2016/17

Department of
Mathematics

MA103

Introduction to Abstract Mathematics

Second part, Analysis and Algebra

Amol Sasane

Revised by Jozef Skokan, Konrad Swanepoel, and Graham Brightwell

Contents

Chapter 1

Analysis

Analysis is the *theory* behind real numbers, sequences, and functions. The word 'theory' is important. You might, for example, have a good idea of what we mean by a 'limit' of a convergent sequence of numbers, or the notion of a 'continuous' function, but in this part of the course we aim to formalize such notions.

1.1 The real numbers

The rational number system is inadequate for many purposes. For instance, there is no rational number q such that $q^2 = 2$, and the set

$$S = \left\{ q \in \mathbb{Q} \;\middle|\; q^2 \leq 2 \right\}$$

does not have a largest element in $\mathbb{Q}$. So we see that the rational number system has "gaps". The real number system $\mathbb{R}$ includes numbers to fill all of these gaps. Thus the set

$$T = \left\{ x \in \mathbb{R} \;\middle|\; x^2 \leq 2 \right\}$$

has a largest element. This is a consequence of a very important property of the real numbers, called the least upper bound property. Before we state this property of $\mathbb{R}$, we need a few definitions.

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

1. *An element $u \in \mathbb{R}$ is said to be an **upper bound of S** if, for all $x \in S$, $x \leq u$. If S has an upper bound, then we also say that S is **bounded above**.*

2. *An element $l \in \mathbb{R}$ is said to be a **lower bound of S** if, for all $x \in S$, $l \leq x$. If S has a lower bound, then we also say that S is **bounded below**.*

3. *The set S is said to be **bounded** if it is bounded above and bounded below.*

Examples.

1. The set $S = \{x \in \mathbb{R} \mid 0 \leq x < 1\}$ is bounded. Any real number y satisfying $y \geq 1$ (for instance 1, π, or 100) is an upper bound of S, and any real number z satisfying $z \leq 0$ (for instance 0, or -1) is a lower bound of S.

2. The set $S = \{n \mid n \in \mathbb{N}\}$ is bounded below; indeed, any real number $x \leq 1$ serves as a lower bound. It has no upper bound – we shall give a formal proof of this shortly – and so it is not bounded above, and not bounded.

3. The set $S = \{(-1)^n \mid n \in \mathbb{N}\}$ is bounded. Note that $S = \{-1, 1\}$. It is bounded above by 1 and bounded below by -1.

   More generally, any finite set S is bounded.

4. The set $S = \left\{\frac{1}{n} \;\middle|\; n \in \mathbb{N}\right\}$ is bounded. Any real number x satisfying $1 \leq x$ is an upper bound, and 0 is a lower bound.

5. The sets $\mathbb{Z}$ and $\mathbb{R}$ are neither bounded above nor bounded below.

6. The set $T = \{x \in \mathbb{R} \mid x^2 \leq 2\}$ is bounded. Any real number x satisfying $x^2 \leq 2$ also satisfies $x^2 < 4$. Therefore, $x^2 - 4 < 0$, that is, $(x-2)(x+2) < 0$. It follows that

   either $x - 2 < 0$ and $x + 2 > 0$,

   or $x - 2 > 0$ and $x + 2 < 0$.

   The second case is impossible. Thus the first case has to hold, that is $-2 < x < 2$. It follows that T is bounded above by 2 and bounded below by -2.

7. The set $\emptyset$ is bounded. For instance, 1 is an upper bound for $\emptyset$, since the condition "for every element x of $\emptyset$, $x \leq 1$" is satisfied: there is certainly no element x of $\emptyset$ that **doesn't** satisfy $x \leq 1$. Indeed, every real number is an upper bound for $\emptyset$, and similarly every real number is a lower bound for $\emptyset$. □

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 $\mathbb{R}$.

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

1. *An element $u_* \in \mathbb{R}$ is said to be a **least upper bound of S** (or a **supremum of S**) if*

   a) *u_* is an upper bound of S, and*

   b) *for each upper bound u of S, it holds that $u_* \leq u$.*

2. *An element $l_* \in \mathbb{R}$ is said to be a **greatest lower bound of S** (or an **infimum of S**) if*

   a) *l_* is a lower bound of S, and*

   b) *for each lower bound l of S, it holds that $l \leq l_*$.*

Example. *Set* $S = \{x \in \mathbb{R} \mid x < 1\}$. *Show that the supremum of* S *is* 1.

Solution. Clearly 1 is an upper bound of S.

Now we show that if u is another upper bound, then $1 \leq u$. Suppose not, that is, suppose that there is an upper bound u of S with $u < 1$. Then we have

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

where both inequalities follow using $u < 1$. From the second inequality, it follows that the number $\frac{u+1}{2}$ belongs to S. The first inequality above then shows that u is not an upper bound for S, a contradiction. Hence 1 is a supremum.

Next we show that 1 is the only supremum. Indeed, if u_* is another supremum, then in particular u_* is also an upper bound, and the above argument shows that $1 \leq u_*$. But $1 < u_*$ is not possible, since 1 is an upper bound for S, and as u_* is a supremum, u_* must be less than or equal to the upper bound 1. So it follows that $u_* = 1$. □

In the above example, there was a unique supremum of the set S. In fact, this is always the case and we have the following result.

Theorem 1.1.1. *If the least upper bound of a subset* S *of* $\mathbb{R}$ *exists, then it is unique.*

Proof. The statement of the theorem means that the set S cannot have two different least upper bounds. Suppose then that u_* and u'_* are two least upper bounds of S. Then in particular u_* and u'_* are also upper bounds of S. Now since u_* is a least upper bound of S and u'_* is an upper bound of S, it follows that

$$u_* \leq u'_*. \tag{1.1}$$

Furthermore, since u'_* is a least upper bound of S and u_* is an upper bound of S, it follows that

$$u'_* \leq u_*. \tag{1.2}$$

From (1.1) and (1.2) we obtain $u_* = u'_*$. □

Thus it makes sense to talk about *the* least upper bound, or *the* supremum, of a set. Similarly, the infimum of a set S (if it exists) is also unique (exercise).

Definitions.

1. *The least upper bound of a set* S *(if it exists) is denoted by* $\mathbf{sup}\, \boldsymbol{S}$.

2. *The greatest lower bound of a set* S *(if it exists) is denoted by* $\mathbf{inf}\, \boldsymbol{S}$.

When the supremum and the infimum of a set belong to the set, we give them the following familiar special names:

Definitions.

1. *If* $\sup S \in S$*, then* $\sup S$ *is called a* ***maximum of S****, denoted by* $\mathbf{max}\, \boldsymbol{S}$.

2. *If* $\inf S \in S$*, then* $\inf S$ *is called a* ***minimum of S****, denoted by* $\mathbf{min}\, \boldsymbol{S}$.

Examples.

1. If $S = \{x \in \mathbb{R} \mid 0 \leq x < 1\}$, then $\sup S = 1 \notin S$ and so $\max S$ does not exist. But $\inf S = 0 \in S$, and so $\min S = 0$.

2. If $S = \mathbb{N}$, then $\sup S$ does not exist, $\inf S = 1$, $\max S$ does not exist, and $\min S = 1$.

3. If $S = \{(-1)^n \mid n \in \mathbb{N}\}$, then $\sup S = 1$, $\inf S = -1$, $\max S = 1$, $\min S = -1$.

4. If $S = \left\{\frac{1}{n} \;\middle|\; n \in \mathbb{N}\right\}$, then $\sup S = 1$ and $\max S = 1$. We show below (after Theorem 1.1.2) that $\inf S = 0$. So $\min S$ does not exist.

5. For the sets $\mathbb{Z}$ and $\mathbb{R}$, sup, inf, max, min do not exist.

6. For the set $\emptyset$, sup, inf, max, min do not exist. □

In the above examples, we note that if S is non-empty 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 now state:

The Least Upper Bound Property.
If S *is a subset of* $\mathbb{R}$ *that is non-empty and bounded above, then* S *has a least upper bound.*

In other words, for a subset $S \subset \mathbb{R}$, if

1. $S \neq \emptyset$, and

2. S has an upper bound,

then $\sup S$ exists.

Examples.

1. *Use the least upper bound property to show that there exists a number* $s \in \mathbb{R}$ *such that* $s > 0$ *and* $s^2 = 2$.

   Solution. Let $S = \{x \in \mathbb{R} \mid x^2 < 2\}$. We'll show that S has a least upper bound $s = \sup S$ and that this s is the real number we want: $s > 0$ and $s^2 = 2$.

   In order to apply the least upper bound property, we first have to show that $S \neq \emptyset$ and that S has an upper bound. This is easy: since $1^2 \leq 2$, $1 \in S$ and therefore $S \neq \emptyset$. Also, since for any $x \in S$, $x^2 \leq 2 < 4$, it follows that $x < 2$, which shows that 2 is an upper bound of S.

By the least upper bound property, S has a least upper bound $s = \sup S$. Since we have already seen that $1 \in S$, and we know that s is an upper bound of S, $s \geq 1 > 0$. It remains to prove that $s^2 = 2$. We do this by showing that each of the two alternatives $s^2 > 2$ and $s^2 < 2$ leads to a contradiction.

Suppose first that $s^2 > 2$. In this case, our plan is to show that, for a suitably small $\varepsilon > 0$, the real number $s - \varepsilon < s$ is also an upper bound for S. In order to find a suitable ε, calculate

$$(s-\varepsilon)^2 = s^2 - 2s\varepsilon + \varepsilon^2 > s^2 - 2s\varepsilon.$$

Since $s^2 > 2$, the right-hand side will be greater than 2 if ε is chosen sufficiently small that $s^2 - 2s\varepsilon > 2$, which can be rearranged as $\varepsilon < \frac{s^2-2}{2s}$. Thus if we choose $\varepsilon = \frac{s^2-2}{4s}$, for instance, then we have $\varepsilon < \frac{s^2-2}{2s}$ which gives $(s-\varepsilon)^2 > s^2 - 2s\varepsilon > 2 > x^2$ for all $x \in S$. Thus $s - \varepsilon > x$ for all $x \in S$, which means that $s - \varepsilon$ is an upper bound of S smaller than s. This is a contradiction to the choice of s as the *least* upper bound.

Suppose next that $s^2 < 2$. In this case we show that by adding a suitably small $\varepsilon > 0$, we obtain that $s + \varepsilon \in S$, which contradicts that s is an upper bound. To find a suitable ε, calculate as follows:

$$(s+\varepsilon)^2 = s^2 + 2s\varepsilon + \varepsilon^2 < s^2 + (2s+1)\varepsilon \quad \text{if } \varepsilon < 1.$$

(Note that in this case we couldn't throw away the positive number ε^2, but instead we bounded it using $\varepsilon^2 < \varepsilon$, which holds as long as $\varepsilon < 1$.) The right-hand side $s^2 + (2s+1)\varepsilon$ is less than 2 if $\varepsilon < \frac{2-s^2}{2s+1}$. Choosing $\varepsilon = \min\{\frac{2-s^2}{2(2s+1)}, \frac{1}{2}\}$, we have $\varepsilon < 1$ and $\varepsilon < \frac{2-s^2}{2s+1}$, which implies $(s+\varepsilon)^2 < s^2 + (2s+1)\varepsilon < 2$. Thus $s + \varepsilon \in S$. Since $s + \varepsilon > s$, s is not an upper bound of S, which is a contradiction.

The only remaining possibility is that $s^2 = 2$. We have shown the existence of $s \in \mathbb{R}$ such that $s > 0$ and $s^2 = 2$. □

2. *Prove the 'greatest lower bound property': if S is a non-empty subset of $\mathbb{R}$ that is bounded below, then S has a greatest lower bound.*

Solution. See Exercise 6. □

Formally, we treat $\mathbb{R}$ as a number system, satisfying all the usual rules of arithmetic, together with a total order $<$. We also treat the least upper bound property as a fundamental property of $\mathbb{R}$. Given these principles, we now deduce some other familiar properties of $\mathbb{R}$.

The following theorem is called the *Archimedean property* of the real numbers.

Theorem 1.1.2 (Archimedean property)**.** *If $x, y \in \mathbb{R}$ and $x > 0$, then there exists an $n \in \mathbb{N}$ such that $y < nx$.*

Proof. Suppose that the conclusion is false, that is, that there does not exist an $n \in \mathbb{N}$ such that $y < nx$. This means that for all $n \in \mathbb{N}$, $y \geq nx$. In other words, y is an upper bound of the non-empty set $S = \{nx \mid n \in \mathbb{N}\}$. By the least upper bound property of the

reals, S has a least upper bound u_*. Note that $u_* - x$ is not an upper bound of S, since $u_* - x$ is smaller than u_* (x is positive) and u_* is the *least* upper bound. Hence there exists an element $mx \in S$ (with $m \in \mathbb{N}$) such that $u_* - x < mx$, that is, $u_* < (m+1)x$. Then $(m+1)x$ is also an element of S with $(m+1)x > u_*$: this contradicts the fact that u_* is an upper bound of S. □

As a consequence of the Archimedean property we are now able to *prove* that the set $\mathbb{N}$ of natural numbers is not bounded above.

Examples.

1. *Show that the set* $\mathbb{N}$ *is not bounded above.*

   Solution. Suppose that $\mathbb{N}$ is bounded above. Then $\mathbb{N}$ has an upper bound $y \in \mathbb{R}$. Since $1 \in \mathbb{N}$, $1 \leq y$, and in particular, $y > 0$. Let $x = 1$. By the Archimedean property (Theorem 1.1.2), there exists an $n \in \mathbb{N}$ such that $y < nx = n$. This contradicts the fact that y is an upper bound of $\mathbb{N}$. □

2. *Set* $S = \left\{ \frac{1}{n} \;\middle|\; n \in \mathbb{N} \right\}$. *Show that* $\inf S = 0$.

   Solution. We know that 0 is a lower bound of S. Suppose that l is a lower bound of S such that $l > 0$. By the Archimedean property (with the real numbers x and y taken as $x = 1$ (> 0) and $y = \frac{1}{l}$), there exists $n \in \mathbb{N}$ such that $\frac{1}{l} = y < nx = n \cdot 1 = n$, and so $\frac{1}{n} < l$, 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. □

1.1.1 Intervals and absolute values

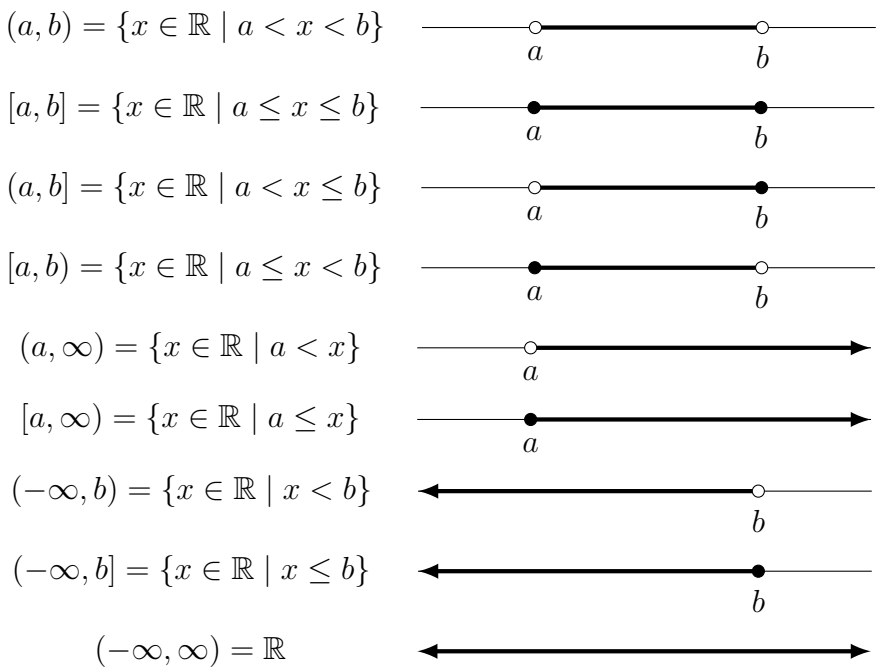
Definition. *An* ***interval*** *(in* $\mathbb{R}$*) is a non-empty subset* S *of* $\mathbb{R}$ *with the property: if* $x, y \in S$ *and* $x \leq z \leq y$*, then* $z \in S$*.*

For instance, the set $S = \{x \in \mathbb{R} \mid x < 2\}$ is an interval: if x and y are both in S, and $x \leq z \leq y$, then in particular $z \leq y < 2$ so $z \in S$.

An interval may or may not have an upper bound: if it does have an upper bound, then it has a supremum which may or may not be in the interval. Similarly, an interval may or may not have a lower bound: if it does have a lower bound, then it has an infimum which may or may not be in the interval. In the example above, S has an upper bound, and the supremum is not in S, while S has no lower bound. There are thus three possible forms for the "lower" end of an interval, and three possible forms for the "upper end", making nine forms in all. These are listed in the figure below, along with the notation for each type of interval.

An interval of the form $(-\infty, b)$, (a, b) or (a, ∞) is called an **open interval**. An interval of the form $(-\infty, b]$, $[a, b]$ or $[a, \infty)$ is called a **closed interval**.

Thus in the notation for intervals used in the Figure, a parenthesis '(' or ')' means that the respective endpoint is not included, and a square bracket '[' or ']' means that the endpoint is included. For example, $[0, 1)$ is 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 be taken as suggesting that there is a real number ∞.) We do not give any special name to intervals of the form $[a, b)$ or $(a, b]$.

In analysis, in order to talk about notions such as *convergence* and *continuity*, we will need a notion of 'closeness' 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.

Definitions.

1. *For a real number x, the **absolute value** $|x|$ of x is defined as follows:*

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

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

Note that $|x| \geq 0$ for all real numbers x, and that $|x| = 0$ iff[1] $x = 0$. Thus $|1| = 1$, $|0| = 0$, $|-1| = 1$, and the distance between the real numbers -1 and 1 is equal to $|-1 - 1| = |-2| = 2$. 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 $x_0 \in \mathbb{R}$ in terms of inequalities in absolute values, that is, by demanding that the distance of the points of the region, to the point x_0, is less than a certain positive number δ, say $\delta = 0.01$ or $\delta = 0.0000001$, and so on. See Exercise 12 and Figure 1.1.

[1]The word "iff" is in common use in mathematics as an abbreviation for "if and only if".

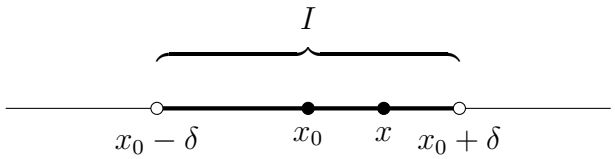


Figure 1.1: The interval $I = (x_0 - \delta, x_0 + \delta) = \{x \in \mathbb{R} \mid |x - x_0| < \delta\}$ is the set of all points in $\mathbb{R}$ whose distance to the point x_0 is strictly less than δ (> 0).

The following properties of the absolute value will be useful.

Theorem 1.1.3. *If x, y are real numbers, then*

$$|xy| = |x|\,|y| \quad \textit{and} \tag{1.3}$$
$$|x + y| \leq |x| + |y|. \tag{1.4}$$

Proof. We prove (1.3) by exhausting all possible cases:

$x = 0$ 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 $|xy| = 0$.

$x > 0$ 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 $|xy| = xy$.

$x > 0$ 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 $|xy| = -xy$.

$x < 0$ and $y > 0$. This follows from 1.1.1 above by interchanging x and y.

$x < 0$ 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 (1.3).

Next we prove (1.4). 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 (1.3), 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:

$x + y \geq 0$. Then $|x + y| = x + y$. As $|x| \geq x$ and $|y| \geq y$, we obtain $|x| + |y| \geq x + y = |x + y|$.

$x + y < 0$. 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 (1.4). □

1.2 Sequences and limits

1.2.1 Sequences

The notion of a sequence occurs in ordinary conversation. An example is the phrase "an unfortunate sequence of events". In this case, we envision one event causing another, which in turn causes another event and so on. We can identify a *first* event, a *second* event, etcetera.

A sequence of real numbers is a list

$$a_1,\, a_2,\, a_3,\, \ldots$$

of real numbers, where there is the *first* number (namely a_1), the *second* number (namely a_2), and so on. For example,

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

is a sequence of real numbers. The first number is 1, the second number is $\frac{1}{2}$ and so on. (There may not be a connection between the numbers appearing in a sequence.) If we think of a_1 as $f(1)$, a_2 as $f(2)$, and so on, then it becomes clear that a sequence of real numbers is a special type of function, namely one with domain $\mathbb{N}$ and co-domain $\mathbb{R}$. This leads to the following formal definition.

Definition. *A* ***sequence*** *of real numbers is a function* $f\colon \mathbb{N} \to \mathbb{R}$.

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 written in any one of the following ways:

$$(a_n)_{n\in\mathbb{N}},\quad (a_n)_{n=1}^{\infty},\quad (a_n)_{n\geq 1},\quad (a_n).$$

In $(a_n)_{n=1}^{\infty}$, the ∞ symbol indicates that the assignment process $1 \mapsto a_1, 2 \mapsto a_2, \ldots$ continues indefinitely. In these notes, we shall normally use the notation $(a_n)_{n\in\mathbb{N}}$. In general, the terms of a sequence need not be real numbers, but in this guide we shall only be dealing with sequences whose entries are real numbers, so we shall simply refer to them as *sequences* from now on.

The nth 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.)

Examples.

1. $\left(\frac{1}{n}\right)_{n\in\mathbb{N}}$ is a sequence with the nth term given by $\frac{1}{n}$, for $n \in \mathbb{N}$. This is the sequence
$$1, \frac{1}{2}, \frac{1}{3}, \ldots.$$

2. $\left(1+\frac{1}{n}\right)_{n\in\mathbb{N}}$ is a sequence with the nth term given by $1+\frac{1}{n}$, for $n \in \mathbb{N}$. This is the sequence
$$2, \frac{3}{2}, \frac{4}{3}, \frac{5}{4}, \frac{6}{5}, \frac{7}{6}, \ldots.$$

3. $\left((-1)^n\left(1+\frac{1}{n}\right)\right)_{n\in\mathbb{N}}$ is a sequence with the nth term given by $(-1)^n\left(1+\frac{1}{n}\right)$, for $n \in \mathbb{N}$. This is the sequence
$$-2, \frac{3}{2}, -\frac{4}{3}, \frac{5}{4}, -\frac{6}{5}, \frac{7}{6}, \ldots.$$

4. $((-1)^n)_{n\in\mathbb{N}}$ is a sequence with the nth term given by $(-1)^n$, for $n \in \mathbb{N}$. This sequence is simply
$$-1, 1, -1, 1, -1, 1, \ldots$$
with the nth term equal to -1 if n is odd, and 1 if n is even.

5. $(1)_{n\in\mathbb{N}}$ is a sequence with the nth term given by 1, for $n \in \mathbb{N}$. This is the constant sequence
$$1, 1, 1, \ldots.$$

6. $(n)_{n\in\mathbb{N}}$ is a sequence with the nth term given by n, for $n \in \mathbb{N}$. This is the strictly increasing sequence
$$1, 2, 3, \ldots.$$

7. $\left(\frac{1}{1^1}+\frac{1}{2^2}+\frac{1}{3^3}+\cdots+\frac{1}{n^n}\right)_{n\in\mathbb{N}}$ is a sequence with the nth term given by $\frac{1}{1^1}+\frac{1}{2^2}+\frac{1}{3^3}+\cdots+\frac{1}{n^n}$, for $n \in \mathbb{N}$. This is the sequence of 'partial sums'
$$\frac{1}{1^1}, \frac{1}{1^1}+\frac{1}{2^2}, \frac{1}{1^1}+\frac{1}{2^2}+\frac{1}{3^3}, \ldots.$$

1.2.2 Limit of a convergent sequence

A sequence can be graphed. For instance, the first 7 points of the graph of the sequence $\left(\frac{1}{n}\right)_{n\in\mathbb{N}}$ are displayed in Figure 1.2. This portion of the graph suggests that the terms of the sequence $\left(\frac{1}{n}\right)_{n\in\mathbb{N}}$ "tend toward 0" as n increases. This is consistent with the idea of convergence that you might have encountered before: 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". Symbolically, this is represented using the notation
$$\lim_{n\to\infty} a_n = L,$$

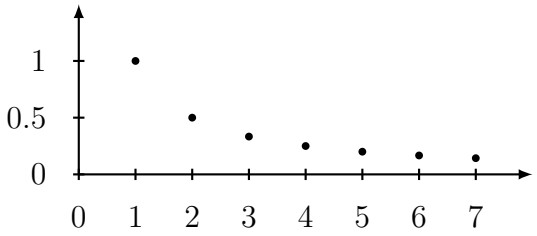


Figure 1.2: First 7 points of the graph of the sequence $\left(\frac{1}{n}\right)_{n\in\mathbb{N}}$.

where L denotes the limit of the sequence. 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.

The problem with this characterization is its imprecision. Exactly what does it mean for the terms of a sequence to get "closer and closer", or "as close as we like", or "arbitrarily close" to some number L? Even if we accept this apparent ambiguity, how would one use the definition given in the preceding paragraph to prove theorems that involve sequences? Since sequences are used throughout analysis, the concepts of their convergence and divergence must be carefully defined.

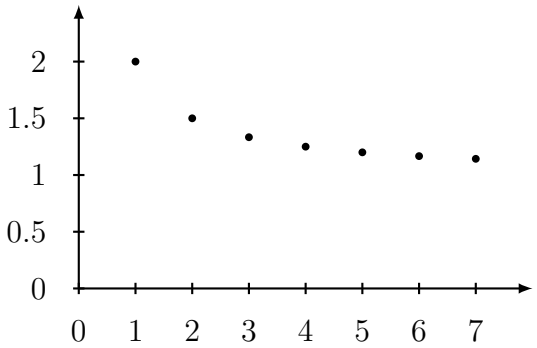


Figure 1.3: First 7 points of the graph of the sequence $\left(1+\frac{1}{n}\right)_{n\in\mathbb{N}}$.

For example, the terms of $\left(1+\frac{1}{n}\right)_{n\in\mathbb{N}}$ get "closer and closer" to 0 (indeed the distance to 0 keeps decreasing), but its limit is 1. See Figure 1.3.

The terms of $\left((-1)^n\left(1+\frac{1}{n}\right)\right)_{n\in\mathbb{N}}$ get "as close as we like" or "arbitrarily close" to 1, but the sequence has no limit. See Figure 1.4.

Definition. *The sequence $(a_n)_{n\in\mathbb{N}}$ is said to* ***converge to L*** *if for every real number $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ (possibly depending on ε) such that for all $n > N$,*

$$|a_n - L| < \varepsilon.$$

Then we say that $(a_n)_{n\in\mathbb{N}}$ is ***convergent with limit L*** *and write*

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

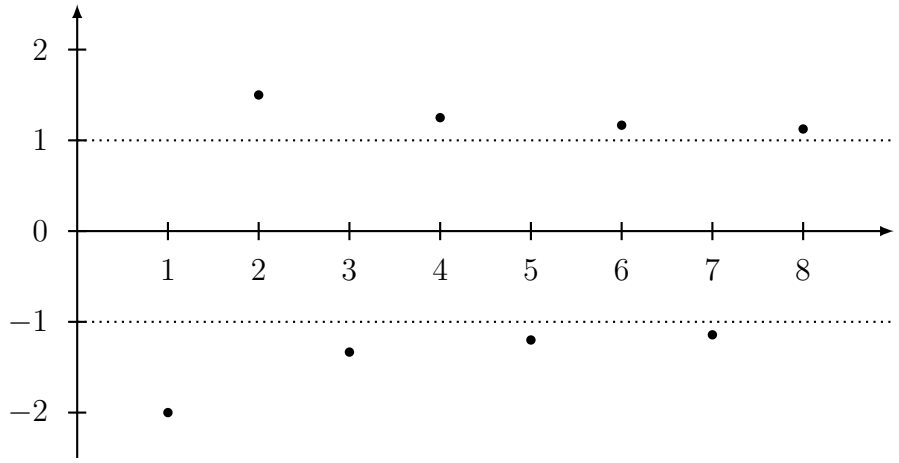


Figure 1.4: First eight points of the graph of the sequence $\left((-1)^n\left(1+\frac{1}{n}\right)\right)_{n\in\mathbb{N}}$.

If there does not exist a number L such that $\lim_{n\to\infty} a_n = L$, then the sequence $(a_n)_{n\in\mathbb{N}}$ is said to be ***divergent****.*

Note that $|a_n - L| < \varepsilon$ iff $a_n \in (L-\varepsilon, L+\varepsilon)$. Hence pictorially, for a convergent sequence with limit L, this definition means the following, as illustrated in Figure 1.5: Pick any $\varepsilon > 0$, and consider the shaded strip of width ε around the horizontal line passing through L. Then one can find an $N \in \mathbb{N}$, large enough, such that all the terms a_n of the sequence, for $n > N$, lie in the shaded strip.

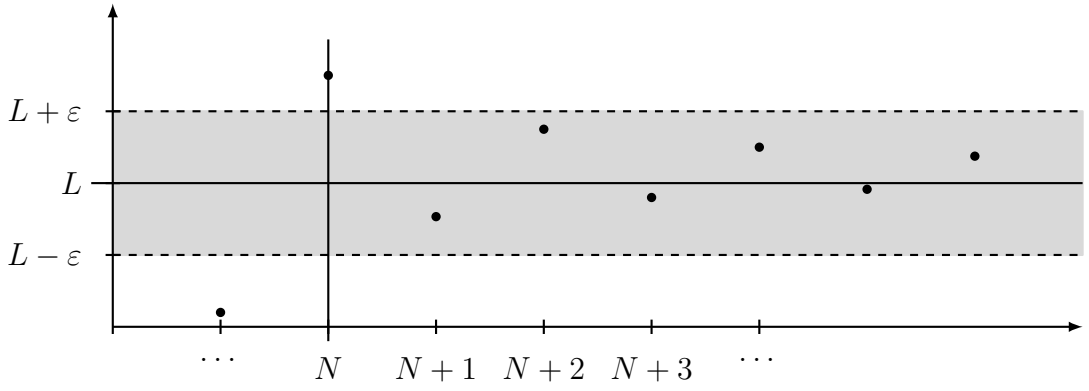


Figure 1.5: Convergence of a sequence with limit L.

Examples.

1. *Use the definition of the limit of a sequence to show that $\left(\frac{1}{n}\right)_{n\in\mathbb{N}}$ is a convergent sequence with limit 0.*

   Solution. Given $\varepsilon > 0$, we need to find N such that, for all $n > N$,

$$|a_n - L| = \left|\frac{1}{n} - 0\right| = \frac{1}{n} < \varepsilon.$$

If we choose $N \in \mathbb{N}$ such that $N > \frac{1}{\varepsilon}$ (such an N exists by the Archimedean property!), then for $n > N$ ($\Leftrightarrow \frac{1}{n} < \frac{1}{N}$), we have

$$|a_n - L| = \left|\frac{1}{n} - 0\right| = \frac{1}{n} < \frac{1}{N} < \varepsilon.$$

Hence $\lim_{n\to\infty} \frac{1}{n} = 0$. □

2. *Use the definition of the limit of a sequence to show that* $\left(1 + \frac{1}{2n^2-n}\right)_{n\in\mathbb{N}}$ *is a convergent sequence with limit* 1.

Solution. Given $\varepsilon > 0$, we need to find N such that for all $n > N$,

$$|a_n - L| = \left|1 + \frac{1}{2n^2 - n} - 1\right| = \frac{1}{2n^2 - n} < \varepsilon.$$

We do not need to be very precise in our choice of ε: if we note that, for all $n \in \mathbb{N}$, $2n^2 - n \geq n$, then we see that we may choose $N \in \mathbb{N}$ such that $N > \frac{1}{\varepsilon}$. So, for $n > N$, we have

$$|a_n - L| = \frac{1}{2n^2 - n} \leq \frac{1}{n} < \frac{1}{N} < \varepsilon.$$

Hence $\lim_{n\to\infty} \left(1 + \frac{1}{2n^2 - n}\right) = 1$. □

3. *Use the definition to show that* $\left((-1)^n \left(1 + \frac{1}{n}\right)\right)_{n\in\mathbb{N}}$ *is a divergent sequence.*

Solution. The idea is as follows. If $L \geq 0$, then for each *odd* n, $(-1)^n \left(1 + \frac{1}{n}\right) < -1$, so these terms are not "close to L" and L is not a limit of the sequence. Similarly, if $L < 0$, then the even terms of the sequence are not close to L, and so again L is not the limit of the sequence. We now proceed to make this informal argument into a formal proof.

Let $a_n = (-1)^n \left(1 + \frac{1}{n}\right)$ for $n \in \mathbb{N}$.

In order to prove that

$$(a_n)_{n\in\mathbb{N}} \text{ is divergent,}$$

we have to show that

$$\neg\left[(a_n)_{n\in\mathbb{N}} \text{ is convergent}\right],$$

that is,

$$\neg\left[\exists L \in \mathbb{R} \text{ such that } \forall \varepsilon > 0\ \exists N \in \mathbb{N} \text{ such that } \forall n > N,\ |a_n - L| < \varepsilon\right],$$

that is,

$$\forall L \in \mathbb{R}\ \exists \varepsilon > 0 \text{ such that } \forall N \in \mathbb{N}\ \exists n > N \text{ such that } |a_n - L| \geq \varepsilon.$$

Let $L \in \mathbb{R}$. Now we need to prove

$$\exists \varepsilon > 0 \text{ such that } \forall N \in \mathbb{N}\ \exists n > N \text{ such that } |a_n - L| \geq \varepsilon.$$

Fix $\varepsilon = 1$. (It is not *obvious* that $\varepsilon = 1$ will work, but it has been found by trial and error.) Now we will show that

$$\forall N \in \mathbb{N}\ \exists n > N \text{ such that } |a_n - L| \geq \varepsilon.$$

So let $N \in \mathbb{N}$ be given. If $L \geq 0$, then choose n to be any odd number $> N$. Then we have

$$|a_n - L| = \left|(-1)^n \left(1 + \frac{1}{n}\right) - L\right| = \left|-1 - \frac{1}{n} - L\right| = 1 + \frac{1}{n} + L > 1 = \varepsilon.$$

If $L < 0$, then choose n to be any even number $> N$. Then we have

$$|a_n - L| = \left|(-1)^n \left(1 + \frac{1}{n}\right) - L\right| = \left|1 + \frac{1}{n} - L\right| = 1 + \frac{1}{n} - L > 1 = \varepsilon.$$

Thus we have shown that for all $L \in \mathbb{R}$, there exists a $\varepsilon > 0$ (namely $\varepsilon = 1$) such that for all $N \in \mathbb{N}$, there exists a $n > N$ (namely any odd number $> N$ if $L \geq 0$, and any even number $> N$ if $L < 0$) such that $|a_n - L| \geq \varepsilon$. Thus the sequence is divergent. □

The notation $\lim_{n\to\infty} a_n$ suggests that the limit of a convergent sequence is unique. Indeed this is the case, and we prove this below.

