'd$.

Proposition 4.16. \sim is an equivalence relation on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$.

Proof.

Reflexivity: For $(n, d) \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$, clearly $(n, d) \sim (n, d)$ (as $nd = nd$).

Symmetry: Let $(n, d), (n', d') \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ be such that $(n, d) \sim (n', d')$, so that $nd' = n'd$. Then $n'd = nd'$, and so $(n', d') \sim (n, d)$.

Transitivity: Let $(n, d), (n', d'), (n'', d'') \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ be such that $(n, d) \sim (n', d')$ and $(n', d') \sim (n'', d'')$. Then $nd' = n'd$ and $n'd'' = n''d'$.

1° $n' = 0$. From $nd' = n'd = 0d = 0$, we conclude that $n = 0$ (because $d' \neq 0$).

Similarly, $n''d' = n'd'' = 0d'' = 0$ implies that $n'' = 0$ (because $d' \neq 0$). But then $nd'' = 0d'' = 0 = 0d = n''d$, so that $(n, d) \sim (n'', d'')$.

2° $n' \neq 0$. We have $(nd')(n'd'') = (n'd)(n''d')$, that is, $(nd'')(n'd') = (n''d)(n'd')$. As $n' \neq 0$ and $d' \neq 0$, by the cancellation rule $nd'' = n''d$, i.e., $(n, d) \sim (n'', d'')$. \square

Exercise 4.36. Let $(n, d) \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ and $m \in \mathbb{Z} \setminus \{0\}$. Show that $(n, d) \sim (mn, md)$.

Definition 4.9 (The set \mathbb{Q} of rational numbers).

Each equivalence class of $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ under the equivalence relation \sim is called a *rational number*. The set of all rational numbers is denoted by \mathbb{Q} . The equivalence class $[(n, d)]$ of $(n, d) \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ is denoted by $\frac{n}{d}$.

Remark 4.4.

- (a) The choice of the letter \mathbb{Q} is motivated by thinking of $[\frac{n}{d}]$ as a ‘quotient’.
- (b) If $(n, d) \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$, then $(-n)d = ((-1)n)d = n((-1)d) = n(-d)$, and so $(n, d) \sim (-n, -d)$. Thus $\mathbf{r} = \frac{n}{d} = \frac{-n}{-d}$. Consequently, for $\mathbf{r} \in \mathbb{Q}$, we can always write $\mathbf{r} = \frac{n}{d}$, where $n \in \mathbb{Z}$ and $d \in \{1, 2, 3, \dots\}$. *

Addition and multiplication of rationals. In elementary school, we learn that the sum of fractions is defined by

$$\frac{n}{d} + \frac{p}{q} = \frac{nq + pd}{dq}.$$

We formalise this below.

Definition 4.10 (Addition in \mathbb{Q}). For $\mathbf{r} := \frac{n}{d} \in \mathbb{Q}$ and $\mathbf{s} := \frac{p}{q} \in \mathbb{Q}$, we define

$$\mathbf{r} + \mathbf{s} = \frac{n}{d} + \frac{p}{q} := \frac{nq + pd}{dq}.$$

Note that if $d \neq 0$ and $q \neq 0$, then $dq \neq 0$.

Exercise 4.37 (Addition is well-defined).

Show that if $(n, d), (n', d'), (p, q), (p', q')$ in $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ are such that $(n, d) \sim (n', d')$ and $(p, q) \sim (p', q')$, then $(nq + pd, dq) \sim (n'q' + p'd', d'q')$.

It can be shown that addition in \mathbb{Q} is commutative and associative. Define the zero element $\mathbf{0} = \frac{0}{1} \in \mathbb{Q}$. Let $\mathbf{r} = \frac{n}{d} \in \mathbb{Q}$. Then $\mathbf{r} + \mathbf{0} = \mathbf{r} = \mathbf{0} + \mathbf{r}$. Indeed,

$$\mathbf{r} + \mathbf{0} = \frac{n}{d} + \frac{0}{1} = \frac{n \cdot 1 + 0 \cdot d}{d \cdot 1} = \frac{n + 0}{d} = \frac{n}{d} = \mathbf{r}.$$

Exercise 4.38. For $\mathbf{r} = \frac{n}{d} \in \mathbb{Q}$, define $-\mathbf{r} = \frac{-n}{d} \in \mathbb{Q}$. Show that this is well-defined. Prove that $\mathbf{r} + (-\mathbf{r}) = \mathbf{0}$.

Definition 4.11 (Multiplication in \mathbb{Q}). For $\mathbf{r} := \frac{n}{d}$ and $\mathbf{s} := \frac{p}{q}$ in \mathbb{Q} , we define

$$\mathbf{r} \cdot \mathbf{s} = \frac{n}{d} \cdot \frac{p}{q} := \frac{np}{dq}.$$

Exercise 4.39 (Multiplication is well-defined).

Show that if $(n, d), (n', d'), (p, q), (p', q')$ in $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ are such that $(n, d) \sim (n', d')$ and $(p, q) \sim (p', q')$, then $(np, dq) \sim (n'p', d'q')$.

It can be checked that multiplication is commutative and associative.

Exercise 4.40 (Distributive law in \mathbb{Q}). Let $\mathbf{r}, \mathbf{r}', \mathbf{s} \in \mathbb{Q}$. Show that $(\mathbf{r} + \mathbf{r}') \cdot \mathbf{s} = \mathbf{r} \cdot \mathbf{s} + \mathbf{r}' \cdot \mathbf{s}$.

The rational number $\mathbf{1} := \frac{1}{1} \in \mathbb{Q}$ serves as the multiplicative identity in \mathbb{Q} . Indeed, for all $\mathbf{r} = \frac{n}{d} \in \mathbb{Q}$, we have $\mathbf{r} \cdot \mathbf{1} = \mathbf{r} = \mathbf{1} \cdot \mathbf{r}$:

$$\frac{n}{d} \cdot \frac{1}{1} = \frac{n \cdot 1}{d \cdot 1} = \frac{n}{d}.$$

Every nonzero rational number has a ‘reciprocal/multiplicative inverse’.

Theorem 4.17. Let $\mathbf{r} \in \mathbb{Q} \setminus \{\mathbf{0}\}$. Then there exists a unique rational number, denoted by $\mathbf{r}^{-1} \in \mathbb{Q}$, such that $\mathbf{r} \cdot \mathbf{r}^{-1} = \mathbf{1} = \mathbf{r}^{-1} \cdot \mathbf{r}$.

Proof.

Existence: Let $\mathbf{r} = \frac{n}{d}$. As $\mathbf{r} \neq \mathbf{0}$, $n \neq 0$. (Otherwise $\mathbf{r} = \frac{n}{d} = \frac{0}{d} = \frac{0}{1} = \mathbf{0}$, as $0 \cdot 1 = 0 = d \cdot 0$.) Set $\mathbf{r}^{-1} = \frac{d}{n} \in \mathbb{Q}$. Then

$$\mathbf{r} \cdot \mathbf{r}^{-1} = \frac{n}{d} \cdot \frac{d}{n} = \frac{nd}{dn} \stackrel{(*)}{=} \frac{1}{1} = \mathbf{1},$$

where $(*)$ holds since $(nd, dn) \sim (1, 1)$. By commutativity, also $\mathbf{r}^{-1} \cdot \mathbf{r} = \mathbf{1}$.

Uniqueness: Suppose $\mathbf{s} \in \mathbb{Q}$ is such that $\mathbf{r} \cdot \mathbf{s} = \mathbf{1} = \mathbf{s} \cdot \mathbf{r}$. Then

$$\mathbf{r}^{-1} = \mathbf{1} \cdot \mathbf{r}^{-1} = (\mathbf{s} \cdot \mathbf{r}) \cdot \mathbf{r}^{-1} = \mathbf{s} \cdot (\mathbf{r} \cdot \mathbf{r}^{-1}) = \mathbf{s} \cdot \mathbf{1} = \mathbf{s}.$$

□

Next we show that \mathbb{Q} does the job that prompted its construction.

Theorem 4.18. *Let $\mathbf{a} \in \mathbb{Q} \setminus \{\mathbf{0}\}$ and $\mathbf{b} \in \mathbb{Q}$. Then there exists a unique solution $\mathbf{x} \in \mathbb{Q}$ to the equation $\mathbf{a} \cdot \mathbf{x} = \mathbf{b}$.*

Proof. Set $\mathbf{x} := \mathbf{a}^{-1} \cdot \mathbf{b} \in \mathbb{Q}$. Then $\mathbf{a} \cdot \mathbf{x} = \mathbf{a} \cdot (\mathbf{a}^{-1} \cdot \mathbf{b}) = (\mathbf{a} \cdot \mathbf{a}^{-1}) \cdot \mathbf{b} = \mathbf{1} \cdot \mathbf{b} = \mathbf{b}$. Also, if \mathbf{x}' is a solution, then $\mathbf{x}' = \mathbf{1} \cdot \mathbf{x}' = (\mathbf{a}^{-1} \cdot \mathbf{a}) \cdot \mathbf{x}' = \mathbf{a}^{-1} \cdot (\mathbf{a} \cdot \mathbf{x}') = \mathbf{a}^{-1} \cdot \mathbf{b}$. \square

Finally, we show that \mathbb{Z} can be thought of as a subset of \mathbb{Q} .

Theorem 4.19. *The map $\mathbb{Z} \ni n \mapsto \frac{n}{1} \in \mathbb{Q}$ is injective. Moreover, for $m, n \in \mathbb{Z}$,*

$$\frac{m}{1} + \frac{n}{1} = \frac{m+n}{1}, \text{ and } \frac{m}{1} \cdot \frac{n}{1} = \frac{mn}{1}.$$

Proof. Suppose for $n, n' \in \mathbb{Z}$, $\frac{n}{1} = \frac{n'}{1}$. Then $(n, 1) \sim (n', 1)$. So $n = n \cdot 1 = n' \cdot 1 = n'$. Hence the map $\mathbb{Z} \ni n \mapsto \frac{n}{1} \in \mathbb{Q}$ is injective. Moreover,

$$\frac{m}{1} + \frac{n}{1} = \frac{m \cdot 1 + n \cdot 1}{1 \cdot 1} = \frac{m+n}{1}, \text{ and } \frac{m}{1} \cdot \frac{n}{1} = \frac{m \cdot n}{1 \cdot 1} = \frac{mn}{1}. \quad \square$$

Order. We now extend the order relation from \mathbb{Z} to \mathbb{Q} . Given rational numbers

$$\mathbf{r} = \frac{m}{n} \text{ and } \mathbf{s} = \frac{p}{q},$$

where m, n, p, q are integers and n, q are *positive* integers, then $\mathbf{r} < \mathbf{s}$ if $mq < pn$. If $\mathbf{r} < \mathbf{s}$, we write equivalently $\mathbf{s} > \mathbf{r}$. Let us check the notion is well-defined. Let

- $\frac{m'}{n'} = \frac{m}{n}$, where n' is a positive integer, and
- $\frac{p'}{q'} = \frac{p}{q}$, where q' is a positive integer.

We want to show $m'q' < p'n'$. As $mq < np$, and as n', q' are positive integers, it follows from Exercise 4.29 that $mqn'q' < npn'q'$, that is, $(mn')qq' < (pq')nn'$. Using $m'n = mn'$ and $pq' = p'q$, we obtain $(m'n)qq' < (p'q)nn'$, i.e., $(m'q')nq < (p'n')nq$. Hence $m'q' < p'n'$. (Otherwise, by the Trichotomy Law $m'q' \geq p'n'$. As n, q are positive integers, Exercise 4.29 implies $(m'q')nq \geq (p'n')nq$, a contradiction.)

It is clear that for $m, n \in \mathbb{Z}$, $m < n$ in \mathbb{Z} if and only if $\frac{m}{1} < \frac{n}{1}$ in \mathbb{Q} .

Exercise 4.41 (Transitivity of $<$ in \mathbb{Q}).

Let $\mathbf{r}, \mathbf{s}, \mathbf{t} \in \mathbb{Q}$ be such that $\mathbf{r} < \mathbf{s}$ and $\mathbf{s} < \mathbf{t}$. Show that $\mathbf{r} < \mathbf{t}$.

Exercise 4.42 (Trichotomy law).

Show that for any rational numbers $\mathbf{r}, \mathbf{s} \in \mathbb{Q}$, one and exactly one of the following holds:

$$1^\circ \mathbf{r} < \mathbf{s} \qquad 2^\circ \mathbf{r} = \mathbf{s} \qquad 3^\circ \mathbf{r} > \mathbf{s}.$$

Exercise 4.43. Let $\mathbf{r}, \mathbf{s}, \mathbf{t} \in \mathbb{Q}$. Prove that if $\mathbf{r} < \mathbf{s}$, then $\mathbf{r} + \mathbf{t} < \mathbf{s} + \mathbf{t}$.

Exercise 4.44. Let $\mathbf{r}, \mathbf{s} \in \mathbb{Q}$ and $\mathbf{r} < \mathbf{s}$. Show that if $\mathbf{t} \in \mathbb{Q}$ is such that $\mathbf{t} > \mathbf{0}$, then $\mathbf{rt} < \mathbf{st}$.

Exercise 4.45. Let $\mathbf{r}, \mathbf{s} \in \mathbb{Q}$ and $\mathbf{r} < \mathbf{s}$. Show that there exists a $\mathbf{t} \in \mathbb{Q}$ such that $\mathbf{r} < \mathbf{t} < \mathbf{s}$.

In the light of Exercise 4.36, we know that given any rational number $\mathbf{r} \in \mathbb{Q}$, we have a unique pair $n, d \in \mathbb{Z}$ such that d is a positive integer, n, d have no common divisor other than 1, and

$$\mathbf{r} = \frac{n}{d}.$$

We have also seen that

- \mathbb{Q} does not have the least upper bound property (Example 1.13), and
- not all Cauchy sequences in \mathbb{Q} are convergent with a limit in \mathbb{Q} (Exercise 4.5).

But in order to do Analysis, it is convenient to work with a number system which has these two properties listed above. This ‘(analytical) flaw’ of \mathbb{Q} is remedied by the set of real numbers. From now onwards, we will denote rational numbers simply using ordinary font letters such as r, s, t, \dots (instead of boldface $\mathbf{r}, \mathbf{s}, \mathbf{t}, \dots$).

4.5. Real numbers

Finally we have reached the point where we can learn about the construction of the most important number system from the point of view of Mathematical Analysis, namely the real number system \mathbb{R} . Roughly speaking, the set of real numbers are the numbers to which Cauchy sequences in \mathbb{Q} ‘want to converge to’. As these limits may not be rational, we just name/label these numbers by the whole Cauchy sequence in \mathbb{Q} itself! But then two Cauchy sequences in \mathbb{Q} might want to converge to the same thing (e.g. think of $(a_n)_{n \in \mathbb{N}}$ and $(a_n + \frac{1}{n})_{n \in \mathbb{N}}$), and so we ought not to distinguish between such two Cauchy sequences. So we must build an equivalence relation \sim on Cauchy sequences (so that

$$(a_n)_{n \in \mathbb{N}} \sim (b_n)_{n \in \mathbb{N}} \text{ if } \lim_{n \rightarrow \infty} (a_n - b_n) = 0,$$

and consider the real numbers as equivalence classes of Cauchy sequences under this equivalence relation. We had met this equivalence relation \sim in Exercise 4.5, where we checked that this is indeed an equivalence relation. However, over there we viewed the convergence using the notion of convergence of a sequence of real numbers (so that the $\epsilon > 0$ was a real number). But since we are trying to construct the reals, we are only allowed to use rational numbers. So we need to restrict ourselves to ϵ that are rational in the definition of convergence. We do this carefully below.

Definition 4.12.

- A *Cauchy sequence in \mathbb{Q}* is a sequence $(a_n)_{n \in \mathbb{N}}$ of rational numbers such that for every rational $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that whenever $m, n > N$, we have $|a_n - a_m| < \epsilon$. The set of all Cauchy sequences in \mathbb{Q} is denoted by \mathcal{C} .
- Let $r \in \mathbb{Q}$. A sequence $(a_n)_{n \in \mathbb{N}}$ in \mathbb{Q} *converges to r in \mathbb{Q}* if for every rational $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that for all $n > N$, $|a_n - r| < \epsilon$.
- The relation \sim on \mathcal{C} is defined as follows: $(a_n)_{n \in \mathbb{N}} \sim (b_n)_{n \in \mathbb{N}}$ if the sequence $(a_n - b_n)_{n \in \mathbb{N}}$ converges to 0 in \mathbb{Q} .

Exercise 4.46. Let $(a_n)_{n \in \mathbb{N}}, (b_n)_{n \in \mathbb{N}}$ be Cauchy sequences in \mathbb{Q} . Show that $(a_n + b_n)_{n \in \mathbb{N}}$ is Cauchy sequence in \mathbb{Q} too.

Proposition 4.20. *Every Cauchy sequence in \mathbb{Q} is bounded.*

Proof. Let $(a_n)_{n \in \mathbb{N}}$ be a Cauchy sequence in \mathbb{Q} . Choose a rational $\epsilon > 0$, say $\epsilon = 1$. Then there exists an $N \in \mathbb{N}$ such that for all $n, m > N$, we have $|a_n - a_m| < \epsilon = 1$. In particular, with $m = N+1 > N$, and $n > N$, $|a_n - a_{N+1}| < 1$. Hence by the Triangle Inequality⁴ in \mathbb{Q} , for all $n > N$,

$$|a_n| = |a_n - a_{N+1} + a_{N+1}| \leq |a_n - a_{N+1}| + |a_{N+1}| < 1 + |a_{N+1}|.$$

On the other hand, for $n \leq N$, $|a_n| \leq \max\{|a_1|, \dots, |a_N|, |a_{N+1}| + 1\} =: M > 0$. Consequently, $|a_n| \leq M$ ($n \in \mathbb{N}$), that is, the sequence $(a_n)_{n \in \mathbb{N}}$ is bounded. \square

Exercise 4.47. Let $(a_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}}$ be Cauchy sequences in \mathbb{Q} . Show that $(a_n b_n)_{n \in \mathbb{N}}$ is Cauchy sequence in \mathbb{Q} too.

Exercise 4.48. Suppose that $(a_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}}$ are sequences in \mathbb{Q} such that $(a_n)_{n \in \mathbb{N}}$ (respectively $(b_n)_{n \in \mathbb{N}}$) converges in \mathbb{Q} to $r_a \in \mathbb{Q}$ (respectively $r_b \in \mathbb{Q}$).

- (1) Show that the limit is unique: If $(a_n)_{n \in \mathbb{N}}$ converges in \mathbb{Q} to $r'_a \in \mathbb{Q}$, then $r_a = r'_a$.
- (2) Show that $(-a_n)_{n \in \mathbb{N}}$ converges to $-r_a$.
- (3) Show that $(a_n + b_n)_{n \in \mathbb{N}}$ converges to $r_a + r_b$.

Exercise 4.49. Show that \sim is an equivalence relation on \mathcal{C} .

Definition 4.13 (The set of real numbers).

A *real number* is an equivalence class of \mathcal{C} under the relation \sim . If $(a_n)_{n \in \mathbb{N}} \in \mathcal{C}$, $[(a_n)_{n \in \mathbb{N}}]$ denotes the real number which is the equivalence class of \mathcal{C} containing the sequence $(a_n)_{n \in \mathbb{N}}$. The set of all real numbers is denoted by \mathbb{R} .

The set of real numbers is supposed to be an *extension* of the rational numbers \mathbb{Q} , that is, we want to see that $\mathbb{Q} \subset \mathbb{R}$. Given a rational number $r \in \mathbb{Q}$, the constant sequence r, r, r, \dots , that is, $(r)_{n \in \mathbb{N}}$, is a Cauchy sequence in \mathbb{Q} . Thus $[(r)_{n \in \mathbb{N}}]$ is a real number. We have the following.

Proposition 4.21. *The map $\mathbb{Q} \ni r \mapsto [(r)_{n \in \mathbb{N}}] \in \mathbb{R}$ is injective.*

Proof. Let $r, s \in \mathbb{Q}$ be such that $[(r)_{n \in \mathbb{N}}] = [(s)_{n \in \mathbb{N}}]$. Then $(r)_{n \in \mathbb{N}} \sim (s)_{n \in \mathbb{N}}$. So

$$\lim_{n \rightarrow \infty} (r - s) = 0.$$

But the constant sequence $r - s, r - s, r - s, \dots$ converges in \mathbb{Q} to $r - s$. By the uniqueness of limits, $r - s = 0$, that is, $r = s$. \square

⁴The proof of the Triangle Inequality is exactly the same, replacing ‘real/ \mathbb{R} ’ everywhere by ‘rational/ \mathbb{Q} ’. Note that we are not allowed to use reals yet, and so we can’t just specialise the Triangle Inequality for \mathbb{R} to the rationals.

