$(\mathfrak{X}, \varrho)$

Metric spaces

Kalle Kytölä

Contents

Foreword

These lecture notes are primarily intended for the core B.Sc. level mathematics course **MS-E1541** *Metric Spaces* at Aalto University.

The structure of these notes is largely based on an earlier version of the course taught by Pekka Alestalo, and parts of the textbook [**Väi99**]. The notes are still in an entirely preliminary and incomplete form, and I plan to frequently update them during the current course. You could help me — and perhaps more importantly the students who will use this material — by sending comments about mistakes, misprints, needs for clarification, etc., to me via the course's Zulip chat forum or by email (`kalle.kytola@aalto.fi`).

Lecture I

Foundations: set theory and logic

Mathematics, as it has been practiced especially in the 20th and 21st centuries, is almost entirely founded on the notions of *sets* and their *elements*. To illustrate the expressive power of set theory we mention that, e.g., the notion of a function can be reduced to that of sets[1], and constructions of real numbers[2] and even of natural numbers[3] can be given based on sets. Such flexibility of set theory justifies taking it as the foundations of mathematics.

Historical reasons for looking for solid foundations of this type include the discovery of numerous apparent paradoxes, questions about the validity of certain arguments about numbers and functions, and even disagreements about what can be considered legitimate topics of mathematics in the first place. The late 19th century and early 20th century featured particularly notable controversies of this kind, and lead to the adoption of set theory as the (more or less) universally agreed on foundations. Later in this course, we will take a look at a few of those examples that had been considered controversial before the foundations were clarified: for example the Cantor set and the Weierstrass function, and countable and uncountable infinities.

A more general perspective may be that being precise about the foundations became necessary when the level of abstraction increased sufficiently so that one could no longer rely on too informal or intuitive reasoning. The historical rationale therefore has its parallel in mathematical studies: there comes a point at which the level of abstraction requires the student to make peace with foundational questions. Given that mathematics is founded on set theory, a student of mathematics must become familiar and comfortable with set theoretical notation and reasoning. To be fair, in this course we use set theory in a rather boring way as a language which is necessary for precise definitions and statements. Set theory and foundations of mathematics themselves lead to fascinating questions, but we will not pursue such directions here.

As a pragmatic motivation for the student, set theory forms the common language in which virtually all standard topics in university level mathematics are phrased: measure and integration, probability theory, group theory and algebra, differential geometry and algebraic geometry, real analysis, complex analysis, functional analysis, etc. etc. The topic of the present course, in fact, is even often labeled as *point set topology* (as opposed to *algebraic topology*). To unravel this terminology a bit for context: the expression *an element of a set* is used exactly parallel to the expression

[1] A function *is* its "graph", a subset of the Cartesian product of the domain and codomain of the function. For more details, we refer to more serious treatments of set theory, e.g. [**Hal74**].

[2] One noteworthy construction of the real numbers is as Dedekind cuts of the set of rational numbers, see for example Wikipedia `https://en.wikipedia.org/wiki/Dedekind_cut`.

[3] Indeed, in a foundational crisis, mathematicians felt the need to *construct* natural numbers 1, 2, 3, . . . , instead of assuming their existence! Again we refer to serious treatments of set theory about the details.

a point in a space; with the only difference that a space is a set that usually has also some additional structure besides merely being a set.[4]

I.1. Basics of set theory

Sets and elements

Sets consist of elements — a **set** is the collection of all **elements** that belong to it. It is customary to denote sets by upper case letters and their elements by lower case letters, as we mostly do also below. However, this is merely a typographical convention commonly followed because it has the potential to help the reader, but it will be discarded when appropriate.[5]

We denote $a \in A$, if a is an element of a set A. If a is not an element of A, we denote $a \notin A$. The *binary relation* denoted by the symbol $\in$ (whose negation is denoted by the symbol $\notin$) is at the heart of set theory. In natural language, "$\in$" can be read as "to belong to"[6], for example

$$\begin{aligned} a \in A \quad &: \quad \text{(the element) } a \text{ belongs to (the set) } A, \\ a \notin A \quad &: \quad \text{(the element) } a \text{ does not belong to (the set) } A. \end{aligned} \tag{I.1}$$

An often used informal (not exactly well-defined) example of a set is the set of all fruits, which has as elements apples, kiwis, mangos, etc. More mathematical examples of sets include

- the set $\mathbb{Z}$ of all integers, having $\ldots, -2, -1, 0, -1, 2, \ldots$ as its elements;
- the open interval $(-\pi, \pi)$, whose elements are the real numbers x such that $-\pi < x < \pi$;
- the set of all continuous real-valued functions on the real line;
- the set of all subsets of the plane $\mathbb{R}^2$;
- $\ldots$.

In particular sets may have numbers as their elements (probably a familiar case already), but they may also have much more general types of elements: functions, other sets, etc. etc.

Remark I.1 (What does it take to specify a set?)**.**
A set is known when its elements are known — i.e., when for every possible object we are able to decide whether the object is an element of the set or not.

In particular two sets A and B are equal, denoted $A = B$, if they have exactly the same elements.

A few more examples of sets are:

[4]In fact, we will not precisely define the notions of *a point* and *a space* as such, but rather we use them as synonymous to *an element* and *a set* whenever we wish to draw attention to the fact that the set is furnished with some structure. Various specific structures will be defined precisely later on: a vector space, and inner product space, a normed space, a metric space, a topological space, ... — but the general term *space* will not be used in a definite mathematical sense.

[5]Indeed, for example for sets whose elements themselves are sets, it becomes impossible to strictly adhere to such a typographical "rule".

[6]Suomeksi "$\in$" luetaan "kuuluu (joukkoon)".

- the **empty set** $\emptyset$ is the set which does not contain any elements;
- a **singleton** $\{a\}$ is the set which contains a single element, a.

Remark I.2 (Do not confuse a singleton with its element!).
Note that

$$a \in \{a\}, \qquad \text{but} \qquad a \neq \{a\}.$$

Exercise I.1 (Empty set vs. the set consisting of the empty set, etc.).
How many elements are in the following sets:

$$\text{(a):} \quad \emptyset, \qquad \text{(b):} \quad \{\emptyset\}, \qquad \text{(c):} \quad \{\{\emptyset\}\} \quad ?$$

The simplest method of specifying a set is to list its elements:

- the notation $\{a_1, a_2, \ldots, a_n\}$ stands for the set consisting of $a_1, a_2, \ldots, a_n$.

Note that we have, e.g., $\{9, 1, 1\} = \{1, 9\}$ (this set has two elements: 1 and 9). A set is only the collection of elements belonging to it; the order of elements has no significance, and repetition is redundant.

Another common method of specifying a set is to use a logical condition as a criterion for whether an element belongs to the set:

- if A is a set and $P(x)$ is a logical proposition depending on a variable x, then the notation
$$\big\{x \in A \;\big|\; P(x)\big\} \tag{I.2}$$
stands for the set consisting of those elements x of the set A for which $P(x)$ is true.