Theorem 1.2.1. *A convergent sequence has a unique limit.*

Proof. The statement means that both the following are true.
(i) A convergent sequence has a limit;
(ii) A convergent sequence cannot have two different limits.
Here, (i) is true by the definition of convergence, so we only have to prove (ii). Accordingly, consider a convergent sequence $(a_n)_{n\in\mathbb{N}}$ and suppose that it has two limits L_1 and L_2, where $L_1 \neq L_2$. Let

$$\varepsilon = \frac{|L_1 - L_2|}{2} > 0,$$

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 for all $n > N_1$,

$$|a_n - L_1| < \varepsilon.$$

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

$$|a_n - L_2| < \varepsilon.$$

Consequently for $n > \max\{N_1, N_2\}$,

$$\begin{aligned} 2\varepsilon = |L_1 - L_2| &= |L_1 - a_n + a_n - L_2| \\ &\leq |L_1 - a_n| + |a_n - L_2| = |a_n - L_1| + |a_n - L_2| \\ &< \varepsilon + \varepsilon = 2\varepsilon, \end{aligned}$$

a contradiction. Therefore the sequence $(a_n)_{n\in\mathbb{N}}$ does not have two different limits, as required. □

1.2.3 The sequence $(x^n)_{n\in\mathbb{N}}$.

In this section, we prove that, whenever $|x| < 1$, the sequence $(x^n)_{n\in\mathbb{N}}$ is convergent, with limit 0. This is a basic, and perhaps "obvious" result, but we need to establish it "from first principles".

We proceed by first proving a useful result known as **Bernoulli's Inequality**.

Theorem 1.2.2. *For all real $x \geq -1$ and all $n \in \mathbb{N}$,*

$$(1+x)^n \geq 1 + nx.$$

Proof. We prove this result by induction on n. Note first that, for $n = 1$, the inequality states that $1 + x \geq 1 + x$, which is certainly true.

Suppose now that, for some $n \in \mathbb{N}$, $(1+x)^n \geq 1 + nx$. Now we have $(1+x)^{n+1} = (1+x)(1+x)^n \geq (1+x)(1+nx) = 1 + nx + x + nx^2 \geq 1 + (n+1)x$. So the inequality holds for $n+1$. Hence, by induction, the inequality holds for all $n \in \mathbb{N}$. □

Now we use Bernoulli's Inequality to show that, whenever $|x| < 1$, $(x_n)_{n\in\mathbb{N}}$ is convergent with limit 0.

Theorem 1.2.3. *Let x be any real number with $-1 < x < 1$. Then $(x^n)_{n\in\mathbb{N}}$ is a convergent sequence with limit 0.*

Proof. First note that, if $x = 0$, then $x^n = 0$ for every $n \in \mathbb{N}$, and so certainly $\lim_{n\to\infty} x^n = 0$. So we may assume that $x \neq 0$. In this case, we have that $1 < \frac{1}{|x|}$, and so $h = \frac{1}{|x|} - 1 > 0$. Now, by Bernoulli's Inequality, Theorem 1.2.2, we have

$$\frac{1}{|x|^n} = (1+h)^n \geq 1 + nh > nh,$$

and so

$$0 \leq |x|^n \leq \frac{1}{nh}.$$

Now we use the definition of convergence. Given any $\varepsilon > 0$, we take $N \geq \frac{1}{\varepsilon h}$. Now, for $n > N$,

$$|x^n - 0| = |x|^n \leq \frac{1}{nh} < \frac{1}{Nh} \leq \varepsilon.$$

Hence $\lim_{n\to\infty} x^n = 0$. □

Example.

Use Bernoulli's Inequality to show that $\lim\limits_{n\to\infty} 2^{\frac{1}{n}} = 1$.

Solution. $2 > 1$ and so $2^{\frac{1}{n}} > 1$ (for otherwise $2 = (2^{\frac{1}{n}})^n \leq 1$, a contradiction). Let $a_n := 2^{\frac{1}{n}} - 1 \geq 0$. Then $2 = (1 + a_n)^n \geq 1 + na_n$ Hence

$$0 \leq a_n \leq \frac{1}{n}.$$

Now, given any $\varepsilon > 0$, choose $N \geq 1/\varepsilon$. Then, for $n > N$,

$$|2^{\frac{1}{n}} - 1| = |a_n| = a_n \leq \frac{1}{n} < \frac{1}{N} \leq \varepsilon.$$

Therefore the sequence $2, \sqrt{2}, \sqrt[3]{2}, \sqrt[4]{2}, \sqrt[5]{2}, \ldots$ is convergent with limit 1. □

1.2.4 Bounded and monotone sequences

It is cumbersome to check from the definition if a sequence is convergent or not. In this section, we will study a condition under which we can conclude that a sequence is convergent even without knowing its limit! We will prove that if a sequence is both 'bounded' as well as 'monotone', then it is always convergent.

Definition. *A sequence* $(a_n)_{n\in\mathbb{N}}$ *is said to be* ***bounded*** *if there exists a real number* $M > 0$ *such that*

$$\text{for all } n \in \mathbb{N}, \quad |a_n| \leq M. \tag{1.5}$$

Note that a sequence is bounded iff the set $S = \{a_n \mid n \in \mathbb{N}\}$ is bounded. (See Exercise 15.)

Examples.

a) $(1)_{n\in\mathbb{N}}$ is bounded, since $|1| = 1 \leq 1$ for all $n \in \mathbb{N}$.

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

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

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

e) *Show that the sequence* $(a_n)_{n\in\mathbb{N}}$ *defined 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 bounded.

Solution. We have to find an upper bound of $|a_n|$ that is independent of n:

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

(Write down a detailed proof using induction on n.) Thus all the terms are bounded by $\frac{3}{2}$, and so the sequence is bounded. □

f) *Show that the sequence* $(a_n)_{n\in\mathbb{N}}$ *given by* $a_n = n$ *for* $n \in \mathbb{N}$*, is not bounded.*

Solution. Given any $M > 0$, there exists an $N \in \mathbb{N}$ such that $M < N$ (Archimedean property with $y = M$ and $x = 1$). Thus

$$\neg[\exists M > 0 \text{ such that for all } n \in \mathbb{N},\ \ |a_n| = |n| = n \leq M],$$

and so $(n)_{n\in\mathbb{N}}$ is not bounded. □

The sequences $(1)_{n\in\mathbb{N}}$, $\left(\frac{1}{n}\right)_{n\in\mathbb{N}}$, $\left(1+\frac{1}{n}\right)_{n\in\mathbb{N}}$ are all convergent, and we have shown above that these 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. See Figure 1.6.

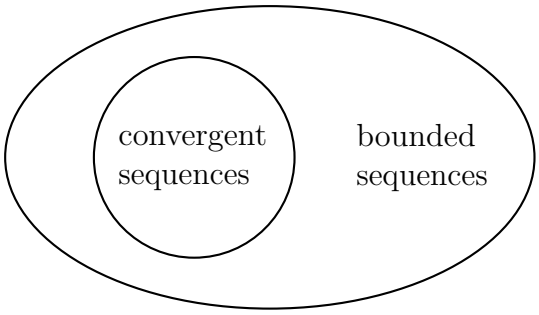


Figure 1.6: All convergent sequences are bounded.

Theorem 1.2.4. *If a sequence is convergent, then it is bounded.*

Proof. Let $(a_n)_{n\in\mathbb{N}}$ be a convergent sequence with limit L. Set $\varepsilon = 1$. Then, using the definition of convergence for this value of ε, we see that there exists $N \in \mathbb{N}$ such that, for all $n > N$,

$$|a_n - L| < 1.$$

Hence for all $n > N$,

$$|a_n| = |a_n - L + L| \leq |a_n - L| + |L| < 1 + |L|.$$

Let $M = \max\{|a_1|, \ldots, |a_N|, 1 + |L|\}$. Then for all $n \in \mathbb{N}$

$$|a_n| \leq M$$

and so $(a_n)_{n\in\mathbb{N}}$ is bounded. □

The above theorem can be used to prove the *divergence* of sequences. Indeed, in contrapositive form it asserts that unbounded sequences are not convergent. Thus, one way to prove that a sequence is divergent, would be to try to prove that it is unbounded.

Example. *Show that the sequence* $(n)_{n\in\mathbb{N}}$ *is divergent.*

Solution. We have seen above that $(n)_{n\in\mathbb{N}}$ is unbounded. It follows from Theorem 1.2.4 that $(n)_{n\in\mathbb{N}}$ is not convergent. □

Definitions. *A sequence $(a_n)_{n\in\mathbb{N}}$ is said to be*

monotonically increasing *(or simply* **increasing***) 'if for all $n \in \mathbb{N}$, $a_n \leq a_{n+1}$,*

strictly increasing *if for all $n \in \mathbb{N}$, $a_n < a_{n+1}$,*

monotonically decreasing *(or simply* **decreasing***) if for all $n \in \mathbb{N}$, $a_n \geq a_{n+1}$,*

strictly decreasing *if for all $n \in \mathbb{N}$, $a_n > a_{n+1}$,*

monotone *if it is either monotonically increasing or monotonically decreasing.*

Thus a sequence $(a_n)_{n\in\mathbb{N}}$ is monotonically increasing if

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

strictly increasing if

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

monotonically decreasing if

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

and strictly decreasing if

$$a_1 > a_2 > a_3 > \dots.$$

Examples.

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

The following theorem can be useful in showing that sequences converge when one does not know the limit beforehand.

Theorem 1.2.5. *If a sequence is monotone and bounded, then it is convergent.*

Proof. Let $(a_n)_{n\in\mathbb{N}}$ be a monotonically increasing sequence. Since $(a_n)_{n\in\mathbb{N}}$ is bounded, it follows that the set

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

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 $\varepsilon > 0$ be given. Since $\sup S - \varepsilon < \sup S$, it follows that $\sup S - \varepsilon$ is not an upper bound for S and so $\exists a_N \in S$ such that $\sup S - \varepsilon < a_N$, that is

$$\sup S - a_N < \varepsilon.$$

Since $(a_n)_{n\in\mathbb{N}}$ is a monotonically increasing sequence, for $n > N$, we have $a_N \leq a_n$. Since $\sup S$ 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

$$|a_n - \sup S| = \sup S - a_n \leq \sup S - a_N < \varepsilon.$$

Hence $(a_n)_{n\in\mathbb{N}}$ is a convergent sequence with limit $\sup S$, as required.
On the other hand, if $(a_n)_{n\in\mathbb{N}}$ is a monotonically decreasing sequence, then clearly $(-a_n)_{n\in\mathbb{N}}$ is a monotonically increasing sequence. 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

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

(see Exercise 6). So given $\varepsilon > 0$, $\exists N \in \mathbb{N}$ such that for all $n > N$, $|-a_n - (-\inf S)| < \varepsilon$, that is, $|a_n - \inf S| < \varepsilon$. Thus $(a_n)_{n\in\mathbb{N}}$ is convergent with limit $\inf S$. □

Examples.

1. We have shown that the sequence $(a_n)_{n\in\mathbb{N}}$ defined 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 strictly increasing since

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

for all $n \in \mathbb{N}$) and bounded (see Example e on page 16). Thus it follows from Theorem 1.2.5 that this sequence[2] is convergent.

[2] Although it is known that this sequence is convergent to some limit $L \in \mathbb{R}$, it is so far not even known if the limit L is rational or irrational, and this is still an open problem in mathematics! Also associated with this sequence is the interesting identity $\sum_{n=1}^{\infty} \frac{1}{n^n} = \int_0^1 \frac{1}{x^x} dx$, the proof of which is beyond the scope of this course.

2. The following table gives a summary of the valid implications, and gives counterexamples to implications which are not true.

Question	Answer	Reason/Counterexample
Is every convergent sequence bounded?	Yes	Theorem 1.2.4
Is every bounded sequence convergent?	No	$((-1)^n)_{n\in\mathbb{N}}$ is bounded, but not convergent.
Is every convergent sequence monotone?	No	$\left(\frac{(-1)^n}{n}\right)_{n\in\mathbb{N}}$ is convergent, but not monotone: $-1 < \frac{1}{2}$ and $\frac{1}{2} > -\frac{1}{3}$.
Is every monotone sequence convergent?	No	$(n)_{n\in\mathbb{N}}$ is not convergent.
Is every bounded AND monotone sequence convergent?	Yes	Theorem 1.2.5

1.2.5 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

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

where the terms $\frac{9}{n^2}$, $\frac{7}{n}$, $\frac{11}{n^2}$ all have limit 0, and by a repeated application of Theorem 1.2.6 given below, we obtain that

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

Theorem 1.2.6. *If $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ are convergent sequences, then the following hold:*

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

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

1. Let $\varepsilon > 0$ be given. By the definition of $\lim_{n\to\infty} a_n = L_a$, there exists an $N \in \mathbb{N}$ such that for all $n > N$,
$$|a_n - L_a| < \frac{\varepsilon}{|\alpha| + 1}.$$
(We added 1 to $|\alpha|$ in the denominator to deal with the possibility that $\alpha = 0$.) Then
$$|\alpha a_n - \alpha L_a| = |\alpha|\,|a_n - L_a| \leq |\alpha|\frac{\varepsilon}{|\alpha| + 1} < \varepsilon,$$
and (again by the definition of a convergent sequence) $(\alpha a_n)_{n\in\mathbb{N}}$ is convergent with limit αL_a, that is,
$$\lim_{n\to\infty} \alpha a_n = \alpha L_a = \alpha \lim_{n\to\infty} a_n.$$
2. Given $\varepsilon > 0$, let $N \in \mathbb{N}$ be such that for all $n > N$,
$$|a_n - L_a| < \varepsilon.$$
Then (by Exercise 14) we have for all $n > N$:
$$||a_n| - |L_a|| \leq |a_n - L_a| < \varepsilon.$$
Hence $(|a_n|)_{n\in\mathbb{N}}$ is convergent with limit $|L_a|$, that is,
$$\lim_{n\to\infty} |a_n| = |L_a| = \left|\lim_{n\to\infty} a_n\right|.$$

3. Given $\varepsilon > 0$, let $N_1 \in \mathbb{N}$ be such that for all $n > N_1$,

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

Let $N_2 \in \mathbb{N}$ be such that for all $n > N_2$,

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

Then for all $n > N := \max\{N_1, N_2\}$, we have by the triangle inequality:

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

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

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

4. Note that

$$\begin{aligned} |a_n b_n - L_a L_b| &= |a_n b_n - L_a b_n + L_a b_n - L_a L_b| \\ &\leq |a_n b_n - L_a b_n| + |L_a b_n - L_a L_b| \\ &= |a_n - L_a|\,|b_n| + |L_a|\,|b_n - L_b|. \end{aligned} \tag{1.6}$$

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

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

This can be achieved by finding an N such that each of the summands in (1.6) is less than $\frac{\varepsilon}{2}$ for $n > N$. This can be done as follows.

Step 1. Since $(b_n)_{n\in\mathbb{N}}$ is convergent, by Theorem 1.2.4 it follows that it is bounded: $\exists M > 0$ such that for all $n \in \mathbb{N}$, $|b_n| \leq M$. Let $N_1 \in \mathbb{N}$ be such that, for all $n > N_1$,

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

Step 2. Let $N_2 \in \mathbb{N}$ be such that for all $n > N_2$,

$$|b_n - L_b| < \frac{\varepsilon}{2|L_a| + 1}.$$

(Note that it is possible that $L_a = 0$.) Thus, for $n > N := \max\{N_1, N_2\}$, we have

$$\begin{aligned} |a_n b_n - L_a L_b| &\leq |a_n - L_a|\,|b_n| + |L_a|\,|b_n - L_b| \\ &< \frac{\varepsilon}{2M} M + |L_a| \frac{\varepsilon}{2|L_a| + 1} \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon. \end{aligned}$$

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

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

5. This can be shown by using induction on k and from part 4 above. It is trivially true with $k = 1$. Suppose that it holds for some k, then $(a_n^k)_{n\in\mathbb{N}}$ is convergent and

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

Hence by part 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

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

Thus $(a_n^{k+1})_{n\in\mathbb{N}}$ is convergent and

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

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 $n > N_1$,

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

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

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

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

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

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

7. Exercise 38.

8. Since $a_n \ge 0$ for each n, we have $L \ge 0$ (see Exercise 24). The case $L = 0$ is left as an exercise.

In the case $L > 0$, we use a technique called "rationalising the numerator": note that

$$\sqrt{a}_n - \sqrt{L} = (\sqrt{a}_n - \sqrt{L})\frac{\sqrt{a}_n + \sqrt{L}}{\sqrt{a}_n + \sqrt{L}} = \frac{a_n - L}{\sqrt{a}_n + \sqrt{L}}.$$

Now, given $\varepsilon > 0$, choose $N \in \mathbb{N}$ so that, for $n > N$, $|a_n - L| < \varepsilon\sqrt{L}$. Then we have, for $n > N$,

$$|\sqrt{a}_n - \sqrt{L}| = \frac{|a_n - L|}{\sqrt{a}_n + \sqrt{L}} \le \frac{\varepsilon\sqrt{L}}{\sqrt{L}} = \varepsilon.$$

Thus $(\sqrt{a_n})_{n\in\mathbb{N}}$ is convergent, with limit $\sqrt{L}$, □

Examples. 1. *Determine whether the following sequence is convergent and find its limit.*

$$\left(\frac{n^2 - 24n^3 + 3n^4 - 12}{1 + 7n + 21n^4}\right)_{n\in\mathbb{N}}$$

Solution. By Example 1 on page 12 we know that $\lim_{n\to\infty} \frac{1}{n} = 0$. We now use Theorem 1.2.6, we obtain

$$\begin{aligned}
\lim_{n\to\infty} \frac{n^2 - 24n^3 + 3n^4 - 12}{1 + 7n + 21n^4} &= \lim_{n\to\infty} \frac{n^4}{n^4} \cdot \frac{\frac{1}{n^2} - \frac{24}{n} + 3 - \frac{12}{n^4}}{\frac{1}{n^4} + \frac{7}{n^3} + 21} \\
&= \frac{\left(\lim_{n\to\infty} \frac{1}{n}\right)^2 - 24 \lim_{n\to\infty} \frac{1}{n} + 3 - 12\left(\lim_{n\to\infty} \frac{1}{n}\right)^4}{\left(\lim_{n\to\infty} \frac{1}{n}\right)^4 + 7\left(\lim_{n\to\infty} \frac{1}{n}\right)^3 + 21} \\
&= \frac{0^2 - 24 \cdot 0 + 3 - 12 \cdot 0^4}{0^4 + 7 \cdot 0^3 + 21} \\
&= \frac{3}{21} = \frac{1}{7}.
\end{aligned}$$

□

2. *Determine whether the following sequence is convergent and find its limit.*

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

Solution. Divide top and bottom by the fastest growing term appearing, which is 3^n, and use that $\lim_{n\to\infty} x^n = 0$ for $|x| < 1$:

$$\begin{aligned}
\lim_{n\to\infty} \frac{2^n + 3^n + 1}{3^{n+1} + 3} &= \frac{(2/3)^n + 1 + (1/3)^n}{3 + 3(1/3)^n} \\
&= \frac{\lim_{n\to\infty}(2/3)^n + \lim_{n\to\infty} 1 + \lim_{n\to\infty}(1/3)^n}{\lim_{n\to\infty} 3 + 3\lim_{n\to\infty}(1/3)^n} \\
&= \frac{0 + 1 + 0}{3 + 0} = \frac{1}{3}.
\end{aligned}$$

□

Remark. It follows from Theorem 1.2.6.3 that if we have three convergent sequences $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$, $(c_n)_{n\in\mathbb{N}}$, then their sum $(a_n + b_n + c_n)_{n\in\mathbb{N}}$ is also convergent with limit

$$\lim_{n\to\infty}(a_n + b_n + c_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n + \lim_{n\to\infty} c_n.$$

This is also true for the sum of four convergent sequences, the sum of five convergent sequences, and by an easy induction proof, the sum of any number of convergent sequences. However, the following reasoning is clearly incorrect (because it would show that $1 = 0$):

$$1 = \underbrace{\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}}_{n \text{ terms}}$$

and therefore,

$$\begin{aligned} 1 = \lim_{n\to\infty} 1 &= \lim_{n\to\infty} \left(\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}\right) \\ &\overset{!}{=} \underbrace{\lim_{n\to\infty} \frac{1}{n} + \lim_{n\to\infty} \frac{1}{n} + \cdots + \lim_{n\to\infty} \frac{1}{n}}_{n \text{ limits}} \\ &= \underbrace{0 + 0 + \cdots + 0}_{n \text{ terms}} \\ &= 0. \end{aligned}$$

The incorrect step is marked with a '!'. The problem here is that the number of terms is not a fixed value, but *depends on* n, and therefore Theorem 1.2.6.3 cannot be applied. □

1.2.6 The Sandwich Theorem

Another theorem that is useful in 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 limits with the *same* limit, then the sandwiched sequence is also convergent with the same limit.

Theorem 1.2.7 (Sandwich theorem)**.** *Let* $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$ *be convergent sequences with the same limit, that is,*

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

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,*

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

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

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

Given $\varepsilon > 0$, let $N_1 \in \mathbb{N}$ be such that for all $n > N_1$, $|a_n - L| < \varepsilon$. Hence for $n > N_1$,

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

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

$$L - \varepsilon < a_n.$$

Still for the same given $\varepsilon > 0$, let $N_2 \in \mathbb{N}$ be such that for all $n > N_2$, $|b_n - L| < \varepsilon$. So for $n > N_2$, $b_n - L < \varepsilon$, that is,

$$b_n < L + \varepsilon.$$

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

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

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

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

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

Examples.

1. *Use the Sandwich theorem to show that* $\lim\limits_{n\to\infty} \dfrac{n}{10^n} = 0$.

   Solution. It can be shown by induction that for all $n \in \mathbb{N}$, $n^2 < 10^n$.

   Consequently, we have

   $$0 \leq \frac{n}{10^n} \leq \frac{n}{n^2} = \frac{1}{n}.$$

   Since $\lim\limits_{n\to\infty} 0 = 0 = \lim\limits_{n\to\infty} \dfrac{1}{n}$, from the Sandwich theorem it follows that the sequence $\left(\frac{n}{10^n}\right)_{n\in\mathbb{N}}$ is convergent and

   $$\lim_{n\to\infty} \frac{n}{10^n} = 0.$$

   Thus the sequence $\frac{1}{10}, \frac{2}{100}, \frac{3}{1000}, \frac{4}{10000}, \ldots$ is convergent with limit 0. □

2. *Use the Sandwich theorem to show that for any* $a, b \in \mathbb{R}$, $\lim\limits_{n\to\infty}(|a|^n + |b|^n)^{\frac{1}{n}} = \max\{|a|, |b|\}$.

   Solution. Clearly,

   $$(\max\{|a|, |b|\})^n \leq |a|^n + |b|^n \leq (\max\{|a|, |b|\})^n + (\max\{|a|, |b|\})^n$$

   and so

   $$\max\{|a|, |b|\} \leq (|a|^n + |b|^n)^{\frac{1}{n}} \leq 2^{\frac{1}{n}} \max\{|a|, |b|\}.$$

   So using the Sandwich theorem, it follows that $\lim\limits_{n\to\infty}(|a|^n + |b|^n)^{\frac{1}{n}} = \max\{|a|, |b|\}$.

   In particular, with $a = 24$ and $b = 2005$, we have that $\lim\limits_{n\to\infty}(24^n + 2005^n)^{\frac{1}{n}} = 2005$, that is, the sequence

   $$2029, 2005.1436, 2005.001146260873, \ldots$$

   is convergent with limit 2005. □

3. *Show that* $\lim_{n\to\infty}\left(\frac{n}{n^2+1}+\frac{n}{n^2+2}+\cdots+\frac{n}{n^2+n}\right)=1$.

   Solution. There are n terms in the sum: the smallest is $\frac{n}{n^2+n}$ and the largest is $\frac{n}{n^2+1}$. Thus, for all $n\in\mathbb{N}$, we have

$$\frac{n^2}{n^2+n}\leq\frac{n}{n^2+1}+\frac{n}{n^2+2}+\cdots+\frac{n}{n^2+n}\leq\frac{n^2}{n^2+1},$$

   and since

$$\lim_{n\to\infty}\frac{n^2}{n^2+n}=1=\lim_{n\to\infty}\frac{n^2}{n^2+1}.$$

   it follows from the Sandwich theorem that

$$\lim_{n\to\infty}\left(\frac{n}{n^2+1}+\frac{n}{n^2+2}+\cdots+\frac{n}{n^2+n}\right)=1.$$ □

1.2.7 Subsequences

In this section we prove an important result in analysis, known as 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. *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence and let $(n_k)_{k\in\mathbb{N}}$ be a strictly increasing sequence of natural numbers. Then $(a_{n_k})_{k\in\mathbb{N}}$ is called a* ***subsequence of*** $(a_n)_{n\in\mathbb{N}}$.

Examples.

1. $\left(\frac{1}{2n}\right)_{n\in\mathbb{N}}$, $\left(\frac{1}{n^2}\right)_{n\in\mathbb{N}}$, $\left(\frac{1}{n!}\right)_{n\in\mathbb{N}}$ and $\left(\frac{1}{n^n}\right)_{n\in\mathbb{N}}$ are all subsequences of $\left(\frac{1}{n}\right)_{n\in\mathbb{N}}$.

2. Let p_n be the n^{th} prime number. (Thus $p_1=2$, $p_2=3$, $p_3=5$, $p_4=7$, etc.) Then the sequence $(a_n)_{n\in\mathbb{N}}$ defined by $a_n=\frac{1}{p_n}$ is a subsequence of $\left(\frac{1}{n}\right)_{n\in\mathbb{N}}$.

3. The sequence

$$\frac{1}{2},\ 1,\ \frac{1}{3},\ \frac{1}{4},\ \ldots$$

   is not a subsequence of $\left(\frac{1}{n}\right)_{n\in\mathbb{N}}$.

4. The sequence $\left((-1)^{2n}\right)_{n\in\mathbb{N}}$, that is, the constant sequence

$$1,\ 1,\ 1,\ \ldots$$

   and the sequence $\left((-1)^{2n-1}\right)_{n\in\mathbb{N}}$, that is, the constant sequence

$$-1,\ -1,\ -1,\ \ldots$$

   are both subsequences of $\left((-1)^n\right)_{n\in\mathbb{N}}$. □

Theorem 1.2.8. *If $(a_n)_{n\in\mathbb{N}}$ is a convergent sequence with limit L, then any subsequence of $(a_n)_{n\in\mathbb{N}}$ is also convergent with the limit L.*

Proof. Let $(a_{n_k})_{k\in\mathbb{N}}$ be a subsequence of $(a_n)_{n\in\mathbb{N}}$. Given $\varepsilon > 0$, let $N \in \mathbb{N}$ be such that for all $n > N$, $|a_n - L| < \varepsilon$. Since the sequence $n_1 < n_2 < n_3 < \dots$ is not bounded above, it follows that there exists a $K \in \mathbb{N}$ such that $n_K > N$. Then for all $k > K$, $n_k > n_K > N$. Hence for $k > K$, $|a_{n_k} - L| < \varepsilon$, and so $(a_{n_k})_{k\in\mathbb{N}}$ is convergent with limit L. □

Examples.

1. $\left(\frac{1}{2n}\right)_{n\in\mathbb{N}}$, $\left(\frac{1}{n^2}\right)_{n\in\mathbb{N}}$, $\left(\frac{1}{n!}\right)_{n\in\mathbb{N}}$ and $\left(\frac{1}{n^n}\right)_{n\in\mathbb{N}}$ are convergent sequences with limit 0.

2. The sequence $((-1)^n)_{n\in\mathbb{N}}$ is divergent since the subsequence $1, 1, 1, \dots$ has limit 1, while the subsequence $-1, -1, -1, \dots$ has limit -1. □

The following theorem is a very important step in the proof of the Bolzano-Weierstrass theorem. It gives us a fact about *arbitrary* sequences of real numbers.

Theorem 1.2.9. *Every sequence has a monotone subsequence.*

Since we do not assume anything about the sequence, other than that it is a sequence of real numbers, there is virtually no obvious way to start the proof. Nevertheless, even though it is very difficult to *discover* such a proof, it is not so difficult to *understand* it, once found.

Before doing the formal proof, we first illustrate the idea behind it. Let $(a_n)_{n\in\mathbb{N}}$ be the given sequence. Imagine that a_n is the height of a hotel with number n, which is followed by hotel $n + 1$, and so on, along an infinite line, where the sea is in the far distance. A hotel is said to have the **seaview property** if it is higher than all hotels following it. See Figure 1.7.

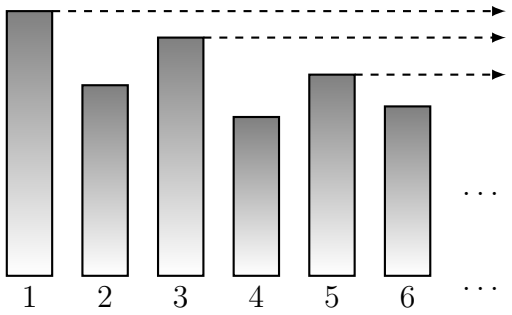


Figure 1.7: The seaview property.

Now there are only two possibilities:

First possibility. There are *infinitely many* hotels with the seaview property. Then their heights form a strictly decreasing subsequence.

Second possibility. There is *only a finite number* of hotels with the seaview property. 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 a monotonically increasing subsequence.

Proof of Theorem 1.2.9. Let

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

that is, S is the set of all natural numbers m such that $a_n < a_m$ for all $n > m$. Then we have the following two cases.

***S* is infinite.** Arrange the elements of S in strictly increasing order: $n_1 < n_2 < n_3 < \dots$. Then $(a_{n_k})_{k\in\mathbb{N}}$ is a strictly decreasing subsequence of $(a_n)_{n\in\mathbb{N}}$.

***S* is finite.** If S is empty, then define $n_1 = 1$, and otherwise let $n_1 = \max S + 1$. Then since n_1 is larger than any element of S, $n_1 \notin S$, which means that there exists an $n_2 > n_1$ such that $a_{n_2} \geq a_{n_1}$. Again, n_2 is larger than any element of S, thus $n_2 \notin S$, which means there exists an $n_3 > n_2$ such that $a_{n_3} \geq a_{n_2}$. Continuing in this way, for any $k \in \mathbb{N}$ we choose an $n_{k+1} > n_k$ such that $a_{n_{k+1}} \geq a_{n_k}$. Then $(a_{n_k})_{k\in\mathbb{N}}$ is a monotonically increasing subsequence of $(a_n)_{n\in\mathbb{N}}$. □

The Bolzano-Weierstrass theorem is now a simple consequence of the above theorem.

Theorem 1.2.10. (Bolzano[3]-Weierstrass[4] theorem.) *Every bounded sequence has a convergent subsequence.*

Proof. Let $(a_n)_{n\in\mathbb{N}}$ be a bounded sequence. Then there exists $M > 0$ such that for all $n \in \mathbb{N}$, $|a_n| \leq M$. From Theorem 1.2.9 above, it follows that the sequence $(a_n)_{n\in\mathbb{N}}$ has a monotone subsequence $(a_{n_k})_{k\in\mathbb{N}}$. 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 1.2.5 that it is convergent. □

Example. Consider the sequence $(a_n)_{n\in\mathbb{N}}$ of fractional parts of integral multiples of $\sqrt{2}$, defined by

$$a_n = n\sqrt{2} - \lfloor n\sqrt{2} \rfloor, \text{ for } n \in \mathbb{N},$$

where, for $x \in \mathbb{R}$, $\lfloor x \rfloor$ is the *floor function* of x, as defined in Exercise 9.

[3]Bernhard Bolzano (1781–1848)
[4]Karl Weierstrass (1815–1897)

The terms of the sequence $(a_n)_{n\in\mathbb{N}}$ are as follows:

$$\begin{aligned}
\sqrt{2} &= 1.414213\ldots & a_1 &= 0.414213\ldots \\
2\sqrt{2} &= 2.828427\ldots & a_2 &= 0.828427\ldots \\
3\sqrt{2} &= 4.242640\ldots & a_3 &= 0.242640\ldots \\
4\sqrt{2} &= 5.656854\ldots & a_4 &= 0.656854\ldots \\
5\sqrt{2} &= 7.071067\ldots & a_5 &= 0.071067\ldots \\
6\sqrt{2} &= 8.485281\ldots & a_6 &= 0.485281\ldots \\
&\vdots
\end{aligned}$$

The sequence $(a_n)_{n\in\mathbb{N}}$ is bounded: indeed, $0 \leq a_n < 1$ for every $n \in \mathbb{N}$. So by the Bolzano-Weierstrass theorem this sequence has a convergent subsequence[5]. □

1.3 Continuity

A function $f\colon \mathbb{R} \to \mathbb{R}$ is a rule of correspondence that assigns to each real number a unique real number. The functions we are most familiar with are actually unusually well-behaved, and most functions are impossible to describe fully, let alone to work with. Many bizarre functions make their appearance in analysis, and in order to avoid falling into pitfalls with simplistic thinking based on our experience with "nice" functions, we need our definitions and the hypotheses (assumptions) of theorems to be stated carefully and clearly.