Addition and multiplication. Clearly, if addition in \mathbb{R} is to respect the addition in \mathbb{Q} , we must have that for $r, s \in \mathbb{Q}$, $[(r)_{n \in \mathbb{N}}] + [(s)_{n \in \mathbb{N}}]$ should equal $[(r + s)_{n \in \mathbb{N}}]$. Similarly, $[(r)_{n \in \mathbb{N}}] \cdot [(s)_{n \in \mathbb{N}}]$ should equal $[(rs)_{n \in \mathbb{N}}]$. This motivates the following.

Definition 4.14. The *sum* of the real numbers $[(a_n)_{n \in \mathbb{N}}]$ and $[(b_n)_{n \in \mathbb{N}}]$ is given by

$$[(a_n)_{n \in \mathbb{N}}] + [(b_n)_{n \in \mathbb{N}}] = [(a_n + b_n)_{n \in \mathbb{N}}].$$

The *product* of the real numbers $[(a_n)_{n \in \mathbb{N}}]$ and $[(b_n)_{n \in \mathbb{N}}]$ is defined by

$$[(a_n)_{n \in \mathbb{N}}] \cdot [(b_n)_{n \in \mathbb{N}}] = [(a_n b_n)_{n \in \mathbb{N}}].$$

As usual, we have to check well-definedness. We leave this as an exercise for addition, but give an argument below for multiplication. Let $[(a_n)_{n \in \mathbb{N}}] = [(a'_n)_{n \in \mathbb{N}}] \in \mathbb{R}$ and $[(b_n)_{n \in \mathbb{N}}] = [(b'_n)_{n \in \mathbb{N}}] \in \mathbb{R}$. The idea is to use the inequality

$$|a'_n b'_n - a_n b_n| = |a'_n b'_n - a'_n b_n + a'_n b_n - a_n b_n| \leq |a'_n| |b'_n - b_n| + |a'_n - a_n| |b_n|$$

and the boundedness of the terms a'_n, b_n to show $(a_n b_n)_{n \in \mathbb{N}} \sim (a'_n b'_n)_{n \in \mathbb{N}}$. We carry out the details below.

As $(a'_n)_{n \in \mathbb{N}}$ is Cauchy, it is bounded, and let $A' > 0$ be a rational number such that $|a'_n| < A'$ for all $n \in \mathbb{N}$. Similarly, $(b_n)_{n \in \mathbb{N}}$ is bounded, and let $B > 0$ be a rational number such that $|b_n| < B$ for all $n \in \mathbb{N}$. Let $\epsilon > 0$ be a rational number. As $(a_n)_{n \in \mathbb{N}} \sim (a'_n)_{n \in \mathbb{N}}$, we have that $(a_n - a'_n)_{n \in \mathbb{N}}$ converges in \mathbb{Q} to 0. So for the rational $\frac{\epsilon}{2B} > 0$, there exists an $N_a \in \mathbb{N}$ such that $|a'_n - a_n| < \frac{\epsilon}{2B}$. Similarly, as $(b_n)_{n \in \mathbb{N}} \sim (b'_n)_{n \in \mathbb{N}}$, we have that for the rational $\frac{\epsilon}{2A'} > 0$, there exists an $N_b \in \mathbb{N}$ such that $|b'_n - b_n| < \frac{\epsilon}{2A'}$. Set $N = N_a + N_b$. For all $n > N$, we have

$$\begin{aligned} |a'_n b'_n - a_n b_n| &= |a'_n b'_n - a'_n b_n + a'_n b_n - a_n b_n| \leq |a'_n| |b'_n - b_n| + |a'_n - a_n| |b_n| \\ &\leq A' |b'_n - b_n| + |a'_n - a_n| B < A' \frac{\epsilon}{2A'} + \frac{\epsilon}{2B} B = \epsilon. \end{aligned}$$

Thus $(a_n b_n)_{n \in \mathbb{N}} \sim (a'_n b'_n)_{n \in \mathbb{N}}$.

Exercise 4.50 (Addition is well-defined).

Let $(a_n)_{n \in \mathbb{N}} \sim (a'_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}} \sim (b'_n)_{n \in \mathbb{N}}$. Show that $(a_n + b_n)_{n \in \mathbb{N}} \sim (a'_n + b'_n)_{n \in \mathbb{N}}$.

Example 4.8 (The real numbers **0** and **1**). We define the real numbers **0** = $[(0)_{n \in \mathbb{N}}]$ and **1** = $[(1)_{n \in \mathbb{N}}]$. Then for every real number $\mathbf{x} \in \mathbb{R}$, we have

$$\begin{aligned} \mathbf{0} + \mathbf{x} &= \mathbf{r} = \mathbf{x} + \mathbf{0}, \text{ and} \\ \mathbf{1} \cdot \mathbf{x} &= \mathbf{r} = \mathbf{x} \cdot \mathbf{1}. \end{aligned}$$

Thus **0** serves as the additive identity and **1** serves as the multiplicative identity. Clearly **1** \neq **0** because the sequence $(1 - 0)_{n \in \mathbb{N}}$ converges in \mathbb{Q} to $1 \neq 0$. \diamond

The set \mathbb{R} , together with the operations $+, \cdot : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ forms a ‘field’, i.e., the following hold.

$$\begin{aligned}
 + \quad & \left\{ \begin{array}{ll} \text{(F1) (Associativity)} & \text{For all } \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}, \mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}. \\ \text{(F2) (Additive identity)} & \text{For all } \mathbf{x} \in \mathbb{R}, \mathbf{x} + \mathbf{0} = \mathbf{x} = \mathbf{0} + \mathbf{x}. \\ \text{(F3) (Inverses)} & \text{For all } \mathbf{x} \in \mathbb{R}, \text{ there exists } -\mathbf{x} \in \mathbb{R} \\ & \text{such that } \mathbf{x} + (-\mathbf{x}) = \mathbf{0} = -\mathbf{x} + \mathbf{x}. \\ \text{(F4) (Commutativity)} & \text{For all } \mathbf{x}, \mathbf{y} \in \mathbb{R}, \mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}. \end{array} \right. \\
 \cdot \quad & \left\{ \begin{array}{ll} \text{(F5) (Associativity)} & \text{For all } \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}, \mathbf{x} \cdot (\mathbf{y} \cdot \mathbf{z}) = (\mathbf{x} \cdot \mathbf{y}) \cdot \mathbf{z}. \\ \text{(F6) (Multiplicative identity)} & \mathbf{1} \neq \mathbf{0} \text{ and for all } x \in \mathbb{R}, \mathbf{x} \cdot \mathbf{1} = \mathbf{x} = \mathbf{1} \cdot \mathbf{x}. \\ \text{(F7) (Inverses)} & \text{For all } \mathbf{x} \in \mathbb{R} \setminus \{\mathbf{0}\}, \text{ there exists } \mathbf{x}^{-1} \in \mathbb{R} \\ & \text{such that } \mathbf{x} \cdot \mathbf{x}^{-1} = \mathbf{1} = \mathbf{x}^{-1} \cdot \mathbf{x}. \\ \text{(F8) (Commutativity)} & \text{For all } \mathbf{x}, \mathbf{y} \in \mathbb{R}, \mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}. \end{array} \right. \\
 +, \cdot \quad & \text{(F9) (Distributivity) For all } \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}, \mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}.
 \end{aligned}$$

In fact, if we replace everywhere \mathbb{R} by \mathbb{Q} (and $\mathbf{1}, \mathbf{0}$ by the rational numbers $1, 0$, respectively), then the set \mathbb{Q} of rational numbers with their addition and multiplication, also satisfy the same properties. We say that $(\mathbb{Q}, +, \cdot)$ is also a field. (However, $(\mathbb{Z}, +, \cdot)$ is not a field, because multiplicative inverses don’t always exist: we had seen that the equation $2x = 1$ has no solution $x \in \mathbb{Z}$.)

We will not check each the above, as they essentially follow by ‘termwise verifications’, and by using the corresponding properties from the field of rationals. We remark that the additive inverse of $\mathbf{x} = [(a_n)_{n \in \mathbb{N}}]$ is $-\mathbf{x} := [(-a_n)_{n \in \mathbb{N}}]$. Let us show the existence of multiplicative inverses for nonzero reals. First we prove the following lemma.

Lemma 4.22. *Let $\mathbf{x} \in \mathbb{R}$ be such that $\mathbf{x} \neq \mathbf{0}$. If $(a_n)_{n \in \mathbb{N}} \in \mathbf{x}$, then there exists a rational $d > 0$ and an $N \in \mathbb{N}$ such that for all $n > N$, $|a_n| > d$.*

Proof. As $[(a_n)_{n \in \mathbb{N}}] = \mathbf{x} \neq \mathbf{0} = [(0)_{n \in \mathbb{N}}]$, we have $\neg((a_n)_{n \in \mathbb{N}} \sim (0)_{n \in \mathbb{N}})$, i.e.,

$$\begin{aligned}
 & \neg((a_n - 0)_{n \in \mathbb{N}} \text{ converges in } \mathbb{Q} \text{ to } 0), \text{ i.e.,} \\
 & \neg(\forall \text{ rational } \epsilon > 0, \exists N \in \mathbb{N} \text{ such that } \forall n > N, |a_n - 0| < \epsilon), \text{ i.e.,}
 \end{aligned}$$

Thus

$$\exists \text{ rational } \epsilon > 0 \text{ such that } \forall N \in \mathbb{N}, \exists n > N \text{ such that } |a_n - 0| \geq \epsilon. \quad (\star)$$

Since $(a_n)_{n \in \mathbb{N}} \in \mathcal{C}$, for the rational $\epsilon/2 > 0$, there exists an $N_* \in \mathbb{N}$ such that for all $n, m > N_*$, $|a_n - a_m| < \frac{\epsilon}{2}$. From (\star) , taking $N = N_*$, there exists $n_* > N_*$ such that $|a_{n_*} - 0| \geq \epsilon$. Hence for $n > N_*$, we have

$$|a_n| = |a_n - a_{n_*} + a_{n_*}| \geq |a_{n_*}| - |a_n - a_{n_*}| \geq \epsilon - \frac{\epsilon}{2} = \frac{\epsilon}{2} =: d. \quad \square$$

Proposition 4.23. *Let the real number $\mathbf{x} \neq \mathbf{0}$. Then there exists an $\mathbf{x}^{-1} \in \mathbb{R}$ such that $\mathbf{x} \cdot \mathbf{x}^{-1} = \mathbf{1} = \mathbf{x}^{-1} \cdot \mathbf{x}$.*

Proof. Let $\mathbf{x} = [(a_n)_{n \in \mathbb{N}}]$. By Lemma 4.22 there exists a rational $d > 0$ and an $N \in \mathbb{N}$ such that $|a_n| > d$ for all $n > N$. In particular, $a_n \neq 0$ for all $n > N$. Set⁵

$$b_n := \begin{cases} 0 & \text{if } 1 \leq n \leq N, \\ a_n^{-1} & \text{if } n > N. \end{cases}$$

Then $(b_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in \mathbb{Q} . Firstly, for $n, m > N$,

$$|b_n - b_m| = \left| \frac{1}{a_n} - \frac{1}{a_m} \right| = \frac{|a_n - a_m|}{|a_n||a_m|} \leq \frac{|a_n - a_m|}{d^2}.$$

Secondly, as $(a_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in \mathbb{Q} , given a rational $\epsilon > 0$, there exists an $M \in \mathbb{N}$ such that for all $n, m > M$, $|a_n - a_m| < \epsilon d^2$. Hence for all $n, m > N + M$,

$$|b_n - b_m| \leq \frac{|a_n - a_m|}{d^2} < \frac{\epsilon d^2}{d^2} = \epsilon.$$

Consequently, $(b_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in \mathbb{Q} . We have⁶

$$a_n b_n := \begin{cases} 0 & \text{if } 1 \leq n \leq N, \\ 1 & \text{if } n > N. \end{cases}$$

Hence $(a_n b_n)_{n \in \mathbb{N}}$ converges in \mathbb{Q} to 1. So $[(a_n b_n)_{n \in \mathbb{N}}] = \mathbf{1}$. Set $\mathbf{x}^{-1} := [(b_n)_{n \in \mathbb{N}}]$. Then we have $\mathbf{x} \cdot \mathbf{x}^{-1} = \mathbf{1} = \mathbf{x}^{-1} \cdot \mathbf{x}$. \square

Exercise 4.51 (Distributive law). Let $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}$. Prove that $\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}$.

Order. To compare real numbers $\mathbf{x} = [(a_n)_{n \in \mathbb{N}}]$ and $\mathbf{y} = [(b_n)_{n \in \mathbb{N}}]$, we would like to use the order relation $<$ on \mathbb{Q} . If we try to define $\mathbf{x} < \mathbf{y}$ by saying that for *all* $n \in \mathbb{N}$, $a_n < b_n$, then this will not be a well-defined notion. Indeed, changing the first few terms of $(a_n)_{n \in \mathbb{N}}$ we could easily violate this, without changing $[(a_n)_{n \in \mathbb{N}}]$. Intuitively, \mathbf{x} is the real number that $(a_n)_{n \in \mathbb{N}}$ converges to. So thinking formally

$$' \mathbf{x} = \lim a_n ', \quad ' \mathbf{y} = \lim b_n ',$$

we would say $\mathbf{x} < \mathbf{y}$ if ' $\lim a_n < \lim b_n$ ', that is,

$$' \lim(b_n - a_n) > 0 '.$$

But from our former intuition with limits, we know that this means that for all *large enough* $n \in \mathbb{N}$, $b_n - a_n$ stays away from 0 by some positive distance d , say. This motivates the following.

Definition 4.15. Let $\mathbf{x} = [(a_n)_{n \in \mathbb{N}}]$ and $\mathbf{y} = [(b_n)_{n \in \mathbb{N}}]$ be real numbers. Then $\mathbf{x} < \mathbf{y}$ if there exists a rational number $d > 0$ and an $N \in \mathbb{N}$ such that for all $n > N$, $b_n - a_n > d$. If $\mathbf{x} < \mathbf{y}$, we write equivalently $\mathbf{y} > \mathbf{x}$.

Let us show that this is a well-defined notion.

⁵Although we set $b_n = 0$ for $1 \leq n \leq N$ here, any arbitrary N rational numbers can be specified here.

⁶Had we specified b_1, \dots, b_N arbitrarily, we would get a bunch of initial terms $a_n b_n$ for $1 \leq n \leq N$, but this won't affect the rest of the proof.

Proposition 4.24. *Let $[(a_n)_{n \in \mathbb{N}}] = [(a'_n)_{n \in \mathbb{N}}] \in \mathbb{R}$ and $[(b_n)_{n \in \mathbb{N}}] = [(b'_n)_{n \in \mathbb{N}}] \in \mathbb{R}$. Suppose that there exists a rational number $d > 0$ and an $N \in \mathbb{N}$ such that for all $n > N$, $b_n - a_n > d$. Then there exists a rational number $d' > 0$ and an $N' \in \mathbb{N}$ such that for all $n > N'$, $b'_n - a'_n > d'$.*

Proof. As $(a_n)_{n \in \mathbb{N}} \sim (a'_n)_{n \in \mathbb{N}}$, we know that $(a_n - a'_n)_{n \in \mathbb{N}}$ converges in \mathbb{Q} to 0. So for the rational $d/4 > 0$, there exists an $N_a \in \mathbb{N}$ such that for all $n > N_a$, $|a_n - a'_n| < d/4$, i.e., $-d/4 < a_n - a'_n < d/4$. In particular

$$a_n - a'_n > -\frac{d}{4} \text{ for all } n > N_a. \quad (*)$$

Similarly, $(b_n)_{n \in \mathbb{N}} \sim (b'_n)_{n \in \mathbb{N}}$ yields the existence of an $N_b \in \mathbb{N}$ such that

$$b'_n - b_n > -\frac{d}{4} \text{ for all } n > N_b. \quad (**)$$

Set $N' = N_a + N_b + N$. Then for all $n > N'$, using $(*)$ and $(**)$, we have

$$b'_n - a'_n = b_n - a_n + b'_n - b_n + a_n - a'_n > d - \frac{d}{4} - \frac{d}{4} = \frac{d}{2} > 0.$$

So with the rational $d' := \frac{d}{2} > 0$, for all $n > N'$, we have $b'_n - a'_n > d'$. \square

Exercise 4.52. Show that if $r, s \in \mathbb{Q}$ and $r < s$, then $[(r)_{n \in \mathbb{N}}] < [(s)_{n \in \mathbb{N}}]$. (In particular, for the real numbers $\mathbf{0}, \mathbf{1}$, we have $\mathbf{0} < \mathbf{1}$.)

Exercise 4.53 (Transitivity of $<$).

Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}$ be such that $\mathbf{x} < \mathbf{y}$ and $\mathbf{y} < \mathbf{z}$. Prove that $\mathbf{x} < \mathbf{z}$.

Theorem 4.25 (Trichotomy Law).

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}$. Then one and exactly one of the following hold:

$$1^\circ \mathbf{x} < \mathbf{y}. \quad 2^\circ \mathbf{x} = \mathbf{y}. \quad 3^\circ \mathbf{x} > \mathbf{y}.$$

Proof. Let $\mathbf{x} = [(a_n)_{n \in \mathbb{N}}]$ and $\mathbf{y} = [(b_n)_{n \in \mathbb{N}}]$. If $\mathbf{x} = \mathbf{y}$, then $(a_n - b_n)_{n \in \mathbb{N}}$ converges in \mathbb{Q} to 0. Let us show that $\neg(\mathbf{x} > \mathbf{y})$. Indeed, otherwise there exists a rational $d > 0$ and an $N \in \mathbb{N}$ such that $a_n - b_n > d$ for all $n > N$. But then taking the rational $\epsilon := d/2 > 0$, we get, thanks to the convergence of $(a_n - b_n)_{n \in \mathbb{N}}$, that there is an $N' \in \mathbb{N}$ such that for all $n > N'$, $|a_n - b_n| < d/2$. So with $n = N + N'$, we arrive at the contradiction that $d < a_n - b_n \leq |a_n - b_n| < d/2$. So if $\mathbf{x} = \mathbf{y}$, then $\neg(\mathbf{x} > \mathbf{y})$. Interchanging the roles of \mathbf{x}, \mathbf{y} , we also have that if $\mathbf{x} = \mathbf{y}$, then $\neg(\mathbf{x} < \mathbf{y})$. Let us also note that if $\mathbf{x} < \mathbf{y}$, then $\neg(\mathbf{y} < \mathbf{x})$: Otherwise there exist rational $d, d' > 0$ and $N, N' \in \mathbb{N}$ such that for all $n > N$ we have $b_n - a_n > d$, and for all $n > N'$, we have $a_n - b_n > d'$, so that with $n := N + N'$, we get $d' < a_n - b_n < -d$, giving $0 > d + d' > d + 0 = d$, a contradiction.