Example I.3 (Examples of sets defined by a condition).
In the following examples, various conditions are used to "extract" from the set $\mathbb{R}$ of all real numbers a subset of those that satisfy the condition in question:

$$\begin{aligned}
\Big\{x \in \mathbb{R} \;\Big|\; \sin(x) = 0\Big\} &= \{\ldots, -2\pi, -\pi, 0, \pi, 2\pi, \ldots\} \\
\left\{x \in \mathbb{R} \;\middle|\; x^4 = \frac{1}{3}\right\} &= \left\{\frac{-1}{\sqrt{\sqrt{3}}}, \frac{1}{\sqrt{\sqrt{3}}}\right\} \\
\Big\{x \in \mathbb{R} \;\Big|\; x^2 < 0\Big\} &= \emptyset.
\end{aligned}$$

We will also take some liberties to modify the above notational conventions in what we hope are reasonably self-explanatory ways, such as:

$$\begin{aligned}
\Big\{a_j \;\Big|\; j \in \{1, 2, \ldots, n\}\Big\} &= \{a_1, a_2, \ldots, a_n\}, \\
\big\{\sqrt{x} \;\big|\; x \in \mathbb{N}\big\} &= \Big\{1, \sqrt{2}, \sqrt{3}, 2, \sqrt{5}, \ldots\Big\}, \qquad \text{etc.}
\end{aligned}$$

Subsets

A set A is said to be a **subset** of another set B if every element of A is also an element of B; we then denote $A \subset B$. Being a subset is a binary relation among sets, denoted by the symbol $\subset$. Its negation is denoted by $\not\subset$, so $A \not\subset B$ means that there exists at least one element in the set A which is not an element of B.

Note the difference between the relations $\in$ and $\subset$.[7]

In natural language, the subset symbol $\subset$ could be read as[8]

$$A \subset B \qquad : \qquad \text{(the set) } A \text{ is contained in (the set) } B, \tag{I.3}$$

or simply as: A is a subset of B. For precise and unambiguous mathematical meaning, it is best to avoid mixing the natural language expressions "belongs to" and "is contained in", which stand for the relations $\in$ and $\subset$, respectively.

Evidently, the subset relation is **transitive**: if $A \subset B$ and $B \subset C$, then $A \subset C$.

Inverted relation symbols

When it is more appropriate to mention, e.g., a set before its element or a set before its subset, the inverted relation symbols $\ni$ and $\supset$ are used so that

$$\begin{aligned} A \ni a \qquad &\text{means} \qquad a \in A, \\ B \supset A \qquad &\text{means} \qquad A \subset B. \end{aligned}$$

The relation $\supset$ can be read as "to contain", so that

$$B \supset A \qquad : \qquad \text{(the set) } B \text{ contains (the set) } A,$$

Remark I.4 (Equality of two sets)**.**

Two sets A and B are equal if they contain exactly the same elements, which occurs if and only if both are subsets of the other: $A \subset B$ and $B \subset A$.

While this may initially seem like a useless remark, it lends itself to a strategy of proof that is very commonly used. To *prove* that $A = B$, it is often practical to separately show $A \subset B$ and $A \supset B$, by first arguing that any element of A must necessarily be also an element of B, and then vice versa.

Familiar sets of numbers

The following examples of sets of numbers should be familiar:

the set of natural numbers	$\mathbb{N} = \{0, 1, 2, 3, \ldots\}$	(I.4)
the set of integers	$\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$	(I.5)
the set of rational numbers	$\mathbb{Q} = \left\{ \frac{n}{m} \;\middle|\; n \in \mathbb{Z},\ m \in \mathbb{N} \right\}$	(I.6)
the set of real numbers[9]	$\mathbb{R}$.	(I.7)

[7] If you have some experience in programming with *Python* or *Mathematica* (or other programming languages which encourage extensive use of nested lists), you are probably familiar with bugs which occur if you mix up a test of whether one list (perhaps containing just one element) is an element rather than a sublist of a given list. A mathematical equivalent of such a programming error is mixing up the relations "$\in$" and "$\subset$". Avoid that bug!

[8] Suomeksi "$\subset$" voidaan lukea "sisältyy" — erotuksena relaatiosta "$\in$", joka luettiin "kuuluu". Toisinaan selkeyden vuoksi on kuitenkin turvallisinta käyttää ilmaisua "on osajoukko".

To give just a few examples, we have

$$2021 \in \mathbb{N}, \qquad -42 \in \mathbb{Z}, \qquad \sqrt{2} \in \mathbb{R},$$
$$0 \notin \mathbb{N}, \qquad -\frac{5}{7} \in \mathbb{Q}, \qquad \sqrt{2} \notin \mathbb{Q}.$$

The last example is classical: it was known to the ancient Greeks.[10] We provide it here as one of our first examples of a careful mathematical proof. It is also a great illustration of the method of proof by contradiction.

Proposition I.5 (Square root of two is irrational)**.**
We have $\sqrt{2} \notin \mathbb{Q}$*.*

Proof. Suppose, by contrary, that $\sqrt{2} \in \mathbb{Q}$. In that case we can write $\sqrt{2} = \frac{n}{m}$ with $n, m \in \mathbb{N}$. If the integers n and m have common factors, they may be cancelled, so we may assume that m and n here are chosen coprime (i.e., no integer greater than one divides both).

Multiplying by m, we obtain that $\sqrt{2}m = n$. Then squaring both sides we see that $2m^2 = n^2$. This shows that n^2 is even, which is only possible if n is even, i.e., $n = 2k$ for some $k \in \mathbb{N}$. But in this case we find $m^2 = \frac{1}{2}n^2 = \frac{1}{2}(2k)^2 = 2k^2$, which similarly implies that m^2 is even, and thus also m is even. Therefore 2 divides both n and m, which contradicts the choice of these coprime integers.

We conclude that $\sqrt{2}$ could not have been rational in the first place, so $\sqrt{2} \notin \mathbb{Q}$. □

Exercise I.2 (★⋆Irrationality of π)**.**
Prove that $\pi \notin \mathbb{Q}$.

From the above sets of numbers we obtain examples of subsets, for example $\mathbb{N} \subset \mathbb{Z}$, $\mathbb{Z} \subset \mathbb{Q}$, and $\mathbb{Q} \subset \mathbb{R}$; or more concisely

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

By transitivity, this gives further subset relations such as $\mathbb{N} \subset \mathbb{Q}$, $\mathbb{Z} \subset \mathbb{R}$, and $\mathbb{N} \subset \mathbb{R}$.

Intervals

Intervals of various types are extremely frequently used subsets of the real line $\mathbb{R}$. For intervals between points $a, b \in \mathbb{R}$, we use the following notation and terminology:

closed interval $$[a,b] = \left\{ x \in \mathbb{R} \;\middle|\; a \leq x \leq b \right\} \tag{I.8}$$

(bounded) **open interval** $$(a,b) = \left\{ x \in \mathbb{R} \;\middle|\; a < x < b \right\} \tag{I.9}$$

$$(a,b] = \left\{ x \in \mathbb{R} \;\middle|\; a < x \leq b \right\} \tag{I.10}$$

$$[a,b) = \left\{ x \in \mathbb{R} \;\middle|\; a \leq x < b \right\}. \tag{I.11}$$

[10] The (cult of) Pythagoreans held it that only rational numbers were of divine origin, and a (probably dubious) legend has it that they killed Hippasus for revealing the irrationality of the length ($\sqrt{2}$) of the diagonal of the unit square to the outside world. As for written records, a proof of this statement is contained in Euclid's "Elements" — almost surely the mathematical textbook that has endured the longest time in bestseller lists (few others are measured in millenia).