Within the huge collection of functions, there is an important subset: the continuous functions. Continuous functions play a prominent role in analysis since they include all the most familiar and useful functions, and because the all continuos functions share many useful properties.

In this section we give the formal definition of a continuous function and prove two of the most important properties of continuous functions: the Extreme value theorem and the Intermediate value theorem.

1.3.1 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 $f\colon \mathbb{R} \to \mathbb{R}$ to be continuous? The common informal definition of this concept states that a function f is continuous if one can sketch its graph without lifting the pencil. In other words, the graph of f has no breaks in it. If a break does occur in the graph, then this break will occur at some point. Thus (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 will be called continuous.

[5] In 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.

If a function has a break at a point, say c, then even if points x are close to c, the points $f(x)$ do not get close to $f(c)$. See Figure 1.8.

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 1.9.

Definitions.

1. *Let I be an interval in $\mathbb{R}$ and let $c \in I$. A function $f\colon I \to \mathbb{R}$ is* ***continuous at c*** *if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $x \in I$ satisfying $|x - c| < \delta$, $|f(x) - f(c)| < \varepsilon$.*

2. *If f is not continuous at c, we say that f is* ***discontinuous at c****.*

3. *A function $f\colon I \to \mathbb{R}$ is* ***continuous on I*** *(or just* ***continuous*** *if I is clear from the context) if for every $c \in I$, f is continuous at c.*

4. *If f is not continuous on I, we say that f is* ***discontinuous on I*** *(or just* ***discontinuous*** *if I is clear from the context).*

Examples.

1. *Show that the function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = x$ for all $x \in \mathbb{R}$ is continuous.*

   Solution. To show that f is continuous, we have to show that it is continuous at c for each $c \in \mathbb{R}$. Accordingly, we start by fixing some $c \in \mathbb{R}$. To show that f is continuous at c, we have to show that, for all $\varepsilon > 0$, there is a $\delta > 0$ such that for all $x \in \mathbb{R}$, if $|x - c| < \delta$, then $|f(x) - f(c)| < \varepsilon$. Suppose then that we are given some $\varepsilon > 0$: we now need to fnd a suitable $\delta > 0$. **A very important point is that δ should be independent of x**, but may (and usually will) depend on ε and c. In this case, the following choice of δ will work: $\delta = \varepsilon$. Then, if $x \in \mathbb{R}$ and $|x - c| < \delta$, we have:

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

   This proves that f is continuous at c. Since the choice of $c \in \mathbb{R}$ was arbitrary, it follows that f is continuous on $\mathbb{R}$. □

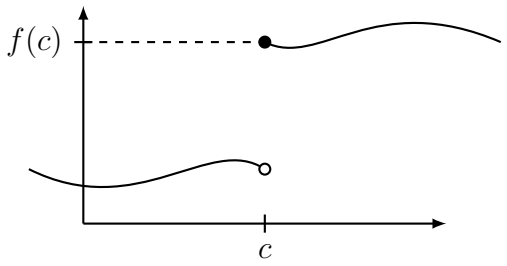


Figure 1.8: A function with a break at c. If x lies to the left of c, then $f(x)$ is not close to $f(c)$, no matter how close x comes to c.

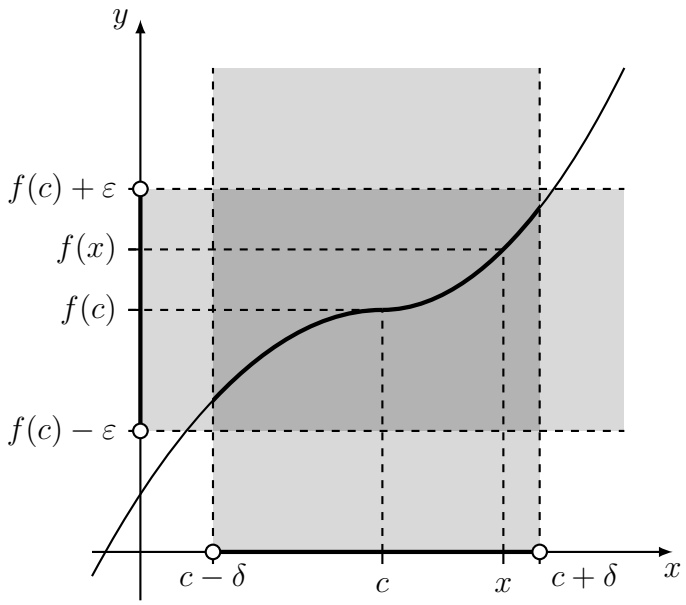


Figure 1.9: The definition of the continuity of a function at point c. If the function is continuous at c, then given any $\varepsilon > 0$ (which determines a strip around the line $y = f(c)$ of width 2ε), there exists a $\delta > 0$ (which determines an interval of width 2δ around the point c) such 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 $f(c) - \varepsilon < f(x) < f(c) + \varepsilon$, that is, $|f(x) - f(c)| < \varepsilon$.

2. *Show that the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 2x + 1$ for all $x \in \mathbb{R}$ is continuous.*

Solution. Fix $c \in \mathbb{R}$. Given $\varepsilon > 0$, we again need to fnd a suitable $\delta > 0$. Here we set $\delta = \varepsilon/2$. Then, if $x \in \mathbb{R}$ and $|x - c| < \delta$, we have:

$$|f(x) - f(c)| = |(2x + 1) - (2c + 1)| = 2|x - c| < 2\delta = \varepsilon.$$

This proves that f is continuous at c. Since the choice of $c \in \mathbb{R}$ was arbitrary, it follows that f is continuous on $\mathbb{R}$. □

3. *Show that the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 1$ for all $x \in \mathbb{R}$ is continuous.*

Solution. Let $c \in \mathbb{R} = (-\infty, \infty)$. We have to prove that f is continuous at c. Let $\varepsilon > 0$ be given. In this case, any positive choice of δ will work; for instance, let $\delta = 1$. Then if $x \in \mathbb{R}$ and $|x - c| < \delta = 1$, we have:

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

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

4. *Show that the function* $f\colon \mathbb{R} \to \mathbb{R}$ *given by*

$$f(x) = \begin{cases} 0 & \text{if } x = 0, \\ 1 & \text{if } x \in \mathbb{R} \setminus \{0\}, \end{cases}$$

is discontinuous at 0 *and continuous at all* $c \in \mathbb{R} \setminus \{0\}$.

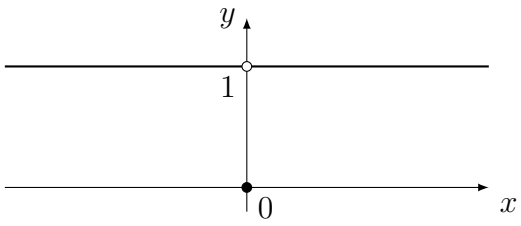


Figure 1.10: A function continuous everywhere except at 0.

Solution. Suppose that f is continuous at 0. Then for any given $\varepsilon > 0$ there exists a $\delta > 0$ such that whenever $|x - 0| < \delta$, $|f(x) - f(0)| < \varepsilon$. In particular, setting $\varepsilon = \frac{1}{2}$, we see that there is a $\delta > 0$ such that for all $x \in \mathbb{R}$, if $|x| = |x - 0| < \delta$, then $|f(x) - f(0)| = |f(x) - 0| = |f(x)| < \varepsilon = \frac{1}{2}$. Take $x = \frac{\delta}{2} \in \mathbb{R}$: then $|x| = \left|\frac{\delta}{2}\right| = \frac{\delta}{2} < \delta$, but $|f(x)| = \left|f\left(\frac{\delta}{2}\right)\right| = |1| = 1 > \frac{1}{2} = \varepsilon$, which is a contradiction. So f is not continuous at 0.

Next we show that for all $c \in \mathbb{R} \setminus \{0\}$, f is continuous at c. Let $\varepsilon > 0$ be given. Take $\delta = \frac{|c|}{2} > 0$. Then if $x \in \mathbb{R}$ and $|x - c| < \delta$, we have

$$|c| - |x| \leq ||c| - |x|| \leq |c - x| = |x - c| < \delta = \frac{|c|}{2}$$

and so

$$|x| > \frac{|c|}{2} > 0.$$

Thus $x \neq 0$ and so $f(x) = 1$. Hence if $x \in \mathbb{R}$ and $|x - c| < \delta$, we obtain

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

Consequently f is continuous at c. □

5. *Show that the function* $f\colon (0, \infty) \to \mathbb{R}$ *given by* $f(x) = \frac{1}{x}$ *for all* $x \in \mathbb{R}$ *is continuous.*

Solution. Let $c \in (0, \infty)$. Given $\varepsilon > 0$, let $\delta = \min\left\{\frac{c}{2}, \frac{\varepsilon c^2}{2}\right\}$ (> 0). Then if $x \in (0, 1)$ and $|x - c| < \delta$, we have

$$c - x \leq |c - x| < \delta \leq \frac{c}{2}, \text{ and so } x > \frac{c}{2} > 0.$$

Consequently, if $x \in (0,\infty)$ and $|x-c| < \delta$,

$$\left|\frac{1}{x} - \frac{1}{c}\right| = \frac{|c-x|}{x\,c} = |x-c| \cdot \frac{1}{x} \cdot \frac{1}{c} < \delta \cdot \frac{2}{c} \cdot \frac{1}{c} = \frac{2\delta}{c^2} \leq \varepsilon.$$

So f is continuous at c. Since the choice of $c \in (0,\infty)$ was arbitrary, it follows that f is continuous on $(0,\infty)$.

NOTE: In practice, the choice of δ is the *last* part of the proof to be filled in. As we go through the rest of the proof, we see the need to have (a) $\delta \leq |c|/2$, and then (b) $\frac{2\delta}{c^2} \leq \varepsilon$. Only after we have collected all the conditions do we make the choice of δ. □

1.3.2 Continuous functions preserve convergent sequences

We now give an alternative characterisation of continuity.

Theorem 1.3.1. *Let I be an interval in $\mathbb{R}$ and let $c \in I$. Suppose that $f\colon I \to \mathbb{R}$ is a function. Then f is continuous at c iff*

$$\boxed{\begin{array}{c}\text{for every convergent sequence } (x_n)_{n\in\mathbb{N}} \text{ contained in } I \text{ with limit } c,\\ (f(x_n))_{n\in\mathbb{N}} \text{ is convergent and } \lim_{n\to\infty} f(x_n) = f(c).\end{array}} \tag{1.7}$$

Proof. "Only if" ($\Longrightarrow$) direction: Assume that f is continuous at $c \in I$ and let $(x_n)_{n\in\mathbb{N}}$ be a convergent sequence contained in I with limit c. We have to show that $(f(x_n))_{n\in\mathbb{N}}$ converges to $f(c)$.

Since f is continuous at $c \in I$, given any $\varepsilon > 0$, $\exists \delta > 0$ such that for all $x \in I$ satisfying $|x-c| < \delta$, $|f(x) - f(c)| < \varepsilon$. Since $(x_n)_{n\in\mathbb{N}}$ is convergent with limit c, $\exists N \in \mathbb{N}$ such that for all $n > N$, $|x_n - c| < \delta$.

Consequently for all $n > N$, $|f(x_n) - f(c)| < \varepsilon$. Thus $(f(x_n))_{n\in\mathbb{N}}$ is convergent with limit $f(c)$.

"If" ($\Longleftarrow$) direction: Suppose that (1.7) holds. We have to show that f is continuous at c. We prove this by contradiction. Assume that f is not continuous at c, that is,

$$\neg\,[\forall \varepsilon > 0\ \exists \delta > 0 \text{ such that } \forall x \in I \text{ such that } |x-c| < \delta,\, |f(x) - f(c)| < \varepsilon,]$$

or equivalently,

$$\exists \varepsilon > 0 \text{ such that } \forall \delta > 0\ \exists x \in I \text{ such that } |x-c| < \delta \text{ but } |f(x) - f(c)| \geq \varepsilon.$$

By choosing $\delta = \frac{1}{n}$, we find $x_n \in I$ such that $|x_n - c| < \delta = \frac{1}{n}$ and $|f(x_n) - f(c)| \geq \varepsilon$.
Claim 1: The sequence $(x_n)_{n\in\mathbb{N}}$ is contained in I and is convergent with limit c.

Indeed, we have for all $n \in \mathbb{N}$, $x_n \in I$. Furthermore, given any $\zeta > 0$, we can find $N \in \mathbb{N}$ such that $\frac{1}{\zeta} < N$ (Archimedean property), that is, $\frac{1}{N} < \zeta$. Hence for $n > N$, $|x_n - c| < \frac{1}{N} < \zeta$. So $(x_n)_{n\in\mathbb{N}}$ is convergent with limit c.

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 $|f(x_n) - f(c)| \geq \varepsilon$. Thus for instance $\frac{\varepsilon}{2} > 0$, but it is not possible to find a large enough $N \in \mathbb{N}$ such that for all $n > N$, $|f(x_n) - f(c)| < \frac{\varepsilon}{2}$ (for if this were possible, then we would arrive at the contradiction $\varepsilon \leq |f(x_n) - f(c)| < \frac{\varepsilon}{2}$).

Claims 1 and 2 show that (1.7) does not hold, a contradiction. Hence f is continuous at c. □

The above theorem allows us to easily prove the following useful Theorem 1.3.2. Before stating it, we introduce some convenient notation.

Definitions. *Let I be an interval in $\mathbb{R}$. Given functions $f\colon I \to \mathbb{R}$ and $g\colon I \to \mathbb{R}$, we define the following:*

1. *If $\alpha \in \mathbb{R}$, then we define the function $\alpha f\colon I \to \mathbb{R}$ by $(\alpha f)(x) = \alpha \cdot f(x)$, $x \in I$.*
2. *We define the **absolute value of f** to be the function $|f|\colon I \to \mathbb{R}$ given by $|f|(x) = |f(x)|$, $x \in I$.*
3. *The **sum of f and g** is the function $f+g\colon I \to \mathbb{R}$ defined by $(f+g)(x) = f(x)+g(x)$, $x \in I$.*
4. *The **product of f and g** is the function $fg\colon I \to \mathbb{R}$ defined by $(fg)(x) = f(x)g(x)$, $x \in I$.*
5. *If $k \in \mathbb{N}$, then we define the **k^{th} power of f**, to be the function $f^k\colon I \to \mathbb{R}$ given by $f^k(x) = (f(x))^k$, $x \in I$.*
6. *If for all $x \in I$, $g(x) \neq 0$, then we define the function $\frac{1}{g}\colon I \to \mathbb{R}$ by $\left(\frac{1}{g}\right)(x) = \frac{1}{g(x)}$, $x \in I$.*

Theorem 1.3.2. *Let I be an interval in $\mathbb{R}$ and let $c \in I$. Suppose that $f\colon I \to \mathbb{R}$ and $g\colon I \to \mathbb{R}$ 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 that $(x_n)_{n\in\mathbb{N}}$ is a convergent sequence contained in I, with limit c. Since f and g are continuous at c, from Theorem 1.3.1, 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. Each one of the statements now follows easily from Theorem 1.2.6 and a second application of Theorem 1.3.1. For example, consider statement number 4 (the other cases may be proved as exercises). By Theorem 1.2.6, $(f(x_n)g(x_n))_{n\in\mathbb{N}}$ is convergent with limit $f(c)g(c)$, that is, $((fg)(x_n))_{n\in\mathbb{N}}$ is convergent with limit $(fg)(c)$. Thus by Theorem 1.3.1, fg is continuous at c. □

Example. Since $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = x$ for $x \in \mathbb{R}$ is continuous (see Example 1 on p31), it follows that for all $k \in \mathbb{N}$, x^k is continuous. Thus given arbitrary scalars $a_0, a_1, \ldots, a_N$ in $\mathbb{R}$, it follows that the functions $a_0 \cdot 1, a_1 \cdot x, \ldots, a_N \cdot x^N$ are continuous. Consequently the **polynomial function** $p\colon \mathbb{R} \to \mathbb{R}$ defined by

$$p(x) = a_0 + a_1 x + \cdots + a_N x^N, \quad x \in \mathbb{R}$$

is continuous. □

1.3.3 Restrictions and compositions of functions

Definition. *If $f\colon I \to \mathbb{R}$ is a function on an interval I, and J is an interval contained in I (that is, $J \subseteq I$), then the **restriction of f to J**, is the function $f\big|_J\colon J \to \mathbb{R}$ defined by*

$$f\big|_J(x) = f(x), \quad x \in \mathbb{R}.$$

The following theorem implies (completely as expected) that the restriction of a continuous function is continuous.

Theorem 1.3.3. *Let I and J be intervals such that $J \subseteq I$ and let $c \in J$. If f is continuous at c, then $f\big|_J$ is continuous at c.*

Proof. Exercise 71. □

The converse of the above theorem is not true unless J is an open interval (see Exercise 71).

Definition. *If $f\colon I \to \mathbb{R}$ and $g\colon J \to \mathbb{R}$ are functions such that for all $x \in I$, $f(x) \in J$, then the **composition of g with f**, is the function $g \circ f\colon I \to \mathbb{R}$ defined by*

$$(g \circ f)(x) = g(f(x)), \quad x \in \mathbb{R}.$$

The following theorem implies that the composition of continuous functions is continuous.

Theorem 1.3.4. *If $f\colon I \to \mathbb{R}$ is continuous at $c \in I$ and $g\colon J \to \mathbb{R}$ is continuous at $f(c)$ (with $f(x) \in J$ for all $x \in I$), then $g \circ f$ is continuous at c.*

Proof. Take $\varepsilon > 0$. Then, as g is continuous at $f(c)$, there is some $\eta > 0$ such that $y \in J$ and $|y - f(c)| < \eta$ imply that $|g(y) - g(f(c))| < \varepsilon$. As also f is continuous at c, there exists $\delta > 0$ such that $x \in I$ and $|x - c| < \delta$ imply $|f(x) - f(c)| < \eta$.

Now we assemble what we know. Suppose that $x \in I$ and $|x - c| < \delta$. Then $f(x) \in J$ and $|f(x) - f(c)| < \eta$, from which we deduce that $|g(f(x)) - g(f(c))| < \varepsilon$, i.e., $|(g \circ f)(x) - (g \circ f)(c)| < \varepsilon$. We conclude that $g \circ f$ is continuous at c. □

Example. *Show that the function $f\colon (1, \infty) \to \mathbb{R}$ defined by*

$$f(x) = \sqrt{\frac{x^2 - 1}{x^2 + 1}}, \quad x > 1,$$

is continuous on $(1, \infty)$.

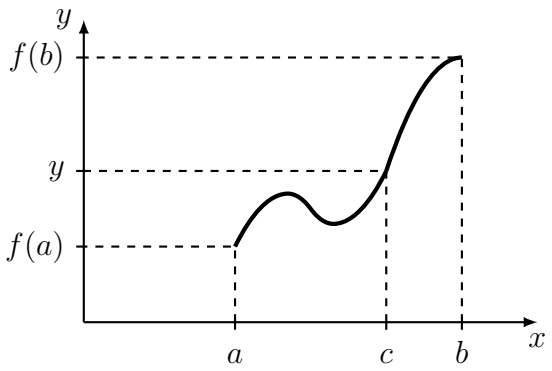
Solution. Let $g\colon (0,\infty) \to \mathbb{R}$ be defined by $g(x) = \sqrt{x}$. By Exercise 68, g is continuous. Let $h\colon \mathbb{R} \to (0,\infty)$ be defined by

$$h(x) = \frac{x^2 - 1}{x^2 + 1}, \quad x \in \mathbb{R}.$$

We have already seen that all polynomial functions are continuous. Therefore, $p_1, p_2\colon \mathbb{R} \to \mathbb{R}$ defined by $p_1(x) = x^2 - 1$ and $p_2(x) = x^2 + 1$ are continuous. Since $p_2(x) \neq 0$ for all $x \in \mathbb{R}$, it follows by Theorem 1.3.2(6) that $\dfrac{1}{p_2}$ is continuous. Then by Theorem 1.3.2(4), $p_1 \cdot \dfrac{1}{p_2} = h$ is continuous. Note that the restriction of h to $(1,\infty)$, $h|_{(1,\infty)}$ assumes only positive values: if $x \in (1,\infty)$ then $x > 1$, therefore $x^2 - 1 > 0$, and since $x^2 + 1 > 0$ anyway, $h(x) > 0$. This shows that the composition $g \circ h|_{(1,\infty)} = f$ is properly defined. By Theorem 1.3.3, $h|_{(1,\infty)}$ is continuous. Finally, by Theorem 1.3.4, $g \circ h|_{(1,\infty)} = f$ is continuous at each $c > 1$, and therefore, continuous. □

1.3.4 Intermediate value theorem

We now prove one of the most fundamental (and obvious!) theorems on continuous functions: a continuous function 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 Intermediate value theorem was first proved by Bernhard Bolzano in 1817.

Theorem 1.3.5. (Intermediate value theorem). *If $f\colon [a,b] \to \mathbb{R}$ is continuous and y is such that $f(a) \leq y \leq f(b)$ or $f(b) \leq y \leq f(a)$, then there exists $c \in [a,b]$ such that $f(c) = y$.*

Proof. Suppose first that $f(a) \leq y \leq f(b)$. The case $f(b) \leq y \leq f(a)$ is similar, and will be discussed at the end of the proof below.

Let $S = \{x \in [a,b] \mid f(x) \leq y\}$. We want to prove that

1. $\sup S$ exists, and

2. if we set $c = \sup S$, then $c \in [a,b]$ and $f(c) = y$.

Since $f(a) \leq y$, it follows that $a \in S$, so $S \neq \emptyset$. Secondly, S is clearly bounded above by b. Therefore, by the l.u.b. property, $\sup S$ exists.

Write $c = \sup S$. Since $a \in S$, $a \leq c$. Since b is an upper bound of S, $c \leq b$. Therefore, $c \in [a, b]$, and it remains to prove that $f(c) = y$. We do this in two steps: first we show that $f(c) \leq y$ and then that $f(c) \geq y$.

Proof that $f(c) \leq y$: For each $n \in \mathbb{N}$, $c - \frac{1}{n}$ is not an upper bound of S, since it is less than the *least* upper bound c. Therefore, there exists $x_n \in S$ such that $c - \frac{1}{n} < x_n \leq c$. By the Sandwich theorem, $\lim_{n\to\infty} x_n = c$. By Theorem 1.3.1, $\lim_{n\to\infty} f(x_n) = f(c)$. Since $x_n \in S$, $f(x_n) \leq y$. Therefore, $f(c) = \lim_{n\to\infty} f(x_n) \leq \lim_{n\to\infty} y = y$ (see Exercise 44.)

Proof that $f(c) \geq y$: The sequence (x_n) defined by

$$x_n = c + \frac{b - c}{n}$$

lies in $[a, b]$, and $\lim_{n\to\infty} x_n = c$. By Theorem 1.3.1, $\lim_{n\to\infty} f(x_n) = f(c)$. Since $x_n > c$ and c is an upper bound of S, $x_n \notin S$, and therefore, $f(x_n) > y$. As before, it follows that $f(c) = \lim_{n\to\infty} f(x_n) \geq \lim_{n\to\infty} y = y$.

We have shown that $f(c) \leq y$ and $f(c) \geq y$. In conclusion, we have found $c \in [a, b]$ such that $f(c) = y$.

Suppose now that $y \in \mathbb{R}$ is such that $f(b) \leq y \leq f(a)$. To prove the theorem in this case, it is easy to modify the above proof. Alternatively, we could use what we have just proved, as follows. Note that we have

$$(-f)(a) \leq -y \leq (-f)(b),$$

so we may apply the above proof to the function $-f$ (which is continuous by Theorem 1.3.2(1) with $\alpha = -1$). Then we obtain the existence of $c \in [a, b]$ such that $(-f)(c) = -y$, that is, $f(c) = y$. □

Examples.

1. *Show that every polynomial of odd degree with real coefficients has at least one real root.*

   Solution. Suppose that p is a polynomial with degree $2m + 1$, $m \in \mathbb{N} \cup \{0\}$. Let

   $$p(x) = a_0 + a_1 x + \cdots + a_{2m} x^{2m} + a_{2m+1} x^{2m+1},$$

   where $a_{2m+1} \neq 0$. If we divide p by its leading coefficient a_{2m+1} then the resulting polynomial will still have the same roots (if any). We may therefore assume without loss of generality that $a_{2m+1} = 1$.

   In order to show that p has a real root, we will choose a large enough $n \in \mathbb{N}$ such that $p(n) > 0$ and $p(-n) < 0$ and restrict our attention to an interval $[-n, n]$. Then appealing to the Intermediate Value Theorem, we can conclude that p must vanish at some point in this interval, that is, for some real $c \in [-n, n]$, $p(c) = 0$.

It remains to find an $n \in \mathbb{N}$ such that $p(n) > 0 > p(-n)$. We expect such an n to be large in general. Intuitively, the term in the polynomial that dominates when x is large is the term with highest exponent x^{2m+1}. We now compare $p(x)$ with x^{2m+1} by considering the quotient:

$$\frac{p(x)}{x^{2m+1}} = \frac{a_0}{x^{2m+1}} + \frac{a_1}{x^{2m}} + \cdots + \frac{a_{2m}}{x} + 1 \tag{1.8}$$

(recall we assumed $a_{2m+1} = 1$). If we make the substitution $x = n$ and let $n \to \infty$, we obtain

$$\lim_{n\to\infty} \frac{p(n)}{n^{2m+1}} = \lim_{n\to\infty} \left(\frac{a_0}{n^{2m+1}} + \frac{a_1}{n^{2m}} + \cdots + \frac{a_{2m}}{n} + 1 \right) = 1.$$

By choosing $\varepsilon = 1/2$ in the definition of a limit, we obtain that there exists $N_1 \in \mathbb{N}$ such that for all $n > N_1$,

$$\left| \frac{p(n)}{n^{2m+1}} - 1 \right| < \varepsilon = \frac{1}{2},$$

and since $|A| \geq -A$ for any $A \in \mathbb{R}$, it follows that

$$1 - \frac{p(n)}{n^{2m+1}} < \frac{1}{2},$$

or $p(n) > \frac{n^{2m+1}}{2} > 0$ as long as $n > N_1$.

To obtain the inequality $p(-n) < 0$ for suitably large n, we next make the substitution $x = -n$ in (1.8) and take the limit as $n \to \infty$ to obtain

$$\lim_{n\to\infty} \frac{p(-n)}{(-n)^{2m+1}} = \lim_{n\to\infty} \left(\frac{a_0}{(-n)^{2m+1}} + \frac{a_1}{(-n)^{2m}} + \cdots + \frac{a_{2m}}{-n} + 1 \right) = 1.$$

Again applying the definition of a limit with $\varepsilon = 1/2$, we obtain as before that there exists an $N_2 \in \mathbb{N}$ such that for all $n > N_2$,

$$1 - \frac{p(-n)}{(-n)^{2m+1}} < \frac{1}{2}.$$

Multiplying by n^{2m+1} we obtain (careful with the minus!)

$$n^{2m+1} + p(-n) < \frac{n^{2m+1}}{2},$$

or $p(-n) < -\frac{n^{2m+1}}{2} < 0$ as long as $n > N_2$.

Thus if we choose $n = 1 + \max\{N_1, N_2\}$, we obtain both inequalities $p(n) > 0$ and $p(-n) < 0$, which finishes the proof. □

For example, the polynomial $p(x) = x^3 - x^{2014} + \frac{1}{399}$ has a real root in $[-1, 1]$: indeed $p(1) = \frac{1}{399} > 0$ and $p(-1) = -2 + \frac{1}{399} < 0$ and so $\exists c \in [-1, 1]$ such that $p(c) = 0$. □

2. *Show that at any given time, there exists a pair of diametrically opposite points on the equator which have the same temperature.*

Solution. Let $T(\Theta)$ denote the surface temperature at the point at longitude Θ. See Figure 1.11. (Note that $\Theta(0) = \Theta(2\pi)$.) Assuming that Θ is a continuous function of

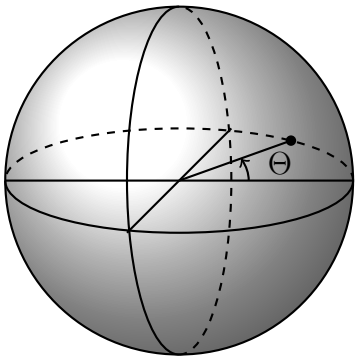
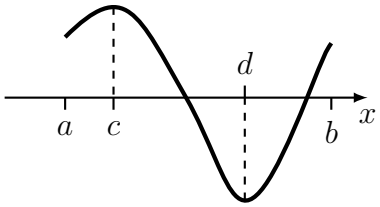


Figure 1.11: The point on the equator with longitude Θ.

Θ, it follows that the function $f\colon [0,\pi] \to \mathbb{R}$ defined by $f(\Theta) = T(\Theta) - T(\Theta + \pi)$ is continuous as well. If $f(0) = 0$, then it follows that the temperatures at 0 and 180° longitude are the same. If $f(0) \neq 0$, then since $f(0)$ and $f(\pi) = -f(0)$ have opposite signs, by the intermediate value theorem, it follows that f must vanish at some point, and so the claim follows. □

1.3.5 Extreme value theorem

The next theorem states that a continuous function on an interval $[a, b]$ attains its maximum and minimum values.



Theorem 1.3.6 (Extreme value theorem)**.** *Let $[a, b]$ be any closed and bounded interval and let $f\colon [a, b] \to \mathbb{R}$ be a continuous function. Then there exists $c \in [a, b]$ such that*

$$f(c) = \sup\{f(x) \mid x \in [a, b]\}, \tag{1.9}$$

and there exists $d \in [a, b]$ such that

$$f(d) = \inf\{f(x) \mid x \in [a, b]\}. \tag{1.10}$$

Since $c, d \in [a, b]$ in the above theorem, the supremum and infimum in (1.9) and (1.10) are in fact the maximum and minimum, respectively. This proof is a beautiful application of the Bolzano-Weierstrass theorem.

Proof. We prove the first half of the theorem, leaving the second half as an exercise.

Let $S = \{f(x) \mid x \in [a, b]\}$. We have to prove that there exists $c \in [a, b]$ such that $f(c) = \sup S$. The plan of the proof is as follows.

1. First we show that $\sup S$ exists.

2. Then we show that $\sup S \in S$.

Then, by the definition of S, we may conclude that $\sup S = f(c)$ for some $c \in [a, b]$.

To show that $\sup S$ exists, we use the l.u.b. property of $\mathbb{R}$. First, $S \neq \emptyset$, since for example $f(a) \in S$. Secondly, S is bounded above, as shown by the following proof by contradiction.

Suppose S is not bounded above. For each $n \in \mathbb{N}$, choose an $f(x_n)$ for some $x_n \in [a, b]$ such that $f(x_n) > n$. Then (x_n) is a bounded sequence. By the Bolzano-Weierstrass Theorem, (x_n) has a convergent subsequence $(x_{n_k})_{k\in\mathbb{N}}$. Let its limit be $c = \lim_{k\to\infty} x_{n_k}$. Since $a \leq x_n \leq b$ and limits preserve inequalities, $a \leq c \leq b$. By Theorem 1.3.1, $(f(x_{n_k})_{k\in\mathbb{N}})$ is convergent (we assumed that f is continuous). By Theorem 1.2.4, $(f(x_{n_k})_{k\in\mathbb{N}})$ is bounded. On the other hand, for each $k \in \mathbb{N}$, $f(x_{n_k}) > n_k \geq k$ (since $1 \leq n_1 < n_2 < n_3 < \dots$). Therefore, $(f(x_{n_k})_{k\in\mathbb{N}})$ is unbounded, a contradiction.

Next we show that $\sup S \in S$. Recall that $S = \{f(x) \mid x \in [a, b]\}$. Write $M = \sup S$. We have to prove that $M \in S$. Since M is the *least* upper bound of S, it follows that for all $n \in \mathbb{N}$, $M - \frac{1}{n}$ is not an upper bound of S. Therefore, there exists $f(x_n) \in S$ (with $x_n \in [a, b]$) such that $M - \frac{1}{n} < f(x_n)$. The sequence (x_n) is bounded. As before, by the Bolzano-Weierstrass Theorem it has a convergent subsequence $(x_{n_k})_{k\in\mathbb{N}}$ with limit $c \in [a, b]$, say. We next show that this c satisfies $f(c) = M = \sup S$.

By Theorem 1.3.1, $(f(x_{n_k})_{k\in\mathbb{N}}$ is convergent with limit $f(c)$. Since $c \in [a, b]$, $f(c) \in S$. Finally, because

$$M - \frac{1}{n_k} < f(x_{n_k}) \leq M \text{ and } \lim_{k\to\infty} M - \frac{1}{n_k} = M,$$

the Sandwich Theorem shows that $f(c) = \lim_{k\to\infty} f(x_{n_k}) = M$.

We have found $c \in [a, b]$ such that $f(c) = M = \sup S$. This finishes the proof of the first half of the theorem.

The proof that there exists $d \in [a, b]$ such that $f(d) = \inf S$ is left as an exercise (Exercise 80). □

Example. *Show that the statement of Theorem 1.3.6 does not hold if* $[a, b]$ *is replaced by* (a, b).