Let $[(a_n)_{n \in \mathbb{N}}] := \mathbf{x} \neq \mathbf{y} = [(b_n)_{n \in \mathbb{N}}]$. Then it is not the case that the sequence $(a_n - b_n)_{n \in \mathbb{N}}$ converges in \mathbb{Q} to 0. Thus there exists a rational $\epsilon_* > 0$ such that

$$\text{for all } N \in \mathbb{N}, \text{ there exists an } n > N \text{ such that } |a_n - b_n| \geq \epsilon_*. \quad (\star)$$

As $(a_n)_{n \in \mathbb{N}}$ is Cauchy, there exists an $N_a \in \mathbb{N}$ such that for all $m, n > N_a$, we have $|a_n - a_m| < \epsilon_*/4$, i.e., $-\epsilon_*/4 < a_n - a_m < \epsilon_*/4$. In particular, $a_n - a_m > -\epsilon_*/4$ for all $n, m > N_a$. Similarly, as $(b_n)_{n \in \mathbb{N}}$ is Cauchy, there exists an $N_b \in \mathbb{N}$ such that for all $m, n > N_b$, $|b_n - b_m| < \epsilon_*/4$, giving in particular $b_n - b_m > -\epsilon_*/4$. Now take $N = N_a + N_b$ in (\star) . Then there exists an $n_* > N$ such that $|a_{n_*} - b_{n_*}| \geq \epsilon_* > 0$. In particular, $a_{n_*} - b_{n_*} \neq 0$. So by the trichotomy law for $<$ in \mathbb{Q} , we have the following two mutually exclusive possible cases:

1° $a_{n_*} - b_{n_*} > 0$. Then $a_{n_*} - b_{n_*} = |a_{n_*} - b_{n_*}| \geq \epsilon_*$. For all $m > n_*$ ($> N = N_a + N_b$),

$$a_m - b_m = a_{n_*} - b_{n_*} + a_m - a_{n_*} + b_{n_*} - b_m > \epsilon_* - \frac{\epsilon_*}{4} - \frac{\epsilon_*}{4} = \frac{\epsilon_*}{2} =: d.$$

So for all $m > n_*$, we have that $a_m - b_m > d$, showing $\mathbf{x} > \mathbf{y}$.

2° $a_{n_*} - b_{n_*} < 0$. Then $b_{n_*} - a_{n_*} = |a_{n_*} - b_{n_*}| \geq \epsilon_*$. For all $m > n_*$ ($> N = N_a + N_b$),

$$b_m - a_m = b_{n_*} - a_{n_*} + b_m - b_{n_*} + a_{n_*} - a_m > \epsilon_* - \frac{\epsilon_*}{4} - \frac{\epsilon_*}{4} = \frac{\epsilon_*}{2} =: d.$$

So for all $m > n_*$, we have that $b_m - a_m > d$, showing $\mathbf{y} > \mathbf{x}$. \square

Definition 4.16 (The set \mathbb{P} of positive reals). We define $\mathbb{P} := \{\mathbf{x} \in \mathbb{R} : \mathbf{x} > \mathbf{0}\}$.

Exercise 4.54. Let $\mathbf{x}, \mathbf{y} \in \mathbb{P}$. Show that $\mathbf{x} + \mathbf{y} \in \mathbb{P}$ and $\mathbf{x} \cdot \mathbf{y} \in \mathbb{P}$.

Exercise 4.55. Let $\mathbf{x} \in \mathbb{R}$ be such that $\mathbf{x} > \mathbf{0}$. Prove that there exists an $r \in \mathbb{Q}$ such that $\mathbf{0} < [(r)_{n \in \mathbb{N}}] < \mathbf{x}$. We write this succinctly as $\mathbf{0} < r < \mathbf{x}$.

Exercise 4.56. Let $(a_n)_{n \in \mathbb{N}}$ be a Cauchy sequence in \mathbb{Q} . Suppose there exists an $N \in \mathbb{N}$ such that for all $n > N$, we have $a_n \geq 0$. Show that the real number $\mathbf{x} = [(a_n)_{n \in \mathbb{N}}] \geq \mathbf{0}$.

Exercise 4.57. Find all positive real x, y such that

$$\begin{aligned} \log_3 x + \log_2 y &= 2, \\ 3^x - 2^y &= 23. \end{aligned}$$

Hint: One solution is $(x, y) = (3, 2)$. Use the second equation to show that $x > 3$ forces $y > 2$, violating the first equation. The case $x < 3$ is handled similarly.

Exercise 4.58 (No order for \mathbb{C}). A field \mathbb{F} is called *ordered* if there is a subset $P \subset \mathbb{F}$, called the *set of positive elements of \mathbb{F}* , satisfying the following:

(P1) For all $x, y \in P$, $x + y \in P$.

(P2) For all $x, y \in P$, $x \cdot y \in P$.

(P3) For each $x \in P$, one and only one of the following three cases is true:

$$1^\circ \quad x = 0. \qquad 2^\circ \quad x \in P. \qquad 3^\circ \quad -x \in P.$$

(Once one has an ordered set of elements in a field, one can compare the elements of \mathbb{F} by defining a relation $>_P$ in \mathbb{F} by setting $y >_P x$ for $x, y \in \mathbb{F}$ if $y - x \in P$.)

Show that \mathbb{C} is not an ordered field. *Hint:* Consider $x := i$, and first look at $x \cdot x$.

(*) The least upper bound property of \mathbb{R} . Finally we are ready to prove the ultimate goal, namely the least upper bound property of \mathbb{R} . This is a bit technical, and so we relegate the proof to an appendix to this chapter. This part can be skipped (as it is non-examinable), but we give the proof here for the sake of ‘completeness’. Thus in the appendix on pages 123-126, the interested student will find the proof of the following important result.

Theorem 4.26 (Least upper bound property of \mathbb{R}).

Every nonempty subset of \mathbb{R} which is bounded above has a supremum.

We reiterate that in Example 1.13 we had seen that \mathbb{Q} does not possess the Least Upper Bound Property. So this ‘analytical flaw’ of the rational number system is remedied by the set of real numbers. Moreover, we had seen that not all Cauchy sequences in \mathbb{Q} converge in \mathbb{Q} . In contrast, we have the following happy situation in \mathbb{R} .

Exercise 4.59. (*) ($\{\text{Cauchy sequences in } \mathbb{R}\} = \{\text{convergent sequences in } \mathbb{R}\}$).

A sequence $(a_n)_{n \in \mathbb{N}}$ of real numbers is said to be a *Cauchy sequence in \mathbb{R}* if for every real $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that for all $m, n > N$, we have $|a_m - a_n| < \epsilon$.

(1) Every convergent sequence in \mathbb{R} is a Cauchy sequence in \mathbb{R} .

Suppose the real sequence $(a_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in \mathbb{R} .

(2) Show that $(a_n)_{n \in \mathbb{N}}$ is bounded. Proceed as in Proposition 4.20.

(3) Show that $(a_n)_{n \in \mathbb{N}}$ has a convergent subsequence (say $(a_{n_k})_{k \in \mathbb{N}}$, converging to $L \in \mathbb{R}$).

Hint: Use Theorems 2.7 and 2.3.

We claim that by virtue of the fact that $(a_n)_{n \in \mathbb{N}}$ is Cauchy, $(a_n)_{n \in \mathbb{N}}$ is itself convergent, with the same limit L (of its convergent subsequence $(a_{n_k})_{k \in \mathbb{N}}$ from part (3)). Let $\epsilon > 0$. Then there exists an $N \in \mathbb{N}$ such that for all $n, m > N$,

$$|a_n - a_m| < \frac{\epsilon}{2}. \quad (\star)$$

As $(a_{n_k})_{k \in \mathbb{N}}$ converges to L , there exists an $n_K > N$ so that $|a_{n_K} - L| < \frac{\epsilon}{2}$. Taking $m = n_K$ in (\star) , for all $n > N$ we have $|a_n - L| = |a_n - a_{n_K} + a_{n_K} - L| \leq |a_n - a_{n_K}| + |a_{n_K} - L| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$. Thus $(a_n)_{n \in \mathbb{N}}$ is also convergent with limit L .

From now on, we will revert back to the notations we are used to for numbers from the various number systems. Otherwise, to speak even about the rational number $-1/3$ as a real number, we would think of it as ‘the equivalence class of the constant sequence of the rational numbers $-\frac{1}{3}$, and each term here, namely the rational number $-\frac{1}{3}$, is the equivalence class of the ordered pair of integers $(-1, 3)$, where the integer -1 is the equivalence class of the pair of natural numbers $(2, 1)$ and the integer 3 is \cdots ’. Moreover, all these equivalence relations are on different sets, and we have indicated this below using various colours:

$$(\mathbb{R} \ni) -\frac{1}{3} \text{ is } [[[(2, 1)], [(1, 4)]], [[(2, 1)], [(1, 4)]], [[(2, 1)], [(1, 4)]], \dots]. \quad \textcircled{99}$$

We cannot sensibly proceed by insisting on using the more accurate right-hand side notation. The main point of this chapter so far was to learn precisely about the various number systems, their operations and properties. So everything we are allowed to use has been justified, and if challenged, we know how the number systems are defined, and how their properties are proved. Having established the properties possessed by the number systems, we now carry on by relying on the succinct notation, and sticking to the properties we have justified.

4.6. Irrational numbers and the Rational Zeroes Theorem

While $\mathbb{Q} \subset \mathbb{R}$ and is a smaller field inside the bigger field \mathbb{R} , its complement $\mathbb{R} \setminus \mathbb{Q}$ of irrational numbers does not form a field⁷, because the sum/product of two irrational numbers is not necessarily irrational: $\sqrt{2} + (-\sqrt{2}) = 0 \in \mathbb{Q}$, $\sqrt{2} \cdot \sqrt{2} = 2 \in \mathbb{Q}$. Thus the restriction to irrationals of the real number addition/multiplication are not maps $+: (\mathbb{R} \setminus \mathbb{Q}) \times (\mathbb{R} \setminus \mathbb{Q}) \rightarrow (\mathbb{R} \setminus \mathbb{Q})$ and $\cdot: (\mathbb{R} \setminus \mathbb{Q}) \times (\mathbb{R} \setminus \mathbb{Q}) \rightarrow (\mathbb{R} \setminus \mathbb{Q})$. We will soon see that there are ‘many more irrational numbers than rational numbers’ in the following section. In the present section, we will learn a simple tool, called the Rational Zeroes Theorem, which can be useful for proving irrationality. But we begin by showing the irrationality of the Euler’s number e , which is an important constant in mathematics. In Exercise 2.20, we had defined $e \in \mathbb{R}$ as the limit of the convergent sequence $(1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!})_{n \in \mathbb{N}}$.

Theorem 4.27. $e \notin \mathbb{Q}$.

Proof. Set $a_n := 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!}$ for all $n \in \mathbb{N}$. Fix an $m \in \mathbb{N}$. Then $(a_{m+k})_{k \in \mathbb{N}}$ is a subsequence of $(a_n)_{n \in \mathbb{N}}$, and so is convergent with the same limit e . So $(a_{m+k} - a_m)_{k \in \mathbb{N}}$ is convergent with limit $e - a_m$. But

$$\begin{aligned} \frac{1}{(m+1)!} &\leq a_{m+k} - a_m = \frac{1}{(m+1)!} + \frac{1}{(m+2)!} + \cdots + \frac{1}{(m+k)!} \\ &= \frac{1}{(m+1)!} \left(1 + \frac{1}{m+2} + \cdots + \frac{1}{(m+2) \cdots (m+k)} \right) \\ &\leq \frac{1}{(m+1)!} \left(1 + \frac{1}{2} + \cdots + \frac{1}{2^{k-1}} \right) \\ &= \frac{1}{(m+1)!} \frac{1 - \frac{1}{2^k}}{1 - \frac{1}{2}} = \frac{1}{(m+1)!} \left(2 - \frac{1}{2^{k-1}} \right) < \frac{2}{(m+1)!}. \end{aligned}$$

Passing to the limit as $k \rightarrow \infty$, we obtain

$$\frac{1}{(m+1)!} \leq e - \left(1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{m!} \right) \leq \frac{2}{(m+1)!} \quad (*)$$

The choice of $m \in \mathbb{N}$ was arbitrary, and so $(*)$ holds for all $m \in \mathbb{N}$. Suppose that $e \in \mathbb{Q}$, and write $(0 <) e = \frac{p}{q}$, where $p, q \in \mathbb{N}$. Take any natural number $m > \max\{q, 2\}$, and multiply $(*)$ throughout by $m!$. Since $m!$ contains q as a factor (because $m > q$), we obtain

$$0 < \frac{1}{m+1} \leq \boxed{\text{an integer}} \leq \frac{2}{m+1} < \frac{2}{2+1} < 1,$$

a contradiction. So $e \notin \mathbb{Q}$. □

⁷Recall from page 10 that a field is a set \mathbb{F} together with two maps, addition, $+: \mathbb{F} \times \mathbb{F} \rightarrow \mathbb{F}$, and multiplication $\cdot: \mathbb{F} \times \mathbb{F} \rightarrow \mathbb{F}$, such that: (F1) addition is associative, (F2) there exists an additive identity $\mathbf{0} \in \mathbb{F}$, (F3) every element has an additive inverse, (F4) addition is commutative, (F5) multiplication is associative, (F6) there exists a multiplicative identity $\mathbf{1} \in \mathbb{F} \setminus \{\mathbf{0}\}$, (F7) every element $\neq \mathbf{0}$ has a multiplicative inverse, (F8) multiplication is commutative, and (F9) multiplication distributes over addition.

Exercise 4.60. Show that $\log_2 3$ is irrational.

Exercise 4.61. Show that there exist irrational $a, b \in \mathbb{R}$ such that a^b is rational.

Hint: Consider $\sqrt{2}^{\sqrt{2}}$: The two possibilities are $\sqrt{2}^{\sqrt{2}} \in \mathbb{Q}$ or $\sqrt{2}^{\sqrt{2}} \notin \mathbb{Q}$.
Alternatively, use Exercise 4.60.

Exercise 4.62. Show that $\tan 1^\circ$ is irrational.

Hint: Use $\tan(\alpha + \beta) = \frac{\tan \alpha + \tan \beta}{1 - (\tan \alpha)(\tan \beta)}$ for $\alpha, \beta \in [0, \pi/4)$.

In Theorem 1.1, we had given a ‘geometric’ proof of the irrationality of $\sqrt{2}$. We now learn about a useful result, called the Rational Zeroes Theorem, which gives a tool for showing irrationality, in particular for surds⁸. The proof of the Rational Zeroes Theorem relies on the fact that if the integers m, n have no common factor, and m divides nk , where $k \in \mathbb{Z}$, then m must divide k . This is Proposition 5.6, which will be proved in the next chapter.

A *polynomial (function)* is a map $p : \mathbb{R} \rightarrow \mathbb{R}$ that is a ‘linear combination’ of the power functions, that is, there exists an integer $d \geq 0$, and real numbers c_0, \dots, c_d such that for all $x \in \mathbb{R}$, $p(x) = c_0 + c_1x + \dots + c_dx^d$. Some terminology:

- The numbers c_0, \dots, c_d are called the *coefficients of the polynomial*.
- If $c_d \neq 0$, then d is called the *degree of the polynomial*.
- If $c_d = 1$, then p is called a *monic polynomial*.
- A real number ζ a *real zero of the polynomial* if $p(\zeta) = 0$.
- If $c_0, \dots, c_d \in \mathbb{Z}$, then we say the *polynomial has integer coefficients*.
- The set of all polynomials with integer coefficients is denoted by $\mathbb{Z}[x]$.

Theorem 4.28 (Rational Zeroes Theorem).

Let $d \in \mathbb{N}$, and let c_0, c_1, \dots, c_d be integers such that c_0 and c_d are not zero.

Let $r = \frac{m}{n}$, where m, n are integers having no common factor, and $n > 0$.

Suppose r is a real zero of the polynomial $p = c_0 + c_1x + \dots + c_dx^d \in \mathbb{Z}[x]$.

Then n divides c_d and m divides c_0 .

Proof. We have $c_0 + c_1\frac{m}{n} + \dots + c_d\frac{m^d}{n^d} = 0$. Multiplying throughout by n^d ,

$$c_d m^d = -(c_0 n^d + c_1 m n^{d-1} + \dots + c_{d-1} m^{d-1} n). \quad (4.1)$$

As n divides the right-hand side, n divides $c_d m^d$. But n has no common factors with m , and this implies that n divides c_d . Also by rearranging (4.1), we obtain

$$c_0 n^d = -(c_1 m n^{d-1} + \dots + c_{d-1} m^{d-1} n + c_d m^d),$$

and since m divides the right hand side, m must divide $c_0 n^d$. But m and n have no common factor. So m must divide c_0 . \square

⁸‘Surds’ refer to irrational numbers which arise as the n^{th} root of a natural number. The mathematician al-Khwarizmi (around 820 AD) called irrational numbers ‘inaudible’, which was later translated to the Latin *surdus* for ‘mute’.

Example 4.9 ($\sqrt{2} \notin \mathbb{Q}$). We show that $\sqrt{2}$ is irrational using the Rational Zeroes Theorem. Suppose that $\sqrt{2}$ is rational, and let

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

where $m, n \in \mathbb{Z}$, $n > 0$, and m, n have no common factor. Then $\frac{m}{n}$ is a rational zero of the polynomial $x^2 - 2 \in \mathbb{Z}[x]$. By the Rational Zeroes Theorem, m divides -2 and n (> 0) divides 1. So $m \in \{2, -2, 1, -1\}$ and $n = 1$. Hence

$$\frac{m}{n} \in \{2, -2, 1, -1\}.$$

But $\sqrt{2} = \frac{m}{n}$ is not equal to any of the values $2, -2, 1, -1$ (as $(\pm 2)^2 = 4 \neq 2$ and $(\pm 1)^2 = 1 \neq 2$). This contradiction shows that $\sqrt{2} \notin \mathbb{Q}$. \diamond

Exercise 4.63. Show that $\sqrt[3]{6}$ is irrational using the Rational Zeroes Theorem.

Exercise 4.64. Show that $\sqrt{2} + \sqrt{3}$ is irrational using the Rational Zeroes Theorem.

Exercise 4.65. (*) Show that $(2 + \sqrt{5})^{1/3} - (-2 + \sqrt{5})^{1/3}$ is rational. What is its value?
Hint: Calling the number α , and cubing, show that $\alpha^3 + 3\alpha - 4 = 0$.

Factorise the polynomial $x^3 + 3x - 4$ assuming it has a rational root.

Exercise 4.66.

(1) Show that $\sin \frac{\pi}{5}$ is a real zero of $16x^4 - 20x^2 + 5$.

Hint: de Moivre's Formula and the Binomial Theorem.

Conclude that $2 \sin \frac{\pi}{5}$ is a real zero of the polynomial $p := x^4 - 5x^2 + 5$.

(2) Prove that $2 \sin \frac{\pi}{5}$ is irrational using the Rational Zeroes Theorem.

(3) Show that if a regular pentagon is inscribed in a circle with radius 1, then its side has an irrational length.

Remark 4.5. (*) (Algebraic and transcendental numbers).

Zeroes of nonzero polynomials in $\mathbb{Z}[x]$ are called *algebraic numbers*. It can be shown⁹ that the set of algebraic numbers forms a field. Nonalgebraic numbers are called *transcendental*. It can be proved that e is transcendental¹⁰ [N, Theorem 2.12]. In 1900, Hilbert listed 23 open problems, which proved to be quite influential¹¹ in Mathematics. The 7th one is:

Is a^b transcendental, for algebraic $a \neq 0, 1$ and irrational algebraic b ?

This was settled in the mid-1930s by Gelfond, and independently by Schneider. They showed the following.

Theorem 4.29 (Gelfond-Schneider theorem).

If a, b are algebraic numbers with $a \notin \{0, 1\}$, and $b \notin \mathbb{Q}$, then a^b is transcendental.

The proof (e.g. in [N, Chap. X]) is beyond the scope of the course. \ast

⁹[N, Theorem 7.2].

¹⁰Accepting this allows us to show that $\log_e 2 \notin \mathbb{Q}$, which was mentioned in the Overview on page 2: If $(0 <) \log_e 2 = p/q$ for positive integers p, q , then $2^q = e^p$, so that e^p is an algebraic number, and so e is an algebraic number (why?), but e is transcendental, a contradiction.

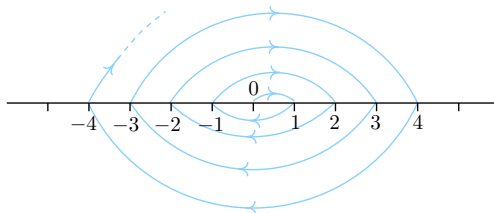
¹¹For example, the 8th problem is the Riemann Hypothesis, a famous unsolved problem at present.

Recall that a set S is *finite* if it is empty or there exists an $n \in \mathbb{N}$ and a bijection $f : \{1, \dots, n\} \rightarrow S$. For finite sets, we can compare sizes by just counting the number of elements, and this is referred to as the *cardinality of the set*: for example, the set $\{A, B, C, \dots, Z\}$ of alphabet letters in the English language has cardinality 26, while the cardinality of $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ is 10. Note that finite sets of the same cardinality can be put in a one-to-one correspondence, that is, we can define a bijection between the two sets. Sets which do not have finite cardinality are called *infinite sets*. For example, the set \mathbb{N} is infinite¹². One can then ask the natural question: Can any two infinite sets always also be put in a one-to-one correspondence? For example, we know that the set \mathbb{N} is infinite, and now suppose we have another infinite set S . Then can we always establish a bijection between the elements of \mathbb{N} and those of S ? In other words can we ‘list’ the elements of S , as the first element of S , the second element of S , and so on? The answer, perhaps surprisingly, is: No! For example, such a bijection fails to exist if we take $S = \mathbb{R}$, and this is the content of Theorem 4.34 below. This motivates the following.