We also use the following notation for unbounded intervals,

$$(-\infty, b] = \big\{x \in \mathbb{R} \;\big|\; x \leq b\big\}, \qquad [a, +\infty) = \big\{x \in \mathbb{R} \;\big|\; a \leq x\big\}, \tag{I.12}$$

$$(-\infty, b) = \big\{x \in \mathbb{R} \;\big|\; x < b\big\}, \qquad (a, +\infty) = \big\{x \in \mathbb{R} \;\big|\; a < x\big\}, \tag{I.13}$$

and sometimes $(-\infty, +\infty) = \mathbb{R}$ is used the whole real axis. In addition to the bounded open intervals (I.9), also the unbounded open intervals (I.13) are considered **open intervals**. By contrast, there are no unbounded closed intervals.

Operations with sets

New sets can be formed from old ones by set theoretic operations. For example, for two sets A and B, we define

$$\text{the } \textbf{union} \text{ of } A \text{ and } B: \quad A \cup B = \big\{x \;\big|\; x \in A \text{ or } x \in B\big\} \tag{I.14}$$

$$\text{the } \textbf{intersection} \text{ of } A \text{ and } B: \quad A \cap B = \big\{x \;\big|\; x \in A \text{ and } x \in B\big\} \tag{I.15}$$

$$\text{the } \textbf{set difference} \text{ of } A \text{ and } B: \quad A \setminus B = \big\{x \;\big|\; x \in A \text{ and } x \notin B\big\}. \tag{I.16}$$

If $B \subset A$, then $A \setminus B$ is also called the **complement** of B in A.

Example I.6 (A few examples of set operations)**.**
If we let

$$A = \{1, 2, 3, 4, 5, 6, \ldots\} \qquad \text{(natural numbers)}$$
$$B = \{\ldots, -6, -4, -2, 0, 2, 4, 6, \ldots\} \qquad \text{(even integers)},$$

then the intersection, union, and set differences are

$$A \cap B = \{2, 4, 6, \ldots\}, \qquad A \setminus B = \{1, 3, 5, \ldots\},$$
$$A \cup B = \{\ldots, -6, -4, -2, 0, 1, 2, 3, 4, \ldots\}, \qquad B \setminus A = \{\ldots, -6, -4, -2, 0\}.$$

The set operations follow more or less obvious rules of calculation: for any sets A, B, C we have

$$\begin{cases} A \cup B = B \cup A \\ A \cap B = B \cap A \end{cases} \qquad \text{(commutativity)} \tag{I.17}$$

$$\begin{cases} A \cup (B \cup C) = (A \cup B) \cup C \\ A \cap (B \cap C) = (A \cap B) \cap C \end{cases} \qquad \text{(associativity)} \tag{I.18}$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \qquad \text{(distributivity)} \tag{I.19}$$

$$\begin{cases} A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C) \\ A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C). \end{cases} \qquad \text{(De Morgan's laws)} \tag{I.20}$$

The properties (I.17), at least, should be obvious from the definitions (I.14) and (I.15). We leave it as an exercise to verify most of the remaining properties, but as an example we prove the first De Morgan's law below.

Proof of the first De Morgan's law. For all of the statements, the most straightforward proof strategy is the following: look at what condition characterizes an element of the set on the left hand side according to the definitions, and then successively rewrite equivalent versions of this condition until we reach the defining condition for elements on the set on the right hand side. In the case of the first De Morgan's law, the chain of equivalent conditions is

$$\begin{aligned}
x \in A \setminus (B \cup C) \quad &\iff \quad x \in A \quad \text{and} \quad x \notin B \cup C \\
&\iff \quad x \in A \quad \text{and} \quad \big(x \notin B \ \text{ and } \ x \notin C\big) \\
&\iff \quad \big(x \in A \ \text{ and } \ x \notin B\big) \quad \text{and} \quad \big(x \in A \ \text{ and } \ x \notin C\big) \\
&\iff \quad x \in (A \setminus B) \cap (A \setminus C).
\end{aligned}$$

Thus an element x belongs to the set $A \setminus (B \cup C)$ if and only if it belongs to the set $(A \setminus B) \cap (A \setminus C)$. This is the asserted equality of these two sets. □

Exercise I.3 (Proofs of set theoretic identities)**.**
Prove the remaining formulas among (I.17) – (I.20).

We can also form unions and intersections of more than two sets — in fact of arbitrary collections of sets. By an indexed collection $(A_j)_{j \in J}$ of sets, we mean that a set A_j is given for each index $j \in J$ (the set J is called the index set of the collection). The **union** of the collection is

$$\bigcup_{j \in J} A_j = \Big\{ x \; \Big| \; x \in A_j \text{ for some } j \in J \Big\}. \tag{I.21}$$

The **intersection** is defined if the collection is non-empty, i.e. if $J \neq \emptyset$, and then it is

$$\bigcap_{j \in J} A_j = \Big\{ x \; \Big| \; x \in A_j \text{ for all } j \in J \Big\}. \tag{I.22}$$

Unions and intersections of two sets, (I.14) – (I.15), are recovered from (I.21) – (I.22) in the special case that the index set has two elements, for example $J = \{1, 2\}$.

When the index set is finite, in particular in the case $J = \{1, 2, \ldots, n\}$, as alternative notations we often use

$$\begin{aligned}
\bigcup_{j \in \{1,\ldots,n\}} A_j \; &= \; \bigcup_{j=1}^{n} A_j \; = \; A_1 \cup \cdots \cup A_n \\
\bigcap_{j \in \{1,\ldots,n\}} A_j \; &= \; \bigcap_{j=1}^{n} A_j \; = \; A_1 \cap \cdots \cap A_n.
\end{aligned}$$

When the index set is $J = \mathbb{N}$, it is conventional to use the alternative notations

$$\bigcup_{j \in \mathbb{N}} A_j = \bigcup_{j=1}^{\infty} A_j \qquad \text{and} \qquad \bigcap_{j \in \mathbb{N}} A_j = \bigcap_{j=1}^{\infty} A_j.$$

Example I.7 (Union and intersection of nested intervals)**.**
Consider the example $A_n = \big[\frac{-1}{n}, \frac{1}{n}\big]$ for $n \in \mathbb{N}$. Then the union and intersection of the collection $(A_n)_{n \in \mathbb{N}}$ are

$$\begin{aligned}
\bigcup_{n=1}^{\infty} \Big[\frac{-1}{n}, \frac{1}{n}\Big] &= [-1, 1] \qquad\qquad \text{(closed interval)} \\
\bigcap_{n=1}^{\infty} \Big[\frac{-1}{n}, \frac{1}{n}\Big] &= \{0\} \qquad\qquad \text{(singleton).}
\end{aligned}$$

The first of the above claims is easy, but let us justify the second claim in detail. Suppose that $x \in \bigcap_{n=1}^{\infty} \big[\frac{-1}{n}, \frac{1}{n}\big]$. According to the definition of an intersection, this means that

$x \in \left[\frac{-1}{n}, \frac{1}{n}\right]$ for all $n \in \mathbb{N}$. At the very least x must then be a real number. But note that if $x \neq 0$, then by choosing a natural number $n_0 > \frac{1}{|x|}$, we have $|x| > \frac{1}{n_0}$, and so $x \notin \left[\frac{-1}{n_0}, \frac{1}{n_0}\right]$. This shows that a non-zero real number $x \neq 0$ can not belong to all of the sets $\left[\frac{-1}{n}, \frac{1}{n}\right]$, $n \in \mathbb{N}$, and thus does not belong to the intersection. On the other hand, $x = 0$ belongs to each of the sets: $0 \in \left[\frac{-1}{n}, \frac{1}{n}\right]$ for all $n \in \mathbb{N}$. By definition of the intersection, then, $0 \in \bigcap_{n=1}^{\infty} \left[\frac{-1}{n}, \frac{1}{n}\right]$. We have thus showed that the intersection consists of the single element 0, i.e., $\bigcap_{n=1}^{\infty} \left[\frac{-1}{n}, \frac{1}{n}\right] = \{0\}$ as claimed.