Solution. The function $f\colon (0,1) \to \mathbb{R}$ given by $f(x) = x$ is continuous on $(0, 1)$. If

$$S = \{f(x) \mid x \in (0,1)\},$$

then $S = \{x \mid x \in (0,1)\} = (0,1)$. Thus $\sup S = 1 \notin S$ and $\inf S = 0 \notin S$, that is, the values 0 and 1 are not attained by the function f. We conclude that the statement of Theorem 1.3.6 does not hold if $[a, b]$ is replaced by (a, b). □

1.4 Exercises

1. Fill in the following table for each of the sets $S = A, B, \ldots, K$ below.

An upper bound	A lower bound	Is S bounded?	$\sup S$	$\inf S$	If $\sup S$ exists, then is $\sup S$ in S?	If $\inf S$ exists, then is $\inf S$ in S?	$\max S$	$\min S$

a) $A = \{x \in \mathbb{R} \mid 0 \leq x < 1\}$

b) $B = \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$

c) $C = \{x \in \mathbb{R} \mid x \in \mathbb{N} \text{ and } x\text{prime}\}$

d) $D = \left\{\frac{1}{n} \;\middle|\; n \in \mathbb{Z} \setminus \{0\}\right\}$

e) $E = \left\{-\frac{1}{n} \;\middle|\; n \in \mathbb{N}\right\}$

f) $F = \{x \in \mathbb{R} \mid x > x^2\}$

g) $G = \{x \in \mathbb{R} \mid x^2 + 2x \leq 1\}$

h) $H = \{0, 2, 5, 2016\}$

i) $I = \left\{(-1)^n \left(1 + \frac{1}{n}\right) \;\middle|\; n \in \mathbb{N}\right\}$

j) $J = \{x^2 \mid x \in \mathbb{R}\}$

k) $K = \left\{\frac{x^2}{1+x^2} \;\middle|\; x \in \mathbb{R}\right\}$

2. Determine whether the following statements are TRUE or FALSE. Give reasons for your answers in each case: either a (brief) proof or a *counterexample*. (For instance, if you think statement (d) is false, you should give an example of a bounded non-empty subset of $\mathbb{R}$ that does not have a maximum, and explain (briefly) why your set has the required properties.)

   a) Let S be a subset of $\mathbb{R}$ such that, for each $x \in S$, there exists $u \in \mathbb{R}$ with $x \leq u$. Then S is bounded above.

   b) If u is an upper bound of a subset S of $\mathbb{R}$, and $u' < u$, then u' is not an upper bound of S.

   c) If u_* is the supremum of a subset S of $\mathbb{R}$, and ε is any positive real number, then $u_* - \varepsilon$ is not an upper bound of S.

   d) Every bounded non-empty subset of $\mathbb{R}$ has a maximum.

   e) Every bounded subset of $\mathbb{R}$ has a supremum.

f) A non-empty subset of $\mathbb{R}$ is bounded above iff [6] it has a supremum.

g) For every set that has a supremum, the supremum belongs to the set.

h) For every non-empty bounded set S, if $\inf S < x < \sup S$, then $x \in S$.

i) For S a non-empty subset of $\mathbb{R}$, $\{x^2 \mid x \in S\}$ is bounded iff S is bounded.

3. For any non-empty bounded set S, prove that $\inf S \leq \sup S$, and that the equality holds iff S is a singleton set (that is a set with exactly one element).

4. Let A and B be non-empty subsets of $\mathbb{R}$ that are bounded above and such that $A \subseteq B$. Prove that $\sup A \leq \sup B$.

5. Let A and B be non-empty subsets of $\mathbb{R}$ that are bounded above and define

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

a) Show that $\sup A + \sup B$ is an upper bound for $A + B$.
Deduce that $\sup(A + B)$ exists and that $\sup(A + B) \leq \sup A + \sup B$.

b) For any real number $\varepsilon > 0$, show that $(\sup A - \varepsilon) + (\sup B - \varepsilon)$ is **not** an upper bound for $A + B$.
Deduce that $\sup(A + B) \geq \sup A + \sup B - 2\varepsilon$, for every $\varepsilon > 0$.

c) Show that $\sup(A + B) = \sup A + \sup B$.

6. Let S be a non-empty subset of real numbers which is bounded below. Let $-S$ denote the set of all real numbers $-x$, where x belongs to S.

a) Show that $-S$ is bounded above, and therefore has a supremum.

b) Prove that $\inf S$ exists, and that $\inf S = -\sup(-S)$.

Note: in this exercise, we deduce the "greatest lower bound principle" from the least upper bound principle. Accordingly, you may *not* assume the greatest lower bound principle in this exercise. (However, once we have established this result, we *can* use the greatest lower bound principle henceforth.)

7. In this exercise it is shown that for any positive real number r there exists a positive real number s such that $s^2 = r$. In other words, any positive real number has a positive square root.

Given $r \in \mathbb{R}$ with $r > 0$, let $S = \{x \in \mathbb{R} \mid x^2 < r\}$.

a) Prove that $S \neq \emptyset$.

b) Prove that S is bounded above.

HINT: Distinguish between the cases $r > 1$ and $r \leq 1$.

c) Conclude that S has a least upper bound s. Show that $s > 0$.

[6] If you think the statement is TRUE, you have two things to prove: (i) if a non-empty set S is bounded above, then it has a supremum; (ii) if a non-empty set S has a supremum, then it is bounded above. If you think the statement is FALSE, you should provide a counterexample to *one* of (i) and (ii).

d) Show that if $s^2 > r$ then there exists $\varepsilon > 0$ such that $s - \varepsilon$ is an upper bound of S.

e) Show that if $s^2 < r$ then there exists $\varepsilon > 0$ such that $s + \varepsilon \in S$.

f) Conclude that $s^2 = r$ by explaining why each of the conclusions in 7d) and 7e) gives a contradiction.

8. Let S be a non-empty set of positive real numbers, and define $S^{-1} = \left\{ \frac{1}{x} \;\middle|\; x \in S \right\}$.

a) Show that, if $\inf S = 0$, then S^{-1} is not bounded above.

b) Show that, if $\inf S > 0$, then S^{-1} is bounded above and $\sup S^{-1} = \frac{1}{\inf S}$.

9. In this exercise we define the **floor** and **ceiling** functions. For any $x \in \mathbb{R}$ define the set $S_x = \{n \in \mathbb{Z} \mid n \leq x\}$.

a) Show that, for any $x \in \mathbb{R}$, the set S_x is non-empty and bounded above.

HINT: To show that $S_x \neq \emptyset$, you will need the Archimedean property.

b) Show that $\sup S_x$ exists and $\sup S_x \in S_x$ for any $x \in \mathbb{R}$. Explain why we obtain as a consequence that the following gives a proper definition of the **floor function** $\lfloor \cdot \rfloor : \mathbb{R} \to \mathbb{Z}$

$$\lfloor x \rfloor = \max \{n \in \mathbb{Z} \mid n \leq x\}, \quad x \in \mathbb{R}.$$

c) Show that $x - 1 < \lfloor x \rfloor \leq x$, for every $x \in \mathbb{R}$.

d) Adapt 9a) and 9b) to explain why the following gives a proper definition of the **ceiling function** $\lceil \cdot \rceil : \mathbb{R} \to \mathbb{Z}$

$$\lceil x \rceil = \min \{n \in \mathbb{Z} \mid n \geq x\}, \quad x \in \mathbb{R}.$$

e) Show that $\lfloor \sqrt{k^2 + k} \rfloor = k$ for all $k \in \mathbb{N}$.

10. Let S be a subset of $\mathbb{R}$ that does not have a maximum. Show that, for every $x \in S$, there is $y \in S$ with $y > x$.

HINT: Deal separately with the two cases: (i) S has a supremum; (ii) S has no supremum.

11. Determine whether the following statements are TRUE or FALSE. Give reasons (brief proofs or counterexamples) for your answers in each case.

a) If $a \leq b$ then $(-\infty, a) \subseteq (-\infty, b)$.

b) For real numbers a and b, $a \leq b$ iff $(b, \infty) \subseteq (a, \infty)$.

c) $(a, b) \subseteq (c, d)$ iff $c \leq a$ and $b \leq d$.

d) $(a, b) \subseteq [c, d]$ iff $c < a$ and $b < d$.

e) For any $x \in \mathbb{R}$, $-|x| \leq x \leq |x|$.

f) For any $x \in \mathbb{R}$, $-x \leq |x| \leq x$.

g) For any $x, y \in \mathbb{R}$, if $x < y$ then $|x| \leq |y|$.

h) For any $x, y \in \mathbb{R}$, if $x^2 \leq y^2$ then $|x| \leq |y|$.

i) If I and J are intervals, then $I \cap J$ is an interval.

j) If I and J are intervals with $0 \in I \cap J$, then $I \cup J$ is an interval.

12. Let $x_0 \in \mathbb{R}$ and $\delta > 0$. Prove that $(x_0 - \delta, x_0 + \delta) = \{x \in \mathbb{R} \mid |x - x_0| < \delta\}$.

13. Prove that the distance satisfies the following properties:

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

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

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

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

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

a) Show that, if S is bounded, then there is some $M \in \mathbb{R}$ such that $|x| \leq M$ for all $x \in S$.

b) Show also the converse: if there is some $M \in \mathbb{R}$ such that $|x| \leq M$ for all $x \in S$, then S is bounded.

Thus we have an alternative definition of what it means for a set S to be bounded: this definition is often more convenient to use.

16. Let a, b, c be real numbers such that $a < c < b$. Show that there exists a $t > 0$ such that $(c - t, c + t) \subseteq (a, b)$.

17. Prove, directly from the definition, that the constant sequence $(1)_{n \in \mathbb{N}}$ is convergent.

18. Prove, directly from the definition, that the sequence $\left(\frac{2}{\sqrt{n}}\right)_{n \in \mathbb{N}}$ is convergent, with limit 0.

19. Prove, directly from the definition, that the sequence $\left(\frac{3n-1}{n+2}\right)_{n \in \mathbb{N}}$ is convergent, and find its limit.

20. a) Let $(a_n)_{n \in \mathbb{N}}$ be a convergent sequence with limit L, and let M be some real number with $M \neq L$. Show that the set $\{n \in \mathbb{N} \mid a_n = M\}$ is bounded above.

b) Prove that the sequence $((-1)^n)_{n \in \mathbb{N}}$ is divergent.

21. Use the definition of limit to prove directly that 1 is not a limit of the sequence $(1/n)_{n \in \mathbb{N}}$.

22. 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.

a) For every $\varepsilon > 0$, there exists an N such that, for infinitely many $n > N$, $|a_n - 1| < \varepsilon$.

b) There exists an $\varepsilon > 0$ and an $N \in \mathbb{N}$ such that for all $n > N$, $|a_n - 1| < \varepsilon$.

23. Let S be a non-empty subset of $\mathbb{R}$ that 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 $n \in \mathbb{N}$) which is convergent with limit equal to $\sup S$.

24. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence such that for all $n \in \mathbb{N}$, $a_n \geq 0$. Prove that if $(a_n)_{n\in\mathbb{N}}$ is convergent with limit L, then $L \geq 0$.

25. A sequence $(a_n)_{n\in\mathbb{N}}$ is said to be a **Cauchy sequence** if for every $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that for all $n, m > N$, $|a_n - a_m| < \varepsilon$.

    Show that every convergent sequence is Cauchy.

    HINT: $|a_n - a_m| = |a_n - L + L - a_m| \leq |a_n - L| + |a_m - L|$.

26. In the proof of Theorem 1.2.2, where did we use the assumption that $x \geq -1$? Is the inequality true for $x = -2$? Is it true for $x = -3$?

27. For a real number x such that $|x| < 1$, show that the sequence $(x^n\sqrt{n})_{n\in\mathbb{N}}$ is convergent, with limit 0.

28. a) Use Bernoulli's Inequality to show that, for all $n \in \mathbb{N}$,

$$1 \leq n^{\frac{1}{n}} < (1+\sqrt{n})^{\frac{2}{n}} \leq \left(1 + \frac{1}{\sqrt{n}}\right)^2 \leq 1 + \frac{3}{\sqrt{n}}.$$

    b) Prove that $(n^{\frac{1}{n}})_{n\in\mathbb{N}}$ is convergent with limit 1.

29. Let a and b be positive real numbers. Show that the sequence $\left(\frac{n+a}{n+b}\right)_{n\in\mathbb{N}}$ is monotone.

30. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence defined by

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

    Prove that $(a_n)_{n\in\mathbb{N}}$ is convergent.

31. Use Bernoulli's Inequality (Theorem 1.2.2) to show that, for x a real number with $x > 1$, the sequence $(x^n)_{n\in\mathbb{N}}$ is divergent.

32. Suppose that the sequence $(a_n)_{n\in\mathbb{N}}$ converges to 0 and the sequence $(b_n)_{n\in\mathbb{N}}$ is bounded. Use the definition to prove that $\lim_{n\to\infty} a_n b_n = 0$.

33. For each $n \in \mathbb{N}$, let $H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$. The sequence $(H_n)_{n\in\mathbb{N}}$ is the sequence of **harmonic numbers**.

    a) Show that for any $m, n \in \mathbb{N}$ with $m > n$,

$$H_m - H_n > \frac{m-n}{m}.$$

    (*Hint: note that* $H_m - H_n = \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{m}$, *a sum of* $m - n$ *terms.*)

    b) Deduce that $H_{2^{k+1}} - H_{2^k} > \frac{1}{2}$ for any $k \in \mathbb{N}$.

    c) Show that $H_2 = \frac{3}{2}$, $H_4 > \frac{4}{2}$, $H_8 > \frac{5}{2}$, $H_{16} > \frac{6}{2}$, $H_{32} > \frac{7}{2}$, and in general, $H_{2^k} > \frac{k+2}{2}$ for all $k \geq 2$.

d) Let $M > 0$ be given. Show that there exists $n \in \mathbb{N}$ such that $H_n > M$.

e) Show that $(H_n)_{n\in\mathbb{N}}$ is divergent.

34. a) Let $(a_n)_{n\in\mathbb{N}}$ be a convergent sequence with limit L. Prove that the sequence $(s_n)_{n\in\mathbb{N}}$, where
$$s_n = \frac{a_1 + \cdots + a_n}{n}, \quad n \in \mathbb{N},$$
is also convergent with limit L.
Hint: use the fact that $(a_n)_{n\in\mathbb{N}}$ is bounded.

b) Give an example of a sequence $(a_n)_{n\in\mathbb{N}}$ such that $(s_n)_{n\in\mathbb{N}}$ is convergent but $(a_n)_{n\in\mathbb{N}}$ is divergent.

35. a) Given a bounded sequence $(a_n)_{n\in\mathbb{N}}$, define
$$\ell_k = \inf\{a_n \mid n \geq k\}, \quad k \in \mathbb{N}.$$
Show that the sequence $(\ell_n)_{n\in\mathbb{N}}$ is bounded above and increasing, and conclude that it is convergent. (Its limit is denoted by $\liminf_{n\to\infty} a_n$.)

b) Find $\liminf_{n\to\infty} (2 + (-1)^n + 1/n)$.

36. Recall the definition of a Cauchy sequence from Exercise 25. Prove that every Cauchy sequence is bounded. (*H*int: follow the proof of Theorem 1.2.2.)

37. Determine whether the following sequences are convergent and, if so, find their limits:
$$\left(\frac{n+\sqrt{n}}{\sqrt{3n^2-1}}\right)_{n\in\mathbb{N}}; \qquad \left(1 + n^{1/2n} + \sqrt{n}(\frac{3}{4})^n\right)_{n\in\mathbb{N}}.$$

38. Let $(a_n)_{n\in\mathbb{N}}$ be a convergent sequence with limit L, and let $k \in \mathbb{N}$. Show that $(a_{n+k})_{n\in\mathbb{N}}$ is also convergent with limit L.

39. Recall the convergent sequence $(a_n)_{n\in\mathbb{N}}$ from Exercise 30 defined by
$$a_1 = 1 \text{ and } a_n = \frac{2n+1}{3n} a_{n-1} \text{ for } n \geq 2.$$
Determine its limit.

HINT: Use the previous exercise.

40. a) Use Theorem 1.2.6 and induction to show that for any fixed $k \in \mathbb{N}$,
$$\lim_{n\to\infty} \left(\frac{1}{n+1} + \frac{1}{n+2} + \frac{1}{n+3} + \cdots + \frac{1}{n+k-1} + \frac{1}{n+k}\right) = 0.$$
Indicate step-by-step which parts of Theorem 1.2.6 you use.

b) Find the mistake in the following calculation:

$$\begin{aligned}
&\lim_{n\to\infty}\left(\frac{1}{n+1}+\frac{1}{n+2}+\frac{1}{n+3}+\cdots+\frac{1}{2n-1}+\frac{1}{2n}\right)\\
&=\lim_{n\to\infty}\frac{1}{n+1}+\lim_{n\to\infty}\frac{1}{n+2}+\lim_{n\to\infty}\frac{1}{n+3}+\cdots+\lim_{n\to\infty}\frac{1}{2n-1}+\lim_{n\to\infty}\frac{1}{2n}\\
&=0+0+0+\cdots+0+0\\
&=0.
\end{aligned}$$

Remark: By Exercise 33a),

$$\frac{1}{n+1}+\frac{1}{n+2}+\frac{1}{n+3}+\cdots+\frac{1}{2n-1}+\frac{1}{2n}=H_{2n}-H_n>\frac{1}{2},$$

and therefore the limit, if it exists, cannot equal 0.[7]

41. Suppose that the sequence $(a_n)_{n\in\mathbb{N}}$ is bounded. Prove that the sequence $(c_n)_{n\in\mathbb{N}}$ defined by
$$c_n=\frac{a_n^3+5n}{a_n^2+n}$$
is convergent, and find its limit.

42. Let $(a_n)_{n\in\mathbb{N}}$ be a convergent sequence with limit 0 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 0.

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

Hint: 'Rationalise the numerator' (as in the proof of Theorem 1.2.6.(8)).

44. Prove that if $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ are convergent sequences such that for all $n\in\mathbb{N}$, $a_n\leq b_n$, then
$$\lim_{n\to\infty}a_n\leq\lim_{n\to\infty}b_n.$$

Hint: Use Exercise 24.

45. Show that, for each $k\in\mathbb{N}$ and each real number x with $|x|<1$, $\lim_{n\to\infty}x^n n^k=0$.

Hint: Write the elements of the sequence as $(y^n\sqrt{n})^{2k}$, for some suitably chosen y, and use Exercise 27 as well as the Algebra of Limits.

46. Prove that the sequence $\left((3+(-1)^n)^{1/n}\right)_{n\in\mathbb{N}}$ is convergent, with limit 1.

47. Recall the *floor function* defined in Exercise 9: $\lfloor x\rfloor=\max\{n\in\mathbb{Z}\mid n\leq x\}$.

Let x be any real number. Set $a_n=\dfrac{\lfloor nx\rfloor}{n}$, for each $n\in\mathbb{N}$. Show that the sequence $(a_n)_{n\in\mathbb{N}}$ is convergent, with limit x.

[7] In fact, $\lim_{n\to\infty}\left(\frac{1}{n+1}+\frac{1}{n+2}+\frac{1}{n+3}+\cdots+\frac{1}{2n-1}+\frac{1}{2n}\right)=\ln 2$, the proof of which is outside the scope of this course.

Is the sequence $(a_n)_{n\in\mathbb{N}}$ always increasing?

Notice that the a_n are all rational numbers, so we have found a sequence of rational numbers converging to the arbitrary real number x.

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

HINT: Observe that $0\le\frac{n!}{n^n}=\frac{1}{n}\cdot\frac{2}{n}\cdot\cdots\cdot\frac{n}{n}\le\frac{1}{n}\cdot 1\cdot\cdots\cdot 1=\frac{1}{n}$.

49. Use the Sandwich Theorem to find $\lim_{n\to\infty} n\sum_{i=n+1}^{2n}\frac{1}{i^2}$.

HINT: $\frac{1}{i}-\frac{1}{i+1}=\frac{1}{i(i+1)}\le\frac{1}{i^2}\le\frac{1}{i(i-1)}=\frac{1}{i-1}-\frac{1}{i}$. Compare with Exercise 40b.

50. Prove that for all $k\in\mathbb{N}$, the sequence

$$\left(\frac{1^k+2^k+3^k+\cdots+n^k}{n^{k+2}}\right)_{n\in\mathbb{N}}$$

is convergent and

$$\lim_{n\to\infty}\frac{1^k+2^k+3^k+\cdots+n^k}{n^{k+2}}=0.$$

51. Suppose that $(a_n)_{n\in\mathbb{N}}$ is a sequence with $\frac{1}{n}\le a_n\le n$ for each $n\in\mathbb{N}$. Show that $\lim_{n\to\infty}a_n^{1/n}=1$.

HINT: See Exercise 28.

52. a) Show that a monotonically increasing sequence always has a constant subsequence or a strictly increasing subsequence.

    b) Show that any sequence of real numbers always has either a constant subsequence or a strictly increasing subsequence or a strictly decreasing subsequence.

53. a) For $x\in\mathbb{R}$, show that $x(1-x)\le\frac{1}{4}$.

    b) Define the sequence $(a_n)_{n\in\mathbb{N}}$ recursively by $a_1=1/3$ and $a_{n+1}=4a_n(1-a_n)$ for each $n\in\mathbb{N}$. Write down a_2, a_3 and a_4. Show that $(a_n)_{n\in\mathbb{N}}$ has a convergent subsequence.

54. a) Let $L\in\mathbb{R}$ and let $(a_n)_{n\in\mathbb{N}}$ be a sequence of real numbers that does not converge to L (that is, it is either divergent or its limit is not equal to L). Use the definition of convergence to L to show that for some $\varepsilon>0$, $(a_n)_{n\in\mathbb{N}}$ has a subsequence $(a_{n_k})_{k\in\mathbb{N}}$ such that $a_{n_k}\notin(L-\varepsilon,L+\varepsilon)$ for all $k\in\mathbb{N}$.

    b) Use 54a) and the Bolzano-Weierstrass theorem to show that given any $L\in\mathbb{R}$, any bounded, divergent sequence has a convergent subsequence with limit not equal to L.

    c) Use 54b) to show that any bounded, divergent sequence has two subsequences which converge to different limits.

d) Give an example of a bounded divergent sequence such that any convergent subsequence converges to either 1 or -1. (This shows that we cannot hope for more than two different limits for subsequences in c).)

55. Recall the definition of a Cauchy sequence from Exercise 25, where we had already seen that every convergent sequence is Cauchy. Use the Bolzano-Weierstrass theorem to prove the converse: if a sequence is Cauchy, then it is convergent.

    HINT: Proceed as follows. Let $(a_n)_{n\in\mathbb{N}}$ be a Cauchy sequence. From Exercise 36, it follows that $(a_n)_{n\in\mathbb{N}}$ is bounded. By the Bolzano-Weierstrass theorem, it follows that $(a_n)_{n\in\mathbb{N}}$ has a convergent subsequence, say $(a_{n_k})_{k\in\mathbb{N}}$ with limit L. Prove (using the fact that $(a_n)_{n\in\mathbb{N}}$ is Cauchy), that then $(a_n)_{n\in\mathbb{N}}$ is itself convergent with limit L.

56. Prove that the following functions are continuous:

    a) $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = 2 - 3x$.

    b) $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = |x|$.

57. Prove, using the definition of continuity, that the function $f\colon [0,\infty) \to \mathbb{R}$ given by $f(x) = \sqrt{x}$ is continuous at 4.

    HINT: First show that $|x - 4| \geq 2|\sqrt{x} - 2|$ for every $x \in [0,\infty)$.

58. Prove that the function $f\colon \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} x & \text{if } x \text{ is irrational,} \\ 0 & \text{if } x \text{ is rational,} \end{cases}$$

    is continuous at 0.

59. Let the function $f\colon \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$.

    a) Prove that f is continuous at 0.

    b) Prove that f is continuous at 1.

    c) Suppose that c is a nonzero real number. Prove that f is continuous at c.

    (In Exercise 64, we will give a simpler proof of the fact that f is continuous on $\mathbb{R}$.)

60. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a function that satisfies $f(x_1 + x_2) = f(x_1) + f(x_2)$ for all $x_1, x_2 \in \mathbb{R}$.

    a) Show that $f(0) = 0$.

    b) Suppose that f is continuous at 0. Prove that f is continuous on $\mathbb{R}$.

    HINT: Since f is continuous at 0, given $\varepsilon > 0$, $\exists \delta > 0$ such that for all $x \in \mathbb{R}$ satisfying $|x| < \delta$, $|f(x)| < \varepsilon$ (why?). 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).

    c) Give an example of such a function.

61. Suppose that $f\colon \mathbb{R} \to \mathbb{R}$ 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: First find $f(0)$.

62. Let $f\colon \mathbb{R} \to \mathbb{R}$ 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 discontinuous 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.

63. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a continuous function. Prove that if for some $c \in \mathbb{R}$, $f(c) > 0$, then there exists a $\delta > 0$ such that for all $x \in (c - \delta, c + \delta)$, $f(x) > 0$.

64. Recall Exercise 59. The function $f\colon \mathbb{R} \to \mathbb{R}$ is given by $f(x) = x^2$. Using the characterization of continuity provided in Theorem 1.3.1, prove that f is continuous on $\mathbb{R}$.

65. Prove that if $f\colon \mathbb{R} \to \mathbb{R}$ is continuous and $f(x) = 0$ whenever x is rational, then $f(x) = 0$ for all $x \in \mathbb{R}$. (Compare this to Exercises 58 and 62.)

HINT: Given any real number c, there exists a sequence of rational numbers $(q_n)_{n\in\mathbb{N}}$ that converges to c: see Exercise 47.

66. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a function that preserves 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 $x_1, x_2, x_1, x_2, \ldots$.

67. Write out a complete proof of statement 6 of Theorem 1.3.2.

68. Show, using Theorem 1.2.6, that the function $f\colon [0, \infty) \to \mathbb{R}$ given by $f(x) = \sqrt{x}$, is continuous on $[0, \infty)$.

69. Show that the **rational function** $f\colon \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \frac{x^2}{1 + x^2}, \quad x \in \mathbb{R},$$

is continuous on $\mathbb{R}$.

70. Prove Theorem 1.3.3 using the definition of continuity.

71. a) Let $J = (a, b)$ be an open interval contained in another interval I. Let $f\colon I \to \mathbb{R}$ be a function. Let $c \in J$ and assume that $f\big|_J$ is continuous at c. Prove that f is continuous at c.

HINT: You'll need Exercise 16.

b) Give an example of intervals J and I with $J \subseteq I$, $c \in J$ and a function $f\colon I \to \mathbb{R}$ such that $f\big|_J$ is continuous at c, but f is not continuous at c.

(This shows that it is necessary to assume that J is an *open* interval in 71a).)

HINT: The floor function.

72. a) Suppose $g : I \to \mathbb{R}$ is continuous, and $g(x) \geq 0$ for $x \in I$. Let $f(x) = \sqrt{g(x)}$. Show that f is continuous on I.

    b) Prove that the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \sqrt{1 + \sqrt{\frac{1 + 2x + x^2}{2 - 2x + x^2}}}, \quad x \in \mathbb{R},$$

    is continuous.

73. Suppose that $f\colon [0,1] \to \mathbb{R}$ 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 continuous function $g(x) = f(x) - x$, and use the intermediate value theorem.

74. 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 $f(t) = u(t) - d(t)$.

75. Show that the polynomial function $p(x) = 2x^3 - 5x^2 - 10x + 5$ has a real root in the interval $[-1, 2]$.

76. Set $p(x) = x^8 + bx^3 - 5$, where b is a real number. Show that $p(x)$ has a root in $[0,1]$ iff $b \geq 4$.

77. Let $f\colon \mathbb{R} \to \mathbb{R}$ be continuous. If $S := \{f(x) \mid x \in \mathbb{R}\}$ 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 $f(x_0) < y < f(x_1)$.

78. Let $f\colon [0,1] \to \mathbb{R}$ be an arbitrary continuous function. Show that there exists a $c \in [0,1]$ such that

$$f(c) - f(1) = (f(0) - f(1))c.$$

    HINT: Find an appropriate function g on which the Intermediate Value theorem can be applied.

79. a) Show that, given any continuous function $f\colon \mathbb{R} \to \mathbb{R}$, there exists $x_0 \in [0,1]$ and $m \in \mathbb{Z}$ such that $f(x_0) = mx_0$. In other words, the graph of f intersects some line $y = mx$ with integer slope, at some point x_0 in $[0,1]$.

    HINT: If $f(0) = 0$, take $x_0 = 0$ and any $m \in \mathbb{Z}$. 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 proceed in a similar manner.

b) Prove that there does not exist a continuous function $f\colon \mathbb{R} \to \mathbb{R}$ such that f assumes rational values at irrational numbers, and irrational values at rational numbers, that is,

$$f(\mathbb{Q}) \subseteq \mathbb{R} \setminus \mathbb{Q} \text{ and } f(\mathbb{R} \setminus \mathbb{Q}) \subseteq \mathbb{Q}.$$

HINT: Show that, if f is such a function, then there do not exist $m \in \mathbb{Z}$ and $x_0 \in \mathbb{R}$ such that $f(x_0) = mx_0$.

80. Complete the proof of Theorem 1.3.6 by proving that if $f\colon [a, b] \to \mathbb{R}$ is continuous, then there exists $d \in [a, b]$ such that $f(d) = \inf\{f(x) \mid x \in [a, b]\}$.

81. Show that the statement of Theorem 1.3.6 does not hold if $[a, b]$ is replaced by $[a, b)$.

82. Give examples of functions $f_1, f_2, f_3\colon [0, 1] \to \mathbb{R}$, continuous on $(0, 1]$, with the following properties.

a) f_1 is not continuous at 0, and $\{f(x) \mid x \in [0, 1]\}$ is not bounded above;

b) f_2 is not continuous at 0, and $\{f(x) \mid x \in [0, 1]\}$ is bounded above, but does not have a maximum;

c) f_3 is not continuous at 0, and $\{f(x) \mid x \in [0, 1]\}$ has a maximum.

83. Let $f\colon [a, b] \to \mathbb{R}$ be an arbitrary continuous function. Let $S = \{f(x) \mid a \leq x \leq b\}$. Show that if S contains more than one element, then S is an interval of the form $[c, d]$.

HINT: First apply the Extreme Value theorem, then the Intermediate Value theorem.

84. The function $f : \mathbb{R} \to \mathbb{R}$ is given by $f(x) = 3 + 4x - x^2 - x^4$. Show that there exists $c \in \mathbb{R}$ such that $f(c) \geq f(x)$ for all $x \in \mathbb{R}$.

HINT: Show that $f(x) \leq 0$ if $|x| > 2$.

85. A function $f\colon \mathbb{R} \to \mathbb{R}$ is a **periodic function** if there exists $T > 0$ such that for all $x \in \mathbb{R}$, $f(x + T) = f(x)$. If $f\colon \mathbb{R} \to \mathbb{R}$ is continuous and periodic, then prove that f is bounded, that is, the set $S = \{f(x) \mid x \in \mathbb{R}\}$ is bounded.

86. Let $f\colon [a, b] \to \mathbb{R}$ be continuous on $[a, b]$, and define f_* as follows:

$$f_*(x) = \begin{cases} f(a) & \text{if } x = a, \\ \max\{f(y) \mid y \in [a, x]\} & \text{if } x \in (a, b]. \end{cases}$$

a) Show that f_* is a well-defined function.

b) Prove that f_* is continuous on $[a, b]$.

c) If $f\colon [-1, 1] \to \mathbb{R}$ is given by $f(x) = x - x^2$, then find f_*.

87. **Sample Exam Question. 2007 Q5.**

(a) What does it mean to say that a sequence $(a_n)_{n\geq 1}$ is convergent? Use this definition to show that if $(a_n)_{n\geq 1}$ is convergent, then $(a_{n+1})_{n\geq 1}$ is also a convergent sequence and $\lim_{n\to\infty} a_n = \lim_{n\to\infty} a_{n+1}$.

Let b be a real number with $2 < b < 3$. We define a sequence $(b_n)_{n\geq 1}$ by

$$b_1 = b \text{ and } b_{n+1} = b_n^2 - 4b_n + 6 \text{ for every } n \in \mathbb{N}.$$

(b) Show that $2 < b_n < 3$ for every $n \in \mathbb{N}$.

(c) Prove that $(b_n)_{n\geq 1}$ is a monotone sequence.

(d) Explain why $\lim_{n\to\infty} b_n$ exists and find its value.

(e) Let $S = \{b_n \mid n \in \mathbb{N}\}$. Find $\sup S, \inf S, \max S, \min S$. Justify your answers.

88. **Sample Exam Question. 2009 Q5.**

(a) Let I be an interval in $\mathbb{R}$ and let $c \in I$.

What does it mean to say that a function $f : I \to \mathbb{R}$ is continuous at c ? What does it mean to say that a function $f : I \to \mathbb{R}$ is continuous on I?

(b) Suppose that $f : \mathbb{R} \to \mathbb{R}$, $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ are functions such that

(1) f and g are continuous on $\mathbb{R}$,

(2) $f(0) = g(0)$, and

(3) for every $x \in \mathbb{R}$, $f(x) \leq h(x) \leq g(x)$.

Show that h is continuous at 0.

(c) State the Intermediate Value Theorem.

Suppose that the function $f : [0, 2] \to \mathbb{R}$ is continuous on the interval $[0, 2]$ and $f(0) = f(2) \geq f(1)$. Prove that there exist numbers $a, b \in [0, 2]$ such that $|a - b| = 1$ and $f(a) = f(b)$.

Hint: Consider the function $g : [0, 1] \to \mathbb{R}$ given by $g(x) = f(x + 1) - f(x)$.

Chapter 2

Algebra

We have already seen many different "algebraic structures", that is, different sets with algebraic operations defined on them. For example, we know of the natural numbers $\mathbb{N}$ on which addition $(+)$ and multiplication $(\cdot)$ are defined, the integers $\mathbb{Z}$ on which subtraction $(-)$ is additionally defined, and the rational numbers $\mathbb{Q}$ on which division $(|)$ is defined in addition to the previous operations. What distinguishes the real numbers $\mathbb{R}$ from $\mathbb{Q}$ is not so much an algebraic property as an order property (the l.u.b. property). On the other hand, the difference between $\mathbb{R}$ and the complex numbers $\mathbb{C}$ is that certain equations with no real solutions, such as $x^2 + 1 = 0$, have solutions in $\mathbb{C}$. Thus $\mathbb{C}$ has an algebraic property not shared by $\mathbb{R}$.