Let S be an infinite set. Then S is said to be *countable* if there is a bijective map from \mathbb{N} onto S . If S is not countable, it is called *uncountable*.

Considering the identity map $n \mapsto n : \mathbb{N} \rightarrow \mathbb{N}$, then we see that \mathbb{N} is countable. \diamond

A nontrivial example is that the set \mathbb{Z} of integers is also countable. This is best seen by means of a picture, as shown in Figure 4.11.



Clearly the resulting map from \mathbb{N} to \mathbb{Z} is injective (since each integer is crossed by the spiral path *only once* ever – having crossed an integer, the subsequent distance of the path to the origin *increases*) and surjective (since every integer will be crossed by the spiral path *sometime*). This argument can be made rigorous¹³, but the idea is clear. We give a different (rigorous) argument in Exercise 4.68 below, which relies on the Fundamental Theorem of Arithmetic (to be proved in the next chapter). \diamond

¹²Firstly, $\mathbb{N} \neq \emptyset$, since $1 \in \mathbb{N}$. Also, if there exists an $n \in \mathbb{N}$ and a bijection $f: \{1, \dots, n\} \rightarrow \mathbb{N}$, then $m := f(1) + \dots + f(n) + 1 > f(i)$ for all $1 \leq i \leq n$. In particular, by the Trichotomy Law, $m \neq f(i)$ for all $1 \leq i \leq n$, showing that $m \in \mathbb{N}$ does not belong to the range of f , contradicting the surjectivity of f .

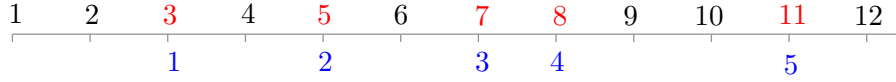
¹³The picture describes the bijective map $f : \mathbb{N} \rightarrow \mathbb{Z}$ given by $f(n) = \begin{cases} \frac{n}{2} & \text{if } n \text{ is even,} \\ \frac{1-n}{2} & \text{if } n \text{ is odd.} \end{cases}$

Exercise 4.67. Let $S \subset \mathbb{N}$ be the set $S = \{m \in \mathbb{N} : \exists n \in \mathbb{N} \text{ such that } m = n^2\}$ of squares. Prove that S is countable.

Next, we will show that the set \mathbb{Q} of rational numbers is countable. To this end, we show the following two auxiliary results, interesting in their own right.

Proposition 4.30. *Every infinite subset of a countable set is countable.*

Proof. Let us first show that any infinite subset S of \mathbb{N} is countable. The idea is quite simple. Imagine the elements of S as coloured red amongst the natural numbers. See the picture below. We start scanning the natural numbers starting from 1 moving rightwards, and the moment we hit a red element, which we label in blue as 1, and we continue scanning till we hit the next red element, which we label in blue as 2, and so on. The map $f : \mathbb{N} \rightarrow S$ is just obtained by sending the blue points to the corresponding read points. We now make this rigorous.



Let $a_1 := \min S$. If $a_1 < \dots < a_k$ have been constructed, then define

$$a_{k+1} := \min(S \setminus \{a_1, \dots, a_k\}).$$

(Since S is not finite, the set $S \setminus \{a_1, \dots, a_k\}$ must be a nonempty subset of \mathbb{N} , and by the Well-ordering Principle, possesses a least element.)

As $S \setminus \{a_1, \dots, a_k\} \subset S \setminus \{a_1, \dots, a_{k-1}\}$ for all $k > 1$, we have $a_{k+1} \geq a_k$. Also, since $a_{k+1} \in S \setminus \{a_1, \dots, a_k\}$, in particular $a_{k+1} \neq a_k$. So $a_{k+1} > a_k$ ($> a_{k-1} > \dots > a_1$). Define $f : \mathbb{N} \rightarrow S$ by $f(n) = a_n$, $n \in \mathbb{N}$. Then f is injective because if $n < m$, then $f(n) < f(m)$.

Also, we claim that f is surjective. Let $m \in S$. As \mathbb{N} is infinite, $f(\mathbb{N})$ cannot be a subset of $\{1, \dots, m\}$ (otherwise, by the Pigeonhole Principle $f(n) = f(n')$ for some $n, n' \in \mathbb{N}$, contradicting injectivity). So there exists an n such that $f(n) > m$. Take the smallest n such that $f(n) \geq m$, and call it n_* . Thus we know $f(n_*) \geq m$ and $f(1), \dots, f(n_* - 1) < m$. Now we show that $f(n_*) \leq m$, which together with $f(n_*) \geq m$ will yield $f(n_*) = m$, establishing surjectivity. As $f(1), \dots, f(n_* - 1) < m$, we have $m \notin \{f(1), \dots, f(n_* - 1)\}$. Thus

$$f(n_*) = a_{n_*} = \min(S \setminus \{a_1, \dots, a_{n_*-1}\}) = \min(S \setminus \{f(1), \dots, f(n_* - 1)\}) \leq m.$$

Justification of the last inequality:

- $m \in S \setminus \{f(1), \dots, f(n_* - 1)\}$ (as $m \in S$ and $m \notin \{f(1), \dots, f(n_* - 1)\}$),
- the minimum $\min(S \setminus \{f(1), \dots, f(n_* - 1)\})$ is less than or equal to each of the members of $S \setminus \{f(1), \dots, f(n_* - 1)\}$, and so in particular the member m .

As $f(n_*) \geq m$ and $f(n_*) \leq m$ together give $f(n_*) = m$. Hence f is surjective. Thus f is bijective. Consequently, any infinite subset S of \mathbb{N} is countable.

Now let S be countable and let T be an infinite subset of S . Let $\varphi : S \rightarrow \mathbb{N}$ be a bijection. There is a bijection from T to the range of $\varphi|_T$.¹⁴ But the range of $\varphi|_T$ is an infinite subset of \mathbb{N} , and so it is countable. Hence T is countable too. \square

Exercise 4.68. Consider $f : \mathbb{Z} \rightarrow \mathbb{N}$ given by

$$f(n) = \begin{cases} 2^n & \text{if } n \geq 0, \\ 3^{-n} & \text{if } n < 0. \end{cases}$$

Prove that f is injective. Conclude that \mathbb{Z} is countable.

Exercise 4.69. An integer n is *even* if there exists a $m \in \mathbb{Z}$ such that $n = 2m$. Show that the set of all even integers is countable.

Exercise 4.70. The aim of this exercise is to show that a finite union of countable sets is countable. It suffices to prove this for just two countable sets, say A and B (Why?). The proof is essentially the same as in Exercise 4.68, where we think of elements of A as the nonnegative integers, and those of B as the negative integers. Here are the details. Let $\alpha : A \rightarrow \mathbb{N}$ and $\beta : B \rightarrow \mathbb{N}$ be bijections. Then consider the map $f : A \cup B \rightarrow \mathbb{N}$ given by

$$f(x) = \begin{cases} 2^{\alpha(x)} & \text{if } x \in A, \\ 3^{\beta(x)} & \text{if } x \in B \setminus (A \cap B). \end{cases}$$

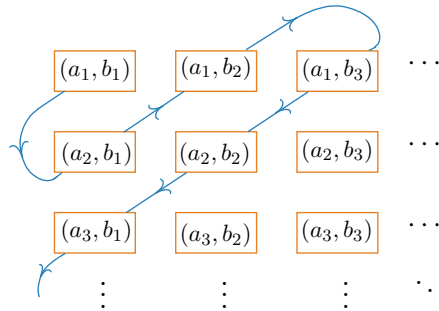
Prove that f is injective, and $f(A \cup B)$ is an infinite set. Conclude that $A \cup B$ is countable.

Proposition 4.31. *If A, B are countable, then $A \times B$ is also countable.*

Proof. Since A and B are countable, we can list their elements:

$$\begin{aligned} A &= \{a_1, a_2, a_3, \dots\}, \\ B &= \{b_1, b_2, b_3, \dots\}. \end{aligned}$$

Arrange the elements of $A \times B$ in an array, and list them following the path shown:



The resulting map from \mathbb{N} to $A \times B$ is clearly surjective, since every element (a_n, b_m) is hit by the zig-zag path *some*time. The map is also injective, since the zig-zag path never hits a point after having crossed it because it moves on to a parallel antidiagonal below. (We will give a different proof in Exercise 4.71 below.) \square

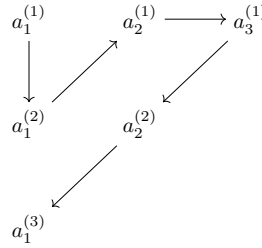
¹⁴Here $\varphi|_T$ denotes the restriction of φ to T . In general, if $f : X \rightarrow Y$ is a function and S is a subset of X , then the *restriction of f to S* is the function $f|_S : S \rightarrow Y$ given by $f|_S(x) = f(x)$ for all $x \in S$.

Exercise 4.71. The aim of this exercise is to give an alternative proof of Proposition 4.31. Let A and B be countable sets, and let $m : A \rightarrow \mathbb{N}$, $n : B \rightarrow \mathbb{N}$ be bijections. Consider the function $f : A \times B \rightarrow \mathbb{N}$ defined by $f(a, b) = 2^{m(a)}3^{n(b)}$, for all $(a, b) \in A \times B$. Show that f is injective. Conclude that $A \times B$ is countable. *Hint:* To show that f is injective, use the Fundamental Theorem of Arithmetic, which will be proved in the following chapter.

Exercise 4.72. (*) The aim of this exercise is to show that a countable union of countable sets is countable: If A_n , $n \in \mathbb{N}$, is a collection of sets such that each A_n is countable, then

$$A := \bigcup_{n \in \mathbb{N}} A_n$$

is countable. Analogous to the proof of Proposition 4.31, it is visually clear that there exists an enumeration of elements using a zig-zag path (where $A_n = \{a_1^{(n)}, a_2^{(n)}, a_3^{(n)}, \dots\}$):



We give a different proof which is more explicit. We proceed as follows. For $n \in \mathbb{N}$, let $f_n : A_n \rightarrow \mathbb{N}$ be a bijection. Now we define a map $f : A \rightarrow \mathbb{N}$ as follows: If $x \in A$, then it belongs to some A_n , and let $n(x)$ be the least/first $n \in \mathbb{N}$ such that $x \in A_n$, and define

$$f(x) = 2^{n(x)}3^{f_{n(x)}(x)}.$$

Prove that f is injective. Conclude that A is countable.

Exercise 4.73. (*) Let A_n , $n \in \mathbb{N}$, be finite sets (some of which may be empty), such that

$$A = \bigcup_{n \in \mathbb{N}} A_n$$

is an infinite set. Prove that A is countable. Proceed as follows. For a nonempty A_n , we denote by m_n its number of elements, and write $A_n = \{a_1^{(n)}, \dots, a_{m_n}^{(n)}\}$. For $x \in A$, let $n(x)$ be the least $n \in \mathbb{N}$ such that $x \in A_n$. As $x \in A$ belongs to $A_{n(x)}$, there exists a unique $k(x) \in \{1, \dots, m_{n(x)}\}$ such that $x = a_{k(x)}^{(n(x))}$. We define $f : A \rightarrow \mathbb{N}$ by

$$f(x) = 2^{n(x)}3^{k(x)} \text{ for all } x \in A.$$

Prove that f is injective. Conclude that A is countable.

Exercise 4.74. (*) Show that $\mathbb{Z}[x]$ is a countable set.

(Since each polynomial of degree d has at most d zeroes, it follows from here that the set of all zeroes of all polynomials in $\mathbb{Z}[x]$ is countable too. Hence the set of algebraic numbers is countable. In particular, the infinite set of all real algebraic numbers is also countable. Since the set of real numbers is uncountable (Theorem 4.34), we conclude that the set of all real transcendental numbers is uncountable.)

We are now ready to show the countability of the rationals.

Theorem 4.32. \mathbb{Q} is countable.

Proof. Each $q \in \mathbb{Q}$ can be written as $q = \frac{n}{d}$, where $n, d \in \mathbb{Z}$, $d > 0$. Among such representations of q , take the smallest possible positive denominator $d =: d_*(q)$, with corresponding numerator $n_*(q)$, that is, $q = \frac{n_*(q)}{d_*(q)}$. So we obtain the map $\mathbb{Q} \ni q \mapsto (n_*(q), d_*(q)) \in \mathbb{Z} \times \mathbb{Z}$, which is injective. But \mathbb{Z} is countable, and so by Proposition 4.31, $\mathbb{Z} \times \mathbb{Z}$ is countable. Consequently, \mathbb{Q} is countable. \square

Exercise 4.75. (*) Show that in the plane \mathbb{R}^2 , the set of all circles whose center lies on \mathbb{Q}^2 (that is, the center is at a point $\mathbf{x} = (r, s)$, whose x - and y -coordinates are rational numbers), and whose radii are rational is a countable set.

Theorem 4.33.

The set $\{f \mid f \text{ is a function from } \mathbb{N} \text{ to } \{0, 1\}\}$, consisting of all $\{0, 1\}$ -valued sequences, is uncountable.

Proof. Suppose there exists an enumeration f_1, f_2, f_3, \dots of these sequences.

(The idea is to arrive at a contradiction by showing that this list misses out on a sequence f , by constructing an f which differs from each of these sequences. A way to construct such an f is to ‘flip/toggle’ the value of the red terms occurring along the diagonal along the diagonal:

$$\begin{array}{rcl} f_1 & = & f_1(1), f_1(2), f_1(3), \dots \\ f_2 & = & f_2(1), f_2(2), f_2(3), \dots \\ f_3 & = & f_3(1), f_3(2), f_3(3), \dots \\ & & \vdots \quad \ddots \end{array}$$

We make this idea precise below.)

We construct an $f : \mathbb{N} \rightarrow \{0, 1\}$ which does not appear in this list. For $n \in \mathbb{N}$, set

$$f(n) = \begin{cases} 0 & \text{if } f_n(n) = 1, \\ 1 & \text{if } f_n(n) = 0. \end{cases}$$

Then

$$\begin{array}{l} f \neq f_1 \text{ since } f(1) \neq f_1(1), \\ f \neq f_2 \text{ since } f(2) \neq f_2(2), \\ f \neq f_3 \text{ since } f(3) \neq f_3(3), \\ \vdots \end{array}$$

showing that f differs from each of f_1, f_2, f_3, \dots , a contradiction. \square

Exercise 4.76. (*) Show that the set of all $\{0, 1\}$ -valued sequences with infinitely many ones is uncountable. *Hint:* Let A (respectively B) be the set of all $\{0, 1\}$ -valued sequences with infinitely many ones (respectively zeroes). What is $A \cup B$?

Exercise 4.77. (*) (How many subsequences for a sequence?)

Let $(a_n)_{n \in \mathbb{N}}$ be a given sequence. Show that $(a_n)_{n \in \mathbb{N}}$ has uncountably many subsequences. *Hint:* Given a subsequence, construct a $\{0, 1\}$ -valued sequence by putting a 1 where the subsequence term appears and 0 whenever it doesn't. Use Exercise 4.76.

Theorem 4.34. \mathbb{R} is uncountable.

Proof. Let $S = \{f \mid f \text{ is a function from } \mathbb{N} \text{ to } \{0, 1\}\}$ be the set of all $\{0, 1\}$ -valued sequences. We will construct an injective map $\varphi : S \rightarrow \mathbb{R}$. Then φ is a bijection from S onto the image $\varphi(S)$. So $\varphi(S)$ is uncountable, since we know from Theorem 4.33 that S is uncountable. But this means that \mathbb{R} is uncountable (for otherwise, its infinite subset $\varphi(S)$ would be countable).

So it remains to construct the promised injective map $\varphi : S \rightarrow \mathbb{R}$. We send an $f \in S$ to the limit of the convergent sequence

$$\left(\frac{f(1)}{3} + \cdots + \frac{f(n)}{3^n} \right)_{n \in \mathbb{N}}.$$

To see that this sequence is convergent, we note that it is increasing and bounded. It is increasing because the difference of the $(n+1)^{\text{st}}$ term and the n^{th} term is $\frac{f(n+1)}{3^{n+1}} \geq 0$. Moreover, boundedness follows from the fact that

$$\frac{f(1)}{3} + \cdots + \frac{f(n)}{3^n} \leq \frac{1}{3} + \cdots + \frac{1}{3^n} = \frac{1}{3} \frac{(1 - \frac{1}{3^n})}{1 - \frac{1}{3}} \leq \frac{1}{3} \frac{1}{\frac{2}{3}} = \frac{1}{2}.$$

Next we show injectivity. Suppose that $f, g \in S$ are distinct sequences. Then there is a smallest $n_* \in \mathbb{N}$ such that $f(n_*) \neq g(n_*)$. Suppose without loss generality that $f(n_*) > g(n_*)$ (otherwise just relabel f, g). Then $f(n_*) = 1$ and $g(n_*) = 0$. If $n_* = 1$, then set $\sigma := 0$, and if $n_* > 1$, set

$$\sigma := \frac{f(1)}{3} + \cdots + \frac{f(n_*-1)}{3^{n_*-1}} = \frac{g(1)}{3} + \cdots + \frac{g(n_*-1)}{3^{n_*-1}}.$$

For $n > n_*$, we have $\frac{f(1)}{3} + \cdots + \frac{f(n_*)}{3^{n_*}} + \frac{f(n_*+1)}{3^{n_*+1}} + \cdots + \frac{f(n)}{3^n} \geq \sigma + \frac{1}{3^{n_*}} + 0$.

Passing to the limit as $n \rightarrow \infty$, we obtain

$$\varphi(f) \geq \sigma + \frac{1}{3^{n_*}}. \quad (\star)$$

For $n > n_*$, we have

$$\begin{aligned} & \frac{g(1)}{3} + \cdots + \frac{g(n_*)}{3^{n_*}} + \frac{g(n_*+1)}{3^{n_*+1}} + \cdots + \frac{g(n)}{3^n} \\ & \leq \sigma + \frac{0}{3^{n_*}} + \frac{1}{3^{n_*+1}} + \cdots + \frac{1}{3^n} \leq \sigma + 0 + \frac{1}{3^{n_*+1}} \frac{(1 - \frac{1}{3^{n-n_*}})}{1 - \frac{1}{3}} \\ & \leq \sigma + \frac{1}{3^{n_*+1}} \frac{1}{\frac{2}{3}} = \sigma + \frac{1}{2 \cdot 3^{n_*}}. \end{aligned} \quad (\star\star)$$

Passing to the limit as $n \rightarrow \infty$, we obtain

$$\varphi(g) \leq \sigma + \frac{1}{2 \cdot 3^{n_*}}. \quad (\star\star\star)$$

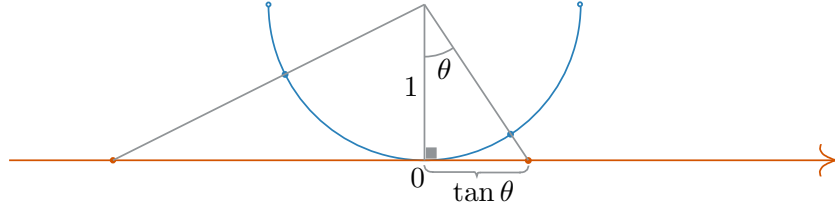
The inequalities (\star) and $(\star\star\star)$ imply that

$$\varphi(f) \geq \sigma + \frac{1}{3^{n_*}} > \sigma + \frac{1}{2 \cdot 3^{n_*}} \geq \varphi(g),$$

so that $\varphi(f) \neq \varphi(g)$. This shows that φ is injective, completing the proof. \square

Example 4.12 (Uncountability of intervals).