Example I.8 (Another union and intersection of nested intervals)**.**
Consider now the example $B_n = \left(0, \frac{1}{n}\right)$ for $n \in \mathbb{N}$. Then the union and intersection of the collection $(B_n)_{n\in\mathbb{N}}$ are

$$\bigcup_{n=1}^{\infty} \left(0, \frac{1}{n}\right) = (0,1) \qquad \text{(open interval)}$$
$$\bigcap_{n=1}^{\infty} \left(0, \frac{1}{n}\right) = \emptyset \qquad \text{(empty set).}$$

The precise justifications of these claims are left as an exercise.

Exercise I.4 (Details of Example I.8)**.**
Provide careful reasoning to justify the claims made in Example I.8.

Example I.9 (Yet another union and intersection of nested intervals)**.**
Consider now the example $C_n = \left[0, 1 - \frac{1}{2n}\right]$ for $n \in \mathbb{N}$. Then the union and intersection of the collection $(C_n)_{n\in\mathbb{N}}$ are

$$\bigcup_{n=1}^{\infty} \left[0, 1 - \frac{1}{2n}\right] = [0,1) \qquad \text{(half-open interval)}$$
$$\bigcap_{n=1}^{\infty} \left[0, 1 - \frac{1}{2n}\right] = \left[0, \frac{1}{2}\right] \qquad \text{(closed interval).}$$

Note in particular that 1 does not belong to the union. The precise justifications are again left as an exercise.

Exercise I.5 (Details of Example I.9)**.**
Provide careful reasoning to justify the claims made in Example I.9.

Cartesian products

Another operation of set theory is forming Cartesian products of sets.

If A and B are sets, then their **Cartesian product** $A \times B$ is the set whose elements are **ordered pairs** (a, b) whose first member belongs to the former set, $a \in A$, and second member to the latter set, $b \in B$. In symbols, the Cartesian product is

$$A \times B = \left\{ (a,b) \;\middle|\; a \in A,\; b \in B \right\}. \tag{I.23}$$

A familiar example is the plane, which (as a set) is the Cartesian product of the real line $\mathbb{R}$ with itself:

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \left\{ (x,y) \;\middle|\; x \in \mathbb{R},\; y \in \mathbb{R} \right\}.$$

Another familiar example is the rectangle, which (as a set) is the Cartesian product of two closed intervals

$$[\alpha,\beta]\times[\gamma,\delta]=\Big\{(x,y)\;\Big|\;\alpha\le x\le\beta,\ \gamma\le y\le\delta\Big\}.$$

Cartesian products of three sets are defined similarly as ordered triples $A\times B\times C=\big\{(a,b,c)\;\big|\;a\in A,\ b\in B,\ c\in C\big\}$, etc. The familiar n-dimensional space is an n-fold Cartesian product of the real line with itself

$$\mathbb{R}^n=\underbrace{\mathbb{R}\times\cdots\times\mathbb{R}}_{n\text{ times}}=\Big\{(x_1,\ldots,x_n)\;\Big|\;x_1,\ldots,x_n\in\mathbb{R}\Big\}.$$

The elements of $\mathbb{R}^n$ are ordered n-tuples[11] of real numbers, consisting of the n coordinates of a point in this space.

One can consider Cartesian products of arbitrary collections of sets, but in this course the finite Cartesian products, such as the ones above, will be quite sufficient.

I.2. Functions

What is a function, precisely? According to the familiar idea, a function f associates a (unique) value $f(x)$ to each argument x. For a precise meaning, we must specify what is the set of acceptable arguments, and what is the set of possible values. The mathematical terms for these two sets are the **domain** of the function and the **codomain** (sometimes also **range**[12]) of the function, respectively.

The notation

$$f\colon X\to Y$$

indicates that f is a **function**, whose **domain** is a set X, and whose **codomain** is a set Y, i.e., that f assigns a unique value $f(x)\in Y$ to each $x\in X$. The notation

$$x\mapsto f(x)\qquad\text{(read: “}x\text{ maps to }f(x)\text{”)}$$

may be used to emphasize this assignment (but the domain X and the codomain Y must also be specified or clear from the context). We may view this as a mapping from the points x of the domain X to points of the codomain Y, and the term **mapping** is considered synonymous to the term **function**. Functions are often defined by giving a formula (but also any other way that ensures that a unique value is assigned to each argument is acceptable).

[11] One uses the terms **pair**, **triple**, **quadruple**, **quintuple**, **sextuple**, . . . for ordered collections of two, three, four, five, six, . . . elements. For general n it has become conventional to refer to an ordered collection of n elements as an n-**tuple**.

[12] Beware, however, as the term **range** is also commonly used to mean the set of actually attained values, which may be smaller than the set of possible values that we are a priori willing to consider.

(The terms *domain* and *codomain* are not as illuminating as I would like, but they are well-established mathematical terminology. Were they not, I would advocate for using *source* and *target* instead. The Finnish terms ”lähtö”/”lähtöjoukko” and ”maali”/”maalijoukko” are in my opinion much better!)

Example I.10 (An example of a function).
We may define a function

$$r\colon \mathbb{R}^3 \to [0,+\infty)$$

by the formula

$$r((x,y,z)) = \sqrt{x^2+y^2+z^2} \qquad \text{for } (x,y,z) \in \mathbb{R}^3.$$

The domain of this function r is (by definition) $\mathbb{R}^3$, and the codomain is (by definition) $[0,+\infty)$. We must only make sure that for each $(x,y,z) \in \mathbb{R}^3$, the formula above is meaningful (as it is), and gives a value in the set $[0,+\infty)$ (as it does).

Often giving a name or a symbol for a function is purposeful: the letters $f, g, \ldots$ are most commonly used, r was chosen in Example I.10, and you have without a doubt seen also

$$\log\colon (0,\infty) \to \mathbb{R}, \qquad \sin\colon [0,2\pi) \to [-1,1], \qquad \text{etc.}$$

Occasionally, it is more meaningful to talk of functions $X \to Y$ (i.e., functions with domain X and codomain Y) without explicitly naming them. The "maps to" notation "$\mapsto$" may then be particularly convenient.

Example I.11 (A few more examples of functions).

- The **exponential function** $\mathbb{R} \to (0,+\infty)$ is given by the formula $x \mapsto e^x$.
- The formula $\theta \mapsto (\cos(\theta), \sin(\theta))$ defines a function $(-\pi,+\pi] \to \mathbb{R}^2$.

Sometimes we simply consider established symbols as the names of functions, and indicate by a "dot" at which slot the argument is to be inserted to get the value.

Example I.12 (Yet a few more examples of functions).

- The **square root function** $\sqrt{\cdot}\colon [0,+\infty) \to [0,+\infty)$ is given by $x \mapsto \sqrt{x}$.
- The **absolute value function** $|\cdot|\colon \mathbb{R} \to [0,+\infty)$ is given by $x \mapsto |x|$.