When we consider the operations on the above sets, we see for example that addition and multiplication are commutative:

$$x + y = y + x \text{ and } xy = yx$$

and associative

$$(x + y) + z = x + (y + z) \text{ and } (xy)z = x(yz).$$

Subtraction and division do not share these properties. On the other hand, they are not really operations in their own right, but are derived from addition and multiplication along with the notion of inverses. Also, addition and multiplication have identity elements (0 and 1, respectively).

We have also seen other algebraic systems, such as matrices, on which addition and multiplication have been defined. We know that matrix addition is commutative and associative ($A + B = B + A$ and $AB = BA$), and matrix multiplication is associative:

$$(AB)C = A(BC),$$

but matrix multiplication is definitely not commutative:

$$AB \neq BA \text{ in general.}$$

Again, there are identity elements (the zero matrix O for addition and the identity matrix I for multiplication).

There are many more different algebraic structures in mathematics. Abstract algebra is the study of general properties of these algebraic structures. In this part of the course, we study two important abstract algebraic structures: groups and vector spaces. We begin with a discussion about groups, one of the *simplest* algebraic structures.

2.1 Groups

In this section, we study one of the most basic algebraic objects, namely a group. A group is a set on which an operation or *law of composition* is defined, such that certain properties hold. The precise definition is given in the next subsection.

2.1.1 Definition of a Group

By a *law of composition* on a set S, we mean a rule for combining pairs a, b of S to get another element, say c, of S. We denote the set of all ordered pairs of elements from a set S by $S \times S$, that is, $S \times S = \{(a, b) \mid a, b \in S\}$.

Definition. *A* ***law of composition on a set*** S *is a function* $f\colon S \times S \to S$.

The functional notation $c = f(a, b)$ is not very convenient for what is going to follow, and so instead, the element obtained by applying the law of composition to a pair (a, b) is usually denoted using a notation resembling that used for addition or multiplication:

$$c = a * b, \quad \text{or} \quad ab, \quad \text{or} \quad a \circ b, \quad \text{or} \quad a + b \quad \text{and so on,}$$

with a fixed choice being made for the particular law in question. In this notation we then have that $a * b$ is a law of composition on S iff $a * b \in S$ for any $a, b \in S$. For later reference, we state the property of $*$ being a law of composition (which any particular $*$ may either have or not have), as an axiom.

G0 (Law of Composition) *For all* $a, b \in G$, $a * b \in G$.

Examples.

1. The addition of integers is a law of composition on $\mathbb{Z}$. Indeed, the sum of two integers is yet another integer, and addition is the function from the set $\mathbb{Z} \times \mathbb{Z}$ to the set $\mathbb{Z}$ that assigns $a + b$ to the pair (a, b), denoted by $(a, b) \mapsto a + b$.

2. The multiplication of real numbers is a law of composition on $\mathbb{R}$.

3. If a, b are rational numbers, then let $a * b = a + b - ab$. The function from $\mathbb{Q} \times \mathbb{Q}$ to $\mathbb{Q}$ given by $(a, b) \mapsto a * b$ is a law of composition on $\mathbb{Q}$.

4. If a, b are real numbers, then define $a * b = \sqrt{a^2 + b^2}$. Then $(a, b) \mapsto a * b$ is not a law of composition on $\mathbb{Q}$, since $1 \in \mathbb{Q}$, but $1 * 1 = \sqrt{2} \notin \mathbb{Q}$. However, $(a, b) \mapsto a * b$ is a law of composition on $\mathbb{R}$.

5. Let $n \in \mathbb{N}$, and let S denote the set of all matrices of size $n \times n$ with real entries. Then matrix multiplication is a law of composition on S.

6. Let $n \in \mathbb{N}$, and let $GL(n, \mathbb{R})$ denote the set of all invertible matrices of size $n \times n$ with real entries. Then matrix multiplication is a law of composition on $GL(n, \mathbb{R})$. Indeed, if $A, B \in GL(n, \mathbb{R})$, then the matrix AB is again a matrix of size $n \times n$ with real entries, and moreover, since A and B are invertible, it follows that AB is also invertible.

7. Let a, b be real numbers such that $a < b$. Let $C[a, b]$ denote the set of all continuous functions on the interval $[a, b]$. Let addition of functions be defined as follows: if f, g belong to $C[a, b]$, then

$$(f + g)(x) = f(x) + g(x), \quad x \in [a, b].$$

   Then addition of functions is a law of composition on $C[a, b]$, since the sum of continuous functions is again continuous; see Theorem 1.3.2.

8. If $a, b \in \mathbb{N}$, then let

$$a * b = \frac{a}{b}.$$

   $*$ is not a law of composition on $\mathbb{N}$, since $1 * 2 = \frac{1}{2} \notin \mathbb{N}$. The function $(a, b) \mapsto \frac{a}{b}$ is also not a law of composition on $\mathbb{Q}$, since $(1, 0)$ is not mapped to any rational number by the function.

On a set there may be many different laws of compositions that can be defined. Some laws of compositions are nicer than others, that is, they possess some desirable properties. A group is a set G together with a law of composition on G that has three such desirable properties, and we give the definition below.

Definition. *A **group** is a set G together with a law of composition $(a, b) \mapsto a*b \colon G \times G \to G$, which has the following properties:*

G1 (Associativity axiom) *For all $a, b, c \in G$, $(a * b) * c = a * (b * c)$.*

G2 (Identity axiom) *There exists an element $e \in G$ such that for all $a \in G$, $a * e = a = e * a$. Such an element e is called an **identity element** of the group G.*

G3 (Inverse axiom) *For every $a \in G$, there exists an element $a^{-1} \in G$ such that $a * a^{-1} = e = a^{-1} * a$. Such an element a^{-1} is called an **inverse** of the element a in the group G.*

*G1, G2, G3 are called the **group axioms**. Sometimes we use the notation $(G, *)$ for the group.*

Remark. Note that hidden in the definition of a group is the axiom G0 *stating that $*$ is actually a law of composition on G.* Hence when checking that a certain set G is a group with respect to a certain operation $*$ that combines pairs of elements from G, **we have to check first of all** that the axiom G0 is satisfied as well: for every $a, b \in G$, $a * b$ belongs to G.

Examples.

1. $\mathbb{Z}$ with addition is a group, as may be seen by checking each of the axioms G0, G1, G2 and G3:

   G0. For all $a, b \in \mathbb{Z}$, $a + b \in \mathbb{Z}$.

   G1. For all $a, b, c \in \mathbb{Z}$, $(a + b) + c = a + (b + c)$.

   G2. 0 serves as an identity element: for all $a \in \mathbb{Z}$, $a + 0 = a = 0 + a$.

   G3. If $a \in \mathbb{Z}$, then $-a \in \mathbb{Z}$ and $a + (-a) = 0 = -a + a$.

2. $\mathbb{R}$ with multiplication is not a group. Indeed, although the multiplication of real numbers satisfies G0, G1 and G2, the group axiom G3 does not hold:

   G0. For all $a, b \in \mathbb{R}$, $ab \in \mathbb{R}$.

   G1. For all $a, b, c \in \mathbb{R}$, $(ab)c = a(bc)$.

   G2. If e is an identity element, then we must have $ae = a = ea$ for all $a \in \mathbb{R}$, and in particular, with $a = 1$, we should have $1e = 1$, and so, $e = 1$. And so if e is an identity element, then it must necessarily be equal to 1.

   We then check that 1 serves as an identity element: for all $a \in \mathbb{R}$, $a1 = a = 1a$.

   G3. Does not hold, since $0 \in \mathbb{R}$, but for any potential inverse $b \in \mathbb{R}$ we find that $0b = b0 = 0 \neq 1 = e$, a contradiction. So there is no inverse of the element $0 \in \mathbb{R}$.

   However, the set $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$ of non-zero real numbers with multiplication forms a group:

   G0. For all $a, b \in \mathbb{R}$, **if in addition** $a, b \neq 0$, then $ab \in \mathbb{R}$ **as well as** $ab \neq 0$.

   G1. For all $a, b, c \in \mathbb{R}^*$, $(ab)c = a(bc)$, since it already holds for all $a, b, c \in \mathbb{R}$.

   G2. $1 \in \mathbb{R}^*$ is the identity element, since for all $a \in \mathbb{R}^*$, $a1 = a = 1a$. (Note that 1 is still in the smaller set $\mathbb{R}^*$.)

   G3. Let $a \in \mathbb{R}^*$. To find out if a has an inverse x, we have to solve the equations $ax = xa = 1$. It is clear that $x = \frac{1}{a}$ (which exists, since $a \neq 0$).

   Thus we have for any $a \in \mathbb{R}^*$ that $a^{-1} = \frac{1}{a}$ serves as an inverse, since $a \cdot \frac{1}{a} = \frac{1}{a} \cdot a = 1$.

   Similarly, the set $(0, \infty)$ of positive real numbers with multiplication is also a group,

3. For $n \in \mathbb{N}$, let $GL(n, \mathbb{R})$ denote the set of all invertible matrices of size $n \times n$ with real entries. Then $GL(n, \mathbb{R})$ is group with matrix multiplication:

   G0. For all $A, B \in GL(n, \mathbb{R})$, since A and B are invertible $n \times n$ matrices, $\det A \neq 0$ and $\det B \neq 0$. Therefore, $\det(AB) = (\det A)(\det B) \neq 0$, which gives that AB is also an invertible $n \times n$ matrix. That is, $AB \in GL(n, \mathbb{R})$.

   G1. For all $A, B, C \in GL(n, \mathbb{R})$, $(AB)C = A(BC)$, since matrix multiplication is associative.

G2. The identity matrix

$$I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

serves as an identity element. I_n is an invertible matrix of size $n \times n$ with real entries, and so it belongs to $GL(n, \mathbb{R})$, and moreover, for all $A \in GL(n, \mathbb{R})$, $AI_n = A = I_n A$.

G3. If $A \in GL(n, \mathbb{R})$, then A is an invertible matrix, and so there exists a matrix A^{-1} such that $AA^{-1} = I_n = A^{-1}A$. The matrix A^{-1} is thus in $GL(n, \mathbb{R})$, and serves as an inverse of A.

This group is called the **general linear group**.

4. Let a, b be real numbers such that $a < b$. Then $C[a, b]$ is a group with addition of functions:

G0. For all $f, g \in C[a, b]$, $f + g \in C[a, b]$ since the sum of continuous functions is continuous.

G1. For all $f, g, h \in C[a, b]$, and any $x \in [a, b]$ we have

$$\begin{aligned} (f + (g + h))(x) &= f(x) + (g + h)(x) \\ &= f(x) + (g(x) + h(x)) \\ &= (f(x) + g(x)) + h(x) \\ &\qquad \text{(since addition is associative in } \mathbb{R}\text{!)} \\ &= (f + g)(x) + h(x) \\ &= ((f + g) + h)(x). \end{aligned}$$

Hence $f + (g + h) = (f + g) + h$.

G2. The constant function $\mathbf{0}$ defined by $\mathbf{0}(x) = 0$ for all $x \in [a, b]$, serves as an identity element. $\mathbf{0}$ is a continuous function on $[a, b]$ and so $\mathbf{0} \in C[a, b]$. Moreover, for all $f \in C[a, b]$, we have for all $x \in [a, b]$:

$$\begin{aligned} (f + \mathbf{0})(x) &= f(x) + \mathbf{0}(x) \\ &= f(x) + 0 \\ &= f(x) \text{ (0 is an identity element for addition in } \mathbb{R}) \\ &= 0 + f(x) \\ &= \mathbf{0}(x) + f(x) \\ &= (\mathbf{0} + f)(x). \end{aligned}$$

Hence $f + \mathbf{0} = f = \mathbf{0} + f$.

G3. If $f \in C[a,b]$, then define $-f$ by $(-f)(x) = -f(x)$, for $x \in [a,b]$. Given $f \in C[a,b]$, we have for all $x \in [a,b]$:

$$\begin{aligned}(f+(-f))(x) &= f(x)+(-f)(x)\\ &= f(x)+(-f(x))\\ &= 0 = \mathbf{0}(x) = 0\\ &= -f(x)+f(x)\\ &= (-f)(x)+f(x)\\ &= (-f+f)(x).\end{aligned}$$

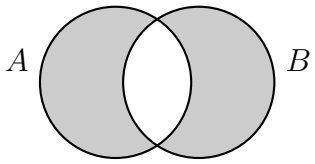
Hence $f+(-f) = \mathbf{0} = -f+f$.

5. For any set S, its **power set** $\mathcal{P}(S)$ is defined to be the set consisting of all subsets of S:

$$\mathcal{P}(S) = \{A \mid A \subseteq S\}.$$

Define the following law of composition on $\mathcal{P}(A)$, called the **symmetric difference operation**:

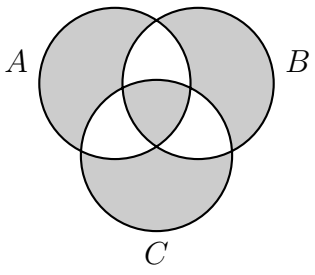
$$A \triangle B = (A \setminus B) \cup (B \setminus A), \quad A, B \in \mathcal{P}(S).$$



Then $\mathcal{P}(S)$ with $\triangle$ is a group:

G0. For any $A, B \in \mathcal{P}(S)$ we have $A, B \subseteq S$, and therefore, $A \setminus B, B \setminus A \subseteq S$ and $A \triangle B = (A \setminus B) \cup (B \setminus A) \subseteq S$. Thus $A \triangle B \in \mathcal{P}(S)$.

G1. For any $A, B, C \in \mathcal{P}(S)$ we have to check that $A \triangle (B \triangle C) = (A \triangle B) \triangle C$. It is easily checked that both expressions correspond to the grey part in the following Venn diagram:



G2. Since $A \triangle \emptyset = A = \emptyset \triangle A$ for all $A \in \mathcal{P}(S)$, the empty set $\emptyset$ serves as an identity element.

G3. Since $A \triangle A = \emptyset$, each element is its own inverse: $A^{-1} = A$ for all $A \in \mathcal{P}(S)$.

6. Given $n \in \mathbb{N}$, let S_n be the set of all bijections from the set $\{1, 2, \ldots, n\}$ to itself, and let $\circ$ denote the composition of functions, so $\alpha \circ \beta$ means the function with $(\alpha \circ \beta)(x) = \alpha(\beta(x))$. Then $(S_n, \circ)$ is a group, called the **symmetric group on** n **symbols**.

   G0. You have seen that the composition of two bijections is a bijection, so indeed if α and β are in S_n then $\alpha \circ \beta$ is also in S_n.

   G1. If α, β, γ are three elements of S_n, and $x \in \{1, \ldots, n\}$, then

   $$\Big((\alpha \circ \beta) \circ \gamma\Big)(x) = (\alpha \circ \beta)(\gamma(x)) = \alpha(\beta(\gamma(x))) = \alpha((\beta \circ \gamma)(x)) = \Big(\alpha \circ (\beta \circ \gamma)\Big)(x),$$

   and so $(\alpha \circ \beta) \circ \gamma = \alpha \circ (\beta \circ \gamma)$ for all $\alpha, \beta, \gamma \in S_n$.

   G2. The identity function ι, with $\iota(x) = x$ for each x, is a bijection, and evidently $\iota \circ \alpha = \alpha = \alpha \circ \iota$ for each $\alpha \in S_n$.

   G3. If $\alpha \in S_n$ is a bijection, then the inverse function α^{-1} is also a bijection, and serves as an inverse to α in $(S_n, \circ)$.

Definitions. *A group $(G, *)$ is said to be a* ***finite group*** *if the set G has finite cardinality. The* ***order of a finite group*** *$(G, *)$ is the cardinality of G. A group is said to be an* ***infinite group*** *if it is not finite.*

Examples.

1. The set $\{-1, 1\}$ with multiplication is a finite group of order 2.

2. The set $\mathbb{Z}$ with addition is an infinite group.

3. The set $\mathcal{P}(S)$ with the symmetric difference operation $\triangle$ is a finite group if S is a finite set, and an infinite group if S is an infinite set.

A finite group can be completely described by writing its *group table*.

Definition. *The* ***group table*** *of a finite group is a table that displays the law of composition as follows: the elements of the group are listed in the first row and the first column. Conventionally, the two lists have the group elements in the same order, with the identity element first. Given $a, b \in G$, the element $a * b$ is entered in the row corresponding to a and the column corresponding to b, as shown below.*

$*$	$\cdots$	b	$\cdots$
$\vdots$			
a		$a * b$	
$\vdots$			

We clarify this further by considering a few examples.

Examples.

1. The finite group $\{-1, 1\}$ with multiplication can be described by the group table given below.

$\cdot$	1	-1
1	1	-1
-1	-1	1

   The table completely describes the law of composition: $1 \cdot 1 = 1$, $1 \cdot (-1) = -1$, $-1 \cdot 1 = -1$ and $-1 \cdot (-1) = 1$.

2. The group S_3 of bijections from the set $\{1, 2, 3\}$ to itself has $3! = 6$ elements, listed as follows:

x	1	2	3
$\iota(x)$	1	2	3

x	1	2	3
$\alpha(x)$	1	3	2

x	1	2	3
$\beta(x)$	3	2	1

x	1	2	3
$\gamma(x)$	2	1	3

x	1	2	3
$\delta(x)$	2	3	1

x	1	2	3
$\varepsilon(x)$	3	1	2

   We can now form the group table for $(S_3, \circ)$. For example, since

$$\begin{aligned}\alpha \circ \beta(1) &= \alpha(\beta(1)) = \alpha(3) = 2,\\ \alpha \circ \beta(2) &= \alpha(\beta(2)) = \alpha(2) = 3,\\ \alpha \circ \beta(3) &= \alpha(\beta(3)) = \alpha(1) = 1,\end{aligned}$$

   we obtain that $\alpha \circ \beta = \delta$. In a similar way, the whole group table may be calculated:

$\circ$	ι	α	β	γ	δ	ε
ι	ι	α	β	γ	δ	ε
α	α	ι	δ	ε	β	γ
β	β	ε	ι	δ	γ	α
γ	γ	δ	ε	ι	α	β
δ	δ	γ	α	β	ε	ι
ε	ε	β	γ	α	ι	δ

3. The finite group $\mathcal{P}(\{1, 2\})$ with the symmetric difference operation $\triangle$ has four elements:

$$\mathcal{P}(\{1, 2\}) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$$

   and has the following group table:

$\triangle$	$\emptyset$	$\{1\}$	$\{2\}$	$\{1, 2\}$
$\emptyset$	$\emptyset$	$\{1\}$	$\{2\}$	$\{1, 2\}$
$\{1\}$	$\{1\}$	$\emptyset$	$\{1, 2\}$	$\{2\}$
$\{2\}$	$\{2\}$	$\{1, 2\}$	$\emptyset$	$\{1\}$
$\{1, 2\}$	$\{1, 2\}$	$\{2\}$	$\{1\}$	$\emptyset$

   Again the table completely describes the law of composition. For example,

$$\{1, 2\} \triangle \{2\} = (\{1, 2\} \setminus \{2\}) \cup (\{2\} \setminus \{1, 2\}) = \{1\} \cup \emptyset = \{1\}.$$

2.1.2 Proving theorems, laws of exponents and solving equations in groups

We now prove a few elementary theorems concerning groups.

Theorem 2.1.1. *There is a unique identity element in a group.*

Proof. Let e and e' be identity elements in $(G, *)$. Since $e \in G$ and e' is an identity, we obtain

$$e = e * e'.$$

Moreover, since $e' \in G$ and e is an identity, we also have

$$e * e' = e'.$$

Consequently, $e = e'$. □

Theorem 2.1.2. *Let $(G, *)$ be a group and let $a \in G$. Then a has a unique inverse.*

Proof. Let the group have the identity e. If a_1 and a_2 are inverses of a, then we have

$$\begin{aligned} a_1 &= a_1 * e \text{ (since } a_1 \in G \text{ and } e \text{ is the identity)} \\ &= a_1 * (a * a_2) \text{ (since } a_2 \text{ is an inverse of } a) \\ &= (a_1 * a) * a_2 \text{ (associativity)} \\ &= e * a_2 \text{ (since } a_1 \text{ is an inverse of } a) \\ &= a_2 \text{ (since } a_2 \in G \text{ and } e \text{ is the identity).} \end{aligned}$$

□

Examples.

1. *Let a and b be elements of a group $(G, *)$. Find all $x \in G$ such that:*

$$a * x = b.$$

Solution. Left-multiply both sides of the equation by a^{-1} (which exists by G3 and satisfies $a * a^{-1} = e = a^{-1} * a$ where e is the identity element of G):

$$\begin{aligned} & & a^{-1} * (a * x) &= a^{-1} * b. \\ \text{Then by G1,} & & (a^{-1} * a) * x &= a^{-1} * b, \\ \text{by G3,} & & e * x &= a^{-1} * b, \\ \text{and finally by G2,} & & x &= a^{-1} * b. \end{aligned}$$

□

We must now check that $x = a^{-1} * b$ *is* a solution to the equation $a * x = b$, i.e., that $a * (a^{-1} * b) = b$. This follows via a very similar calculation.

2. *Use G1 to show that* $(a \circ b) \circ (c \circ d) = (a \circ (b \circ c)) \circ d$ *for any elements* a, b, c, d *of a group* $(G, \circ)$.

Solution. Note in each step how the associativity axiom (G1) is applied:

$$\begin{aligned}(a \circ b) \circ (c \circ d) &= a \circ (b \circ (c \circ d)) \\ &= a \circ ((b \circ c) \circ d) \\ &= (a \circ (b \circ c)) \circ d.\end{aligned}$$

□

In the previous example we saw that by the associativity axiom, two ways of using parentheses on $a \circ b \circ c \circ d$ resulted in equal expressions. It can be shown that in general, for any elements $a_1, a_2, \ldots, a_n$ of a group $(G, *)$, any two ways of parenthesising $a_1 * a_2 * \cdots * a_n$ will result in equal elements in the group. We therefore in general do not show the parentheses. Instead we write $a \circ b \circ c \circ d$ or $a_1 * a_2 * \cdots * a_n$, or even just $abcd$ or $a_1 a_2 \ldots a_n$, when it is clear which law of composition is meant.

Definitions. *Let* G *be a group and let* $a \in G$. *We define*

$$a^0 = e \text{ and } a^n = a^{n-1} * a \text{ for } n \in \mathbb{N}.$$

Moreover, if $n \in \mathbb{N}$, *define*

$$a^{-n} = (a^n)^{-1}.$$

It can be shown that the usual **laws of exponents** hold: for all $m, n \in \mathbb{Z}$,

$$a^m * a^n = a^{m+n} \text{ and } (a^m)^n = a^{mn}.$$

Remark. It is **not necessarily true** that $(a * b)^n = a^n * b^n$.

Example. *Let* b *be an element of a group* G *with identity element* e. *Find all solutions* x *to the following simultaneous equations:*

$$x^2 = b \text{ and } x^5 = e.$$

Solution. Square the first equation: $b^2 = (x^2)^2 = x^4$. Left-multiply by x: $xb^2 = xx^4 = x^5 = e$ by the second equation. Right multiply by b^{-2}: $xb^2b^{-2} = eb^{-2}$. Simplifying, $xb^{2-2} = xb^0 = xe = x$ and $eb^{-2} = b^{-2}$. Therefore, $x = b^{-2}$.

So far, we have shown that, *if there is a solution* x *to the equations*, then that solution is $x = b^{-2}$. We must now ask whether this is indeed a solution: substituting $x = b^{-2}$ into the original equations yields $b^{-4} = b$ and $b^{-10} = e$: the first of these is equivalent to $b^5 = e$, and if this is the case then also $b^{-10} = e$. The conclusion is that, if $b^5 = e$, then $x = b^{-2} = b^3$ is the unique solution to the equations, whereas if $b^5 \neq e$, then the equations have no solutions. □

2.1.3 Abelian groups

The groups in Examples 1, 2, 4 and 5 on page 58 above also satisfy the following axiom:

G4 (Commutativity axiom) *For all $a, b \in G$, $a * b = b * a$.*

On the other hand, the group in Example 3 does not satisfy G4 for any $n \geq 2$. For example, when $n = 2$:

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \neq \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

This gives rise to the following natural definition.

Definition. *A group $(G, *)$ is said to be an **abelian group**[1] if its law of composition is **commutative**, that is, for all $a, b \in G$, $a * b = b * a$.*

Examples.

1. The group $(\mathbb{Z}, +)$ of integers with addition is an abelian group.
2. The group $(\mathbb{R}^*, \times)$ of non-zero real numbers with multiplication is an abelian group.
3. Let $m, n \in \mathbb{N}$. The set of matrices $\mathbb{R}^{m \times n}$ of size $m \times n$ with entries in $\mathbb{R}$ with matrix addition is an abelian group.
4. Let $n \in \mathbb{N}$. The set $GL(n, \mathbb{R})$ with matrix multiplication is a group, but it is not an abelian group if $n \geq 2$.
5. For any S, the group $(\mathcal{P}(S), \triangle)$ is an abelian group.
6. The symmetric group S_3 is not an abelian group. For instance, referring to the group table given earlier, we see that $\alpha \circ \beta = \delta$, but $\beta \circ \alpha = \varepsilon \neq \delta$.

We normally use **additive notation** when referring with abelian groups. We then

use a symbol such as '$+$', '$\oplus$', '$\boxplus$', etc., to denote the law of composition,

use a symbol such as 0, $\boldsymbol{o}$, etc., to denote the identity element,

and denote the inverse of an element x by $-x$.

Furthermore, instead of writing a^m for $a + a + \cdots + a$, we use the more natural-looking notation $m \cdot a$. As before, we extend it also to $m = 0$ by writing $0 \cdot a = 0$ (where the 0 on the right-hand side is the identity element of the group) and to $m < 0$, where we still write $m \cdot a$ instead of a^m. In particular, in this notation we then have $-x = (-1) \cdot x$.

This notation is nothing more than a new notation for a^m in the case of abelian groups where additive notation is used. The laws of exponents look as follows in additive notation:

$$m \cdot a + n \cdot a = (m + n) \cdot a \quad \text{and} \quad m \cdot (n \cdot a) = (mn) \cdot a.$$

[1] after the Norwegian mathematician Niels Henrik Abel (1802–1829).

2.1.4 Subgroups

Definition. *A subset H of a group G, with operation $*$, is called a* ***subgroup*** *of G if it has the following properties:*

H1 (Closure) *If $a, b \in H$, then $a * b \in H$.*

H2 (Identity) *$e \in H$.*

H3 (Inverses) *If $a \in H$, then $a^{-1} \in H$.*

The three properties of a subgroup H ensure that $(H, *)$ is itself a group, as we now explain.

H1. This condition tells us that the law of composition $*$ on the group G can be used to define a law of composition on H, namely, the function from $H \times H$ to H given by $(a, b) \mapsto a * b$, which is called the *induced law of composition*. Since $*$ is associative, it follows that the induced law of composition is associative as well: for all $a, b, c \in H$, $a * (b * c) = (a * b) * c$.

H2. This shows that H, with the induced law of composition, has an identity element. Indeed the identity element from G (which also belongs to H) serves as the identity element also in H: for every $a \in H$, $a * e = a = e * a$.

H3. Finally this shows that every element in H possesses an inverse element in H. Of course, since G is a group, we already knew that a possesses an inverse element $a^{-1} \in G$. But now H3 says that this inverse element is in H. Thus for all $a \in H$, $\exists a^{-1} \in H$ such that $a * a^{-1} = e = a^{-1} * a$.

Thus the conditions H1, H2, H3 imply that the subset H, with the induced law of composition, is a group. Thus, a subgroup is itself a group which sits in a larger group.

Examples.

1. The subset of even integers $\{2m \mid m \in \mathbb{Z}\}$ is a subgroup of the group of integers $\mathbb{Z}$ with addition. Indeed, the sum of even numbers is even (and so H1 holds), 0 is even (and so H2 holds), and finally, given the even number $2m$, $-2m = 2(-m)$ is even as well (and so H3 holds).

2. The group of integers with addition $(\mathbb{Z}, +)$ is a subgroup of the group of rational numbers with addition $(\mathbb{Q}, +)$, which in turn is a subgroup of the group of real numbers with addition $(\mathbb{R}, +)$.

3. If G is a group with identity e, then $\{e\}$ and G are both subgroups of G.

4. The subset of 2×2 **symmetric matrices** with real entries, namely

$$\left\{ \left[\begin{array}{cc} a & b \\ b & d \end{array} \right] \,\middle|\, a, b, d \in \mathbb{R} \right\}$$

is a subgroup of the set of all 2×2 matrices with real entries together with matrix addition. Indeed, given any two symmetric matrices

$$\left[\begin{array}{cc} a & b \\ b & d \end{array} \right] \text{ and } \left[\begin{array}{cc} a' & b' \\ b' & d' \end{array} \right],$$

their sum

$$\left[\begin{array}{cc} a & b \\ b & d \end{array} \right] + \left[\begin{array}{cc} a' & b' \\ b' & d' \end{array} \right] = \left[\begin{array}{cc} a+a' & b+b' \\ b+b' & d+d' \end{array} \right]$$

is also symmetric, and so H1 holds. Clearly the identity element

$$\left[\begin{array}{cc} 0 & 0 \\ 0 & 0 \end{array} \right]$$

is symmetric, and so H2 also holds. Finally, the inverse (with respect to matrix addition) of any symmetric matrix

$$\left[\begin{array}{cc} a & b \\ b & d \end{array} \right],$$

is the element

$$\left[\begin{array}{cc} -a & -b \\ -b & -d \end{array} \right]$$

which is also symmetric, and so H3 holds.

5. The subset of *upper triangular* invertible matrices,

$$UT(2, \mathbb{R}) = \left\{ \left[\begin{array}{cc} a & b \\ 0 & d \end{array} \right] \,\middle|\, a, b, d \in \mathbb{R} \text{ and } ad \neq 0 \right\}$$

is a subgroup of the group $GL(2, \mathbb{R})$ with matrix multiplication. Indeed, if

$$\left[\begin{array}{cc} a & b \\ 0 & d \end{array} \right], \left[\begin{array}{cc} a' & b' \\ 0 & d' \end{array} \right] \in UT(2, \mathbb{R}),$$

then

$$\left[\begin{array}{cc} a & b \\ 0 & d \end{array} \right] \left[\begin{array}{cc} a' & b' \\ 0 & d' \end{array} \right] = \left[\begin{array}{cc} aa' & ab'+bd' \\ 0 & dd' \end{array} \right] \in UT(2, \mathbb{R}),$$

and so H1 holds. Clearly

$$\left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right] \in UT(2, \mathbb{R}),$$

and so H2 also holds. Finally,

$$\text{if } \left[\begin{array}{cc} a & b \\ 0 & d \end{array} \right] \in UT(2, \mathbb{R}), \text{ then } \left[\begin{array}{cc} a & b \\ 0 & d \end{array} \right]^{-1} = \left[\begin{array}{cc} \frac{1}{a} & -\frac{b}{ad} \\ 0 & \frac{1}{d} \end{array} \right] \in UT(2, \mathbb{R}).$$

2.1.5 The order of an element of a group

Definitions. *Let G be a group and suppose that $a \in G$.*

1. *If there exists an $m \in \mathbb{N}$ such that $a^m = e$, then a is said to have* ***finite order****.*

2. *If a has finite order, then the* ***order of*** *a, denoted by* $\mathrm{ord}(a)$*, is*

$$\mathrm{ord}(a) = \min\{m \in \mathbb{N} \mid a^m = e\}.$$

3. *If a does not have finite order, then a is said to have* ***infinite order****.*

Examples.

1. The element -1 has order 2 in the group of non-zero real numbers with multiplication.

2. The element 2 has infinite order in the group of integers with addition.

3. The element $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ is an element of order 3 in the group $GL(3, \mathbb{R})$.

We now prove that given any element from a group, the set of its powers forms a subgroup of the group.

Theorem 2.1.3. *Suppose that $(G, *)$ is a group and $a \in G$. Let $\langle a \rangle = \{a^n \mid n \in \mathbb{Z}\}$. Then:*

1. *$\langle a \rangle$ is a subgroup of G. Moreover, $\langle a \rangle$ is the* smallest *subgroup of G containing a.*

2. *If a is an element with finite order m, then $\langle a \rangle = \{e, a, a^2, a^3, \ldots, a^{m-1}\}$.*

Proof.

1. We prove that H1, H2, H3 hold.

   H1. Given $m, n \in \mathbb{Z}$, clearly $m + n \in \mathbb{Z}$ and so $a^m * a^n = a^{m+n} \in \langle a \rangle$.

   H2. $e = a^0 \in \langle a \rangle$.

   H3. For any $m \in \mathbb{Z}$, $-m \in \mathbb{Z}$ and so $(a^m)^{-1} = a^{-1 \cdot m} = a^{-m} \in \langle a \rangle$.

   So $\langle a \rangle$ is a subgroup of G.

   Conversely, if H is any subgroup of G containing the element a, then:

   (a) $a^0 = e \in H$ by H2.

   (b) Since $a \in H$, so is $a * a = a^2$ by H1, and similarly $a^3, a^4, \ldots$ are all in H.

   (c) By H3, all the inverses $a^{-n} = (a^n)^{-1}$ are also in H.

   So H includes all the elements a^n, for $n \in \mathbb{Z}$, in other words $\langle a \rangle \subseteq H$.