Here is a geometric proof of the uncountability of $(-\frac{\pi}{2}, \frac{\pi}{2})$ based on the uncountability of \mathbb{R} , using the picture below, showing a one-to-one correspondence between points of the semicircular arc of radius 1 and the real line:

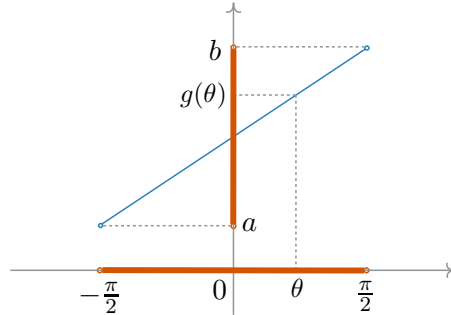


A bijection $f : (-\frac{\pi}{2}, \frac{\pi}{2}) \rightarrow \mathbb{R}$ is given explicitly by $f(\theta) = \tan \theta$ for all $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$. Based on the continuity of \tan on $(-\frac{\pi}{2}, \frac{\pi}{2})$, and the fact that $\tan \theta \rightarrow \pm\infty$ as $\theta \rightarrow \pm\frac{\pi}{2}$, it follows from the Intermediate Value Theorem that f is surjective. It is also injective, because it can be shown that

$$f'(\theta) = \frac{1}{(\cos \theta)^2} > 0,$$

showing that f is strictly increasing on $(-\frac{\pi}{2}, \frac{\pi}{2})$. Hence f is a bijection. Since \mathbb{R} is uncountable, so is $(-\frac{\pi}{2}, \frac{\pi}{2})$.

It follows from here that for any real numbers a, b with $a < b$, the open interval (a, b) is uncountable. This is because there is a bijection $g : (-\frac{\pi}{2}, \frac{\pi}{2}) \rightarrow (a, b)$, e.g. using the following picture:



This bijection g is given explicitly by $g(\theta) = (\theta + \frac{\pi}{2})\frac{(b-a)}{\pi} + a$, $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$. ◇

Exercise 4.78. Show that the set of all irrational numbers is uncountable.

Exercise 4.79. Show that for all $d \in \mathbb{N}$, there is no bijection¹⁵ $\varphi : \mathbb{R} \rightarrow \mathbb{Q}^d$.

Exercise 4.80. One can use the result from Example 4.12 to show that $[0, 1]$ is uncountable. Here we give a different proof. Use

$$\mathbb{R} = \bigcup_{m \in \mathbb{Z}} [m, m+1)$$

and Exercise 4.72 to show that $[0, 1]$ is uncountable.

¹⁵This allows one to conclude that \mathbb{R} is not a ‘finite-dimensional vector space over \mathbb{Q} ’.

Definition 4.18 (Power set of a set).

Let S be a set. The collection of all of the subsets of S is called the *power set* of S . The power set of a set S is denoted by $\mathcal{P}(S)$.

Example 4.13. Let $S = \{a, b, c\}$. Then

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

is the power set of S . If $S = \emptyset$, then $\mathcal{P}(S) = \{\emptyset\}$. ◇

Theorem 4.35 (Cantor).

There does not exist a surjection from a set S onto its power set $\mathcal{P}(S)$.

Proof. Let S be a set and let $f : S \rightarrow \mathcal{P}(S)$ be a surjection. Define the set

$$X = \{x \in S : x \notin f(x)\}.$$

(This makes sense as a set. Indeed, for each $x \in S$, $f(x) \in \mathcal{P}(S)$ is a subset of S , and so we can ask whether or not the element x of S belongs to this subset. X is simply the collection of those elements of S for which $x \notin f(x)$.)

Note that $X \in \mathcal{P}(S)$ as X consists of elements from S (having a certain property).

Now we claim that for all $x \in S$, $f(x) \neq X$. This means that $X \in \mathcal{P}(S)$ fails to be in the range of the map f , showing f is not surjective. We have two possible cases:

- 1° $x \in X$. Then by the definition of X , $x \notin f(x)$. So this element x belongs to X but not in $f(x)$, showing that the two sets $f(x)$ and X can't be the same.
- 2° $x \notin X$. Then $\neg(x \notin f(x))$, i.e., $x \in f(x)$. So this element x doesn't belong to X but belongs to $f(x)$, showing that the two sets $f(x)$ and X can't be the same.

So we have shown that for all $x \in S$, $f(x) \neq X$, completing the proof. □

Corollary 4.36. *There is no bijection between a set and its power set.*

Exercise 4.81. Let S be a set with $n \geq 1$ elements.

- (1) Show that the number of subsets of S with k ($\leq n$) elements is $\binom{n}{k}$.
- (2) Show that $\mathcal{P}(S)$ has 2^n elements. *Hint:* Use the Binomial Theorem to expand $(1 + 1)^n$.
- (3) Deduce that $n < 2^n$ for all $n \in \mathbb{N}$.

Exercise 4.82.

- (1) Show that the set of all two element subsets of \mathbb{N} is countable.
- (2) Show that the set of all nonempty finite subsets of \mathbb{N} is countable.
- (3) Show that the set of all subsets of \mathbb{N} is uncountable.

Remark 4.6. (*) (Continuum hypothesis). For sets A and B , we say that A has *smaller cardinality* than B if there is an injection from A to B , but there is no bijection from A to B . Thus \mathbb{N} has smaller cardinality than \mathbb{R} . It can be shown that there is a bijection between $\mathcal{P}(\mathbb{N})$ and \mathbb{R} . A famous conjecture of set theory¹⁶, advanced by Cantor, was:

¹⁶First problem in Hilbert's list of 23 problems!

There does not exist a set S such that \mathbb{N} has cardinality smaller than S , and S has cardinality smaller than \mathbb{R} .

This is called the *continuum hypothesis* (because \mathbb{R} was called the ‘continuum’). Results by Gödel and Cohen showed that the continuum hypothesis is independent of the ZFC axioms of set theory: It can neither be proved nor disproved within the (best-known and most studied) axiomatic set theory system called the ‘Zermelo-Fraenkel set theory with the axiom of choice’ (ZFC); see e.g. [F]. *

Appendix: Proof of the Least Upper Bound Property (*)

In this appendix, we will show that every nonempty subset of \mathbb{R} which is bounded above has a supremum. This is not examinable material, and may be skipped.

Lemma 4.37 (‘Baby’ Archimedean Principle).

If $\mathbf{x} \in \mathbb{R}$, then there exists a natural number $n \in \mathbb{N}$ such that $n > \mathbf{x}$.

We cannot use the Archimedean Principle to prove the above, since that earlier result was proved using the Least Upper Bound Property of \mathbb{R} , which we haven’t established yet!

Proof. If $\mathbf{x} \leq 0$, then take $n = 1$, since $0 < 1$ gives by transitivity that $\mathbf{x} < 1$.

Let $\mathbf{x} = [(a_n)_{n \in \mathbb{N}}] > 0$. We have seen that every Cauchy sequence in \mathbb{Q} is bounded. So there exists a rational $A > 0$ such that for all $n \in \mathbb{N}$, $a_n \leq A$. This implies $\mathbf{x} < [(A + 1)_{n \in \mathbb{N}}]$ (since $A + 1 - a_n \geq A + 1 - A = 1 > 0$ for all $n \in \mathbb{N}$). Write $A + 1 = [(\frac{p}{q})]$, where $p, q \in \mathbb{N}$. Set $n = p + 1$. Then $A + 1 = [(\frac{p}{q})] < [(\frac{n}{1})]$ (since $p < p + 1 \leq (p + 1)q = nq$). So $\mathbf{x} < [(A + 1)_{n \in \mathbb{N}}] < [(\frac{n}{1})]$. Succinctly, $\mathbf{x} < n$. \square

Lemma 4.38 (Density of \mathbb{Q} in \mathbb{R} redone).

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}$ be such that $\mathbf{x} < \mathbf{y}$. Then there exists an $r \in \mathbb{Q}$ such that $\mathbf{y} < r < \mathbf{x}$.

Proof. As $\mathbf{y} - \mathbf{x} > 0$, we have in particular $\mathbf{y} - \mathbf{x} \neq 0$, and so $(\mathbf{y} - \mathbf{x})^{-1}$ exists in \mathbb{R} . By Lemma 4.37, there exists an $n \in \mathbb{N}$ such that $n > (\mathbf{y} - \mathbf{x})^{-1}$, and so $n(\mathbf{y} - \mathbf{x}) > 1$, i.e., $n\mathbf{x} + 1 < n\mathbf{y}$.

By Lemma 4.37, there exists an $m_1 \in \mathbb{N}$ such that $m_1 > n\mathbf{x}$, and there exists an $m_2 \in \mathbb{N}$ such that $m_2 > -n\mathbf{x}$. So $-m_2 < n\mathbf{x} < m_1$ for some integers m_1, m_2 . Among the finitely many integers $k \in \mathbb{Z}$ such that $-m_2 \leq k \leq m_1$, we take as $\lfloor n\mathbf{x} \rfloor$ the largest one such that it is also $\leq n\mathbf{x}$.

Let $m := \lfloor n\mathbf{x} \rfloor + 1$. Then $\lfloor n\mathbf{x} \rfloor \leq n\mathbf{x} < \lfloor n\mathbf{x} \rfloor + 1$, that is, $m - 1 \leq n\mathbf{x} < m$. So

$$\mathbf{x} < \frac{m}{n} \leq \frac{n\mathbf{x} + 1}{n} < \frac{n\mathbf{y}}{n} = \mathbf{y}.$$

With $r := \frac{m}{n} \in \mathbb{Q}$, the proof is complete. \square

Lemma 4.39. Let $\mathbf{x} = [(a_n)_{n \in \mathbb{N}}] \in \mathbb{R}$. Given any rational $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that for all $n > N$, $a_n + \epsilon \geq \mathbf{x} \geq a_n - \epsilon$.

Proof. Let a rational $\epsilon > 0$ be given. As $(a_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in \mathbb{Q} , there exists an $N \in \mathbb{N}$ such that for all $n, m > N$, $|a_n - a_m| < \epsilon/2$, i.e., $-\frac{\epsilon}{2} < a_n - a_m < \frac{\epsilon}{2}$. Fix an $n > N$. For all $m > N$, $a_n + \epsilon = a_n - a_m + a_m + \epsilon > -\frac{\epsilon}{2} + a_m + \epsilon = a_m + \frac{\epsilon}{2}$, i.e., $(a_n + \epsilon) - a_m > \frac{\epsilon}{2} =: d > 0$. Thus $[(a_n + \epsilon, a_n + \epsilon, a_n + \epsilon, \dots)] > [(a_m)_{m \in \mathbb{N}}] = \mathbf{x}$. For all $m > N$, we also have $a_n - \epsilon = a_n - a_m + a_m - \epsilon < \frac{\epsilon}{2} + a_m - \epsilon = a_m - \frac{\epsilon}{2}$, i.e., $a_m - (a_n - \epsilon) > \frac{\epsilon}{2} > 0$. So $\mathbf{x} = [(a_m)_{m \in \mathbb{N}}] > [(a_n - \epsilon, a_n - \epsilon, a_n - \epsilon, \dots)]$. \square

Lemma 4.40. Let $\mathbf{x} = [(a_n)_{n \in \mathbb{N}}]$. Suppose that there exist $\alpha, \beta \in \mathbb{R}$ and $N \in \mathbb{N}$ such that for all $n > N$, $\alpha \leq a_n \leq \beta$. Then $\alpha \leq \mathbf{x} \leq \beta$.

Proof. By the density of \mathbb{Q} in \mathbb{R} , for each $n \in \mathbb{N}$, there exist $\alpha_n, \beta_n \in \mathbb{Q}$ such that

$$\alpha - \frac{1}{n} < \alpha_n < \alpha, \text{ and } \beta < \beta_n < \beta + \frac{1}{n}.$$

We claim that $(\alpha_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in \mathbb{Q} . Indeed, for any $n, m \in \mathbb{N}$

$$\alpha - \frac{1}{n} < \alpha_n < \alpha, \text{ and } -\alpha < -\alpha_m < -\alpha + \frac{1}{m},$$

which together give $-\frac{1}{n} < \alpha_n - \alpha_m < \frac{1}{m}$. Given any rational $\epsilon > 0$, let $N' \in \mathbb{N}$ be such that $N' > \epsilon^{-1}$. Then for $n, m > N'$,

$$-\frac{1}{N'} < -\frac{1}{n} < \alpha_n - \alpha_m < \frac{1}{m} < \frac{1}{N'},$$

so that $|\alpha_n - \alpha_m| < 1/N' < \epsilon$. So $(\alpha_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in \mathbb{Q} . A similar proof shows that also $(\beta_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in \mathbb{Q} .

We now show that $\alpha = [(\alpha_n)_{n \in \mathbb{N}}]$. To do this we eliminate the other possibilities that $\alpha < [(\alpha_n)_{n \in \mathbb{N}}]$ or $\alpha > [(\alpha_n)_{n \in \mathbb{N}}]$. Let $\alpha = [(\tilde{\alpha}_n)_{n \in \mathbb{N}}]$.

1° Suppose $\alpha < [(\alpha_n)_{n \in \mathbb{N}}]$. Then there exists a rational $d > 0$ and an $M \in \mathbb{N}$ such that for all $n > M$, $\alpha_n - \tilde{\alpha}_n > d$. By Lemma 4.39 with $\epsilon = d/2$, there exists an $M' \in \mathbb{N}$ such that for all $n > M'$, $\tilde{\alpha}_n + \epsilon = \tilde{\alpha}_n + d/2 \geq \alpha$. Thus for $n > M + M'$, we obtain $d < \alpha_n - \tilde{\alpha}_n \leq \alpha_n - \alpha + \frac{d}{2} < 0 + \frac{d}{2} = \frac{d}{2}$, a contradiction.

2° Suppose $\alpha > [(\alpha_n)_{n \in \mathbb{N}}]$. Then there exists a rational $d > 0$ and an $M \in \mathbb{N}$ such that for all $n > M$, $\tilde{\alpha}_n - \alpha_n > d$. By Lemma 4.39 with $\epsilon = d/4$, there exists an $M' \in \mathbb{N}$ such that for all $n > M'$, $\tilde{\alpha}_n - \epsilon = \tilde{\alpha}_n - d/4 \leq \alpha$. Finally, there exists an $M'' \in \mathbb{N}$ such that $M'' > 4/d$. Then for all $n > M + M' + M''$, we have $d < \tilde{\alpha}_n - \alpha_n \leq \frac{d}{4} + \alpha - \alpha_n < \frac{d}{4} + \frac{1}{n} < \frac{d}{4} + \frac{d}{4} = \frac{d}{2}$, a contradiction.

Thus $\alpha = [(\alpha_n)_{n \in \mathbb{N}}]$. In a similar manner, we also have $\beta = [(\beta_n)_{n \in \mathbb{N}}]$.

Since for all $n > N$ we have $a_n - \alpha_n \geq \alpha - \alpha_n > 0$, and $\beta_n - a_n \geq \beta_n - \beta > 0$, it follows from Exercise 4.56 that $[(a_n - \alpha_n)_{n \in \mathbb{N}}] \geq \mathbf{0}$ and $[(\beta_n - a_n)_{n \in \mathbb{N}}] \geq \mathbf{0}$, that is, $\mathbf{x} - \alpha \geq \mathbf{0}$ and $\beta - \mathbf{x} \geq \mathbf{0}$. Rearranging, we obtain $\alpha \leq \mathbf{x} \leq \beta$. \square

Theorem 4.41. Every nonempty subset of \mathbb{R} , bounded above, has a supremum.

Proof. Let $S \subset \mathbb{R}$ be a nonempty subset, which is bounded above. Since S is nonempty, there exists an element $a_0 \in S$, and as S is bounded above, there exists an upper bound $b_0 \in \mathbb{R}$, that is, $a \leq b_0$ for all $a \in S$.



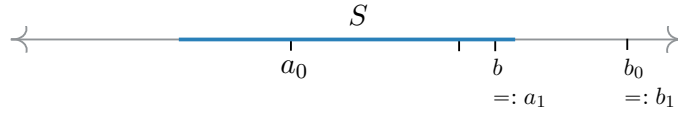
We define a_1, b_1 as follows:

1° If $\frac{a_0+b_0}{2}$ is an upper bound of S , then define $a_1 := a_0$ and $b_1 := \frac{a_0+b_0}{2}$.



Then $a_0 \leq a_1$, $b_0 \geq b_1$, $a_1 \in S$, b_1 is an upper bound of S , $0 \leq b_1 - a_1 \leq \frac{b_0 - a_0}{2}$.

2° If $\frac{a_0+b_0}{2}$ is not an upper bound of S , then there exists a $b \in S$ such that $\frac{a_0+b_0}{2} < b$, and taking any such b , we define $a_1 := b$ and $b_1 = b_0$.



Then $a_0 \leq a_1$, $b_0 \geq b_1$, $a_1 \in S$, b_1 is an upper bound of S , $0 \leq b_1 - a_1 \leq \frac{b_0 - a_0}{2}$.

Suppose for some $n \in \mathbb{N}$,

- $a_0, a_1, \dots, a_{n-1} \in S$ and
- b_0, b_1, \dots, b_{n-1} , upper bounds for S ,

have been constructed such that

- $a_0 \leq a_1 \leq \dots \leq a_{n-1}$
- $b_0 \geq b_1 \geq \dots \geq b_{n-1}$, and
- $0 \leq b_k - a_k \leq \frac{b_0 - a_0}{2^k}$, $k \in \{1, \dots, n-1\}$.

Now we construct a new $a_n \in S$ and a new upper bound b_n of S .

1° If $\frac{a_{n-1}+b_{n-1}}{2}$ is an upper bound of S , then $a_n := a_{n-1}$ and $b_n := \frac{a_{n-1}+b_{n-1}}{2}$.

Then $a_{n-1} \leq a_n$, $b_{n-1} \geq b_n$, $a_n \in S$, b_n is an upper bound of S , and we have

$$0 \leq b_n - a_n = \frac{b_{n-1} - a_{n-1}}{2} \leq \frac{b_0 - a_0}{2 \cdot 2^{n-1}} = \frac{b_0 - a_0}{2^n}.$$

2° If $\frac{a_{n-1}+b_{n-1}}{2}$ is not an upper bound of S , then there exists a $b \in S$ such that $\frac{a_{n-1}+b_{n-1}}{2} < b$, and taking any such b , we define $a_n := b$ and $b_n = b_{n-1}$. Then

$$a_{n-1} = \frac{a_{n-1}+a_{n-1}}{2} \leq \frac{a_{n-1}+b_{n-1}}{2} < b = a_n, \quad b_{n-1} \geq b_n, \quad a_n \in S, \quad b_n \text{ is an upper bound of } S, \text{ and}$$

$$0 \leq b_n - a_n = b_{n-1} - b < b_{n-1} - \frac{a_{n-1}+b_{n-1}}{2} = \frac{b_{n-1} - a_{n-1}}{2} \leq \frac{b_0 - a_0}{2^n}.$$

So we get sequences a_0, a_1, \dots in S , and b_0, b_1, \dots of upper bounds of S , such that

- $a_0 \leq a_1 \leq \dots$,
- $b_0 \geq b_1 \geq \dots$, and
- $0 \leq b_n - a_n \leq \frac{b_0 - a_0}{2^n}$, $n \in \mathbb{N}$.

If for some $n \geq 0$, $a_n = b_n$, then we claim that $u_* := b_n$ is the supremum of S . Indeed, firstly, $u_* = b_n$ is an upper bound of S by construction. Moreover, for any $u < u_* = b_n$, u cannot be an upper bound of S (because $u < u_* = b_n = a_n \in S$).

So we now have to consider the case that for all $n \geq 0$, $a_n < b_n$. By the density of \mathbb{Q} in \mathbb{R} (Lemma 4.38), for each $n \in \mathbb{N}$, there exists an $r_n \in \mathbb{Q}$ such that $a_n < r_n < b_n$. We claim that $(r_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in \mathbb{Q} . To see this, let $\epsilon > 0$ be a given rational number. By the ‘baby’ Archimedean principle (Lemma 4.37), there exists an $N \in \mathbb{N}$ such that $N > \frac{b_0 - a_0}{\epsilon}$, and so

$$\frac{b_0 - a_0}{2^N} \leq \frac{b_0 - a_0}{N} < \epsilon$$

(thanks to the inequality $n < 2^n$ for $n \in \mathbb{N}$: Indeed, we have $1 < 2^1$, and if $n < 2^n$, then $n+1 < 2^n + 1 < 2^n + 2^n = 2 \cdot 2^n = 2^{n+1}$. See also Exercise 4.81.). Now if $n > m > N$, then $a_m < r_m < b_m$, $a_n < r_n < b_n$, $a_n \geq a_m$, which together give

$$r_m - r_n < b_m - r_n < b_m - a_n \leq b_m - a_m.$$

As $b_n \leq b_m$, we have

$$r_m - r_n > a_m - r_n > a_m - b_n \geq a_m - b_m.$$

Hence

$$|r_m - r_n| < b_m - a_m \leq \frac{b_0 - a_0}{2^m} < \frac{b_0 - a_0}{2^N} \leq \frac{b_0 - a_0}{N} < \epsilon.$$

So $(r_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in \mathbb{Q} , and $u_* := [(r_n)_{n \in \mathbb{N}}] \in \mathbb{R}$.