Note that functions defined by the same formula are not the same functions unless also their domains and ranges are the same!

Example I.13 (Different functions given by the same formula).
The following functions are given by the rule $x \mapsto x^2$:

- $f_1\colon \mathbb{R} \to \mathbb{R}$ given by $f_1(x) = x^2$,
- $f_2\colon \mathbb{R} \to [0,+\infty)$ given by $f_2(x) = x^2$,
- $f_3\colon [0,+\infty) \to \mathbb{R}$ given by $f_3(x) = x^2$,
- $f_4\colon (-\infty,0] \to \mathbb{R}$ given by $f_4(x) = x^2$,
- $f_5\colon (-\infty,0] \to [0,+\infty)$ given by $f_5(x) = x^2$.

Nevertheless, each of the above is a different function: $f_1 \neq f_2$, $f_3 \neq f_5$, etc.

The examples we have used so far should appear unintimidating. The concept of a function is nevertheless very general: the domain and range of a function may be arbitrary sets. We will in particular encounter functions whose arguments are functions (i.e., the domain is some set of functions), functions whose values are sets (i.e., the codomain is some set whose elements are sets), etc. Be forewarned.

Surjective, injective, and bijective functions

A few properties of functions are relevant.

Definition I.14 (Surjectivity, injectivity, and bijectivity).

Let $f : X \to Y$ be a function.

We say that f is **surjective** if for every $y \in Y$ there exists an $x \in X$ such that $y = f(x)$.

We say that f is **injective** if for any $x_1, x_2 \in X$ which are different, $x_1 \neq x_2$, also the corresponding function values are different, $f(x_1) \neq f(x_2)$.

We say that f is **bijective** if it is both surjective and injective.

Consider a function $f : X \to Y$ and a given $y \in Y$. If f is injective, there can be at most one $x \in X$ such that $f(x) = y$ (since for any $x' \neq x$ we have $f(x') \neq f(x) = y$). For a surjective function there always exists at least one such x. For bijective functions, therefore, there exists exactly one such x, i.e., the condition $y = f(x)$ gives rise to a well-defined mapping $y \mapsto x$, the **inverse function**

$$f^{-1} : Y \to X.$$

Exercise I.6 (Examples of injectivity and surjectivity).

Consider the functions

$$f_1 : \mathbb{R} \to \mathbb{R}, \qquad f_2 : \mathbb{R} \to [0, +\infty), \qquad f_3 : [0, +\infty) \to \mathbb{R},$$
$$f_4 : (-\infty, 0] \to \mathbb{R}, \qquad f_5 : (-\infty, 0] \to [0, +\infty),$$

from Example I.13, each given by the formula $x \mapsto x^2$. Check that:

(a) The functions f_2 and f_5 are surjective, whereas f_1, f_3, and f_4 are not. are not.
(b) The functions f_3, f_4, and f_5 are injective, whereas f_1 and f_2 are not.
(c) The function f_5 is bijective, whereas f_1, f_2, f_3, and f_4 are not.
(d) The inverse function $f_5^{-1} : [0, +\infty) \to (-\infty, 0]$ of f_5 is given by $f_5^{-1}(y) = \sqrt{-y}$, whereas f_1, f_2, f_3, and f_4 do not have inverse functions.

Restricting the domain

Given a function $f : X \to Y$ and a subset $A \subset X$ of its domain, it is possible to consider a function defined only in the subset by otherwise the same formula. This function $A \to Y$ is denoted by $f|_A$ and is called the **restriction** of f to the subset A; it is given by

$$f|_A : \ A \to Y$$
$$f|_A(a) = f(a) \qquad \text{for } a \in A. \tag{I.24}$$

Example I.15 (Examples of restrictions of functions).

Consider again the functions

$$f_1 : \mathbb{R} \to \mathbb{R}, \qquad f_2 : \mathbb{R} \to [0, +\infty), \qquad f_3 : [0, +\infty) \to \mathbb{R},$$
$$f_4 : (-\infty, 0] \to \mathbb{R}, \qquad f_5 : (-\infty, 0] \to [0, +\infty),$$

from Example I.13, each given by the formula $x \mapsto x^2$. Then we have

$$f_3 = f_1|_{[0,+\infty)}, \qquad f_4 = f_1|_{(-\infty,0]}, \qquad f_5 = f_2|_{(-\infty,0]}$$

Images and preimages under functions

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

If $A \subset X$ is a subset of the domain of f, then the **image** of A under f is the subset

$$f[A] \;=\; \big\{ f(a) \;\big|\; a \in A \big\} \tag{I.25}$$

of the codomain of f consisting of those points $y \in Y$ which are obtained as the value at some point $a \in A \subset X$, i.e., $y = f(a)$.

Example I.16 (Some images).
Consider the function

$$f\colon \mathbb{R} \to \mathbb{R} \qquad \text{given by} \qquad f(x) = x^2 \quad \text{for } x \in \mathbb{R}.$$

The preimage of the (singleton) subset $A = \{-2, 2\} \subset \mathbb{R}$ is

$$f\big[\{-7, -5, 5, 7\}\big] \;=\; \{25, 49\}\,.$$

In particular the image of a subset may contain fewer elements than the subset itself.

Example I.17 (The image of a singleton).
The image of a singleton $\{x\} \subset X$ is $f\big[\{x\}\big] = \{f(x)\}$.

Exercise I.7 (Singleton image).
If the image of $A \subset X$ is a singleton $f[A] = \{y\}$, then what can be said about the behavior of the function f on the subset A?

Example I.18 (Surjectivity by images).
A function $f\colon X \to Y$ is surjective if and only if the image of the whole domain is the whole codomain, $f[X] = Y$.

If $B \subset Y$ is a subset of the codomain of f, then the **preimage** of B under f is the subset

$$f^{-1}[B] \;=\; \big\{ x \in X \;\big|\; f(x) \in B \big\} \tag{I.26}$$

of the domain of f consisting of those points $x \in X$ at which the value belongs to the subset $B \subset Y$. The notation f^{-1} is used for both the inverse function and the preimage, but we try to consistently use square brackets in the latter case.[13] The inverse function only exists if f is a bijection, whereas preimages are always defined.

Example I.19 (Some preimages).
Consider the function

$$f\colon \mathbb{R} \to \mathbb{R} \qquad \text{given by} \qquad f(x) = x^2 \quad \text{for } x \in \mathbb{R}.$$

The preimage of the (singleton) subset $A = \{5\} \subset \mathbb{R}$ is

$$f^{-1}\big[\{5\}\big] \;=\; \big\{ -\sqrt{5},\; \sqrt{5}\,\big\}.$$

[13] In most mathematical literature ordinary parentheses are used also for preimages, or parentheses are altogether omitted. You are in any case expected to figure out which one is which, since the inverse function is applied to *elements* of the codomain, whereas in the preimage is applied to *subsets* of the codomain.

Similarly, we try to consistently use square brackets for the images of subsets under a function. Elsewhere in literature ordinary parentheses or no parentheses at all are used. Again you are in any case expected to figure out which one is which, since the function is applied to *elements* of the domain, whereas in the image is applied to *subsets* of the domain.

In particular the preimage of a singleton can contain more than one element.