2. Clearly $\{e, a, a^2, a^3, \ldots, a^{m-1}\} \subset \langle a \rangle$. Conversely, if $n \in \mathbb{Z}$, then there exist integers q and r, such that $0 \leq r \leq m-1$, and $n = q \cdot m + r$. So

$$\begin{aligned} a^n &= a^{q \cdot m + r} = a^{q \cdot m} * a^r = (a^m)^q * a^r = (e)^q * a^r \\ &= e * a^r = a^r \in \{e, a, a^2, a^3, \ldots, a^{m-1}\}. \end{aligned}$$ □

Example. The element [1] in the group $\mathbb{Z}_5$ with addition modulo 5 has finite order, since

$$[1] \oplus [1] \oplus [1] \oplus [1] \oplus [1] = [0],$$

and the subgroup $\langle [1] \rangle = \{[0], [1], [2], [3], [4]\}$ is in fact the whole group. □

The above example motivates the following definition.

Definition. *A group G is said to be a* ***cyclic group*** *if there exists an element $a \in G$ such that $G = \langle a \rangle$. Such an element a is then called a* generator *of the group G.*

2.1.6 Homomorphisms and isomorphisms

The study of algebra is not just about *structures* with algebraic operations on them, but also about functions between such structures, that in an obvious sense "respect" the algebraic operations. The following definition illustrates this idea in the case of groups.

Definition. *Let $(G, *)$ and $(G', *')$ be groups. A* ***homomorphism*** *$\varphi \colon G \to G'$ is a function such that*

$$\text{for all } a, b \in G, \quad \varphi(a * b) = \varphi(a) *' \varphi(b).$$

Examples.

1. Let G be $\mathbb{R}$ with addition, and G' the set of positive reals with multiplication. The exponential function from G to G', given by $x \mapsto 2^x$, is a homomorphism. Indeed, we have $2^{x+y} = 2^x 2^y$ for all real x, y.

2. Let G be the group $GL(2, \mathbb{R})$ with matrix multiplication, and G' be the group of non-zero real numbers with multiplication. Then the determinant function $\det \colon G \to G'$ is a homomorphism, since for all 2×2 real matrices A, B we have $\det(AB) = \det(A) \det(B)$.

3. Let G be the group $C[0, 1]$ of continuous functions on the interval $[0, 1]$ with addition, and G' be the group $\mathbb{R}$ with addition. Then the function $f \mapsto f\left(\frac{1}{2}\right)$ is a homomorphism. Indeed, if f, g are continuous functions on the interval $[0, 1]$, then $(f + g)\left(\frac{1}{2}\right) = f\left(\frac{1}{2}\right) + g\left(\frac{1}{2}\right)$, by the definition of $f + g$.

4. If G is a group, then the identity map $\iota \colon G \to G$ defined by $\iota(a) = a$ for all $a \in G$, and the trivial map $\zeta \colon G \to G$ defined by $\zeta(a) = e$ for all $a \in G$ are homomorphisms.

Thus a homomorphism is a function between two groups that respects the law of composition. In the next theorem we show that it preserves the identity element and the inverses of elements as well.

Theorem 2.1.4. *Let G be a group with identity e and G' be a group with identity e'. If $\varphi\colon G \to G'$ is a homomorphism, then:*

1. $\varphi(e) = e'$.

2. *If $a \in G$, then $(\varphi(a))^{-1} = \varphi(a^{-1})$.*

Proof. We have

$$e' *' \varphi(e) = \varphi(e) = \varphi(e * e) = \varphi(e) *' \varphi(e),$$

and so cancelling $\varphi(e)$ on both sides, we obtain $e' = \varphi(e)$. Next,

$$\varphi(a^{-1}) *' \varphi(a) = \varphi(a^{-1} * a) = \varphi(e) = e'$$

and similarly,

$$e' = \varphi(e) = \varphi(a * a^{-1}) = \varphi(a) *' \varphi(a^{-1}).$$

Thus $\varphi(a^{-1}) *' \varphi(a) = e' = \varphi(a) *' \varphi(a^{-1})$, and by the uniqueness of the inverse of $\varphi(a)$ in G', we obtain $(\varphi(a))^{-1} = \varphi(a^{-1})$. □

Every group homomorphism φ determines two important subgroups: its image and its kernel.

Definitions. *Let G, G' be groups and let $\varphi\colon G \to G'$ be a group homomorphism.*

1. *The **kernel** of φ is the set* $\ker(\varphi) = \{a \in G \mid \varphi(a) = e'\}$.

2. *The **image** of φ is the set* $\mathrm{im}(\varphi) = \{a' \in G' \mid \exists a \in G \text{ such that } \varphi(a) = a'\}$.

Using Theorem 2.1.4, we now prove the following result.

Theorem 2.1.5. *Let G, G' be groups and let $\varphi\colon G \to G'$ be a group homomorphism. Then:*

1. $\ker(\varphi)$ *is a subgroup of G.*

2. $\mathrm{im}(\varphi)$ *is a subgroup of G'.*

Proof. If a, b belong to $\ker(\varphi)$, then $\varphi(a * b) = \varphi(a) *' \varphi(b) = e' *' e' = e'$, and so H1 holds. Moreover, as $\varphi(e) = e'$, $e \in \ker(\varphi)$ and so H2 holds too. Finally, H3 holds, since if $a \in \ker(\varphi)$, then $\varphi(a^{-1}) = (\varphi(a))^{-1} = (e')^{-1} = e'$, and so $a^{-1} \in \ker(\varphi)$. Hence $\ker(\varphi)$ is a subgroup of G.

We now also check that $\mathrm{im}(\varphi)$ is a subgroup of G'. If a', b' belong to $\mathrm{im}(\varphi)$, then there exist elements a, b in G such that $\varphi(a) = a'$ and $\varphi(b) = b'$. Consequently, $\varphi(a * b) = \varphi(a) *' \varphi(b) = a' *' b'$, and so there exists an element in G, namely $a * b$, such that $\varphi(a * b) = a' *' b'$, that is, $a' *' b' \in \mathrm{im}(\varphi)$. Thus H1 holds. Since $\varphi(e) = e'$, it follows that $e' \in \mathrm{im}(\varphi)$. Finally, if $a' \in \mathrm{im}(\varphi)$, then there exists an $a \in G$ such that $\varphi(a) = a'$, and so $a'^{-1} = (\varphi(a))^{-1} = \varphi(a^{-1})$. Hence $a'^{-1} \in \mathrm{im}(\varphi)$, and H3 holds. So $\mathrm{im}(\varphi)$ is a subgroup of G'. □

Examples.

1. Let G be $\mathbb{R}$ with addition, and G' be the set of positive reals with multiplication. The exponential function from G to G', given by $x \mapsto 2^x$, is a homomorphism with kernel the trivial subgroup comprising the element 0, and the image is the whole group of positive reals with multiplication, since it can be shown that given any $y > 0$, there exists a unique real number (called the *logarithm of y to the base* 2, denoted by $\log_2 y$) such that $y = 2^{\log_2 y}$.

2. Let G be the group $GL(2, \mathbb{R})$ with matrix multiplication, and G' be the group of non-zero real numbers with multiplication. Then the determinant function $\det\colon G \to G'$ is a homomorphism. Its kernel is the set of all invertible matrices with determinant equal to 1, and we denote this subgroup by $SL(2, \mathbb{R})$, and it is called the **special linear group**:
$$SL(2, \mathbb{R}) = \{A \in GL(2, \mathbb{R}) \mid \det(A) = 1\}.$$
The image of this homomorphism is the whole group of non-zero reals: indeed, given any real number a not equal to zero, we have that
$$A := \begin{bmatrix} 1 & 0 \\ 0 & a \end{bmatrix} \in GL(2, \mathbb{R}),$$
and $\det(A) = 1 \cdot a - 0 \cdot 0 = a$.

3. Let G be the group $C[0, 1]$ of continuous functions on the interval $[0, 1]$ with addition, and G' be the group $\mathbb{R}$ with addition. Then the function $f \mapsto f\left(\frac{1}{2}\right)$ is a homomorphism, and its kernel is the set of all continuous functions on the interval $[0, 1]$ that have a root at $\frac{1}{2}$ (for instance the straight line $f(x) = x - \frac{1}{2}$ belongs to the kernel). The image is the set of all real numbers, since given any $a \in \mathbb{R}$, the constant function $f(x) = a$ for all $x \in [0, 1]$ is continuous, and $f\left(\frac{1}{2}\right) = a$.

In Example 1 above, the homomorphism between the two groups was also bijective. We give a special name to such homomorphisms.

Definition. *Let G, G' be groups. A homomorphism $\varphi\colon G \to G'$ is said to be an **isomorphism** if it is bijective.*

Examples.

1. Let G be $\mathbb{R}$ with addition, and G' be the set of positive reals with multiplication. The exponential function from G to G', given by $x \mapsto 2^x$, is an isomorphism.

2. If G is a group, then the identity function $\iota\colon G \to G$ defined by $\iota(a) = a$ for all $a \in G$ is an isomorphism.

3. Let G be the subgroup of $GL(2, \mathbb{R})$ comprising all matrices of the form $\begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix}$, where $x \in \mathbb{R}$. Let G' be the group $\mathbb{R}$ with addition. Then the function from G to G', given by $\begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix} \mapsto x$, is an isomorphism.

Isomorphisms are important because their existence between two groups means that the two groups are essentially the "same", in the sense that as far as algebraic properties go, there is no real difference between them.

Definition. *Two groups G and G' are called* ***isomorphic*** *if there exists an isomorphism $\varphi\colon G \to G'$.*

2.1.7 Cosets and Lagrange's theorem

The main purpose of this final section on group theory is to establish perhaps the most important result on the topic, Lagrange's Theorem. This states that, if H is a subgroup of a finite group, then the order of H divides the order of G. One consequence, very useful in itself, is that the order of any element of a finite group divides the order of the group.

We start by developing a key tool in the proof of Lagrange's Theorem.

Definition. *Given a subgroup H of a group G, define the relation R on G by*

$$a\ R\ b \quad \textit{if} \quad b = a * h \quad \textit{for some} \quad h \in H. \tag{2.1}$$

Then R is an equivalence relation:

E1 (Reflexivity). For all $a \in G$, $a\ R\ a$ since $e \in H$ and $a = a * e$.

E2 (Symmetry). For all a, b in G, if $a\ R\ b$, then there exists a $h \in H$ such that $b = a * h$ and so $b * h^{-1} = a$. Since H is a subgroup, and $h \in H$, it follows that $h^{-1} \in H$. Thus $b\ R\ a$.

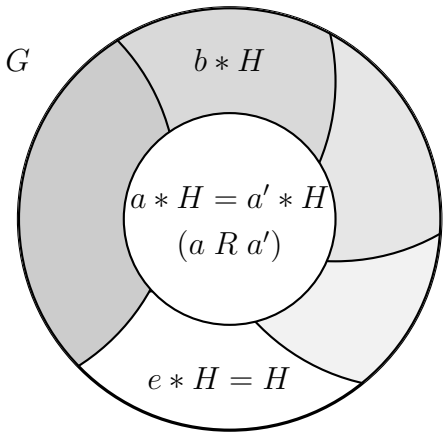
E3 (Transitivity). For all a, b, c in G, if $a\ R\ b$ and $b\ R\ c$, then there exist elements h_1, h_2 in H such that $b = a * h_1$ and $c = b * h_2$. Hence we obtain $c = b * h_2 = (a * h_1) * h_2 = a * (h_1 * h_2)$. Since $h_1, h_2 \in H$ and H is a subgroup, it follows that $h_1 * h_2 \in H$, and so $a\ R\ c$.

Definition. *Let H be a subgroup of a group G, and let R be the equivalence relation given by* (2.1)*. If $a \in G$, then the equivalence class of a, namely the set*

$$\{x \in G \mid a\ R\ x\} = \{x \in G \mid \exists h \in H \textit{ such that } x = a * h\} = \{a * h \mid h \in H\},$$

is called a ***left coset*** *of H, and is denoted by $a * H$.*

We know that the equivalence classes of an equivalence relation partition the set. Recall that by a **partition** of a set S, we mean a subdivision of the set S into nonoverlapping subsets:



Hence we obtain the following result:

Corollary 2.1.6. *Let H be a subgroup of a group G. Then the left cosets of H partition the group G.*

Remarks.

1. The notation $a * H$ denotes a certain subset of G. As with any equivalence relation, different notations may represent the same subset. In fact, we know that $a * H$ is the unique left coset containing a, and so

$$a * H = b * H \text{ iff } a \ R \ b.$$

   Corollary 2.1.6 says that if $a * H$ and $b * H$ have an element in common then they are equal.

2. Indeed, the following are all equivalent:

$$a \ R \ b; \quad b \ R \ a; \quad a * H = b * H; \quad a \in b * H; \quad b \in a * H.$$

3. One can also define the relation R' on G by $a \ R' \ b$ if $b = h * a$ for some $h \in H$. The associated equivalence classes are called **right cosets**.

Examples.

1. Consider the group $(\mathbb{Z}_6 \oplus)$ of integers modulo 6, with addition modulo 6, and let H be the subgroup $\langle [2] \rangle = \{[0], [2], [4]\}$. The left cosets, which we now denote by $a \oplus H$ are

$$[0] \oplus H = [2] \oplus H = [4] \oplus H = \{[0], [2], [4]\}, \text{ and}$$
$$[1] \oplus H = [3] \oplus H = 5 \oplus H = \{[1], [3], [5]\}.$$

   Note that the cosets $\{[0], [2], [4]\}$ and $\{[1], [3], [5]\}$ form a partition of G:

$$G = \{[0], [1], [2], [3], [4], [5]\} = \{[0], [2], [4]\} \cup \{[1], [3], [5]\},$$
$$\text{and } \{[0], [2], [4]\} \cap \{[1], [3], [5]\} = \emptyset.$$

2. Consider the group $(\mathbb{Z}, +)$, and let H be the subgroup of even numbers $\{2m \mid m \in \mathbb{Z}\}$. The left cosets, which we now denote by $a + H$ are

$$\begin{aligned}\cdots &= -4 + H = -2 + H = 0 + H \\ &= \{2m \mid m \in \mathbb{Z}\} = \{\dots, -4, -2, 0, 2, 4, \dots\} \\ &= 2 + H = 4 + H = 6 + H = \cdots,\end{aligned}$$

and

$$\begin{aligned}\cdots &= -3 + H = -1 + H = 1 + H \\ &= \{2m + 1 \mid m \in \mathbb{Z}\} = \{\dots, -3, -1, 1, 3, \dots\} \\ &= 3 + H = 5 + H = 7 + H = \cdots.\end{aligned}$$

Note that the cosets $\{\dots, -4, -2, 0, 2, 4, \dots\}$ and $\{\dots, -3, -1, 1, 3, \dots\}$ do indeed partition the set of all integers.

Note that in Example 1 above, there are only two distinct cosets, and

$$|G| = 6 = 3 \cdot 2 = |H| \cdot (\text{number of cosets of } H).$$

In particular the order of H (namely 3) divides the order of G (namely 6). This is not a coincidence. We now prove an important result concerning the order of a group G and the number of cosets of a subgroup H, due to Lagrange[2].

Theorem 2.1.7 (Lagrange's Theorem)**.** *Let H be a subgroup of a finite group G. Then the order of H divides the order of G.*

Proof. We claim that, for any $a \in G$, the function ϕ from the subgroup H to the coset $a * H$, given by $\phi(h) = a * h$, for $h \in H$, is a bijection. Indeed, it is a surjection by definition of the coset $a * H$; to see that it is an injection, we note that $\phi(h) = \phi(h') \Rightarrow a * h = a * h' \Rightarrow h = h'$.

Consequently, each coset $a * H$ has the same number of elements as H does.

Since G is the union of the cosets of H, and since these cosets do not overlap, we obtain the **counting formula**

$$|G| = |H| \cdot (\text{number of cosets of } H).$$

In particular, $|H|$ divides $|G|$. □

Corollary 2.1.8. *Let G be a finite group and let $a \in G$. Then the order[3] of a divides the order of G. Moreover, $a^{|G|} = e$.*

Proof. Let a have order m. Recall that $\langle a \rangle$ is a subgroup of G, so Theorem 2.1.3 implies that $|\langle a \rangle| = m$ divides $|G|$, and so $|G| = m \cdot k$ for some $k \in \mathbb{N}$. Thus also $a^{|G|} = a^{m \cdot k} = (a^m)^k = e^k = e$. □

[2] Joseph Louis Lagrange (1736–1813)

[3] By Exercise 26, every element in a finite group has a finite order.

The following theorem characterises all groups whose order is a prime number.

Corollary 2.1.9. *If G is a group with prime order p, then G is cyclic, and $G = \langle a \rangle$ for every $a \in G \setminus \{e\}$.*

Proof. If $a \neq e$, then a has order > 1, say m. Since m divides p, and p is prime, it follows that $m = p$. As G itself has order p, it now follows that $G = \langle a \rangle$, and so G is cyclic. □

2.1.8 Exercises

1. For each of the following definitions for $a * b$ and a given set, determine which of the axioms G0, G1, G2, G3 are satisfied by $a * b$. In which cases do we obtain a group?

   a) $\dfrac{a+b}{ab}$ on the set $\mathbb{Z}$.

   b) $\sqrt{|ab|}$ on the set $\mathbb{Q}$.

   c) $a - b$ on the set $\mathbb{Z}$.

   d) $|a - b|$ on the set $\mathbb{N} \cup \{0\}$.

   e) $-ab$ on the set $\mathbb{Q}$.

   f) $-ab$ on the set $\mathbb{Q} \setminus \{0\}$.

   g) $a + b + 1$ on the set $\mathbb{R}$.

   h) $a + 2b + 4$ on the set $\mathbb{R}$.

   i) $a + b - ab$ on the set $\mathbb{R}$.

   j) $\max\{a, b\}$ on the set $\mathbb{N}$.

   k) $\min\{a, b\}$ on the set $\mathbb{Z}$.

   l) $a + b$ on $\mathbb{R}$.

   m) $a + b$ on $[0, \infty)$.

   n) ab on $\{2^n \mid n \in \mathbb{Z}\}$.

   o) $A \cup B$ on $\mathcal{P}(S)$ where S is any set.

   p) $A \cap B$ on $\mathcal{P}(S)$ where S is any set.

2. Consider the set

$$S = \left\{ \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \;\middle|\; a, b \in \mathbb{R} \text{ and } a^2 + b^2 \neq 0 \right\}.$$

   Show that S with the operation of matrix multiplication forms a group.

3. a) Show that the set $\mathbb{Z}_6$ of integers modulo 6, with addition modulo 6 is a group. Write down its group table.

   b) Is $\mathbb{Z}_6$ with multiplication modulo 6 also a group? If, instead, we consider the set $\mathbb{Z}_6^*$ of non-zero integers modulo 6, then is $\mathbb{Z}_6^*$ a group with multiplication modulo 6?

   c) Let m be an integer such that $m \geq 2$, and let $\mathbb{Z}_m^*$ denote the set of non-zero integers modulo m. Prove that $\mathbb{Z}_m^*$ is a group with multiplication modulo m iff m is a prime number.

   HINT: If m is a prime number, then any $r \in \{1, \ldots, m-1\}$ is coprime to m, and so there exist integers s and t such that $sm + tr = 1$.

4. Let $(G, *)$ be a group.
   a) Show that if $a, b, c \in G$ are such that $a * b = a * c$, then $b = c$.
   b) Show that if $a, b \in G$, then the equation $a * x = b$ has a unique solution. Explain why this implies that every row of a group table (of a finite group) contains each element of the group exactly once.
   c) Show that if $a, b \in G$, then $(a * b)^{-1} = b^{-1} * a^{-1}$.

5. Let $(G, *)$ be a group and $a \in G$.
   a) Show that, for all $n \in \mathbb{N}$, $a^n = a * a^{n-1}$. *Use induction.*
   b) Show that, for all $n \in \mathbb{Z}$, $a^n = a^{n-1} * a$.
   c) Show that, for any $m \in \mathbb{Z}$ and $n \in \mathbb{N}$, $a^m * a^n = a^{m+n}$. *Use induction on n, with m fixed.*
   d) Show that, for any $m \in \mathbb{Z}$ and $n \in \mathbb{N}$, $a^m * a^{-n} = a^{m-n}$. *Use the previous part.*
   e) Deduce that $a^m * a^n = a^{m+n}$ for any $m, n \in \mathbb{Z}$.

6. Let G be a group with identity element e, and let a and b be elements of G. For each of the following pairs of equations, find all solutions x, and determine conditions on a and b for there to be any solutions to the equations.
   a) $x^2a = bx$ and $ax = xa$.
   b) $ax^2 = b$ and $x^3 = e$.
   c) $(xax)^3 = bx$ and $x^2a = (xa)^{-1}$.

7. Let a and b be elements of a group $(G, *)$. Show that $(a * b)^2 = a^2 * b^2$ iff $a * b = b * a$.
   Give an example of a group $(G, *)$ and two elements $a, b \in G$ such that $(a*b)^2 \neq a^2*b^2$.

8. Determine if the following statements are TRUE or FALSE: Give reasons for your answers in each case.
   In any group G with identity element e,
   a) for any $x \in G$, if $x^2 = e$ then $x = e$.
   b) for any $x \in G$, if $x^2 = x$ then $x = e$.
   c) for any $x \in G$ there exists $y \in G$ such that $x = y^2$.
   d) for any $x, y \in G$ there exists $z \in G$ such that $y = xz$.

9. Prove the following statements for any group G with identity element e:
   a) For all $a, b, c \in G$, if $abc = e$ then $cab = e$ and $bca = e$.
   b) Suppose that $a = a^{-1}$, $b = b^{-1}$ and $c = c^{-1}$ (that is, each of a, b and c is its own inverse). For all $a, b, c \in G$, if $ab = c$ then $bc = a$ and $ca = b$.
   c) For all $a \in G$, $a = a^{-1}$ iff $a^2 = e$.

10. a) Show that in any group, the inverse of the inverse of any element is itself.

    b) Let $(G, *)$ be a finite group with identity element e. Show that if the order of G is even, then there exists an element $x \in G$ such that $x \neq e$ and $x^{-1} = x$. (That is, in a group of even order there always exists an element other than the identity which is its own inverse.)

    HINT: Use 10a) to show that there is an even number of elements that are not their own inverses.

11. Show that the general linear group $GL(3, \mathbb{R})$ with matrix multiplication is not an abelian group.

12. Prove that a group G is abelian iff $(ab)^{-1} = a^{-1}b^{-1}$ for all $a, b \in G$.

13. Prove that a group G is abelian iff $(ab)^2 = a^2b^2$ for all $a, b \in G$.

14. Show that S_2 is abelian. Show that S_n is non-abelian for all $n \geq 3$.

15. How can you tell whether a finite group is abelian by looking at its group table?

16. Determine if the following statements are TRUE or FALSE. Give reasons for your answers in each case.

    a) The nonnegative integers form a subgroup of $\mathbb{Z}$ with addition.

    b) The odd integers form a subgroup of $\mathbb{Z}$ with addition.

    c) The set $\{4k \mid k \in \mathbb{Z}\}$ is a subgroup of $\mathbb{Z}$ with addition.

    d) If G is abelian and H is a subgroup of G, then H is abelian.

    e) If H and K are subgroups of G and $H \subseteq K$, then H is a subgroup of K.

17. Does there exist an infinite group with a finite subgroup?

18. Show that the set $H = \{2^n \mid n \in \mathbb{Z}\}$ is a subgroup of the group $(0, \infty)$ of positive real numbers with multiplication.

19. For $n \in \mathbb{N}$, show that the set $H = \{\alpha \in S_n \mid \alpha(1) = 1\}$ is a subgroup of $(S_n, \circ)$.

20. Show that the set $H = \{x \in \mathbb{R} \mid \sin x \in \mathbb{Q} \text{ and } \cos x \in \mathbb{Q}\}$ is a subgroup of the group $(\mathbb{R}, +)$ of real numbers with addition.

    You may use any properties of the sine and cosine functions that you know (or can look up).

21. Let S and T be sets such that $S \subseteq T$. Note that then $\mathcal{P}(S) \subseteq \mathcal{P}(T)$, that is, every subset of S is also a subset of T. Show that $\mathcal{P}(S)$ is a subgroup of the group $(\mathcal{P}(T), \triangle)$.

22. a) Let $(G, +)$ be an abelian group. Show that the set $H = \{x \in G \mid x = -x\}$ (that is, the set of all elements of G that are inverses of themselves) is a subgroup of G.

    b) Show that the set $H = \{\theta \in S_3 \mid \theta = \theta^{-1}\}$ is not a subgroup of $(S_3, \circ)$.

23. a) Show that, if H and K are subgroups of a group G, then $H \cap K$ is also a subgroup of G.

    b) Consider the group of integers $\mathbb{Z}$ with addition. Suppose that H is the subgroup of $\mathbb{Z}$ comprising multiples of 4, and let K be the subgroup of $\mathbb{Z}$ comprising multiples of 6. What is $H \cap K$?

    c) If H and K are subgroups of a group G, is $H \cup K$ necessarily a subgroup of G? If true, give a proof. If false, provide a counterexample.

24. Let $C[0,1]$ denote the group comprising the set of continuous functions on the interval $[0,1]$ with addition of functions defined pointwise: if f, g belong to $C[0,1]$, then for all $x \in [0,1]$, $(f+g)(x) = f(x) + g(x)$. Prove that each of the following subsets of $C[0,1]$ are subgroups of $C[0,1]$.

    a) $H_1 = \left\{ f \in C[0,1] \;\middle|\; f\left(\frac{1}{2}\right) = 0 \right\}$.

    b) The set H_2 of all polynomial functions, that is: the set of all functions $p\colon [0,1] \to \mathbb{R}$ such that there exists an $n \in \mathbb{N} \cup \{0\}$ and real numbers $a_0, a_1, a_2, \ldots, a_n$ such that $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$, for all $x \in [0,1]$.

    c) $H_3 = \{ f \in C[0,1] \mid \forall x \in [0,1]\ f(1-x) = -f(x) \}$.

25. Let G be a group. The **centre of** G is the set $Z(G) = \{ z \in G \mid \forall a \in G, \quad z * a = a * z \}$.

    a) Show that $Z(G)$ is not empty.

    b) If G is abelian, then determine $Z(G)$.

    c) Show that $Z(G)$ is a subgroup of G.

    d) If G is the group $GL(2,\mathbb{R})$ with matrix multiplication, then determine $Z(G)$.

    HINT: Consider $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. Which elements Z of $GL(2,\mathbb{R})$ satisfy both $ZA = AZ$ and $ZB = BZ$?

26. If G is a finite group, then show that every element in the group has finite order, which is at most equal to $|G|$.

    HINT: If $a \in G$, then consider the set $S = \{e, a, a^2, a^3, \ldots, a^{|G|}\}$, and use the pigeonhole principle.

27. Determine the order of $\begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix}$ in the group $GL(2,\mathbb{R})$.

28. Determine the orders of all the elements of the following groups:

    a) the group $\mathbb{Z}_6$ with addition modulo 6.

    b) the symmetric group S_3.

    c) the group $(\mathcal{P}(S), \triangle)$.

29. Let H be a subgroup of a group $(G, *)$, and let h be an element of H. Is it necessarily true that the order of h in $(H, *)$ is the same as the order of h in $(G, *)$? Justify your answer, by means of a proof or a counterexample.

30. Prove that in any group $(G, *)$, and for any a, b in G, the orders of $a * b$ and $b * a$ are the same. (Note that it is not given that $(G, *)$ is abelian.)

31. Is the group of integers with addition cyclic? What is a generator of this group? Is it unique?

32. Show that any cyclic group is abelian. Give an example of an abelian group that is not cyclic.

33. Define $\phi : (\mathbb{R}, +) \to (\mathbb{C}, \times)$ by $\phi(x) = e^{ix} = \cos x + i \sin x$. Show that ϕ is a homomorphism, and find its kernel and image. *You may use any properties you know about the function* $x \mapsto e^{ix}$.

34. a) Let G, G', G'' be groups and let $\varphi\colon G \to G'$ and $\psi\colon G' \to G''$ be group homomorphisms. Prove that the composition $\psi \circ \varphi\colon G \to G''$ defined by $(\psi \circ \varphi)(a) = \psi(\varphi(a))$, $a \in G$ is a group homomorphism.

    b) Describe the kernel of $\psi \circ \varphi$.

35. Let $\varphi\colon G \to G'$ be a homomorphism from a group G with identity element e to a group G'. Show that φ is an injection iff $\ker(\varphi) = \{e\}$.

36. Let G be a group and let a be an element of G. Prove that the function from $\mathbb{Z}$ to $\langle a \rangle$, given by $m \mapsto a^m$, is a homomorphism from the group $(\mathbb{Z}, +)$ to G. What is the image of this homomorphism? What is the kernel of the homomorphism: (a) if a has finite order m, (b) if a has infinite order.

37. Show that the function $\varphi(n) = 2n$ is an isomorphism from $(\mathbb{Z}, +)$ to its subgroup of even integers $\{2k \mid k \in \mathbb{Z}\}$.

38. Let G be the group defined on $\mathbb{R}$ with law of composition $a * b = a + b + 1$. Show that G and $(\mathbb{R}, +)$ are isomorphic.

    HINT: Try $\varphi(x) = x + 1$.

39. Let G be the group defined on $\mathbb{R} \setminus \{-1\}$ with law of composition $a * b = a + b - ab$. Show that G is isomorphic to the group $\mathbb{R} \setminus \{0\}$ with multiplication.

    HINT: Try $\varphi(x) = 1 - x$.

40. A subgroup N of a group G is called a **normal subgroup** if for every $a \in N$ and every $b \in G$, then $b * a * b^{-1} \in N$. Prove that if G, G' are groups and $\varphi\colon G \to G'$ is a homomorphism, then $\ker(\varphi)$ is a normal subgroup of G.

41. Show that the function from G to G given by $a \mapsto a^{-1}$ is an isomorphism iff G is abelian.

42. Let G, G' be groups and let $\varphi\colon G \to G'$ be an isomorphism. Prove that the inverse function $\varphi^{-1}\colon G' \to G$ is also an isomorphism.

43. a) Let G and G' be isomorphic groups with isomorphism $\varphi\colon G \to G'$.
    i) Show that for any $a \in G$, a has the same order as $\varphi(a)$ in G'.
    ii) Show that if G is finite, then G' is also finite and $|G| = |G'|$.
    iii) Show that if G is abelian, then G' is also abelian.
    b) Use 43a) to show that $(\mathbb{Z}_6, +)$ and S_3 are not isomorphic.
    c) Use 43a) to show that $(\mathbb{Z}_4, +)$ and $\mathcal{P}(\{1,2\})$ are not isomorphic.

44. Let $(G, *)$ be a group. Define a new law of composition $\circ$ on G by $a \circ b = b * a$. (Thus if $(G, *)$ is abelian, $\circ$ is exactly the same as $*$, but we are not assuming that $(G, *)$ is abelian.)
    a) Show that $(G, \circ)$ is a group.
    b) Show that the groups $(G, *)$ and $(G, \circ)$ are isomorphic.
    HINT: Try $a \mapsto a^{-1}$ for an isomorphism.

45. Determine if the following statements are TRUE or FALSE. Give reasons for your answers in each case.
    a) If H is a subgroup of G and $a, b \in G$ are such that $a \neq b$, then $(a * H) \cap (b * H) = \emptyset$.
    b) If H is a subgroup of G and $a \in G$ is such that $a * H$ has 4 elements, then H has 4 elements.
    c) If H is a subgroup of a finite group G, then for any $a \in G$, the left coset $a * H$ has the same number of elements as the right coset $H * a$.
    d) All groups of order 7 are isomorphic to $(\mathbb{Z}_7, +)$.
    e) All groups of order 6 are isomorphic to $(\mathbb{Z}_6, +)$.
    f) The group S_4 does not have an element of order 13.
    g) The group S_4 does not have an element of order 24.
    h) The group S_4 does not have an element of order 12.

46. a) Verify that
$$G = \left\{ \begin{bmatrix} x & y \\ 0 & 1 \end{bmatrix} \,\middle|\, x, y \in \mathbb{R} \text{ and } x > 0 \right\}$$
    is a group with matrix multiplication.
    b) Show that
$$H = \left\{ \begin{bmatrix} x & 0 \\ 0 & 1 \end{bmatrix} \,\middle|\, x \in \mathbb{R} \text{ and } x > 0 \right\}$$
    is a subgroup of G.
    c) An element $\begin{bmatrix} x & y \\ 0 & 1 \end{bmatrix}$ of G can be represented by the point (x, y) in the *right half-plane* $\{(x, y) \mid x, y \in \mathbb{R} \text{ and } x > 0\}$. Draw the partitions of the right half-plane into left and into right cosets of H.

47. Let
$$G = \left\{ \left[\begin{array}{cc} a & b \\ 0 & a^{-1} \end{array} \right] \;\middle|\; a, b \in \mathbb{Z}_5 \text{ and } a \neq 0 \right\}.$$

(a) Show that G is closed under matrix multiplication, with the addition and multiplication carried out in $\mathbb{Z}_5$.

(b) Find the inverse of the matrix $\left[\begin{array}{cc} a & b \\ 0 & a^{-1} \end{array} \right]$ in $(G, \otimes)$, where $\otimes$ denotes matrix multiplication as in (a).

You may assume without further proof that $(G, \otimes)$ is a group.

(c) What is the order of G?

(d) Let H be the subset of G consisting of the diagonal matrices. Show that H is a subgroup of G.

(e) Let $m = \left[\begin{array}{cc} 2 & 1 \\ 0 & 3 \end{array} \right]$. Find all the elements of the left coset $m \otimes H$.

48. Let H and K be subgroups of a group G of orders 3 and 5 respectively. Prove that $H \cap K = \{e\}$.

49. a) Let p be a prime, and let $(\mathbb{Z}_p^*, \otimes)$ be the group of non-zero integers modulo p, with multiplication modulo p. Show that, if a is an integer such that a is not divisible by p, then $[a]^{p-1} = [1]$.

b) Prove **Fermat's little theorem**: for any integer a, $a^p \equiv a \pmod{p}$.

HINT: If $a \in \mathbb{Z}$ is not divisible by p, then $[a] \neq [0]$, and so by part (49a) above, $[a]^{p-1} = [1]$. Hence $p \mid (a^{p-1} - 1)$, and so $p \mid (a^p - a)$.

c) Show that 7 divides $2222^{5555} + 5555^{2222}$.