We will now show that u_* is the supremum of S . First, for every fixed m , we have for all $n \geq m$ that $a_m \leq a_n < r_n < b_n \leq b_m$, and so by Lemma 4.40,

$$a_m \leq u_* \leq b_m. \quad (\star)$$

Now suppose that u_* is not an upper bound of S . Then there exists an $a \in S$ such that $a > u_*$. By the density of \mathbb{Q} in \mathbb{R} , there exists an $r \in \mathbb{Q}$ such that

$$0 < r < a - u_*. \quad (\star\star)$$

By the ‘baby’ Archimedean Principle, there exists an $m \in \mathbb{N}$ such that $m > \frac{b_0 - a_0}{r}$. So

$$0 \leq b_m - a_m \leq \frac{b_0 - a_0}{2^m} \leq \frac{b_0 - a_0}{m} < r.$$

Hence using (\star) and $(\star\star)$,

$$b_m < a_m + r \leq u_* + r < a,$$

a contradiction to the fact that b_m is an upper bound of S .

Next, suppose that $u < u_*$. Let $r \in \mathbb{Q}$ be such that $0 < r < u_* - u$. In the same manner as above, there exists an $m \in \mathbb{N}$ such that $0 \leq b_m - a_m < r$. Then

$$a_m > b_m - r \stackrel{(\star)}{\geq} u_* - r > u,$$

showing u is not an upper bound of S . So u_* is the least upper bound of S . \square

Chapter 5

The ring of integers

In Chapter 4, we saw that the set of integers \mathbb{Z} with its operations of integer addition and integer multiplication does not form a field (unlike the rationals and the reals). It fails narrowly from making the cut: Nonzero integers do not necessarily have an integer multiplicative inverse. In general, such a structure is called a ‘ring’ in Mathematics. We will not study rings in general in this course, but focus on the integers and its ring structure. The main topics to be studied in this chapter are:

- The Division Algorithm.
- Euclid’s Algorithm.
- Prime numbers and the Fundamental Theorem of Arithmetic.
- The ring \mathbb{Z}_n and modular arithmetic.

We begin with something we are very familiar with right from elementary school, namely the Division Algorithm.

5.1. The Division Algorithm

Theorem 5.1 (Division Algorithm).

Given integers m, d with $d > 0$, there exist unique integers q and r such that

- $m = qd + r$, and
- $0 \leq r < d$.

Proof. Consider the set $S = \{m - nd : n \in \mathbb{Z} \text{ and } m - nd \geq 0\}$. Then $S \neq \emptyset$: Indeed, if $m \geq 0$, then $m - 0d \in S$, and if $m < 0$, then $m - md = (-m)(d - 1) \geq 0$.

1° If $0 \in S$, then $m - qd = 0 =: r$ for some $q \in \mathbb{Z}$. So $m = qd + 0$, and we are done.

2° If $0 \notin S$, then $S \subset \mathbb{N}$. Thus, by the Well-Ordering Principle, S possesses a least element, say r . Hence $r = m - qd$ for some $q \in \mathbb{Z}$, that is, $m = qd + r$. As $r \in S$, $r \geq 0$. It remains to show that $r < d$. Suppose that $r \geq d$. Then $r' := r - d \geq 0$, and $r' = r - d = m - qd - d = m - (q + 1)d \in S$. As r is a lower bound of S , we have $r \leq r' = r - d$, which yields $d \leq 0$, a contradiction.

Uniqueness: Suppose that $q, q' \in \mathbb{Z}$ and integers r, r' are such that $0 \leq r, r' < d$ and $n = qd + r = q'd + r'$. Let $r \neq r'$. Without loss of generality, let $r' > r$. Then $(q - q')d = r' - r > 0$. As $d > 0$, we must have $q - q' > 0$ (Exercise 4.31). Thus $r' - r = (q - q')d \geq 1d = d$, so that $d > r' \geq r + d \geq 0 + d = d$, a contradiction. So we must have $r' = r$. But then $(q - q')d = r' - r = 0$, and as $d \neq 0$, by Theorem 4.14, $q - q' = 0$, that is, $q = q'$. \square

In the above, we had assumed that the integer d was positive. This wasn't essential:

Corollary 5.2. *Given integers m, d with $d \neq 0$, there exist unique integers q and r such that $m = qd + r$, $0 \leq r < |d|$.*

Proof. We only need to show this when $d < 0$. But then $-d > 0$, and so by Theorem 5.1, there exist $q', r \in \mathbb{Z}$ such that $m = q'(-d) + r$, where $0 \leq r < -d = |d|$. Setting $q = -q' \in \mathbb{Z}$, we get $m = qd + r$, as wanted. To show uniqueness, suppose that $m = qd + r = pd + s$, for some $p, q \in \mathbb{Z}$ and $0 \leq r, s < |d| = -d$. Then we have $m = (-q)(-d) + r = (-p)(-d) + s$. By the uniqueness part of Theorem 5.1, $-q = -p$ (implying $p = q$) and $r = s$. \square

Exercise 5.1.

- (1) Show that all perfect squares (i.e., squares of integers) have the form $4k$ or $4k + 1$ for some $k \in \mathbb{Z}$.
- (2) Show that none of the numbers in the sequence $11, 111, 1111, \dots$ is a perfect square.

5.2. Divisibility, gcd, and the Euclid Algorithm

Definition 5.1 (Divisor/multiple).

Let $m, n \in \mathbb{Z}$. Then we say

- n is a divisor/factor of m , or
- n divides m , or
- m is a multiple of n , or
- m is divisible by n

if there exists an integer $d \in \mathbb{Z}$ such that $m = n \cdot d$. We write

- $n | m$ if n is a divisor of m
- $n \nmid m$ if n is not a divisor of m .

Example 5.1. $3 | 12$, as $12 = 3 \cdot 4$. Similarly $4 | 12$. Also $-3 | 12$ since $12 = (-3)(-4)$. But $0 \nmid 1$. For all $n \in \mathbb{Z}$, $n | 0$ because $0 = n \cdot 0$, and $1 | n$ as $n = 1 \cdot n$. \diamond

Exercise 5.2. Let d, a, b are integers, $d | a$ and $d | b$. Show that for all $x, y \in \mathbb{Z}$, $d | ax + by$.

Exercise 5.3. Let $a, b, c \in \mathbb{Z}$ be such that $a | b$ and $b | c$. Show that $a | c$.

Exercise 5.4. Let $a, b \in \mathbb{Z}$, $a | b$ and $b \neq 0$. Show that $|a| \leq |b|$.

Exercise 5.5. Let $a, b \in \mathbb{Z}$, $a | b$ and $b | a$. Show that $a = b$ or $a = -b$.

Exercise 5.6. (*) Find all integers $n \in \mathbb{Z}$ such that $n^2 + 1$ is divisible by $n + 1$.

Hint: Consider $n^2 - 1$.

Exercise 5.7. Suppose $a \in \mathbb{Z}$ is such that $4 \mid a - 1$. Prove that $4 \mid a^2 + 3$, but $8 \nmid a^2 + 3$.

Exercise 5.8. Show that if $n \in \mathbb{Z}$ is odd, then $16 \mid n^4 + 4n^2 + 11$.

Exercise 5.9. Let $a, b \in \mathbb{Z}$, $k \in \mathbb{N}$. Prove that $a + b \mid a^{2k-1} + b^{2k-1}$. *Hint:* Induction on k .

Exercise 5.10. Show that $1^{2021} + 2^{2021} + 3^{2021} + \dots + 2020^{2021}$ is divisible by 2021.

Exercise 5.11. (*) (Infinite descent).

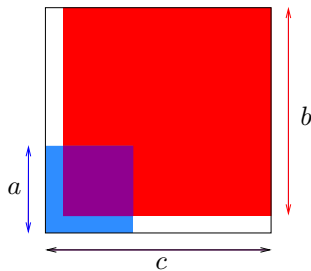
In this exercise we will show that there are no integer solutions to $x^3 + 3y^3 + 9z^3 = 0$ besides the trivial solution $x = y = z = 0$ as an illustration of the ‘method of infinite descent’ (where the idea is that if a statement holds for a bunch of integers, then the same statement is valid for smaller integers, leading to an infinite descent, and ultimately somehow a contradiction to the Well-Ordering Principle). To use the Well-Ordering Principle, we should have a set of integers bounded below, and so we define (with hindsight) the set

$$S := \{x^2 + y^2 + z^2 : (0, 0, 0) \neq (x, y, z) \in \mathbb{Z}^3 \text{ and } x^3 + 3y^3 + 9z^3 = 0\} \subset \mathbb{N}.$$

Suppose S is not empty. By the Well-Ordering Principle, it has a smallest element, say $m \in \mathbb{N}$. Thus $m = x_0^2 + y_0^2 + z_0^2$ for some integers x_0, y_0, z_0 , not all zeros, such that $x_0^3 + 3y_0^3 + 9z_0^3 = 0$. Prove that then x_0, y_0, z_0 are each divisible by 3, allowing a new solution x_1, y_1, z_1 which is also nontrivial and for which $S \ni x_1^2 + y_1^2 + z_1^2 < m$, a contradiction. Conclude that there are no integer solutions to $x^3 + 3y^3 + 9z^3 = 0$ besides the trivial solution $x = y = z = 0$. *Hint:* To show the claimed divisibility by 3, start with x_0 .

Exercise 5.12. (*) (Pythagorean triples). A *Pythagorean triple* is a triple (a, b, c) of natural numbers a, b, c such that $c^2 = a^2 + b^2$. For example, $(3, 4, 5)$, $(5, 12, 13)$ are Pythagorean triples. Pythagorean triples have many interesting number theoretic properties, and the aim of this exercise is to consider two simple ones below.

- (1) Show that if (a, b, c) is a Pythagorean triple, then $(c - b)(c - a)/2$ is a perfect square by considering the following picture.



- (2) Show that in any Pythagorean triple, 3 divides one of the numbers.
- (3) Show that for all $n, m \in \mathbb{N}$ with $n > m$, $(2nm, n^2 - m^2, n^2 + m^2)$ is a Pythagorean triple. Thus we see that there are infinitely many Pythagorean triples. (This is in striking contrast with Fermat's Last Theorem, saying that there are no integer solutions to the equation $x^n + y^n = z^n$ for integer $n > 2$. This statement was mentioned by Fermat in 1637 in the margin of a copy of *Arithmetica* where he claimed he had a proof, but that it was too large to fit in the margin. After over 300 years, a first proof was given by Andrew Wiles in 1995.)

Definition 5.2 (Greatest common divisor).

Let $a, b \in \mathbb{Z}$. A positive integer d is called the *greatest common divisor* of a, b if

- (common divisor) $d \mid a$, $d \mid b$, and
- whenever $d' \in \mathbb{Z}$ is such that $d' \mid a$ and $d' \mid b$, then $d' \mid d$.

We will denote the greatest common divisor of a, b by $\gcd(a, b)$.

Let us show uniqueness, justifying use of the unambiguous notation $\gcd(a, b)$. Suppose that d_1, d_2 both satisfy the conditions demanded. Then as d_2 is a common divisor of a and b , and since d_1 is a greatest common divisor, there holds $d_2 \mid d_1$. Also, since d_1 is a common divisor of a and b , and since d_2 is a greatest common divisor, there holds $d_1 \mid d_2$. It follows from Exercise 5.5 that $d_1 = d_2$ (since both are positive). We now show that the $\gcd(a, b)$ exists whenever not both a, b are zero.

Theorem 5.3. *Suppose that a and b are integers, not both 0. Then $\gcd(a, b)$ exists. Moreover, there exist $x, y \in \mathbb{Z}$ such that¹ $\gcd(a, b) = ax_0 + by_0$.*

Proof. Let $S = \{ax + by : x, y \in \mathbb{Z}\}$. Since at least one of a, b is nonzero, there are nonzero integers on S . Also, if $ax + by \in S$, then $-(ax + by) = a(-x) + b(-y) \in S$. So S contains some positive integers. So the set $S_+ := \{n \in S : n > 0\}$ is nonempty, and by the Well-Ordering Principle, S_+ has a least element d . As $d \in S_+$, we have $d = ax_0 + by_0$ for some $x_0, y_0 \in \mathbb{Z}$, and $d > 0$. We claim that $d = \gcd(a, b)$. (Thus the Bezout equation is then satisfied too: $\gcd(a, b) = d = ax_0 + by_0$.)

For any element $n = ax + by \in S$, by the Division Algorithm, there exist $q \in \mathbb{Z}$ and $r \in \mathbb{Z}$ such that $n = qd + r$ and $0 \leq r < d$. Thus $ax + by = n = qd + r$, that is,

$$0 \leq r = ax + by - q(ax_0 + by_0) = a(x - qx_0) + b(y - qy_0) \in S.$$

If $r \neq 0$, then $0 < r$, so that $r \in S_+$. As $r < d$, we have a contradiction with the definition of d being the least element in S_+ . So $r = 0$. Hence $d \mid n$. So we have shown that d divides each element of S . But as $a = a \cdot 1 + b \cdot 0$ and $b = a \cdot 0 + b \cdot 1$, we have that $a, b \in S$. Hence d divides a and b .

Suppose d' divides a and b . So d' divides $ax + by$ for all $x, y \in \mathbb{Z}$ (Exercise 5.2). In particular, d' divides $ax_0 + by_0 = d$. \square

Suppose $x_0, y_0 \in \mathbb{Z}$ satisfy $\gcd(a, b) = ax_0 + by_0$. Then for any $n \in \mathbb{Z}$, we have that $x := x_0 + bn$, $y := y_0 - an$ also satisfy the Bezout equation:

$$ax + by = a(x_0 + bn) + b(y_0 - an) = ax_0 + abn + by_0 - ban = ax_0 + by_0 = \gcd(a, b).$$

So the coefficients x_0, y_0 of a, b solving the Bezout equation are not unique.

Exercise 5.13. (*) Let $a, b \in \mathbb{Z}$, not both zero, and let $d := \gcd(a, b) = ax_0 + by_0$ for some $x_0, y_0 \in \mathbb{Z}$. Let $x, y \in \mathbb{Z}$. Show that the following are equivalent:

- (1) The integers x, y satisfy $ax + by = d$.
- (2) There exists an $n \in \mathbb{Z}$ such that $x = x_0 + (\frac{b}{d})n$, $y = y_0 - (\frac{a}{d})n$.

¹This equation is called the *Bezout equation*.

Exercise 5.14. Recall the Fibonacci sequence from Exercise 2.19. Show that any two successive Fibonacci numbers are relatively prime. *Hint:* Induction.

Exercise 5.15. Let a, b be integers, not both zero. From the definition of the greatest common divisor, it is clear that $\gcd(a, b) = \gcd(b, a)$ (because the order of a, b doesn't matter). Show also that $\gcd(a, b) = \gcd(-a, b) = \gcd(a, -b) = \gcd(-a, -b)$.

Euclid's Algorithm. To compute the greatest common divisor of integers a, b , not both zero, there is no loss of generality (by Exercise 5.15), in assuming that one of them, say $a > 0$. Moreover, it is clear that for $a > 0$, we have $\gcd(a, 0) = a$, and $\gcd(a, a) = a$. We now learn an algorithm for finding the greatest common divisor supposing $b > a > 0$. The key result is the following, allowing us to use the Division Algorithm successively to keep reducing the numbers.

Lemma 5.4. Let $a, b \in \mathbb{Z}$, with $a \neq 0$. For any $x \in \mathbb{Z}$, $\gcd(a, b) = \gcd(a, b + ax)$.

Proof. As $a \neq 0$, $d := \gcd(a, b)$ and $d' := \gcd(a, b + ax)$ exist. As $d = \gcd(a, b)$, we have $d|a$ and $d|b$. So $d|a$ and $d|b + ax$. Since $d' = \gcd(a, b + ax)$, we have

$$d|d'. \quad (*)$$

As $d' = \gcd(a, b + ax)$, we have $d'|a$ and $d'|b + ax$. So $d'|(b + ax) + a(-x)$, that is, $d'|b$. Since $d = \gcd(a, b)$, and as d' divides a and b , we have

$$d'|d. \quad (**)$$

As $d, d' > 0$, and they divide each other, it follows from Exercise 5.5 that $d = d'$. \square

We now describe Euclid's Algorithm for determining the greatest common divisor. Let a, b be integers, where $b > a > 0$. We divide b by a obtaining integers q_1, r_1 , such that $b = q_1a + r_1$, and $0 \leq r_1 < a$.

1° If $r_1 = 0$, then $\gcd(a, b) = \gcd(a, b - q_1a) = \gcd(a, r_1) = \gcd(a, 0) = a$.

2° If $r_1 > 0$, we repeat the process with $a' := r_1 > 0$ and $b' := a > r_1 = a' > 0$ replacing a, b , respectively.

We note that if $a + b =: N \in \mathbb{N}$, then in the case 2° eventually, we have

$$a' + b' = r_1 + a \leq a - 1 + b - 1 = a + b - 2 = N - 2.$$

So, by the Well-Ordering Principle, we cannot keep obtaining case 2° forever. Hence we must end up in case 1° at some point, and then the previous remainder (that is, the current a) is the wanted gcd. To see this more explicitly, let us suppose that we keep getting case 2° $n - 1$ times for an $n \in \mathbb{N}$ and then in the n^{th} step, we get remainder 0, that is:

$$\begin{array}{ll} b = q_1a + r_1, & 0 < r_1 < a, \\ a = q_2r_1 + r_2, & 0 < r_2 < r_1, \\ \vdots & \vdots \\ r_{n-3} = q_{n-1}r_{n-2} + r_{n-1}, & 0 < r_{n-1} < r_{n-2}, \\ r_{n-2} = q_nr_{n-1}. & \end{array}$$

We have by a repeated application of Lemma 5.4 that

$$\begin{aligned}
 \gcd(a, b) &= \gcd(a, b - q_1 a) = \gcd(a, r_1) = \gcd(r_1, a) \\
 &= \gcd(r_1, a - q_2 r_1) = \gcd(r_1, r_2) = \gcd(r_2, r_1) \\
 &\quad \dots \\
 &= \gcd(r_{n-2}, r_{n-3} - q_{n-1} r_{n-2}) = \gcd(r_{n-2}, r_{n-1}) = \gcd(r_{n-1}, r_{n-2}) \\
 &= \gcd(r_{n-1}, r_{n-2} - q_n r_{n-1}) = \gcd(r_{n-1}, 0) = r_{n-1}.
 \end{aligned}$$

What about expressing $\gcd(a, b)$ as a ‘linear combination’ of a, b as guaranteed to be possible (the Bezout equation)? By writing each remainder in a later step using the ones from the previous step, and working backwards, this can be done. Rather than explain this abstractly, we consider an example.

Example 5.2. Let us find $\gcd(1976, 2375)$. We have

$$\begin{aligned}
 2375 &= 1 \cdot 1976 + 399, \\
 1976 &= 4 \cdot 399 + 380, \\
 399 &= 1 \cdot 380 + 19, \\
 380 &= 20 \cdot 19.
 \end{aligned}$$

Thus $\gcd(1976, 2375) = 19$. To find a solution to the Bezout equation, we start with the penultimate remainder (i.e., $19 = \gcd(1976, 2375)$), and work backwards, each time writing the remainder in terms of the other data, until we reach a ‘linear combination’ with integer coefficients of the given numbers $a = 1976$ and $b = 2375$:

$$\begin{aligned}
 19 &= 399 - 1 \cdot 380 \\
 &= 399 - 1 \cdot (1976 - 4 \cdot 399) \\
 &= -1976 + 5 \cdot 399 \\
 &= -1976 + 5 \cdot (2375 - 1 \cdot 1976) \\
 &= 5 \cdot 2375 - 6 \cdot 1976.
 \end{aligned}$$

Thus $1976 \cdot (-6) + 2375 \cdot 5 = \gcd(1976, 2375)$. In light of Exercise 5.13, we then know *all* the (infinitely many) solutions to the Bezout equation. \diamond

Exercise 5.16. Show that for every $n \in \mathbb{N}$, $\frac{21n+4}{14n+3}$ is irreducible (i.e., the numerator and denominator have no common factor besides 1 or -1). *Hint:* Show $\gcd(21n+4, 14n+3) = 1$.