The preimage of the (singleton) subset $A = \{-5\} \subset \mathbb{R}$ is the empty set,

$$f^{-1}\big[\{-5\}\big] \;=\; \emptyset.$$

In particular the preimage of a nonempty set can be empty.

Example I.20 (Bijectivity by preimages)**.**
A function $f\colon X \to Y$ is bijective if and only if the preimage of every singleton $\{y\} \subset Y$ is a singleton.

Example I.21 (Injectivity by preimages)**.**
A function $f\colon X \to Y$ is injective if and only if the preimage of every singleton $\{y\} \subset Y$ is either a singleton or the empty set.

Exercise I.8 (Images and preimages of unions and intersections)**.**
Let X and Y be sets and $f\colon X \to Y$ a function.

(a) Show that for any $C, D \subset Y$, the preimages satisfy
$$f^{-1}[C \cup D] = f^{-1}[C] \cup f^{-1}[D].$$
(b) Show that for any $C, D \subset Y$, the preimages satisfy
$$f^{-1}[C \cap D] = f^{-1}[C] \cap f^{-1}[D].$$
(c) Show that for any $A, B \subset X$, the images satisfy
$$f[A \cup B] = f[A] \cup f[B].$$
(d) Give an example in which for the images of subsets $A, B \subset X$ we have
$$f[A \cap B] \neq f[A] \cap f[B].$$

Hint: *In parts (a)–(c) it is possible to argue by a chain of equivalent conditions*

$$x \in \textit{left hand side set} \quad \Longleftrightarrow \quad \cdots \quad \Longleftrightarrow \quad x \in \textit{right hand side set.}$$

A comparison of (b) and (d) above indicates that preimages behave better with set operations than images. Another example is the following.

Exercise I.9 (Images and preimages of complements)**.**
Let X and Y be sets and $f\colon X \to Y$ a function.

(a) Suppose that $C \subset D \subset Y$. Show that the preimages satisfy
$$f^{-1}[D \setminus C] \;=\; f^{-1}[D] \setminus f^{-1}[C].$$
(b) Give an example in which for the images of subsets $A \subset B \subset X$ we have
$$f[B \setminus A] \;\neq\; f[B] \setminus f[A].$$

The formulas of Exercise I.8(a)–(c) generalize to arbitrary unions and intersections (more than two sets), and it is easy to modify the proofs to account for this general case. Specifically, if $f\colon X \to Y$ is a function and $B_j \subset Y$ for $j \in J$, then the preimages satisfy

$$f^{-1}\Big[\bigcup_{j\in J} B_j\Big] \;=\; \bigcup_{j\in J} f^{-1}[B_j], \qquad\qquad f^{-1}\Big[\bigcap_{j\in J} B_j\Big] \;=\; \bigcap_{j\in J} f^{-1}[B_j].$$

Likewise if $A_j \subset X$ for $j \in J$, then the images satisfy

$$f\Big[\bigcup_{j\in J} A_j\Big] \;=\; \bigcup_{j\in J} f[A_j].$$

I.3. Logic and related notation

Mathematics in general concerns with statements that can be logically proven to be true, starting from some specified assumptions. In other words, we want to deduce conclusions of interest from known premises. The rules for valid deduction are the subject of logic. Logic itself could be formalized, but for our purposes it suffices to introduce a little bit of notation and typical examples. The main message is that in this course and in all mathematics, we must have unambiguous definitions and statements, and we must provide valid arguments to justify the statements made.

Quantifiers

Statements involving free variables

Commonly, statements (predicates) involve one or more variables, and whether the statements are true or false depends on those variables. It may be thought of as a function of those variables, whose possible values are true and false.

Example I.22 (An example statement depending on a variable)**.**
The statement

$$\text{“}k \text{ is even”}$$

is true for if the variable k has the value $k = -8$, but it is false if $k = 33$.

Those variables of which the truth value of a statement depends are called free variables; without specifying values for the free variables, the statement itself is not yet true nor false. In mathematical text we often indicate that we fix the values of (some) free variables by expressions such as: “Fix $n \in \mathbb{N}$” (after which the varible n is no longer thought of as free, but having some a priori arbitrary but fixed natural number value) or “Let $\varepsilon > 0$” (after which the varible ε is no longer thought of as free, but fixed to some a priori arbitrary positive real value).

Statements involving quantifiers

The following two example statements (predicates),

$$\text{“for all real numbers } x \text{ we have } x^2 \geq 0\text{”}$$

and

$$\text{“there exists a real number } x \text{ such that } x^2 = -1\text{”},$$

both involve a variable x which has not been fixed, but the statements have definite truth values: the former statement is true and the latter is false. The variable x in the above formulas is not free, because a *quantifier* is applied on it: each statement above makes a claim about the *quantity* of values of x for which another statement, which involves x as a free variable, is true.

There are two quantifiers in ordinary logic. The **universal quantifier** denoted by the symbol $\forall$ means that the statement that follows it is true *for all values of the variable* (which appears immediately after the quantifier); in natural language its symbol $\forall$ is read as “for all”. The **existential quantifier** denoted by the symbol $\exists$ means that the statement that follows it is true *for at least one value of the variable* (which appears immediately after the quantifier); in natural language its symbol $\exists$ is read as "there exists". Both quantifiers must be followed first by the variable which they quantify, and they should (in careful usage) also be followed by a specification of the set in which the variable is allowed to take values, and then another statement (possibly) involving the variable. For clarity we usually also separate the other statement (in which the variable appears as free) by a colon (:). For example the two example statements that we started with should be written as

$$\forall\, x \in \mathbb{R}: \quad x^2 \geq 0$$

and

$$\exists\, x \in \mathbb{R}: \quad x^2 = -1.$$

Example I.23 (Revisiting subsets).
The meaning of the subset relation $A \subset B$ is, in concise logical notation with quantifiers,

$$\forall\, x \in A: \quad x \in B.$$

In other words, the statement above is true if $A \subset B$, and false if $A \not\subset B$.

Example I.24 (Revisiting unions and intersections).
Let $(A_j)_{j \in J}$ be a collection of sets. Then the meaning of the statement that an element x belongs to the union of this collection, $x \in \bigcup_{j \in J} A_j$, can be expressed in concise logical notation with quantifiers as

$$\exists\, j \in J: \quad x \in A_j.$$

Similarly, if the collection is nonempty ($J \neq \emptyset$), then the meaning of the statement that an element x belongs to the intersection of this collection, $x \in \bigcap_{j \in J} A_j$, can be expressed as

$$\forall\, j \in J: \quad x \in A_j.$$

Note that the quantified variable must appear as a free variable in the statement that follows, but that statement may itself have quantifiers for other variables. This is in fact very common in, e.g., definitions related to metric space topology (continuity, limits of sequences, etc.). An example statement with more than one quantifier is

$$\forall\, \alpha \in \mathbb{R}: \quad \exists\, C > 0: \quad \forall\, x \in \mathbb{R}: \quad \left| e^{3+5\alpha x - x^2} \right| \leq C$$

(this statement is true, by the way). Literally this statement reads "for all real numbers α there exists a positive number C such that for all real numbers x we have $|e^{3+5\alpha x-x^2}| \leq C$", and allowing for a bit more liberties of expression and interpretation could also be read "whatever the value of the real paramenter α, the expression $e^{3+5\alpha x-x^2}$ involving a real variable x is bounded in absolute value by some positive constant C". The use of logical notation makes the statements more concise and unambiguous. By contrast, in natural language one has to be quite careful to archieve an unambiguous meaning.[14]

Note that the order of quantifiers matters! For instance the statement

$$\forall\, x \in \mathbb{R}: \quad \exists\, C > 0: \quad \forall\, \alpha \in \mathbb{R}: \quad \left| e^{3+5\alpha x - x^2} \right| \leq C$$

[14]This does not mean that natural language is strictly worse. Among advantages of natural language are that it allows to emphasize certain aspects of the statement and to draw attention to interpretations (for example interpreting α as a parameter, C as a constant, and x as a variable in the above expression). None of this affects the validity of logical statements, but to claim that such things are irrelevant would be rather extreme. . .

is false, although it "only" differs from the previous example (a true statement) by the order of quantifiers.