HINT: Note that $2222 = 7 \cdot 317 + 3$, so that in $\mathbb{Z}_7$, $[2222^{5555}] = [3]^{5555}$. Now use the fact that $[3]^6 = [1]$ to conclude that $[2222^{5555}] = [3^5]$. Now simplify $[5555^{2222}]$ in a similar fashion.

50. **Sample Exam Question. 2012 Q7.**

(a) Let $G = \mathbb{R} \setminus \{0\}$, and for each $a, b \in G$ define $a * b = -ab$.

(i) Show that $(G, *)$ is an Abelian group.

(ii) Define the function $\varphi : (G, *) \to (G, *)$ by
$$\varphi(x) = \frac{x}{|x|}.$$
Show that φ is a homomorphism. Determine $\ker(\varphi)$ and $\mathrm{im}(\varphi)$.

(iii) Give an example of an infinite subgroup of $(G, *)$ not equal to G, and an example of a finite subgroup of $(G, *)$ of order greater than 1.

(b) Let G be a group, and a an element of G.

(i) State the definition of the *order* of a.

(ii) Show that, if the order of a is m^2, then the order of a^m is m.

(iii) Suppose that the order of G is 25. Use (ii) and the theorem of Lagrange to show that G contains an element of order 5.

2.2 Vector spaces

In this section we introduce our second important example of an algebraic structure, called a vector space. Roughly speaking it is a set of elements, called "vectors". Any two vectors can be "added", resulting in a new vector, and any vector can be multiplied by an element from $\mathbb{R}$ so as to give a new vector. The precise definition is given in the next subsection.

2.2.1 Definition of a vector space

Definition. *A **vector space** V is a set together with two functions, $+\colon V \times V \to V$, called **vector addition**, and $\cdot\colon \mathbb{R} \times V \to V$, called **scalar multiplication**, such that $(V, +)$ is an abelian group, and the following hold:*

V1. *For all $\boldsymbol{v} \in V$, $1 \cdot \boldsymbol{v} = \boldsymbol{v}$.*

V2. *For all $\alpha, \beta \in \mathbb{R}$ and all $\boldsymbol{v} \in V$, $\alpha \cdot (\beta \cdot \boldsymbol{v}) = (\alpha\beta) \cdot \boldsymbol{v}$.*

V3. *For all $\alpha, \beta \in \mathbb{R}$ and all $\boldsymbol{v} \in V$, $(\alpha + \beta) \cdot \boldsymbol{v} = \alpha \cdot \boldsymbol{v} + \beta \cdot \boldsymbol{v}$.*

V4. *For all $\alpha \in \mathbb{R}$ and all $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$, $\alpha \cdot (\boldsymbol{v}_1 + \boldsymbol{v}_2) = \alpha \cdot \boldsymbol{v}_1 + \alpha \cdot \boldsymbol{v}_2$.*

*V3 and V4 are called the **distributive laws**. The elements of a vector space are called* vectors.

We observe that since $(V, +)$ is an abelian group, a vector space also has the following properties built into its definition: (note the use of additive notation)

G1. For all $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3 \in V$, $\boldsymbol{v}_1 + (\boldsymbol{v}_2 + \boldsymbol{v}_3) = (\boldsymbol{v}_1 + \boldsymbol{v}_2) + \boldsymbol{v}_3$.

G2. There exists an element $\boldsymbol{o} \in V$ (called the[4] **zero vector**) such that for all $\boldsymbol{v} \in V$, $\boldsymbol{v} + \boldsymbol{o} = \boldsymbol{v} = \boldsymbol{o} + \boldsymbol{v}$.

G3. For every $\boldsymbol{v} \in V$, there exists a unique[5] element in V, denoted by $-\boldsymbol{v}$, such that $\boldsymbol{v} + (-\boldsymbol{v}) = \boldsymbol{o} = -\boldsymbol{v} + \boldsymbol{v}$.

A. For all $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$, $\boldsymbol{v}_1 + \boldsymbol{v}_2 = \boldsymbol{v}_2 + \boldsymbol{v}_1$.

Examples.

[4] Since there is a unique identity element in a group, the zero vector is unique!
[5] In a group, every element has a unique inverse!

1. Let $m, n \in \mathbb{N}$, and consider the set $\mathbb{R}^{m\times n}$ of $m \times n$ matrices having real entries. It is easy to check that $\mathbb{R}^{m\times n}$, with matrix addition, is an abelian group: we omit the verification of G1, G2, G3 and A. Define scalar multiplication as follows: if $\alpha \in \mathbb{R}$ and

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \in \mathbb{R}^{m\times n},$$

then

$$\alpha \cdot A = \begin{bmatrix} \alpha a_{11} & \dots & \alpha a_{1n} \\ \vdots & & \vdots \\ \alpha a_{m1} & \dots & \alpha a_{mn} \end{bmatrix}. \tag{2.2}$$

Then $\alpha \cdot A \in \mathbb{R}^{m\times n}$, and moreover V1, V2, V3, V4 are satisfied:

V1. If $A \in \mathbb{R}^{m\times n}$, then clearly

$$\begin{aligned} 1 \cdot A = 1 \cdot \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} &= \begin{bmatrix} 1a_{11} & \dots & 1a_{1n} \\ \vdots & & \vdots \\ 1a_{m1} & \dots & 1a_{mn} \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} = A. \end{aligned}$$

V2. For all $\alpha, \beta \in \mathbb{R}$ and all $A \in \mathbb{R}^{m\times n}$,

$$\begin{aligned} \alpha \cdot (\beta \cdot A) = \alpha \cdot \left(\beta \cdot \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \right) &= \alpha \cdot \begin{bmatrix} \beta a_{11} & \dots & \beta a_{1n} \\ \vdots & & \vdots \\ \beta a_{m1} & \dots & \beta a_{mn} \end{bmatrix} \\ = \begin{bmatrix} \alpha(\beta a_{11}) & \dots & \alpha(\beta a_{1n}) \\ \vdots & & \vdots \\ \alpha(\beta a_{m1}) & \dots & \alpha(\beta a_{mn}) \end{bmatrix} &= \begin{bmatrix} (\alpha\beta) a_{11} & \dots & (\alpha\beta) a_{1n} \\ \vdots & & \vdots \\ (\alpha\beta) a_{m1} & \dots & (\alpha\beta) a_{mn} \end{bmatrix} \\ = (\alpha\beta) \cdot \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} &= (\alpha\beta) \cdot A. \end{aligned}$$

V3. For all $\alpha, \beta \in \mathbb{R}$ and all $A \in \mathbb{R}^{m\times n}$,

$$\begin{aligned}
(\alpha+\beta)\cdot A &= (\alpha+\beta)\cdot \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \\
&= \begin{bmatrix} (\alpha+\beta)a_{11} & \dots & (\alpha+\beta)a_{1n} \\ \vdots & & \vdots \\ (\alpha+\beta)a_{m1} & \dots & (\alpha+\beta)a_{mn} \end{bmatrix} \\
&= \begin{bmatrix} \alpha a_{11}+\beta a_{11} & \dots & \alpha a_{1n}+\beta a_{1n} \\ \vdots & & \vdots \\ \alpha a_{m1}+\beta a_{m1} & \dots & \alpha a_{mn}+\beta a_{mn} \end{bmatrix} \\
&= \begin{bmatrix} \alpha a_{11} & \dots & \alpha a_{1n} \\ \vdots & & \vdots \\ \alpha a_{m1} & \dots & \alpha a_{mn} \end{bmatrix} + \begin{bmatrix} \beta a_{11} & \dots & \beta a_{1n} \\ \vdots & & \vdots \\ \beta a_{m1} & \dots & \beta a_{mn} \end{bmatrix} \\
&= \alpha\cdot \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} + \beta\cdot \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \\
&= \alpha\cdot A + \beta\cdot A.
\end{aligned}$$

V4. For all $\alpha \in \mathbb{R}$ and all $A, B \in \mathbb{R}^{m\times n}$,

$$\begin{aligned}
\alpha\cdot(A+B) &= \alpha\cdot\left(\begin{bmatrix} a_{11} & \dots a_{1n} \\ \vdots & \vdots \\ a_{m1} & \dots a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & \dots b_{1n} \\ \vdots & \vdots \\ b_{m1} & \dots b_{mn} \end{bmatrix}\right) \\
&= \alpha\cdot \begin{bmatrix} a_{11}+b_{11} & \dots a_{1n}+b_{1n} \\ \vdots & \vdots \\ a_{m1}+b_{m1} & \dots a_{mn}+b_{mn} \end{bmatrix} \\
&= \begin{bmatrix} \alpha(a_{11}+b_{11}) & \dots \alpha(a_{1n}+b_{1n}) \\ \vdots & \vdots \\ \alpha(a_{m1}+b_{m1}) & \dots \alpha(a_{mn}+b_{mn}) \end{bmatrix} \\
&= \begin{bmatrix} \alpha a_{11}+\alpha b_{11} & \dots \alpha a_{1n}+\alpha b_{1n} \\ \vdots & \vdots \\ \alpha a_{m1}+\alpha b_{m1} & \dots \alpha a_{mn}+\alpha b_{mn} \end{bmatrix} \\
&= \begin{bmatrix} \alpha a_{11} & \dots \alpha a_{1n} \\ \vdots & \vdots \\ \alpha a_{m1} & \dots \alpha a_{mn} \end{bmatrix} + \begin{bmatrix} \alpha b_{11} & \dots \alpha b_{1n} \\ \vdots & \vdots \\ \alpha b_{m1} & \dots \alpha b_{mn} \end{bmatrix} \\
&= \alpha\cdot \begin{bmatrix} a_{11} & \dots a_{1n} \\ \vdots & \vdots \\ a_{m1} & \dots a_{mn} \end{bmatrix} + \alpha\cdot \begin{bmatrix} b_{11} & \dots b_{1n} \\ \vdots & \vdots \\ b_{m1} & \dots b_{mn} \end{bmatrix}
\end{aligned}$$

$$= \alpha \cdot A + \alpha \cdot B.$$

Hence $\mathbb{R}^{m\times n}$ is a vector space with matrix addition and with scalar multiplication defined by (2.2). If $n = 1$, then we denote the vector space of column vectors $\mathbb{R}^{m\times 1}$ by $\mathbb{R}^m$.

2. Let a, b be real numbers with $a < b$. The set $C[a, b]$ of continuous functions on the interval $[a, b]$ with addition of functions is an abelian group. Let scalar multiplication be defined as follows:

$$\text{if } \alpha \in \mathbb{R} \text{ and } f \in C[a, b], \text{ then } (\alpha \cdot f)(x) = \alpha f(x), \quad x \in [a, b]. \tag{2.3}$$

(We say that scalar multiplication, and addition, are defined "pointwise".) Then $\alpha \cdot f \in C[a, b]$, and moreover V1, V2, V3, V4 are satisfied:

V1. Let $f \in C[a, b]$. For all $x \in [a, b]$, we have

$$(1 \cdot f)(x) = 1f(x) = f(x),$$

and so $1 \cdot f = f$.

V2. Let $\alpha, \beta \in \mathbb{R}$ and $f \in C[a, b]$. For all $x \in [a, b]$, we have

$$\begin{aligned}(\alpha \cdot (\beta \cdot f))(x) &= \alpha(\beta \cdot f)(x)\\ &= \alpha(\beta f(x))\\ &= (\alpha\beta)f(x)\\ &= ((\alpha\beta) \cdot f)(x),\end{aligned}$$

and so $(\alpha \cdot (\beta \cdot f) = (\alpha\beta) \cdot f$.

V3. Let $\alpha, \beta \in \mathbb{R}$ and $f \in C[a, b]$. For all $x \in [a, b]$, we have

$$\begin{aligned}((\alpha + \beta) \cdot f)(x) &= (\alpha + \beta)f(x)\\ &= \alpha f(x) + \beta f(x)\\ &= (\alpha \cdot f)(x) + (\beta \cdot f)(x)\\ &= (\alpha \cdot f + \beta \cdot f)(x),\end{aligned}$$

and so $(\alpha + \beta) \cdot f = \alpha \cdot f + \beta \cdot f$.

V4. Let $\alpha \in \mathbb{R}$ and $f, g \in C[a, b]$. For all $x \in [a, b]$, we have

$$\begin{aligned}(\alpha \cdot (f + g))(x) &= \alpha(f + g)(x)\\ &= \alpha(f(x) + g(x))\\ &= \alpha f(x) + \alpha g(x)\\ &= (\alpha \cdot f)(x) + (\alpha \cdot g)(x)\\ &= (\alpha \cdot f + \alpha \cdot g)(x),\end{aligned}$$

and so $\alpha \cdot (f + g) = \alpha \cdot f + \alpha \cdot g$.

Hence $C[a, b]$ with addition and scalar multiplication is a vector space, called the **vector space of continuous functions** on $[a, b]$.

We now prove a few elementary properties of vector spaces.

Theorem 2.2.1. *Let V be a vector space. Then the following hold:*

1. *For all $\boldsymbol{v} \in V$, $0 \cdot \boldsymbol{v} = \boldsymbol{o}$.*
2. *For all $\alpha \in \mathbb{R}$, $\alpha \cdot \boldsymbol{o} = \boldsymbol{o}$.*
3. *If $\boldsymbol{v} \in V$, then $(-1) \cdot \boldsymbol{v} = -\boldsymbol{v}$.*

Proof. 1. To see this, we use the distributive law to write

$$0 \cdot \boldsymbol{v} + 0 \cdot \boldsymbol{v} = (0 + 0) \cdot \boldsymbol{v} = 0 \cdot \boldsymbol{v}.$$

Now add $-(0 \cdot \boldsymbol{v})$ on both sides, to obtain $0 \cdot \boldsymbol{v} = \boldsymbol{o}$.

2. Similarly, $\alpha \cdot \boldsymbol{o} + \alpha \cdot \boldsymbol{o} = \alpha \cdot (\boldsymbol{o} + \boldsymbol{o}) = \alpha \cdot \boldsymbol{o}$, and hence $\alpha \cdot \boldsymbol{o} = \boldsymbol{o}$.

3. Finally, we have

$$\begin{aligned} \boldsymbol{v} + (-1) \cdot \boldsymbol{v} &= 1 \cdot \boldsymbol{v} + (-1) \cdot \boldsymbol{v} \text{ (since } 1 \cdot \boldsymbol{v} = \boldsymbol{v}) \\ &= (1 + (-1)) \cdot \boldsymbol{v} \text{ (distributive law)} \\ &= 0 \cdot \boldsymbol{v} \text{ (since } 1 + (-1) = 0) \\ &= \boldsymbol{o} \text{ (since } 0 \cdot \boldsymbol{v} = \boldsymbol{o}), \end{aligned}$$

and so $\boldsymbol{v} + (-1) \cdot \boldsymbol{v} = \boldsymbol{o} = (-1) \cdot \boldsymbol{v} + \boldsymbol{v}$. Hence $(-1) \cdot \boldsymbol{v}$ is an inverse of $\boldsymbol{v}$. But the inverse of an element from a group is unique, and so $(-1) \cdot \boldsymbol{v} = -\boldsymbol{v}$. □

2.2.2 Subspaces and linear combinations

Definition. *Let V be a vector space. A subset U is called a* subspace of V *if*

S1. *$\boldsymbol{o} \in U$.*

S2. *If $\boldsymbol{v}_1, \boldsymbol{v}_2 \in U$, then $\boldsymbol{v}_1 + \boldsymbol{v}_2 \in U$.*

S3. *If $\boldsymbol{v} \in U$ and $\alpha \in \mathbb{R}$, then $\alpha \cdot \boldsymbol{v} \in U$.*

Subspaces of a vector space are analogous to subgroups of a group in the sense that a subspace of a vector space is itself a vector space with the same addition and scalar multiplication as with V (this is easy to check). So a subspace is really a smaller vector space sitting inside a larger vector space.

Examples.

1. If V is a vector space, then the subset U comprising only the zero vector, namely $U = \{\boldsymbol{o}\}$, is a subspace of V.

   Also, the entire vector space, that is $U = V$, is a subspace of V.

   If a subspace U of V is neither $\{\boldsymbol{o}\}$ nor V, then it is called a **proper subspace** of V.

2. Consider the vector space $\mathbb{R}^{2\times 2}$ with matrix addition and scalar multiplication defined by (2.2). Then the set of **upper triangular matrices**

$$U_1 = \left\{ \begin{bmatrix} a & b \\ 0 & d \end{bmatrix} \;\middle|\; a, b, d \in \mathbb{R} \right\}$$

   is a subspace of $\mathbb{R}^{2\times 2}$.

   Also, the set of **symmetric matrices**

$$U_2 = \left\{ \begin{bmatrix} a & b \\ b & d \end{bmatrix} \;\middle|\; a, b, d \in \mathbb{R} \right\}$$

   is a subspace of $\mathbb{R}^{2\times 2}$.

3. Let $a, b \in \mathbb{R}$ and $a < b$. Consider the set of all polynomial functions

$$P[a,b] = \left\{ p\colon [a,b] \to \mathbb{R} \;\middle|\; \begin{array}{l} \exists n \in \mathbb{N} \cup \{0\} \text{ and} \\ a_0, a_1, a_2, \ldots, a_n \in \mathbb{R} \text{ such that} \\ p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \\ \text{for all } x \in [a,b] \end{array} \right\}.$$

   Then $U = P[a,b]$ is a subspace of the vector space $C[a,b]$ with addition of functions and scalar multiplication defined by (2.3).

Definitions. *Let V be a vector space.*

1. *If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are vectors in V and $\alpha_1, \ldots, \alpha_n$ belong to $\mathbb{R}$, then the vector $\alpha_1 \cdot \boldsymbol{v}_1 + \cdots + \alpha_n \cdot \boldsymbol{v}_n$ is called a **linear combination of the vectors** $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$.*

2. *Let S be a non-empty subset of a vector space V. The **span** of S, denoted by* $\operatorname{lin}(S)$*, is defined as the set of all possible linear combinations[6] of vectors from S:*

$$\operatorname{lin}(S) = \{\alpha_1 \cdot \boldsymbol{v}_1 + \cdots + \alpha_n \cdot \boldsymbol{v}_n \mid n \in \mathbb{N},\ \boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in S,\ \alpha_1, \ldots, \alpha_n \in \mathbb{R}\}.$$

   *We also define the **span of the empty set** as* $\operatorname{lin}(\emptyset) = \{\boldsymbol{o}\}$.

[6] Note that although S might be infinite, a linear combination, by definition, is always a linear combination of a *finite* set of vectors from S.

Examples.

1. Let $m \in \mathbb{N}$. Any vector in the vector space $\mathbb{R}^m$ is a linear combination of the vectors

$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ \ldots, \ \boldsymbol{e}_m = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

   Hence $\operatorname{lin}(\{\boldsymbol{e}_1, \boldsymbol{e}_2 \ldots, \boldsymbol{e}_m\}) = \mathbb{R}^m$.

2. Let $a, b \in \mathbb{R}$ with $a < b$. Any polynomial p on the interval $[a, b]$ is a linear combination of the functions from the set $S = \{1, x, x^2, \ldots\}$. Hence $\operatorname{lin}(S) = P[a, b]$ in the vector space $C[a, b]$.

The span of a set of vectors turns out to be a special subspace of the vector space.

Theorem 2.2.2. *Let V be a vector space and S be a subset of V. Then $\operatorname{lin}(S)$ is a subspace of V, and moreover $\operatorname{lin}(S)$ is the smallest subspace of V that contains S.*

Proof. The result is true in the case where $S = \emptyset$, as $\operatorname{lin}(\emptyset) = \{\boldsymbol{o}\}$ is the smallest subspace of V.

Suppose that S is non-empty, and take any $\boldsymbol{v} \in S$. Then $\boldsymbol{o} = 0 \cdot \boldsymbol{v} \in \operatorname{lin}(S)$. If $\boldsymbol{u}, \boldsymbol{v} \in \operatorname{lin}(S)$, then we know that

$$\boldsymbol{u} = \alpha_1 \cdot \boldsymbol{u}_1 + \cdots + \alpha_n \cdot \boldsymbol{u}_n \text{ and } \boldsymbol{v} = \beta_1 \cdot \boldsymbol{v}_1 + \cdots + \beta_m \cdot \boldsymbol{v}_m$$

for some vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_m \in S$ and scalars $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_m \in \mathbb{R}$. Consequently,

$$\boldsymbol{u} + \boldsymbol{v} = \alpha_1 \cdot \boldsymbol{u}_1 + \cdots + \alpha_n \cdot \boldsymbol{u}_n + \beta_1 \cdot \boldsymbol{v}_1 + \cdots + \beta_m \cdot \boldsymbol{v}_m \in \operatorname{lin}(S).$$

Finally, if $\boldsymbol{v} \in \operatorname{lin}(S)$, then we know that $\boldsymbol{v} = \alpha_1 \cdot \boldsymbol{v}_1 + \cdots + \alpha_n \cdot \boldsymbol{v}_n$ for some $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in S$ and scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, and so for any $\beta \in \mathbb{R}$, we have

$$\begin{aligned} \beta \cdot \boldsymbol{v} &= \beta \cdot (\alpha_1 \cdot \boldsymbol{v}_1 + \cdots + \alpha_n \cdot \boldsymbol{v}_n) \\ &= \beta \cdot (\alpha_1 \cdot \boldsymbol{v}_1) + \cdots + \beta \cdot (\alpha_n \cdot \boldsymbol{v}_n) \\ &= (\beta\alpha_1) \cdot \boldsymbol{v}_1 + \cdots + (\beta\alpha_n) \cdot \boldsymbol{v}_n \in \operatorname{lin}(S). \end{aligned}$$

So $\operatorname{lin}(S)$ satisfies S1,S2,S3, and so it is a subspace of V. Moreover, if $\boldsymbol{v} \in S$, then $\boldsymbol{v} = 1 \cdot \boldsymbol{v} \in \operatorname{lin}(S)$, and so $S \subset \operatorname{lin}(S)$. Thus $\operatorname{lin}(S)$ contains S.

If U is another subspace that contains the vectors from S, then from S2 and S3 it follows that all linear combinations of vectors from S belong to U, and so $\operatorname{lin}(S) \subset U$.

Hence $\operatorname{lin}(S)$ is the smallest subspace of V containing S. □

Definitions. *Let V be a vector space, and suppose that $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are vectors that belong to V.*

1. *The vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are called **linearly dependent** if the following condition holds:*

$\exists \alpha_1, \ldots, \alpha_n \in \mathbb{R}$ *such that*[7] $(\alpha_1, \alpha_2, \ldots, \alpha_n) \neq (0, 0, \ldots, 0)$ *and* $\alpha_1 \cdot \boldsymbol{v}_1 + \cdots + \alpha_n \cdot \boldsymbol{v}_n = \boldsymbol{o}$.

An arbitrary subset S of vectors from V is said to be a ***linearly dependent set*** *if there exists a non-empty finite set of dependent vectors from S.*

2. *The vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are called* ***linearly independent*** *if they are not linearly dependent, that is,*

   $\forall \alpha_1, \ldots, \alpha_n \in \mathbb{R}$, *if* $\alpha_1 \cdot \boldsymbol{v}_1 + \cdots + \alpha_n \cdot \boldsymbol{v}_n = \boldsymbol{o}$, *then* $\alpha_1 = \cdots = \alpha_n = 0$.

   An arbitrary subset S of vectors from V is said to be a ***linearly independent set*** *if every non-empty finite set of vectors from S is an independent set of vectors.*

Examples.

1. Let V be a vector space. Then any finite set of vectors from V containing the zero vector is linearly dependent. Indeed if $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in V$ and $\boldsymbol{v}_k = \boldsymbol{o}$, then

$$0 \cdot \boldsymbol{v}_1 + \cdots + 0 \cdot \boldsymbol{v}_{k-1} + 1 \cdot \boldsymbol{v}_k + 0 \cdot \boldsymbol{v}_{k+1} + \cdots + 0 \cdot \boldsymbol{v}_n = \boldsymbol{o}.$$

2. The vectors

$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ \ldots, \ \boldsymbol{e}_m = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

   are linearly independent in $\mathbb{R}^m$. Indeed if $\alpha_1 \cdot \boldsymbol{e}_1 + \boldsymbol{e}_2 + \cdots + \alpha_m \cdot \boldsymbol{e}_m = \boldsymbol{o}$, then

$$\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix} = \alpha_1 \cdot \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \alpha_2 \cdot \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \alpha_m \cdot \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix},$$

   and so $\alpha_1 = \cdots = \alpha_m = 0$.

3. Let $a, b \in \mathbb{R}$ with $a < b$. The functions $1, x$ on the interval $[a, b]$ are linearly independent. Indeed, if for all $x \in [a, b]$, $\alpha \cdot 1 + \beta \cdot x = \mathbf{0}(x)$, then in particular, we have

$$\alpha + \beta a = 0 \text{ and } \alpha + \beta b = 0,$$

   and since $a \neq b$, it follows that $\alpha = \beta = 0$.

4. It follows from the definition that the empty set is not linearly dependent, since it does not contain a non-empty subset. Thus the empty set is a linearly independent subset of any vector space.

[7] Note that $(\alpha_1, \alpha_2, \ldots, \alpha_n) \neq (0, 0, \ldots, 0)$ iff not all of $\alpha_1, \ldots, \alpha_n$ are equal to 0, that is, iff at least one of them is non-zero.

2.2.3 Basis of a vector space

Definition. *Let V be a vector space. Then a set B of vectors is said to be a* ***basis*** *of V if*

B1. $\lin(B) = V$, *and*

B2. *B is linearly independent.*

Examples.

1. The vectors
$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ \ldots, \ \boldsymbol{e}_m = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$
form a basis of $\mathbb{R}^m$.

2. The empty set $\emptyset$ is a basis of the subspace $\{\boldsymbol{o}\}$ of any vector space.

Theorem 2.2.3. *Let V be a vector space and B be a basis of V such that B has n elements. Then for any linearly independent set S of $m \leq n$ vectors there exists a subset $B' \subset B$ of $n - m$ vectors such that $S \cup B'$ is also a basis of V.*

Proof. We use induction on $m \geq 0$.

The case $m = 0$ is simple: $S = \emptyset$ and for B' we take the whole of B. Then $S \cup B' = \emptyset \cup B = B$, a basis of V.

Now assume as induction hypothesis that $m \in \mathbb{N}$ is fixed, then for any linearly independent set S_1 of $m - 1 \leq n$ vectors there exists a subset $B_1' \subset B$ of $n - (m - 1)$ vectors such that $S_1 \cup B_1'$ is a basis of V.

We have to prove that for any linearly independent set S of $m \leq n$ vectors there exists a subset $B' \subset B$ of $n - m$ vectors such that $S \cup B'$ is also a basis of V.

Write $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$. To apply the induction hypothesis we choose $S_1 = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{m-1}\} = S \setminus \{\boldsymbol{v}_m\}$. Then S_1 contains $m-1$ vectors, and because $m \leq n$, also $m-1 \leq n-1 < n$, so we may apply the induction hypothesis to obtain a subset $B_1' \subset B$ of $n-(m-1) = n-m+1$ vectors such that $S_1 \cup B_1'$ is a basis of V. Write $B_1' = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{n-m+1}\}$.

We now want to remove a vector from B_1' to obtain the required B', that is, such that $S \cup B'$ will end up as a basis. Since
$$S_1 \cup B_1' = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{m-1}, \boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{n-m+1}\}$$
is a basis, there exist numbers
$$\alpha_1, \alpha_2, \ldots, \alpha_{m-1}, \beta_1, \beta_2, \ldots, \beta_{n-m+1} \in \mathbb{R}$$
such that
$$\boldsymbol{v}_m = \alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_{m-1} \boldsymbol{v}_{m-1} + \beta_1 \boldsymbol{u}_1 + \beta_2 \boldsymbol{u}_2 + \cdots + \beta_{n-m+1} \boldsymbol{u}_{n-m+1}. \qquad (2.4)$$

Suppose that $\beta_1 = \beta_2 = \cdots = \beta_{n-m+1} = 0$. Then $\boldsymbol{v}_m = \alpha_1\boldsymbol{v}_1 + \alpha_2\boldsymbol{v}_2 + \cdots + \alpha_{m-1}\boldsymbol{v}_{m-1}$, which can be rewritten as

$$\alpha_1\boldsymbol{v}_1 + \alpha_2\boldsymbol{v}_2 + \cdots + \alpha_{m-1}\boldsymbol{v}_{m-1} - \boldsymbol{v}_m = \boldsymbol{o},$$

which implies that $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$ is linearly dependent, a contradiction.

Therefore, some $\beta_j \neq 0$. We may then write $\boldsymbol{u}_j$ as a linear combination of $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{m-1}$ and the remaining $\boldsymbol{u}_i$, $i \neq j$, as follows:

$$\begin{aligned}\beta_j\boldsymbol{u}_j = &-\alpha_1\boldsymbol{v}_1 - \cdots - \alpha_{m-1}\boldsymbol{v}_{m-1} + \boldsymbol{v}_m \\ &- \beta_1\boldsymbol{u}_1 - \cdots - \beta_{j-1}\boldsymbol{u}_{j-1} - \beta_{j+1}\boldsymbol{u}_{j+1} - \cdots - \beta_{n-m+1}\boldsymbol{u}_{n-m+1},\end{aligned}$$

which gives

$$\begin{aligned}\boldsymbol{u}_j = &-\frac{\alpha_1}{\beta_j}\boldsymbol{v}_1 - \cdots - \frac{\alpha_{m-1}}{\beta_j}\boldsymbol{v}_{m-1} + \frac{1}{\beta_j}\boldsymbol{v}_m - \frac{\beta_1}{\beta_j}\boldsymbol{u}_1 - \ldots \\ &- \frac{\beta_{j-1}}{\beta_j}\boldsymbol{u}_{j-1} - \frac{\beta_{j+1}}{\beta_j}\boldsymbol{u}_{j+1} - \cdots - \frac{\beta_{n-m+1}}{\beta_j}\boldsymbol{u}_{n-m+1}.\end{aligned} \tag{2.5}$$

Now remove $\boldsymbol{u}_j$ from B_1' to obtain $B' = B_1' \setminus \{\boldsymbol{u}_j\}$. Then B' contains $n - m$ vectors. It remains to prove that $S \cup B'$ is a basis of V. For this we have to show that

1. $\operatorname{lin}(S \cup B') = V$, and

2. $S \cup B'$ is linearly independent.

First we show that $\operatorname{lin}(S \cup B') = V$. Let $\boldsymbol{v} \in V$. Since $S_1 \cup B_1'$ is a basis, there exist $\lambda_1, \ldots, \lambda_{m-1}, \mu_1, \ldots, \mu_{n-m+1} \in \mathbb{R}$ such that

$$\boldsymbol{v} = \lambda_1\boldsymbol{v}_1 + \cdots + \lambda_{m-1}\boldsymbol{v}_{m-1} + \mu_1\boldsymbol{u}_1 + \cdots + \mu_{n-m+1}\boldsymbol{u}_{n-m+1}.$$

Use (2.5) to eliminate $\boldsymbol{u}_j$ in the right-hand side. It is not necessary to calculate this explicitly; it is sufficient to know that this will give $\boldsymbol{v}$ as a linear combination of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_{n-m+1}$ except for $\boldsymbol{u}_j$, and then we may conclude that $\boldsymbol{v} \in \operatorname{lin}(S \cup B')$.

Thus we have shown that $V \subset \operatorname{lin}(S \cup B')$. The opposite inclusion $\operatorname{lin}(S \cup B') \subset V$ is obvious. Thus $\operatorname{lin}(S \cup B') = V$.

Next we show that $S \cup B'$ is a linearly independent set. To do this, write $\boldsymbol{o}$ as a linear combination of the vectors in $S \cup B'$:

$$\begin{aligned}\boldsymbol{o} = &\gamma_1\boldsymbol{v}_1 + \cdots + \gamma_m\boldsymbol{v}_m + \delta_1\boldsymbol{u}_1 + \cdots + \delta_{j-1}\boldsymbol{u}_{j-1} \\ &+ \delta_{j+1}\boldsymbol{u}_{j+1} + \cdots + \delta_{n-m+1}\boldsymbol{u}_{n-m+1}.\end{aligned} \tag{2.6}$$

We may eliminate $\boldsymbol{v}_m$ by using (2.4):

$$\begin{aligned}\boldsymbol{o} &= \gamma_1\boldsymbol{v}_1 + \cdots + \gamma_m(\alpha_1\boldsymbol{v}_1 + \cdots + \alpha_{m-1}\boldsymbol{v}_{m-1} + \beta_1\boldsymbol{u}_1 + \cdots + \beta_{n-m+1}\boldsymbol{u}_{n-m+1}) \\ &\quad + \delta_1\boldsymbol{u}_1 + \cdots + \delta_{j-1}\boldsymbol{u}_{j-1} + \delta_{j+1}\boldsymbol{u}_{j+1} + \cdots + \delta_{n-m+1}\boldsymbol{u}_{n-m+1} \\ &= (\gamma_1 + \gamma_m\alpha_1)\boldsymbol{v}_1 + (\gamma_2 + \gamma_m\alpha_2)\boldsymbol{v}_2 + \cdots + (\gamma_{m-1} + \gamma_m\alpha_{m-1})\boldsymbol{v}_{m-1} \\ &\quad + (\gamma_m\beta_1 + \delta_1)\boldsymbol{u}_1 + (\gamma_m\beta_2 + \delta)2)\boldsymbol{u}_2 + \cdots + (\gamma_m\beta_{j-1} + \delta_{j-1})\boldsymbol{u}_{j-1} \\ &\quad + \gamma_m\beta_j\boldsymbol{u}_j + (\gamma_m\beta_{j+1} + \delta_{j+1})\boldsymbol{u}_{j+1} + \cdots + \delta_{n-m+1}\boldsymbol{u}_{n-m+1}.\end{aligned}$$

Since $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{m-1}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_{n-m+1}\} = S_1 \cup B_1'$ is a basis, all the coefficients in the latter linear combination equal 0:

$$\begin{aligned}
\gamma_1 + \gamma_m\alpha_1 = 0, \qquad & \gamma_2 + \gamma_m\alpha_2 = 0, \quad \ldots \qquad & \gamma_{m-1} + \gamma_m\alpha_{m-1} = 0,\\
\gamma_m\beta_1 + \delta_1 = 0, \qquad & \gamma_m\beta_2 + \delta_2 = 0, \quad \ldots \qquad & \gamma_m\beta_{j-1} + \delta_{j-1} = 0,\\
\gamma_m\beta_j = 0, & & \\
\gamma_m\beta_{j+1} + \delta_{j+1} = 0, \qquad & \ldots \qquad & \gamma_m\beta_{n-m+1} + \delta_{n-m+1} = 0.
\end{aligned}$$

Since $\beta_j \neq 0$ it follows that $\gamma_m = 0$. It then follows that all

$$\gamma_1 = \cdots = \gamma_{m-1} = 0 = \delta_1 = \cdots = \delta_{j-1} = \delta_{j+1} = \cdots = \delta_{n-m+1}.$$

Thus all the coefficients in (2.6) are equal to 0, which shows that $S \cup B'$ is linearly independent.