Definition 5.3 (Relatively prime/copprime integers).

Integers a, b are called *relatively prime* or *coprime* if $\gcd(a, b) = 1$.

Proposition 5.5. Let $a, b \in \mathbb{Z}$. The following are equivalent:

- (1) a, b are relatively prime.
- (2) There exist integers x_0, y_0 such that $ax_0 + by_0 = 1$.

Proof. (1) \Rightarrow (2) follows immediately from Theorem 5.3.

(2) \Rightarrow (1): Suppose that there exist $x_0, y_0 \in \mathbb{Z}$ such that $ax_0 + by_0 = 1$. Then not both a and b are zero (otherwise $1 = ax_0 + by_0 = a0 + b0 = 0$, a contradiction). So $\gcd(a, b)$ exists. Since $\gcd(a, b) \mid a$ and $\gcd(a, b) \mid b$, also $\gcd(a, b) \mid ax_0 + by_0 = 1$. By Exercise 5.4, $(0 <) \gcd(a, b) \leq 1$. So $\gcd(a, b) = 1$, i.e., a, b are relatively prime. \square

Proposition 5.6. *Let $a, b, n \in \mathbb{Z}$ be such that $\gcd(a, n) = 1$ and $n \mid ab$. Then $n \mid b$.*

Proof. Since $\gcd(a, n) = 1$, Proposition 5.5 implies that there exist $x, y \in \mathbb{Z}$ such that $ax + ny = 1$. Multiplying by b , $abx + nyb = b$. As $n \mid ab$, we can write $ab = nm$ for some $m \in \mathbb{Z}$. Hence $b = abx + nyb = nm x + nyb = n(mx + yb)$, and so $n \mid b$. \square

Exercise 5.17. Let $m \in \mathbb{Z}$, $n, c_0 \in \mathbb{Z} \setminus \{0\}$, $d \in \mathbb{N}$. If $\gcd(m, n) = 1$ and $m \mid c_0 n^d$, then $m \mid c_0$.

Exercise 5.18. Let $a, b, n \in \mathbb{Z}$, and $a \mid n$, $b \mid n$ and $\gcd(a, b) = 1$. Show that $ab \mid n$.

Hint: $n = ad$ for a $d \in \mathbb{Z}$, $b \mid n = ad$. Use Proposition 5.6.

Exercise 5.19. Show that if $a, b \in \mathbb{Z}$ are relatively prime, then every $n \in \mathbb{Z}$ can be written as a ‘linear combination of a, b with integer coefficients’, i.e., $n = ax + by$ for some $x, y \in \mathbb{Z}$.

Exercise 5.20. Let $a, b, n \in \mathbb{Z}$ satisfy $\gcd(a, n) = 1 = \gcd(b, n)$. Prove that $\gcd(ab, n) = 1$.

Exercise 5.21. Let $a, b \in \mathbb{Z}$, not both zero. Set $d := \gcd(a, b)$. Show that $\gcd(\frac{a}{d}, \frac{b}{d}) = 1$.

Exercise 5.22. Let a, b be relatively prime. Show that a^2, b^2 are relatively prime too.
Hint: First show that $\gcd(a^2, b) = 1$ by squaring the Bezout equation.

Exercise 5.23. Show that $\gcd(n! + 1, (n+1)! + 1) = 1$ for all $n \in \mathbb{N}$.

Exercise 5.24.

(1) Show that if $a, b, a', b' \in \mathbb{Z}$ and $a + b\sqrt{2} = a' + b'\sqrt{2}$, then $a = a'$ and $b = b'$.

(2) Define a_n, b_n via the relation $a_n + b_n\sqrt{2} = (1 + \sqrt{2})^n$ for all $n \in \mathbb{N}$. Show that a_n, b_n are well-defined, they belong to \mathbb{N} , and that $\gcd(a_n, b_n) = 1$ for all $n \in \mathbb{N}$.

Hint: Use induction.

Exercise 5.25. (*) (Catalan numbers).

(1) Show that for all $n \in \mathbb{N}$, $\gcd(2n+1, n+1) = 1$. *Hint:* $2(n+1) - (2n+1) = 1$.

(2) The numbers $C_n := \frac{1}{n+1} \binom{2n}{n}$, $n \in \mathbb{N}$, are called the *Catalan numbers*. Prove that for all $n \in \mathbb{N}$, $C_n \in \mathbb{N}$. *Hint:* Start by showing that $\frac{2n+1}{n+1} \binom{2n}{n}$ is an integer.

Exercise 5.26. (*) Given an angle of 11° , show that one can divide it into 11 equal parts using a straight edge and a compass. *Hint:* It is enough to construct 1° . Let a° be a known constructible (with straightedge and ruler) angle where $a \in \mathbb{N}$ is relatively prime to 11. There exist $x, y \in \mathbb{Z}$ such that $ax + 11y = 1$, allowing the construction of 1° .

Exercise 5.27. (*) Let $n, N \in \mathbb{N}$ and $n < N$. Prove that $\gcd(2^{2^n} + 1, 2^{2^N} + 1) = 1$.

Exercise 5.28. (*) (Bezout equation in $C[a, b]$). Let $C[a, b]$ denote the set of all continuous functions on the interval $[a, b]$. For an $f \in C[a, b]$, let the *zero set of f* be defined by $Z_f = \{x \in [a, b] : f(x) = 0\}$. Let $\mathbf{1} \in C[a, b]$ be the constant function taking value 1 everywhere. Let $f_1, f_2 \in C[a, b]$. Show that the following are equivalent:

(1) There exist $g_1, g_2 \in C[a, b]$ such that $f_1 g_1 + f_2 g_2 = \mathbf{1}$.

(2) $Z_{f_1} \cap Z_{f_2} = \emptyset$. (That is, f_1, f_2 have no common zero.)

Hint: For (2) \Rightarrow (1), $f_1^2 + f_2^2 > 0$ everywhere, and so $g_i := \frac{f_i}{f_1^2 + f_2^2} \in C[a, b]$, $i = 1, 2$.

5.3. Prime numbers and the Fundamental Theorem of Arithmetic

Definition 5.4 (Prime numbers).

An integer p is called a *prime* if $p > 1$ and if the only positive divisors of p are 1 and p . Natural numbers that are not prime are called *composite numbers*.

Thus 2 is a prime number². This is the only even prime number. The first few terms of the subsequent list of primes are 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, \dots . (We will see below that there are infinitely many primes.)

Proposition 5.7.

If p is a prime number, and $n \in \mathbb{Z}$ is such that $p \nmid n$, then $\gcd(p, n) = 1$.

Proof. Let $d := \gcd(p, n)$. Then $d \mid p$ and $d \mid n$. But p is prime, and so $d = 1$ or $d = p$. But we know that $p \nmid n$, so that d cannot be p . Consequently, $d = 1$. \square

In particular, distinct primes are coprime. Also, for any $n \in \mathbb{Z}$, and any prime p , we have the dichotomy that either $p \mid n$ or otherwise $\gcd(p, n) = 1$.

Exercise 5.29. (*) If p and $8p - 1$ are prime, then show that $8p + 1$ is composite.

Hint: Consider the remainder left by p when divided by 3.

Exercise 5.30.

(1) Let $a, b \in \mathbb{Z}$. Let p be a prime such that $p \mid ab$. Show that $p \mid a$ or $p \mid b$.

Hint: Proposition 5.6

(2) Let $a_1, \dots, a_n \in \mathbb{Z}$ for some $n \geq 2$. Let p be a prime such that $p \mid a_1 \cdots a_n$.

Prove that p divides at least one of the factors a_1, \dots, a_n . *Hint:* Induction on n .

Theorem 5.8 (Fundamental Theorem of Arithmetic).

Every integer $n \neq 0$ can be written as a product $n = cp_1 \cdots p_k$, where $c \in \{1, -1\}$, each p_i is a prime number, and $k \geq 0$. (If $k = 0$, the ‘empty product’ $p_1 \cdots p_k$ is defined to be 1.) This expression is unique except for the ordering of the primes.

Proof.

Existence: It is enough to consider $n \in \mathbb{N}$. We use induction on n . If $n = 1$, then take $c = 1$ and $k = 0$. The claim is true for $n = 2$, since $n = 2 = 1 \cdot 2$, and 2 is a prime number. Suppose each natural number $\leq n$ for some $n \in \mathbb{N}$ possesses the claimed factorisation. If $n + 1$ is a prime number, then we are done. Otherwise, it has a divisor $d \in \mathbb{N}$ which is not 1 or $n + 1$. So $n + 1 = md$ for some $m \in \mathbb{N}$. Then m is not 1 or $n + 1$ either. But then $1 < m = m \cdot 1 < m \cdot d = n + 1$. Also, $1 < d = d \cdot 1 < d \cdot m = n + 1$. Hence $1 < m, d \leq n$. By the induction hypothesis³, $(0 <) m = p_1 \cdots p_k$ and $(0 <) d = q_1 \cdots q_\ell$, where $p_1, \dots, p_k, q_1, \dots, q_\ell$ are primes. So $n + 1 = md = p_1 \cdots p_k q_1 \cdots q_\ell$ is the desired factorisation of $n + 1$. By induction, the proof is complete.

²If $2 = m \cdot n$ for positive integers m and n , then they can't both be equal to 1. Let $n > 1$. Then $1 \leq m = m \cdot 1 < m \cdot n = 2$, giving $1 \leq m < 2$. So $m = 1$ and $n = 2$.

³As $m > 0$, and all primes are > 0 , if $m = cp_1 \cdots p_k$, where $c \in \{-1, 1\}$, then $c = +1$.

Uniqueness: Let S be the set of all integers that fail to have a unique factorisation. Suppose $S \neq \emptyset$. We claim that S contains natural numbers. Indeed, if $0 > m \in S$ has two different factorisations $cp_1 \cdots p_k = m = c'q_1 \cdots q_\ell$, then (as the primes are all positive) $c = c' = -1$, and so $n := -m \in \mathbb{N}$ has two different factorisations $p_1 \cdots p_k = q_1 \cdots q_\ell$. Thus the set S_+ of all natural numbers that fail to have a unique factorisation is not empty as well. By the Well-Ordering principle, S_+ has a least element, say n_* . Let $cp_1 \cdots p_k = n_* = c'q_1 \cdots q_\ell$ where $c, c' \in \{-1, 1\}$ and $p_1, \dots, p_k, q_1, \dots, q_\ell$ are primes, be distinct factorisations of n_* . As $n_* > 0$, $c = c' = 1$. Applying Exercise 5.30 with $p = p_1$, we conclude that p_1 divides some q_i . But as q_i is prime, $p_1 = q_i$. Cancelling p_1 and q_i , we obtain that

$$\mathbb{N} \ni n' := \frac{n_*}{p_1} = \frac{n_*}{q_i} < n_*,$$

showing that n' has the distinct factorisations $p_2 \cdots p_k = n' = q_1 \cdots q_{i-1}q_{i+1} \cdots q_\ell$, contradicting the minimality of n_* . Consequently $S = \emptyset$. \square

Thus every nonzero integer n can be expressed as a product

$$n = cp_1^{\alpha_1} \cdots p_k^{\alpha_k}$$

where $c \in \{-1, 1\}$, $k \geq 0$, $p_1 < p_2 < \cdots < p_k$ are primes, and $\alpha_1, \dots, \alpha_k \geq 0$. By using the exponent zero to raise a prime, we can expand any two numbers using the same list of primes, for example, $20 = 2^2 3^0 5^1$ and $18 = 2^1 3^2 5^0$. We now give a different way of computing the greatest common divisor of two nonzero integers.

Lemma 5.9. *Let $d, a \in \mathbb{N}$, and suppose that $d = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$, $a = p_1^{\beta_1} \cdots p_k^{\beta_k}$, where $p_1 < p_2 < \cdots < p_k$ are primes, and $\alpha_1, \dots, \alpha_k$ and β_1, \dots, β_k are nonnegative integers. If $d|a$, then $\alpha_i \leq \beta_i$ for all $1 \leq i \leq k$.*

Proof. As $p_1^{\alpha_1} \cdots p_k^{\alpha_k} = d|a = p_1^{\beta_1} \cdots p_k^{\beta_k}$, in particular, $p_1^{\alpha_1} | p_1^{\beta_1} \cdots p_k^{\beta_k}$. Suppose $\alpha_1 > \beta_1$. Then $p_1^{\alpha_1 - \beta_1} | p_2^{\beta_2} \cdots p_k^{\beta_k}$, and so in particular, $p_1 | p_2^{\beta_2} \cdots p_k^{\beta_k}$, a contradiction to Exercise 5.30. Hence $\alpha_1 \leq \beta_1$. Similarly, $\alpha_i \leq \beta_i$ for all $2 \leq i \leq k$. \square

Proposition 5.10. *Let $a, b \in \mathbb{Z} \setminus \{0\}$ be given by $a = cp_1^{\alpha_1} \cdots p_k^{\alpha_k}$ and $b = c'p_1^{\beta_1} \cdots p_k^{\beta_k}$, where $c, c' \in \{-1, 1\}$, $k \geq 0$, $p_1 < p_2 < \cdots < p_k$ are primes, and $\alpha_1, \dots, \alpha_k$ and β_1, \dots, β_k are nonnegative integers. Then $\gcd(a, b) = p_1^{\min\{\alpha_1, \beta_1\}} \cdots p_k^{\min\{\alpha_k, \beta_k\}}$.*

Proof. Let $d = \gcd(a, b)$. If p is a prime that divides d , then p also divides a and b . Hence p can only be one of p_1, \dots, p_k by Exercise 5.30. So we can write $d = p_1^{\gamma_1} \cdots p_k^{\gamma_k}$ for some nonnegative $\gamma_1, \dots, \gamma_k$. Lemma 5.9, $\gamma_i \leq \alpha_i$, and $\gamma_i \leq \beta_i$ for $1 \leq i \leq k$. Hence $\gamma_i \leq \min\{\alpha_i, \beta_i\}$ for $1 \leq i \leq k$.

Next, let $d' = p_1^{\min\{\alpha_1, \beta_1\}} \cdots p_k^{\min\{\alpha_k, \beta_k\}}$. Then $d' | a$ (as $\min\{\alpha_i, \beta_i\} \leq \alpha_i$ for all i) and $d' | b$ (as $\min\{\alpha_i, \beta_i\} \leq \beta_i$ for all i). As $d = \gcd(a, b)$, we have $d' | d$. It follows from Lemma 5.9, that $\min\{\alpha_i, \beta_i\} \leq \gamma_i$ for $1 \leq i \leq k$.

Thus $\gamma_i = \min\{\alpha_i, \beta_i\}$ for $1 \leq i \leq k$. So $\gcd(a, b) = p_1^{\min\{\alpha_1, \beta_1\}} \cdots p_k^{\min\{\alpha_k, \beta_k\}}$. \square

The above is mostly of theoretical value, and the method is good to use when we have small numbers, so that we can check if the number has a certain prime factor easily. For example, since $20 = 2^2 3^0 5^1$ and $18 = 2^1 3^2 5^0$, it is immediate that $\gcd(20, 18) = 2^1 3^0 5^0 = 2$.

Example 5.3. Let us recalculate $\gcd(1976, 2375)$ using the above. We have

$$\begin{aligned} 2375 &= 5 \cdot 475 = 5^2 \cdot 95 = 5^3 \cdot 19, \\ 1976 &= 2 \cdot 988 = 2^2 \cdot 494 = 2^3 \cdot 247 = 2^3 \cdot 13 \cdot 19, \end{aligned}$$

and so $\gcd(1976, 2375) = 19^1 = 19$. \diamond

Theorem 5.11. *There are infinitely many primes.*

Proof. Suppose that p_1, \dots, p_n are the only prime numbers. Consider the integer $N = p_1 \cdots p_n + 1$. Then N is not divisible by any of p_1, \dots, p_n . But N cannot be a prime itself, since it is strictly bigger than each of p_1, \dots, p_n . Thus N is not a prime number. By the Fundamental Theorem of Arithmetic, N has a factorisation into primes, and in particular $N (> 1)$ must be divisible by some prime p' . This p' cannot be any of p_1, \dots, p_n , because none of them divide N . This contradicts our assumption that p_1, \dots, p_n were the only primes. \square

Exercise 5.31. (*) Use Exercise 5.27 to show that there are infinitely many primes.

Hint: If not, then for some distinct $n, N \in \mathbb{N}$, $2^{2^n} + 1$, $2^{2^N} + 1$ would share a prime factor.

Exercise 5.32 (Twin primes: Maybe. Prime triples: No!).

(A *twin prime* is an ordered pair $(n, n+2)$, where n and $n+2$ are both primes. For example, $(3, 5)$, $(5, 7)$, $(11, 13)$ are twin primes. A famous open problem is the *Twin Prime Conjecture* stating that there are infinitely many twin primes.) Motivated by this, we define a *prime triple* as a triple $(n, n+2, n+4)$, where n , $n+2$ and $n+4$ are all primes. For example $(3, 5, 7)$ is a prime triple. Show that there are no others! *Hint:* Divide n by 3.

Exercise 5.33. Can a right angled triangle with integer sides have the nonhypotenuse sides equal to twin primes?

Exercise 5.34 (There exist arbitrarily large gaps between primes). Let $n \in \mathbb{N}$. Show that the list of n consecutive numbers given by $(n+1)! + 2, \dots, (n+1)! + n+1$ has no primes.

Exercise 5.35 (Fermat primes).

If $2^n + 1$ is prime for an $n \in \mathbb{N}$, show that n is a power of 2, that is, $n = 2^m$ for some nonnegative integer m . *Hint:* Use Exercise 5.9. (A prime number of the form $2^{2^m} + 1$ is called a *Fermat prime*. Not all numbers of the form $f_m := 2^{2^m} + 1$, $m \in \mathbb{Z}$, are primes. We have $f_0 = 3$, $f_1 = 5$, $f_2 = 17$, $f_3 = 257$, $f_4 = 65537$ are all primes. But Euler showed that $f_5 = 2^{32+1} = 4924967297 = 641 \times 6700417$. It is conjectured that f_0, f_1, f_2, f_3, f_4 are the only Fermat primes.)

Exercise 5.36. (*) To show that $n \in \mathbb{N}$ is a prime number, prove that it is sufficient to check that n is not divisible by all prime numbers p satisfying $p \leq \sqrt{n}$.

Exercise 5.37. Show that if a, b are relatively prime, then ab , $a+b$ are relatively prime.

Hint: Let p be a prime which is a common divisor of ab and $a+b$.

Exercise 5.38. (*) (Least common multiple).

Let $a, b \in \mathbb{N}$. A *least common multiple* of a, b is a positive integer m such that

- $a|m$, $b|m$, and
- whenever $m' \in \mathbb{N}$ is such that $a|m'$ and $b|m'$, we have $m|m'$.

(1) Show that the least common multiple always exists and is unique (denoted by $\text{lcm}(a, b)$).

Hint: Consider the set $S = \{m \in \mathbb{N} : a|m \text{ and } b|m\}$.

(2) Show that $\text{gcd}(a, b) \cdot \text{lcm}(a, b) = ab$.

Hint: First show that $\frac{ab}{\text{gcd}(a, b)} =: m_*$ is an integer, and that both a, b divide m_* . Next if a, b divide an $m \in \mathbb{N}$, then use the Bezout equation (for a, b) to show that $m_*|m$.

(3) Show that if $a = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$ and $b = p_1^{\beta_1} \cdots p_k^{\beta_k}$, where $p_1 < \cdots < p_k$ are primes and $\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_k \in \mathbb{Z}$ are nonnegative, then $\text{lcm}(a, b) = p_1^{\max\{\alpha_1, \beta_1\}} \cdots p_k^{\max\{\alpha_k, \beta_k\}}$.

Exercise 5.39. (*)

(1) Show that if $a, b \in \mathbb{Z}$ and $2a^2 = 3b^2$, then $a = b = 0$.