Exercise I.10 (Check the truth values of the above examples).
Prove that the statement $\forall\, \alpha \in \mathbb{R} : \exists\, C > 0 : \forall\, x \in \mathbb{R} : |e^{3+5\alpha x - x^2}| \leq C$ is true and that the statement $\forall\, x \in \mathbb{R} : \exists\, C > 0 : \forall\, \alpha \in \mathbb{R} : |e^{3+5\alpha x - x^2}| \leq C$ is false.

Implications and equivalences

Implications

Suppose that P and Q are two statements (predicates) possibly involving free variables (the same variables in both). Then one can form a new statement out of them, called an **implication** and denoted by

$$P \implies Q,$$

which in natural language can be read in any of the following ways

P implies Q ;
if P, then Q ;
P only if Q .

The natural language descriptions (the simplest of which is probably "if P, then Q") already should explain exactly how the truth value of an implication is determined: $P \Rightarrow Q$ is true if either P and Q are both true, or if P is false.[15]

Example I.25 (Revisiting subsets again).
In concise logical notation, the meaning of the subset relation $A \subset B$ is the implication

$$x \in A \implies x \in B.$$

In other words, the above logical statement is true if $A \subset B$, and false if $A \not\subset B$.

A typical use of an implication is in theorem statements: P may represent the assumptions of a theorem (the premises) and Q the conclusion, so the theorem statement itself is the implication $P \Rightarrow Q$ ("if the assumptions hold, then also the conclusion holds"). An implication in a theorem statement should of course be true (the role of the *proof* of the theorem is exactly to verify the validity of the asserted implication).

Note, however, that the implication $P \Rightarrow Q$ is merely a statement about the relationship of the truth values of the two statements P and Q — it does not signify any *causal* relationship.

Instead of $P \Rightarrow Q$ we also occasionally switch the places of P and Q and invert the arrow direction, i.e., we alternatively write

$$Q \impliedby P,$$

[15]Especially from the natural language phrasing "if P, then Q" it should be clear that the only situation in which the implication is false is if P is true but Q is nevertheless false. In particular, if P does not hold, the implication is claiming nothing whatsoever about the truth or falsity of Q.

which in natural language can be read in any of the following ways

$$Q \text{ is implied by } P \text{ ;}$$
$$Q \text{ if } P \text{ .}$$

So we have seen altogether five different phrases in natural language for the same implication — the logical meaning of each is exactly the same, but they may be used to express a slightly different emphasis.

Equivalences

Two statements P and Q, possibly depending on some free variables (same in both), are said to be **logically equivalent** if they have the same truth value for all values of the free variables. This precisely amounts to requiring that both implications $P \Rightarrow Q$ and $P \Leftarrow Q$ hold (Q is true if P is true and vice versa). We denote **logical equivalence** by

$$P \iff Q,$$

and in natural language read it as either

$$P \text{ and } Q \text{ are (logically) equivalent ;}$$
$$P \text{ if and only if } Q \text{ .}$$

Example I.26 (Revisiting the equality of sets)**.**
The meaning of the equality of two sets $A = B$ is

$$x \in A \iff x \in B.$$

Remark I.27 (Warning: a convention used in definitions)**.**
When a new term is defined, we usually say that the term is used if some conditions are satisfied. In this context we actually mean that the term is used *if and only if* the conditions are satisfied! But it is conventional in definitions to only state the "if" part, leaving it implicit that the newly introduced term is *not* used unless the conditions are satisfied.

For example our definition of a bijective function was: "we say that f is bijective if it is both injective and surjective". We of course *also* meant that we *do not* call f bijective *unless* it is both injective and surjective.

So while the natural language word "if" in usual mathematical text signifies just one implication, in definitions it is conventionally used to signify equivalence.

Negations

If P is a logical statement (predicate), its **negation**

$$\texttt{not } P$$

is the logical statement which always has the opposite truth value: if P is true, then `not` P is false and if P is false, then `not` P is true. In natural language, `not` is read as "not".

Clearly the statement `not` (`not` P) is logically equivalent to P (a double negation).

We have already introduced specific symbols for the negations of a few common statements, e.g.,

$$\begin{aligned} \texttt{not}\ \big(a = b\big) \qquad & \text{is denoted} \qquad a \neq b \\ \texttt{not}\ \big(a \in A\big) \qquad & \text{is denoted} \qquad a \notin A \\ \texttt{not}\ \big(A \subset B\big) \qquad & \text{is denoted} \qquad A \not\subset B. \end{aligned}$$

Contrapositives

The implications

$$\begin{aligned} P \quad & \Longrightarrow \quad Q \\ \texttt{not}\ Q \quad & \Longrightarrow \quad \texttt{not}\ P \end{aligned} \tag{I.27}$$

are logically equivalent — you should make sure that you understand why![16] The second implication is known as the **contrapositive** of the first.

Let us take an example from Section I.2.

Example I.28 (Revisiting injectivity)**.**
Recall definition of an injective function $f\colon X \to Y$ from Definition I.14. In logical symbols, injectivity means

$$\forall\, x_1, x_2 \in X : \qquad x_1 \neq x_2 \implies f(x_1) \neq f(x_2). \tag{I.28}$$

An equivalent form of this condition can be obtained by taking a contrapositive of the implication above. This is particularly helpful, since the contrapositive will involve negations of the propositions $x_1 \neq x_2$ and $f(x_1) \neq f(x_2)$, which are simply $x_1 = x_2$ and $f(x_1) = f(x_2)$, respectively. Using the contrapositive form, we find that injectivity equivalently means

$$\forall\, x_1, x_2 \in X : \qquad f(x_1) = f(x_2) \implies x_1 = x_2. \tag{I.29}$$

The original formulation (I.28) of injectivity can be read: "At any two different points, the function has different values". The formulation (I.29) obtained by contrapositive can be read: "Whenever the values at two points are the same, the points themselves must be the same". With a bith of thinking, one easily convinces oneself about the equivalence of these, but routine use of logical symbols clearly facilitates this.

The equivalence of the two implications in (I.27) in particular underlies the idea of a *proof by contrapositive* or and *indirect proof* — of which the *proof by contradiction* is essentially a special case.

To elaborate, suppose our goal is to prove that from assumption P, conclusion Q follows. Another way of doing that is to prove that if we assume that the desired conclusion Q does not hold, then it implies that the assumption P cannot hold, either. This is the proof by contrapositive.