It follows that $S \cup B'$ is a basis, and the induction step is complete. □

Corollary 2.2.4. *Let V be a vector space and B be a basis of V such that B has n elements. Then any linearly independent set S of n vectors is also a basis of V.*

Proof. This corollary follows from Theorem 2.2.3 by taking $m = n$. Then $B' = \emptyset$ because it has $n - m = n - n = 0$ elements and $S \cup B' = S$ is a basis of V. □

Given a vector space, there are of course many bases. However, the next result says that the cardinality of the basis is unique for any given vector space.

Corollary 2.2.5. *If a vector space V has a basis with n elements, then every basis of V has the same number of elements.*

Proof. Suppose that B is a basis of V with n elements and suppose that B' is another basis of V.

Let B' have more than n elements. Then take any n distinct elements $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ from B'. From the previous corollary, it follows that these span V, and so if $\boldsymbol{v} \in B' \setminus \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$, it can be written as a linear combination of $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$, which contradicts the independence of B'.

If B' has fewer than n elements, then by interchanging the roles of B and B' and proceeding as above, we once again arrive at a contradiction. □

The above result motivates the following natural definitions.

Definitions. *Let V be a vector space.*

1. *If there exists a basis B of V such that B has n elements, then n is called the* ***dimension of*** *V, and it is denoted by* $\dim(V)$.
2. *If a vector space has a basis with a finite number of elements, then it is called a* ***finite dimensional vector space****.*
3. *If a vector space is not finite dimensional, then it is called an* ***infinite dimensional vector space****.*

Examples.

1. It follows from Example 1 on p. 90 that $\dim(\mathbb{R}^m) = m$.
2. It follows from Example 2 on p. 90 that $\dim(\{\boldsymbol{o}\}) = 0$.

2.2.4 Linear transformations

Definition. *Let U, V be two vector spaces. A function $T\colon U \to V$ is called a* ***linear transformation*** *if it satisfies the following two properties:*

L1. *For all $\boldsymbol{u}_1, \boldsymbol{u}_2 \in U$, $T(\boldsymbol{u}_1 + \boldsymbol{u}_2) = T(\boldsymbol{u}_1) + T(\boldsymbol{u}_2)$.*

L2. *For all $\boldsymbol{u} \in U$, and all $\alpha \in \mathbb{R}$, $T(\alpha \cdot \boldsymbol{u}) = \alpha \cdot T(\boldsymbol{u})$.*

Just as group homomorphisms are functions that respect group operations, linear transformations are functions that respect vector space operations.

Examples.

1. Let $m, n \in \mathbb{N}$. Let

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \in \mathbb{R}^{m \times n}.$$

Then the function $T_A\colon \mathbb{R}^n \to \mathbb{R}^m$ defined by

$$T_A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n} a_{1k}x_k \\ \vdots \\ \sum_{k=1}^{n} a_{mk}x_k \end{bmatrix} \tag{2.7}$$

for all $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$, is a linear transformation from the vector space $\mathbb{R}^n$ to the vector space $\mathbb{R}^m$. Indeed, we have

$$\begin{aligned} T_A\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}\right) &= T_A \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix} \\ &= \begin{bmatrix} \sum_{k=1}^{n} a_{1k}(x_k + y_k) \\ \vdots \\ \sum_{k=1}^{n} a_{mk}(x_k + y_k) \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} \sum_{k=1}^{n}(a_{1k}x_k + a_{1k}y_k) \\ \vdots \\ \sum_{k=1}^{n}(a_{mk}x_k + a_{mk}y_k) \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n} a_{1k}x_k + \sum_{k=1}^{n} a_{1k}y_k \\ \vdots \\ \sum_{k=1}^{n} a_{mk}x_k + \sum_{k=1}^{n} a_{mk}y_k \end{bmatrix}$$
$$= \begin{bmatrix} \sum_{k=1}^{n} a_{1k}x_k \\ \vdots \\ \sum_{k=1}^{n} a_{mk}x_k \end{bmatrix} + \begin{bmatrix} \sum_{k=1}^{n} a_{1k}y_k \\ \vdots \\ \sum_{k=1}^{n} a_{mk}y_k \end{bmatrix}$$
$$= T_A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + T_A \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},$$

for all

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n,$$

and so L1 holds. Moreover,

$$T_A\left(\alpha \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right) = T_A \begin{bmatrix} \alpha x_1 \\ \vdots \\ \alpha x_n \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n} a_{1k}\alpha x_k \\ \vdots \\ \sum_{k=1}^{n} a_{mk}\alpha x_k \end{bmatrix}$$
$$= \begin{bmatrix} \alpha \sum_{k=1}^{n} a_{1k}x_k \\ \vdots \\ \alpha \sum_{k=1}^{n} a_{mk}x_k \end{bmatrix} = \alpha \cdot \begin{bmatrix} \sum_{k=1}^{n} a_{1k}x_k \\ \vdots \\ \sum_{k=1}^{n} a_{mk}x_k \end{bmatrix} = \alpha \cdot T_A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix},$$

for all $\alpha \in \mathbb{R}$ and all vectors

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n,$$

and so L2 holds as well. Hence T_A is a linear transformation.

2. The function $T\colon C[0,1] \to \mathbb{R}$ given by

$$Tf = f\left(\frac{1}{2}\right) \text{ for all } f \in C[0,1],$$

is a linear transformation from the vector space $C[0,1]$ to the vector space $\mathbb{R}$. Indeed,

$$T(f+g) = (f+g)\left(\frac{1}{2}\right) = f\left(\frac{1}{2}\right) + g\left(\frac{1}{2}\right) = T(f) + T(g)$$

for all $f, g \in C[0,1]$, and so L1 holds. Furthermore,

$$T(\alpha \cdot f) = (\alpha \cdot f)\left(\frac{1}{2}\right) = \alpha f\left(\frac{1}{2}\right) = \alpha T(f)$$

for all $\alpha \in \mathbb{R}$ and all $f \in C[0,1]$, and so L2 holds too. Thus T is a linear transformation.

Just as in the case of homomorphisms between groups, there are two important subsets associated with a linear transformation between vector spaces, whose definitions are given below.

Definitions. *Let U, V be vector spaces and $T\colon U \to V$ a linear transformation.*

1. *The **kernel** of T is defined to be the set* $\ker(T) = \{\boldsymbol{u} \in U \mid T(\boldsymbol{u}) = \boldsymbol{o}_V\}$, *where $\boldsymbol{o}_V$ denotes the zero vector of V.*
2. *The **image** of T is defined to be the set* $\operatorname{im}(T) = \{\boldsymbol{v} \in V \mid \exists \boldsymbol{u} \in U \text{ such that } T(\boldsymbol{u}) = \boldsymbol{v}\}$.

Examples.

1. Let $A \in \mathbb{R}^{m\times n}$, and let $T_A\colon \mathbb{R}^n \to \mathbb{R}^m$ be the linear transformation defined by (2.7). The kernel of the linear transformation is the set of all vectors

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$$

such that the system of linear equations

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= 0 \end{aligned}$$

is simultaneously satisfied.

The range of T_A is the set of all vectors

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \in \mathbb{R}^m$$

such that there exists a vector

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$$

such that

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= y_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= y_m. \end{aligned}$$

2. The function $T\colon C[0,1] \to \mathbb{R}$ given by $Tf = f\left(\frac{1}{2}\right)$ for all $f \in C[0,1]$, is a linear transformation from the vector space $C[0,1]$ to the vector space $\mathbb{R}$, with kernel equal to the set of continuous functions on the interval $[0,1]$ that vanish at the point $\frac{1}{2}$. The range of T is the whole vector space $\mathbb{R}$.

Analogous to Theorem 2.1.5 for homomorphisms between groups, we now prove the following result.

Theorem 2.2.6. *Let U, V be vector spaces and $T\colon U \to V$ a linear transformation. Then*

1. $\ker(T)$ *is a subspace of U.*

2. $\mathrm{im}(T)$ *is a subspace of V.*

Proof. In each of the cases, we check that S1, S2, S3 hold.

Let $\boldsymbol{o}_U, \boldsymbol{o}_V$ denote the zero vectors in the vector spaces U, V, respectively. It is easy to check that $\ker(T)$ is a subspace of U. Indeed, as $T(\boldsymbol{o}_U) = T(\boldsymbol{o}_U + \boldsymbol{o}_U) = T(\boldsymbol{o}_U) + T(\boldsymbol{o}_U)$, it follows that $T(\boldsymbol{o}_U) = \boldsymbol{o}_V$, and so $\boldsymbol{o}_U \in \ker(T)$. Thus S1 holds. If $\boldsymbol{u}_1, \boldsymbol{u}_2$ belong to $\ker(T)$, then $T(\boldsymbol{u}_1 + \boldsymbol{u}_2) = T(\boldsymbol{u}_1) + T(\boldsymbol{u}_2) = \boldsymbol{o}_V + \boldsymbol{o}_V = \boldsymbol{o}_V$, and so S2 holds as well. Finally, S3 holds, since if $\alpha \in \mathbb{R}$ and $\boldsymbol{u} \in \ker(T)$, then $T(\alpha \cdot \boldsymbol{u}) = \alpha \cdot T(\boldsymbol{u}) = \alpha \cdot \boldsymbol{o}_V = \boldsymbol{o}_V$. Hence $\ker(T)$ is a subspace of U.

We now also check that $\mathrm{im}(T)$ is a subspace of V. Since $T(\boldsymbol{o}_U) = \boldsymbol{o}_V$, it follows that $\boldsymbol{o}_V \in \mathrm{im}(T)$, and so S1 holds. If $\boldsymbol{v}_1, \boldsymbol{v}_2$ belong to $\mathrm{im}(T)$, then there exist elements $\boldsymbol{u}_1, \boldsymbol{u}_2$ in U such that $T(\boldsymbol{u}_1) = \boldsymbol{v}_1$ and $T(\boldsymbol{u}_2) = \boldsymbol{v}_2$. Consequently, $T(\boldsymbol{u}_1 + \boldsymbol{u}_2) = T(\boldsymbol{u}_1) + T(\boldsymbol{u}_2) = \boldsymbol{v}_1 + \boldsymbol{v}_2$, and so there exists an element in U, namely $\boldsymbol{u}_1 + \boldsymbol{u}_2$, such that $T(\boldsymbol{u}_1 + \boldsymbol{u}_2) = \boldsymbol{v}_1 + \boldsymbol{v}_2$, that is, $\boldsymbol{v}_1 + \boldsymbol{v}_2 \in \mathrm{im}(T)$. Thus S2 holds. Finally, if $\alpha \in \mathbb{R}$ and $\boldsymbol{v} \in \mathrm{im}(T)$, then there exists a $\boldsymbol{u} \in U$ such that $T(\boldsymbol{u}) = \boldsymbol{v}$, and so $\alpha \cdot \boldsymbol{v} = \alpha \cdot T(\boldsymbol{u}) = T(\alpha \cdot \boldsymbol{u})$. Hence $\alpha \cdot \boldsymbol{v} \in \mathrm{im}(T)$, and S3 holds. So $\mathrm{im}(T)$ is a subspace of V. □

Definition. *Let U, V be vector spaces and $T\colon U \to V$ a linear transformation. Then*

1. *the* ***nullity*** *of T is defined to be the dimension of the kernel of T:*

$$\mathrm{nullity}(T) = \dim(\ker(T)),$$

2. *and the* ***rank*** *of T is defined to be the dimension of the image of T:*

$$\mathrm{rank}(T) = \dim(\mathrm{im}(T)).$$

Our final result shows a connection between the nullity and the rank of a linear transformation.

Theorem 2.2.7. *Let U be a finite dimensional vector space, V be a vector space, and let $T\colon U \to V$ be a linear transformation. Then:*

$$\text{nullity}(T) + \text{rank}(T) = \dim(U).$$

Proof. Let U be a vector space with dimension n and basis B, V be a vector space, and $T\colon U \to V$ be a linear transformation. By the previous theorem, $\ker(T)$ is a subspace of U, thus it is finite dimensional (see Exercise 16), with dimension $k \leq n$. Let $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$ be a basis of $\ker(T)$. By Theorem 2.2.3, there exists a subset B' of B with $n-k$ elements such that $S \cup B$ is a basis of U. Denote the elements of B' by $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-k}$. We are going to prove that $W := \{T(\boldsymbol{v}_1), \ldots, T(\boldsymbol{v}_{n-k})\}$ is a basis of $\text{im}(T)$ and, hence,

$$\dim(\ker(T)) + \dim(\text{im}(T)) = k + (n-k) = n = \dim(U).$$

First, we verify B2. Let $\alpha_1, \ldots, \alpha_{n-k} \in \mathbb{R}$ be real numbers such that

$$\alpha_1 \cdot T(\boldsymbol{v}_1) + \cdots + \alpha_{n-k} \cdot T(\boldsymbol{v}_{n-k}) = \boldsymbol{o}_V.$$

Using L1 and L2 we obtain

$$\boldsymbol{o}_V = \alpha_1 \cdot T(\boldsymbol{v}_1) + \cdots + \alpha_{n-k} \cdot T(\boldsymbol{v}_{n-k}) = T(\alpha_1 \cdot \boldsymbol{v}_1 + \cdots + \alpha_{n-k} \cdot \boldsymbol{v}_{n-k}).$$

In other words, $\alpha_1 \cdot \boldsymbol{v}_1 + \cdots + \alpha_{n-k} \cdot \boldsymbol{v}_{n-k} \in \ker(T)$. Since S is a basis of $\ker(T)$, there are real numbers $\beta_1, \ldots, \beta_k \in \mathbb{R}$ so that

$$\alpha_1 \cdot \boldsymbol{v}_1 + \cdots + \alpha_{n-k} \cdot \boldsymbol{v}_{n-k} = \beta_1 \cdot \boldsymbol{u}_1 + \cdots + \beta_k \cdot \boldsymbol{u}_k.$$

But $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-k}$ are linearly independent vectors because $S \cup B'$ is a basis of U, so we must have $\alpha_1 = \cdots = \alpha_{n-k} = \beta_1 = \cdots = \beta_k = 0$. Hence, W is a set of linearly independent vectors.

To see B1, let $\boldsymbol{v}$ be any vector in $\text{im}(T)$. So there is $\boldsymbol{u} \in U$ such that $T(\boldsymbol{u}) = \boldsymbol{v}$. Since $S \cup B'$ is a basis of U, there exist $\alpha_1, \ldots, \alpha_{n-k}, \beta_1, \ldots, \beta_k \in \mathbb{R}$ such that

$$\boldsymbol{u} = \beta_1 \cdot \boldsymbol{u}_1 + \cdots + \beta_k \cdot \boldsymbol{u}_k + \alpha_1 \cdot \boldsymbol{v}_1 + \cdots + \alpha_{n-k} \cdot \boldsymbol{v}_{n-k}.$$

From this we obtain that

$$\begin{aligned}
\boldsymbol{v} &= T(\boldsymbol{u}) \\
&= T(\beta_1 \cdot \boldsymbol{u}_1 + \cdots + \beta_k \cdot \boldsymbol{u}_k + \alpha_1 \cdot \boldsymbol{v}_1 + \cdots + \alpha_{n-k} \cdot \boldsymbol{v}_{n-k}) \\
&= \beta_1 \cdot T(\boldsymbol{u}_1) + \cdots + \beta_k \cdot T(\boldsymbol{u}_k) + \alpha_1 \cdot T(\boldsymbol{v}_1) + \cdots + \alpha_{n-k} \cdot T(\boldsymbol{v}_{n-k}).
\end{aligned}$$

Since $S \subset \ker(T)$, we have $T(\boldsymbol{u}_1) = \cdots = T(\boldsymbol{u}_k) = \boldsymbol{o}_V$, therefore,

$$\begin{aligned}
\boldsymbol{v} &= \beta_1 \cdot T(\boldsymbol{u}_1) + \cdots + \beta_k \cdot T(\boldsymbol{u}_k) + \alpha_1 \cdot T(\boldsymbol{v}_1) + \cdots + \alpha_{n-k} \cdot T(\boldsymbol{v}_{n-k}) \\
&= \beta_1 \cdot \boldsymbol{o}_V + \cdots + \beta_k \cdot \boldsymbol{o}_V + \alpha_1 \cdot T(\boldsymbol{v}_1) + \cdots + \alpha_{n-k} \cdot T(\boldsymbol{v}_{n-k}) \\
&= \alpha_1 \cdot T(\boldsymbol{v}_1) + \cdots + \alpha_{n-k} \cdot T(\boldsymbol{v}_{n-k}) \in \operatorname{lin} W.
\end{aligned}$$

Hence, $\text{im}(T) = \operatorname{lin} W$. (Why?) □

2.2.5 Exercises

1. a) Is the set of invertible 2×2 matrices having real entries with matrix addition and with scalar multiplication defined by (2.2) a vector space?

   b) Is the set of invertible 2×2 matrices having real entries with matrix multiplication and with scalar multiplication defined by (2.2) a vector space?

2. Let V be a vector space. Prove that if $\alpha \in \mathbb{R}$ and $\boldsymbol{v} \in V$ are such that $\alpha \cdot \boldsymbol{v} = \boldsymbol{o}$, then either $\alpha = 0$ or $\boldsymbol{v} = \boldsymbol{o}$.

   HINT: If $\alpha \neq 0$, how can you manipulate the equation $\alpha \cdot \boldsymbol{v} = \boldsymbol{o}$?

3. Let U and V be vector spaces, and define

$$U \times V = \{(u, v) \mid \boldsymbol{u} \in U, \boldsymbol{v} \in V\}.$$

   Show that $U \times V$, with addition defined by $(\boldsymbol{u}_1, \boldsymbol{v}_1) + (\boldsymbol{u}_2, \boldsymbol{v}_2) = (\boldsymbol{u}_1 + \boldsymbol{u}_2, \boldsymbol{v}_1 + \boldsymbol{v}_2)$ and scalar multiplication by $\alpha \cdot (\boldsymbol{u}, \boldsymbol{v}) = (\alpha \cdot \boldsymbol{u}, \alpha \cdot \boldsymbol{v})$, is a vector space.

4. Consider the set $\mathbb{R}^\infty$ of all sequences with addition defined as follows:

$$\text{if } (a_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}} \in \mathbb{R}^\infty, \text{ then } (a_n)_{n\in\mathbb{N}} + (b_n)_{n\in\mathbb{N}} = (a_n + b_n)_{n\in\mathbb{N}}, \tag{2.8}$$

   and scalar multiplication defined as follows:

$$\text{if } \alpha \in \mathbb{R} \text{ and } (a_n)_{n\in\mathbb{N}} \in \mathbb{R}^\infty, \text{ then } \alpha \cdot (a_n)_{n\in\mathbb{N}} = (\alpha a_n)_{n\in\mathbb{N}}. \tag{2.9}$$

   Prove that $\mathbb{R}^\infty$ is a vector space with the above addition and scalar multiplication.

5. Show that a subspace of a vector space is itself a vector space.

6. Determine if the following statements are TRUE or FALSE. Give reasons for your answers in each case.

   a) The union of two subspaces of a vector space V is a subspace of V.

   b) The intersection of two subspaces of a vector space V is a subspace of V.

   c) $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \in \operatorname{lin}\left(\left\{\begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right\}\right)$ in the vector space $\mathbb{R}^3$.

   d) The vectors $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}$ are linearly independent in the vector space $\mathbb{R}^3$.

   e) If $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$ are linearly independent vectors, then $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ are linearly independent.

   f) If $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$ are linearly dependent vectors, then $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ are linearly dependent.

   g) If $\boldsymbol{v}_1, \boldsymbol{v}_2$ are linearly independent, and $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ are linearly dependent, then $\boldsymbol{v}_3$ is a linear combination of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$.

h) If $\boldsymbol{v}_1, \boldsymbol{v}_2$ are linearly dependent, then there is a real number α such that $\boldsymbol{v}_1 = \alpha \boldsymbol{v}_2$.

7. Let X and Y be two subspaces of a vector space V. Define the set $X + Y$ to be

$$X + Y = \{\boldsymbol{x} + \boldsymbol{y} \mid \boldsymbol{x} \in X, \boldsymbol{y} \in Y\}.$$

Show that $X + Y$ is a subspace of V.

8. Consider the space $C[0, \pi]$ of continuous functions on $[0, \pi]$, and the four functions f_1, f_2, f_3, f_4, where $f(x) = 1$ (i.e., f_1 is the constant function 1), $f_2(x) = \cos x$, $f_3(x) = \cos 2x$ and $f_4(x) = \cos^2 x$.

a) Show that $\{f_1, f_2, f_3, f_4\}$ is linearly dependent. *You may use any properties of the trigonometric functions that you know.*

b) Show that $\{f_1, f_2, f_3\}$ is linearly independent.

HINT: If $\alpha_1 \cdot f_1 + \alpha_2 \cdot f_2 + \alpha_3 \cdot f_3 = \boldsymbol{o}$, then $\alpha_1 f_1(x) + \alpha_2 f_2(x) + \alpha_3 f_3(x) = 0$ for all $x \in [0, 1]$, and so in particular for $x = 0$, $x = \pi/2$ and $x = \pi$.

9. Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$ be linearly independent vectors in a vector space V. Let $V_1 = \text{lin}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\})$ and $V_2 = \text{lin}(\{\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\})$. Find $V_1 \cap V_2$, and justify your answer in detail.

10. a) Prove that if t_1 and t_2 are distinct real numbers in $\mathbb{R}$, then

$$\text{lin}\left(\left\{\begin{bmatrix} 1 \\ t_1 \end{bmatrix}, \begin{bmatrix} 1 \\ t_2 \end{bmatrix}\right\}\right) = \mathbb{R}^2$$

in the vector space $\mathbb{R}^2$.

b) Prove that $\mathbb{R}^2$ is not the union of a finite number of proper subspaces.

HINT: Consider the intersection of any proper subspace with the infinite subset $S = \left\{\begin{bmatrix} 1 \\ t \end{bmatrix} \,\middle|\, t \in \mathbb{R}\right\}$.

11. Let $\mathbb{R}^\infty$ be the vector space of all sequences, with addition and scalar multiplication defined by (2.8) and (2.9), respectively. We define the following subsets of $\mathbb{R}^\infty$:

a) ℓ^∞ is the set of all bounded sequences.

b) c is the set of all convergent sequences.

c) c_0 is the set of all convergent sequences with limit 0.

d) $c_{00} = \{(a_n)_{n\in\mathbb{N}} \in \mathbb{R}^\infty \mid \exists N \in \mathbb{N} \text{ such that } \forall n > N, a_n = 0\}$, is the set of all sequences that are **eventually zero**.

Prove that $c_{00} \subset c_0 \subseteq c \subseteq \ell^\infty \subseteq \mathbb{R}^\infty$, and that each is a subspace of the next one.

12. Consider the vector space $C[0, 1]$ with addition of functions and scalar multiplication defined by (2.3). Let $S(y_1, y_2) = \{f \in C[0, 1] \mid f(0) = y_1 \text{ and } f(1) = y_2\}$. Show that $S(y_1, y_2)$ is a subspace of $C[0, 1]$ iff $y_1 = 0 = y_2$.

13. Consider the vector space $C[-1,1]$ with addition of functions and scalar multiplication defined by (2.3).
    a) Let S_e be the set of all **even functions** in $C[-1,1]$:
    $$S_e = \{f \in C[-1,1] \mid \forall x \in [-1,1]\ f(x) = f(-x)\}\,.$$
    Show that S_e is a subspace of $C[-1,1]$.
    b) Is the set S_o of **odd functions**, defined as follows, also a subspace of $C[-1,1]$?
    $$S_o = \{f \in C[-1,1] \mid \forall x \in [-1,1]\ f(-x) = -f(x)\}\,.$$

14. (∗) Consider the vector space $C[0,1]$, with the usual addition and scalar multiplication. Show that the set $S = \{1, x, x^2, x^3, \dots\}$ of functions in $C[0,1]$ is linearly independent. What is $\mathrm{lin}(S)$?
    HINT: In order to show that a function f in $C[0,1]$ is not equal to the zero function, one approach is to try and show that there is some $\delta > 0$ such that $f(x) > 0$ for all x with $0 < x < \delta$.

15. Prove or disprove that
    $$B = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$
    is a basis for $\mathbb{R}^3$.

16. Prove that if B is a basis of a finite-dimensional vector space V, then every element $\boldsymbol{v} \in V$ can be written as a unique linear combination of the vectors from B.

17. Let V be a vector space, and let U be a subspace of V of dimension n. Let $\boldsymbol{v}$ be a vector in V such that $\boldsymbol{v} \notin U$. Let $W = \{\boldsymbol{u} + k\boldsymbol{v} \mid \boldsymbol{u} \in U, k \in \mathbb{R}\}$. Show that W is a subspace of V of dimension $n+1$.

18. Is the following statement true or false? Justify your answer, by means of a proof or a counterexample.
    Let V be a vector space of dimension 3. Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ be vectors in V such that neither $\boldsymbol{v}$ nor $\boldsymbol{w}$ is in $\mathrm{lin}(\{\boldsymbol{u}\})$. Then $\{\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}\}$ is a basis of V.

19. a) Suppose that S is a linearly independent set of vectors in a vector space V such that $\mathrm{lin}(S) \neq V$. Show that there is a vector $\boldsymbol{v}$ in V such that $S \cup \{\boldsymbol{v}\}$ is linearly independent.
    b) Suppose U is an infinite-dimensional vector space. Show that, for each $n \in \mathbb{N}$, U has an independent set with n elements.
    c) Suppose that V is a finite-dimensional vector space and that U is a subspace of V. Show that U does not contain an independent set with more than $\dim(V)$ elements. Deduce that U is finite-dimensional, and $\dim(U) \leq \dim(V)$.
    Show also that, if $\dim(U) = \dim(V)$, then $U = V$.

20. Show that $C[0,1]$ is infinite dimensional.

    HINT: See Exercise 13.

21. a) (∗) For $k \in \mathbb{N}$, let e_k denote the sequence with the k^{th} term equal to 1, and all other terms equal to zero:

$$e_k = (a_n)_{n\in\mathbb{N}}, \text{ where } a_n = \begin{cases} 1 & \text{if } n = k, \\ 0 & \text{if } n \neq k, \end{cases}$$

    and let $B = \{e_k \mid k \in \mathbb{N}\}$. Prove that B is a basis for the vector space c_{00}, comprising all sequences that are eventually zero.

    b) (∗) Is $B = \{e_k \mid k \in \mathbb{N}\}$ also a basis for the vector space $\mathbb{R}^\infty$?

    HINT: Consider the constant sequence $(1)_{n\in\mathbb{N}}$.

22. Find the kernel and image of the linear transformations $T_A \colon \mathbb{R}^2 \to \mathbb{R}^2$, where $A \in \mathbb{R}^{2\times 2}$ is given by:

    a) $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

    b) $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$.

    c) $A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

    Each vector $\begin{bmatrix} x \\ y \end{bmatrix}$ in $\mathbb{R}^2$ represents a point in the plane. In each of the cases, draw a picture of the subspaces $\ker(T_A)$ and $\operatorname{im}(T_A)$ in the plane.

23. Consider the vector space $\mathbb{R}^2$ with matrix addition and the usual scalar multiplication defined by (2.2). Define the function $T \colon \mathbb{R}^2 \to \mathbb{R}^2$ as follows:

$$\text{if } \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2, \text{ then } T\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{cases} \begin{bmatrix} 0 \\ x_2 \end{bmatrix} & \text{if } x_2 \neq 0, \\ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} & \text{if } x_2 = 0. \end{cases}$$

    Show that T satisfies L2, but not L1, and hence it is not a linear transformation.

24. Let $P[0,1]$ denote the space of polynomial functions on $[0,1]$, with pointwise addition and scalar multiplication. For a function $p \in P[0,1]$, with $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$, define $p^*(x) = a_1 + a_2x + \cdots + a_nx^{n-1}$.

    Show that the function $T : P[0,1] \to P[0,1]$ given by $T(p) = p^*$ is a linear transformation.

25. Let c denote the vector space of all convergent sequences. Consider the function $T\colon c \to \mathbb{R}$ given by
$$T((a_n)_{n\in\mathbb{N}}) = \lim_{n\to\infty} a_n.$$
   a) Prove that T is a linear transformation from the vector space c to the vector space $\mathbb{R}$.
   b) What is the kernel of T?
   c) Show that $\operatorname{im}(T) = \mathbb{R}$.

26. Recall the definitions of the subspaces of even functions S_e and odd functions S_o of $C[-1,1]$ in Exercise 12.

    a) Define, for each function $f \in C[-1,1]$, the function $f_e\colon [-1,1] \to \mathbb{R}$ by

    $$f_e(x) = \frac{f(x) + f(-x)}{2}, \quad x \in [-1,1].$$

    Show that $f_e \in S_e$ for each $f \in C[-1,1]$.

    b) Define the function $T\colon C[-1,1] \to C[-1,1]$ by $T(f) = f_e$. Show that T is a linear transformation.

    c) Find $\ker(T)$ and $\operatorname{im}(T)$.

27. In this exercise, the associativity of matrix multiplication $A(BC) = (AB)C$ is proved using linear transformations.

    a) Show that for any linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^m$ there is a unique $m \times n$ matrix A such that $T(\boldsymbol{x}) = A\boldsymbol{x}$ for all $\boldsymbol{x} \in \mathbb{R}^n$.

    HINT: If $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ is the standard basis of $\mathbb{R}^n$, then explain why the only possibility for A is

    $$A = \begin{bmatrix} \vdots & \vdots & \cdots & \vdots \\ T(\boldsymbol{e}_1) & T(\boldsymbol{e}_2) & \cdots & T(\boldsymbol{e}_n) \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix}.$$

    b) Let $T_1\colon \mathbb{R}^n \to \mathbb{R}^m$ and $T_2\colon \mathbb{R}^k \to \mathbb{R}^n$ be linear transformations. Denote the matrices corresponding to T_1 and T_2 by A_1 and A_2, respectively.

        i) Show that the composition $T_1 \circ T_2\colon \mathbb{R}^k \to \mathbb{R}^m$, which is defined (in the usual way) by $(T_1 \circ T_2)(\boldsymbol{v}) = T_1(T_2(\boldsymbol{v}))$ for all $\boldsymbol{v} \in \mathbb{R}^k$, is also a linear transformation.

        ii) Show that the matrix that corresponds to $T_1 \circ T_2$ is the product A_1A_2.

    c) Let $T_1\colon \mathbb{R}^n \to \mathbb{R}^m$, $T_2\colon \mathbb{R}^k \to \mathbb{R}^n$ and $T_3\colon \mathbb{R}^p \to \mathbb{R}^k$ be linear transformations.

        i) Show that $(T_1 \circ T_2) \circ T_3$ and $T_1 \circ (T_2 \circ T_3)$ are the same transformations.

        ii) Denote the matrices corresponding to T_1, T_2 and T_3 by A_1, A_2, A_3, respectively. Show that the matrix corresponding to $(T_1 \circ T_2) \circ T_3$ is $(A_1A_2)A_3$ and the matrix corresponding to $T_1 \circ (T_2 \circ T_3)$ is $A_1(A_2A_3)$.

        Conclude that $(A_1A_2)A_3 = A_1(A_2A_3)$ for any $m \times n$ matrix A_1, $n \times k$ matrix A_2 and $k \times p$ matrix A_3.

    d) Suppose that U, V and W are vector spaces, and that $T : U \to V$ and $S : V \to W$ are linear transformations. Consider the composition $R : U \to W$ given by $R(\boldsymbol{u}) = S(T(\boldsymbol{u}))$.

        a) Show that R is a linear transformation.

        b) Show that $\operatorname{im}(R) \subseteq \operatorname{im}(S)$ and that $\ker(T) \subseteq \ker(S)$. Deduce that $r(R) \leq \min(r(S), r(T))$.

Exam Question. 2010 Q8 (a) Define the following terms: *linear transformation*, *kernel* and *image* of a linear transformation.

(b) Define what it means for vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots \boldsymbol{v}_k$ to be *linearly independent*.

Let V and W be two vector spaces, $T : V \to W$ a linear transformation, and $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ vectors in V. Suppose that $T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \ldots, T(\boldsymbol{v}_k)$ are linearly independent vectors in W.

Show that $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ are linearly independent in V.

(c) Consider a vector space V with basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$. Let $T : V \to V$ be a linear transformation such that

$$T(\boldsymbol{v}_1) = \boldsymbol{v}_1 - \boldsymbol{v}_2; \quad T(\boldsymbol{v}_1 + \boldsymbol{v}_2) = \boldsymbol{v}_2 - \boldsymbol{v}_1. \qquad (*)$$

Explain why there is a unique linear transformation that satisfies (*).

Find $\ker(T)$ and $\mathrm{im}(T)$.