Hint: Write $a = 2^\alpha 3^\beta A$ and $b = 2^{\alpha'} 3^{\beta'} B$ for nonnegative integers $\alpha, \beta, \alpha', \beta'$ and A, B integers with prime factorisations not containing the primes 2, 3.

(2) Let $a, b, c \in \mathbb{Z}$ be such that $a\sqrt{2} + b\sqrt{3} + c = 0$. Prove that $a = b = c = 0$.

Hint: Square both sides of $a\sqrt{2} + c = -b\sqrt{3}$.

5.4. Modular arithmetic and the ring \mathbb{Z}_n

Definition 5.5 (mod n).

Let $n \in \mathbb{N}$ be fixed. Integers a, b are said to be *congruent modulo n* , if $n|a - b$.

We then write $a \equiv b \pmod{n}$.

Exercise 5.40. Let $n \in \mathbb{N}$. Let $a, b, c, d \in \mathbb{Z}$ be such that $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$. Prove that $(a + c) \equiv (b + d) \pmod{n}$ and $ac \equiv bd \pmod{n}$.

Exercise 5.41. Let $n \in \mathbb{N}$. Let $a, b, c, d \in \mathbb{Z}$ be such that $ab \equiv ac \pmod{n}$ and $\text{gcd}(a, n) = 1$. Show that $b \equiv c \pmod{n}$. Give an example to show that the hypothesis $\text{gcd}(a, n) = 1$ is not superfluous.

Proposition 5.12. Let $n \in \mathbb{N}$.

(1) *Congruent modulo n is an equivalence relation on \mathbb{Z} .*

(2) *For this equivalence relation the number of distinct equivalence classes is n .*

Proof.

(1) Congruence modulo n is reflexive, symmetric, and transitive:

Reflexivity: For all $a \in \mathbb{Z}$, $a \equiv a \pmod{n}$ because $n|0 = a - a$.

Symmetry: Let $a, b \in \mathbb{Z}$ be such that $a \equiv b \pmod{n}$. Then $n|a - b$.

Thus $n|-(a - b) = b - a$. Hence $b \equiv a \pmod{n}$.

Transitivity: Let $a, b, c \in \mathbb{Z}$ be such that $a \equiv b \pmod{n}$ and $b \equiv c \pmod{n}$.

Thus $n|a - b$ and $n|b - c$. So $n|(a - b) + (b - c) = a - c$. Hence $a \equiv c \pmod{n}$.

(2) We claim that $[0], [1], \dots, [n-1]$ are the only equivalence classes, and that these are distinct. If $a \in \mathbb{Z}$, then dividing a by n , by the Division Algorithm, there exist integers q and r such that $a = qn + r$, where $0 \leq r < n$. Because $n \mid qn = a - r$, we have $a \equiv r \pmod{n}$, i.e., $[a] = [r] \in \{[0], [1], \dots, [n-1]\}$.

Now we will show that the equivalence classes $[0], \dots, [n-1]$ are distinct. Suppose on the contrary that $0 \leq r < r' \leq n-1$ and $[r] = [r']$. Then we have $r' \equiv r \pmod{n}$, and so $n \mid r' - r$. So there exists a $q \in \mathbb{Z}$ such that $0 < r' - r = qn$. Thus $q > 0$, i.e., $q \geq 1$. So $r' - r = qn \geq n$. Hence $n \leq r' - r \leq (n-1) - 0 = n-1$, which gives $1 \leq 0$, a contradiction. \square

Exercise 5.42. Let $n \in \mathbb{N}$ and $a \in \mathbb{Z}$. Show that $[a] = \{a + qn : q \in \mathbb{Z}\}$.

Let \mathbb{Z}_n denote the set of equivalence classes under the relation of congruency modulo n on \mathbb{Z} , that is,

$$\mathbb{Z}_n = \{[0], [1], \dots, [n-1]\}.$$

Note that each of the members is actually a set:

$$\begin{aligned} [a] &= \{b \in \mathbb{Z} : b \equiv a \pmod{n}\} \\ &= \{a + qn : q \in \mathbb{Z}\} \\ &= \{\dots, a-3n, a-2n, a-n, a, a+n, a+2n, a+3n, \dots\}. \end{aligned}$$

For example, if $n = 1$, then $\mathbb{Z}_1 = \{[0]\}$ and

$$[0] = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\} = \mathbb{Z}.$$

If $n = 2$, then $\mathbb{Z}_2 = \{[0], [1]\}$, where

$$\begin{aligned} [0] &= \{2q : q \in \mathbb{Z}\} = \{\text{even integers}\} = \{\dots, -6, -4, -2, 0, 2, 4, 6, \dots\}, \\ [1] &= \{2q + 1 : q \in \mathbb{Z}\} = \{\text{odd integers}\} = \{\dots, -5, -3, -1, 1, 3, 5, \dots\}. \end{aligned}$$

We now define addition and multiplication in \mathbb{Z}_n making it a ‘ring’.

Definition 5.6 (Addition and multiplication modulo n).

Let $n \in \mathbb{N}$. We define $+: \mathbb{Z}_n \times \mathbb{Z}_n \rightarrow \mathbb{Z}_n$ and $\cdot: \mathbb{Z}_n \times \mathbb{Z}_n \rightarrow \mathbb{Z}_n$ by

$$\begin{aligned} [a] + [b] &= [a + b] \text{ and} \\ [a] \cdot [b] &= [ab] \end{aligned}$$

for all $[a], [b] \in \mathbb{Z}_n$.

We need to check the well-definition, that is, if $[a] = [a']$ and $[b] = [b']$, then $[a + b] = [a' + b']$ and $[ab] = [a'b']$. We have $a' \equiv a \pmod{n}$ and $b' \equiv b \pmod{n}$, and so by Exercise 5.40, we have $(a' + b') \equiv (a + b) \pmod{n}$ and $a'b' \equiv ab \pmod{n}$. Thus $[a + b] = [a' + b']$ and $[ab] = [a'b']$.

Using the arithmetic properties of addition and multiplication in \mathbb{Z} , it is easy to see that the following properties hold for all $[a], [b], [c] \in \mathbb{Z}_n$:

- $[a] + ([b] + [c]) = ([a] + [b]) + [c]$.
- $[a] + [0] = [a] = [0] + [a]$.
- $[a] + [-a] = [0] = [-a] + [a]$.
- $[a] + [b] = [b] + [a]$.
- $[a] \cdot ([b] \cdot [c]) = ([a] \cdot [b]) \cdot [c]$.
- $[a] \cdot [1] = [a] = [1] \cdot [a]$.
- $[a] \cdot [b] = [b] \cdot [a]$.
- $[a] \cdot ([b] + [c]) = [a] \cdot [b] + [a] \cdot [c]$.

As an example, let us show the last one, namely the distributive property:

$$\begin{aligned} [a] \cdot ([b] + [c]) &= [a] \cdot [b + c] = [a(b + c)] = [ab + ac] \\ &= [ab] + [ac] = [a] \cdot [b] + [a] \cdot [c]. \end{aligned}$$

In the above list, while we have stated that additive inverses always exist, we have not said anything about the existence of multiplicative inverses. Not every element in \mathbb{Z}_n may have a multiplicative inverse. But if a is relatively prime to n , then $[a] \in \mathbb{Z}_n$ does possess a multiplicative inverse. In particular, if $n = p$, a prime, then every nonzero element of \mathbb{Z}_p has a multiplicative inverse in \mathbb{Z}_p ; see Corollary 5.14.

Proposition 5.13. *Suppose that $n \in \mathbb{N}$, and that $a \in \mathbb{Z}$ is relatively prime to n . Then there exists an $x \in \mathbb{Z}$ such that $[a] \cdot [x] = [1] = [x] \cdot [a]$ in \mathbb{Z}_n .*

Proof. As a is relatively prime to n , i.e., $\gcd(a, n) = 1$, by Proposition 5.5, there exist $x, y \in \mathbb{Z}$ such that $ax + ny = 1$. As $n \equiv 0 \pmod n$, $[0] = [n]$ in \mathbb{Z}_n . Hence

$$\begin{aligned} [1] &= [ax + ny] = [ax] + [ny] \\ &= [a] \cdot [x] + [n] \cdot [y] \\ &= [a] \cdot [x] + [0] \cdot [y] \\ &= [a] \cdot [x] + [0y] \\ &= [a] \cdot [x] + [0] \\ &= [a] \cdot [x]. \end{aligned}$$

As multiplication in \mathbb{Z}_n is commutative, also $[x] \cdot [a] = [1]$. □

If a multiplicative inverse of $[a] \in \mathbb{Z}_n$ exists, then it must be unique: Indeed, if $[x], [y] \in \mathbb{Z}_n$ are such that $[a] \cdot [x] = [x] \cdot [a] = [1]$ and $[a] \cdot [y] = [y] \cdot [a] = [1]$, then

$$[x] = [1] \cdot [x] = ([y] \cdot [a]) \cdot [x] = [y] \cdot ([a] \cdot [x]) = [y] \cdot [1] = [y].$$

Thus $[x] = [y]$ in \mathbb{Z}_n . We denote the unique multiplicative inverse of $[a]$ in \mathbb{Z}_n , if it exists, by the symbol $[a]^{-1}$.

Corollary 5.14. *If p is a prime, then \mathbb{Z}_p is a field.*

Proof. We only have to show that every nonzero element in \mathbb{Z}_p has a multiplicative inverse. Suppose $a \in \mathbb{Z}$ is such that $[a] \neq [0]$ in \mathbb{Z}_p . Then $p \nmid a - 0 = a$. By Proposition 5.7, $\gcd(a, p) = 1$, that is, a, p are relatively prime. From Proposition 5.13, it follows that there exists an $x \in \mathbb{Z}$ such that $[a] \cdot [x] = [1] = [x] \cdot [a]$ in \mathbb{Z}_p . \square

Lemma 5.15. *Let p be a prime, and $a \in \mathbb{Z}$. Then $[a]^2 = [1]$ in \mathbb{Z}_p if and only if $[a] = [1]$ or $[a] = [-1]$ in \mathbb{Z}_p .*

Proof. ‘If’ part: We have $[1]^2 = [1] \cdot [1] = [1]$, and $[-1]^2 = [-1(-1)] = [1]$.

‘Only if’ part: Suppose that $[a]^2 = [1]$, that is, $[a^2] = [1]$. So $a^2 \equiv 1 \pmod{p}$, that is $p \mid a^2 - 1 = (a - 1)(a + 1)$. By Exercise 5.30, $p \mid a - 1$ or $p \mid a + 1$. Consequently, $a \equiv 1 \pmod{p}$ or $a \equiv -1 \pmod{p}$, that is, $[a] = [1]$ or $[a] = [-1]$. \square

Lemma 5.16. *Let $p > 3$ be a prime. Then $[2] \cdot [3] \cdots [p - 3] \cdot [p - 2] = [1]$ in \mathbb{Z}_p .*

Proof. As p is odd, there are an even number of terms in the product on the left-hand side. For each of the elements $[a]$ in $\{[2], \dots, [p - 2]\}$, we have that $[a] \neq [1]$ and $[a] \neq [-1]$, and so, by Lemma 5.15, $[a]^2 \neq [1]$. So each term $[a]$ in $[2], \dots, [p - 2]$ pairs up together with another term (its inverse $[a]^{-1}$) in the same list $[2], \dots, [p - 2]$, such that their product gives $[1]$. This proves the claim. \square

Theorem 5.17. *Let p be a prime. Then $[1] \cdot [2] \cdots [p - 2] \cdot [p - 1] = [-1]$ in \mathbb{Z}_p .*

Proof. If $p = 2$, then $[1] = [-1]$ in \mathbb{Z}_2 :

$$1 \equiv -1 \pmod{2} \text{ (as } 2 \mid 1 - (-1) = 2\text{)}.$$

If $p = 3$, then $[1][2] = [2] = [-1]$ in \mathbb{Z}_3 :

$$2 \equiv -1 \pmod{3} \text{ (as } 3 \mid 2 - (-1) = 3\text{)}.$$

If $p > 3$, then by Lemma 5.16, $[2] \cdots [p - 2] = [1]$. Hence

$$[1] \cdot [2] \cdots [p - 2] \cdot [p - 1] = [1] \cdot [1] \cdot [p - 1] = [p - 1] = [-1].$$

This completes the proof. \square

Corollary 5.18 (Wilson’s Theorem).

Let p be a prime. Then $(p - 1)! \equiv -1 \pmod{p}$, that is, $p \mid (p - 1)! + 1$.

Proof. From Theorem 5.17,

$$\begin{aligned} [(p - 1)!] &= [1 \cdot 2 \cdot 3 \cdots (p - 2) \cdot (p - 1)] \\ &= [1] \cdot [2] \cdots [p - 2] \cdot [p - 1] = [-1]. \end{aligned}$$

So $(p - 1)! \equiv -1 \pmod{p}$, that is, $p \mid (p - 1)! + 1$. \square

Exercise 5.43. In Remark 1.1 (p.29), we had seen that if $n \geq k \geq 0$, then

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \in \mathbb{Z}.$$

The aim of this exercise is to show that the product of any k consecutive integers is divisible by $k!$. In other words, for any $n \in \mathbb{Z}$ and all $k \in \mathbb{N}$, the product $n(n-1)\cdots(n-k+1)$ is divisible by $k!$. Proceed as follows. Choose a large enough $m \in \mathbb{N}$ such that $n+m \cdot (k!) \geq k$. Set $N = n+m \cdot (k!) \in \mathbb{N}$. From Remark 1.1, we have $N(N-1)\cdots(N-k+1) \equiv 0 \pmod{(k!)}$. Conclude that $n(n-1)\cdots(n-k+1) \equiv 0 \pmod{(k!)}$. *Hint:* $(N-\ell) \equiv (n-\ell) \pmod{(k!)}$.

Exercise 5.44 (ISBN numbers). An International Standard Book Number (ISBN) is a number used for the purpose of uniquely identifying books. For instance, the ISBN number of the book *I want to be a Mathematician* by Paul Halmos is 0-387-96078-3. The initial digit 0 indicates that the book is published in an English-speaking region. The next block 387 identifies the publisher. The third block 96078 is assigned by the publisher and identifies this book. The last digit, 3, is called the *check digit*. Sometimes the check digit happens to be 'X', in which case, it has the numerical value 10. A valid ISBN number $d_1d_2\cdots d_9d_{10}$ has a check digit which satisfies $d_{10} \equiv (d_1 + 2d_2 + 3d_3 + \cdots + 9d_9) \pmod{11}$. The check digit is used for 'error detection', to find out if one of the two most common errors, namely

- typing a digit wrong, or
- transposing adjacent digits,

has occurred. For example, 0-387-96078-3 is a valid ISBN number, since we have that $1 \cdot 0 + 2 \cdot 3 + 3 \cdot 8 + 4 \cdot 7 + 5 \cdot 9 + 6 \cdot 6 + 7 \cdot 0 + 8 \cdot 7 + 9 \cdot 8 = 267 = 264 + 3 = 11 \cdot 24 + 3$.

- (1) Show that if any two adjacent digits among $d_1 \cdots d_9$ are swapped, then the check digit flags that an error has been committed.
- (2) Prove that if any one digit $d_1 \cdots d_9$ is incorrect, then the check digit flags that an error has been committed.

Exercise 5.45. (*) What are the last two digits of 3^{1234} ? *Hint:* Show $3^{20} \equiv 1 \pmod{100}$.

Exercise 5.46.

- (1) Show that for any $m \in \mathbb{Z}$, $m^3 \equiv n \pmod{9}$ for some $n \in \{-1, 0, 1\}$.
- (2) Show that there is no solution in integers to $x^3 + y^3 + z^3 = 2020$.

Exercise 5.47. Solve the equation $[2] \cdot x = [3]$ in \mathbb{Z}_5 .

Exercise 5.48. (*) (Chinese Remainder Theorem).

- (1) Show that there is no integer x such that

$$\begin{aligned} x &\equiv 4 \pmod{6}, \\ x &\equiv 5 \pmod{10}. \end{aligned}$$

Hint: Consider the 'parity of x ' (that is, the cases x is even/odd).

- (2) Let m, n be relatively prime, and $a, b \in \mathbb{Z}$. Show that there exists an $x \in \mathbb{Z}$ such that

$$\begin{aligned} x &\equiv a \pmod{m}, \\ x &\equiv b \pmod{n}. \end{aligned}$$

Moreover show that for any two solutions x, x' , we have $x' \equiv x \pmod{mn}$.

Hint: Let $x_0, y_0 \in \mathbb{Z}$ satisfy $mx_0 + ny_0 = 1$. Multiply by $b-a$ to get $mk + n\ell = b-a$, and set $x = a + mk = b - n\ell$.

Bibliography

- [B] Victor Bryant.
Yet another introduction to analysis.
Cambridge University Press, 1990.
- [C] Leon Cohen and Gertrude Ehrlich.
The structure of the real number system.
Van Nostrand Reinhold Company, 1963.
- [F] Solomon Feferman.
Does mathematics need new axioms?
American Mathematical Monthly, 106:99-111, no. 2, 1999.
- [N] Ivan Niven.
Irrational numbers. The Carus Mathematical Monographs, No. 11.
The Mathematical Association of America. Sixth printing, 2006.
- [R] Walter Rudin.
Principles of mathematical analysis. Third edition.
McGraw-Hill, 1976.
- [S] Michael Spivak.
Calculus. Third edition.
Cambridge University Press, 2006.

Index

- $\mathbb{Z}[x]$, 113
- k th power of a function, 72
- mod , 137

- absolute value, 24
- absolute value of a function, 72
- al-Khwarizmi, 113
- Archimedean property, 20

- Bezout equation, 130
- Bolzano-Weierstrass theorem, 61
- bounded above, set, 13
- bounded below, set, 13
- bounded sequence, 41
- bounded set, 14

- Catalan number, 133
- Chinese Remainder Theorem, 141
- compact interval, 24
- composition of functions, 74
- congruent modulo, 137
- continuity of a function at a point, 66
- continuous function, 66
- convergent sequence, 35
- coprime integers, 132
- countable set, 115

- decreasing sequence, 43
- degree of a polynomial, 113
- distance, 24
- divergent sequence, 35
- divides, 128
- Division Algorithm, 127
- divisor, 128

- equivalence class, 86
- equivalence relation, 85

- Euler's number, e , 47, 52
- Extreme Value Theorem, 79

- factor, 128
- Fermat, 129
- Fermat prime, 136
- Fermat's Last Theorem, 129
- Fibonacci sequence, 52
- field axioms, 10
- Fourier, 67
- fractional part of a real number, 61
- Fundamental Theorem of Arithmetic, 134

- geometric progression, 54
- golden ratio, 9
- greatest common divisor, 130
- greatest integer part, 22
- greatest upper bound, 15

- Heaviside function, 68

- increasing sequence, 43
- infimum, 15
- infinite descent, 129
- integers (definition), 94
- intermediate value theorem, 76
- interval, 23
- ISBN number, 141

- least common multiple, 137
- least upper bound, 15
- least upper bound property, 17
- limit inferior, \liminf , 47
- limit superior, \limsup , 47
- lower bound, 13

- maximum, 17

-
- minimum, 17
 - modular arithmetic, 138
 - modulus, 24
 - monic polynomial, 113
 - monotone sequence, 43
 - multiple, 128

 - negative integers, 100
 - number line, 6

 - open interval, 23
 - order axiom, 11
 - order relation, 11
 - ordered field, 110

 - Peano axioms, 88
 - periodic function, 83
 - polynomial function, 73
 - positive definiteness, 26
 - positive integers, 100
 - positive numbers, 11
 - power set of a set, 122
 - prime number, 134
 - product of functions, 72
 - product of integers (definition), 97
 - Pythagorean triple, 129

 - rational function, 73
 - rational number, 101
 - Rational Zeroes Theorem, 113
 - recursive definition, 89
 - reflexive relation, 85
 - relation, 85
 - relatively prime integers, 132
 - restriction of a function, 117

 - Sandwich theorem, 53
 - seaview property, 60
 - sequence, 31
 - subsequence, 56
 - successor, 88
 - successor function, 88
 - sum of functions, 72
 - supremum, 15
 - surds, 113
 - symmetric relation, 85
 - symmetry, 26

 - transitive relation, 85
 - triangle inequality, 25, 26
 - Trichotomy Law, 91, 100, 109

 - uncountable set, 115
 - upper bound, 12

 - vacuous logic, 13