In particular, if there are no (explicit) assumptions, we can consider the assumptions represented by an identically true proposition P. Then this proof strategy is proof by contradiction. Indeed, by showing the contrapositive, we find that if the the desired conclusion Q would not hold, then the negation of the identically true assumption P

[16] The most straightforward way to convince oneself about this is to write a truth table, i.e., to consider the truth value of both implications for all the four possibilities of the truth values of the propositions P and Q (P false and Q false, P false and Q true, P true and Q false, P true and Q true).

would hold, which is absurd; a contradition. The proof of Proposition I.5 is a classic example.

Negations and quantifiers

Since typical definitions in metric space topology involve many logical quantifiers, it is essential to be comfortable working with them. In particular one has to be able to routinely form the negation of a statement involving a quantifier.

Consider a statement involving the universal quantifier, say

$$\forall\, x \in X : \quad P(x).$$

The negation of this statement is

$$\exists\, x \in X : \quad \texttt{not}\ P(x);$$

you should make sure you understand why!

A common use of this observation is that to *disprove* the validity of a claim (for example a belief that we might have) that $P(x)$ holds for all $x \in X$, it is sufficient to find just one example of an $x \in X$ for which $P(x)$ is false. Such an example is called a **counterexample**, because it is sufficient to invalidate the whole original claim (which started with the universal quantifier).

And even if we are seeking to prove rather than to disprove a claim of this form, in an indirect proof (contrapositive) we would first form the negation.

Symmetrically, consider a statement involving the existential quantifier, say

$$\exists\, x \in X : \quad Q(x).$$

The negation of this statement is (think about it)

$$\forall\, x \in X : \quad \texttt{not}\ Q(x).$$

On the surface this does not appear as practical as the negation of a statement with a universal quantifier. But in indirect proofs, or when disproving a claim (say a false belief) of this form, we would have to form the negation. Another typical use of this appears when a statement involves many quantifiers; we may want to consider the negation to get rid of some universal quantifiers, but the existential quantifiers are affected as well.

Example I.29 (There are arbitrarily large natural numbers)**.**
The logical statement

$$\forall\, q \in \mathbb{Q} : \quad \exists\, n \in \mathbb{N} : \quad n > q \tag{I.30}$$

has the interpretation that there are arbitrarily large natural numbers ("for any rational number q there exists a larger natural number n").

The statement (I.30) is true: it is the so called **Archimedean property** of the rational numbers. Let us examine a proof of this statement using a proof by contradiction.

For a proof by contradiction, one would suppose that (I.30) is false, and show that this leads to a contradiction. If (I.30) is false, then its negation

$$\texttt{not}\ \Big(\forall\, q \in \mathbb{Q} : \quad \exists\, n \in \mathbb{N} : \quad n > q\Big).$$

is true. The original statement (I.30) begins with a universal quantifier ($\forall$), and its negation takes the form

$$\exists\, q \in \mathbb{Q} : \quad \texttt{not}\ \Big(\exists\, n \in \mathbb{N} : \quad n > q\Big).$$

In this form, after the first existential quantifier ($\exists$) we have the negation of a statement which begins with an existential quantifier ($\exists$), which we may further unravel. The negation of (I.30) is thus written as

$$\exists\, q \in \mathbb{Q}: \quad \forall\, n \in \mathbb{N}: \quad \texttt{not}\ \big(n > q\big).$$

The negation $\texttt{not}\ (n > q)$ is obviously $n \leq q$, so the negation of (I.30) has been rewritten as

$$\exists\, q \in \mathbb{Q}: \quad \forall\, n \in \mathbb{N}: \quad n \leq q,$$

which in natural language claims that there exists some rational number q such that all natural numbers n are bounded from above by it. This clearly already sounds absurd, but we leave it to the reader to now derive the contradiction.[17] The point here was to illustrate how we typically end up considering negations of statements with quantifiers, and how they can be systematically unraveled.

I.4. ♡ Cardinalities of sets

Let us denote the number of elements of a set A by $\#A$.

Example I.30 (Number of elements in a few example sets)**.**

- For the empty set, we have $\#\emptyset = 0$.
- For a singleton, we have $\#\{a\} = 1$.
- We have $\#\{0,1,2,\ldots,8,9\} = 10$.
- Let $c_0, c_1, \ldots, c_{d-1}, c_d \in \mathbb{R}$ with $c_d \neq 0$, so that $p(x) = c_d\, x^d + c_{d-1}\, x^{d-1} + \cdots + c_1\, x + c_0$ is a polynomial of degree $d \in \mathbb{N}_0$. Then for the set of real zeroes of p we have
$$\#\big\{x \in \mathbb{R} \;\big|\; p(x) = 0\big\} \;\leq\; d.$$

If a set A has infinitely many different elements, we denote $\#A = \infty$.

Example I.31 (Some infinite sets)**.**
The sets of natural numbers, integers, rational numbers, and real numbers are infinite

$$\#\mathbb{N} = \infty, \qquad \#\mathbb{Z} = \infty, \qquad \#\mathbb{Q} = \infty, \qquad \#\mathbb{R} = \infty.$$

Also nontrivial intervals are infinite, for example $\#(\pi, \pi] = \infty$.

Moreover, as we will verify later, a non-empty open interval $(a,b) \subset \mathbb{R}$ contains infinitely many rational numbers and infinitely many irrational numbers

$$\#\Big((a,b) \cap \mathbb{Q}\Big) \;=\; \infty, \qquad \#\Big((a,b) \setminus \mathbb{Q}\Big) \;=\; \infty.$$

A set A is called **finite** if $\#A \in \{0,1,2,\ldots\}$, and **infinite** if $\#A = \infty$. It turns out that among infinite sets, some are nevertheless larger than others. The notion of *cardinality* captures this. To motivate the definition, let us nevertheless begin by some observations about sizes of finite sets.

Observation I.32 (Comparing sizes of sets using surjective functions)**.**
If A, B are two finite sets and $B \neq \emptyset$, then the following are equivalent:[18]

---

[17]**Hint:** *If such a rational number $q \in \mathbb{Q}$ exists, write it as $q = \frac{m}{k}$ with $m \in \mathbb{Z}$ and $k \in \mathbb{N}$. Then use for example the natural number $n = |m| + 1$ as a counterexample to the statement $\forall\, n \in \mathbb{N}: \ n \leq q$.*

[18]For the empty set $B = \emptyset$, the comparison of sizes has to be handled separately (the empty set is smaller than any other set), since from a nonempty set $A \neq \emptyset$ there does not exist any functions $A \to \emptyset$ — let alone surjective functions.

- $\#A \geq \#B$
- there exists a surjective function $A \to B$.[19]

In view of the above observation, it appears meaningful to consider a set A at least as large as a set $B \neq \emptyset$ if there exists a surjective function $A \to B$. In this case we denote $A \succeq B$ — or interchangeably $B \preceq A$. This is the comparison of **cardinalities** of sets. We may observe the following properties, which match with our intuition about sizes of sets:

- If $A \succeq B$ and $B \succeq C$, then $A \succeq C$.[20]
- If $B \subset A$ is a nonempty subset, then $A \succeq B$.[21]

Example I.33 (Cardinality comparisons of some infinite sets).
Because of the subset relations $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$, we have the cardinality comparisons $\mathbb{N} \preceq \mathbb{Z} \preceq \mathbb{Q} \preceq \mathbb{R}$.

We say that two nonempty sets A and B have **equal cardinalities** if $A \succeq B$ and $A \preceq B$. By Observation I.32, (nonempty) finite sets have equal cardinalities if and only if they have the same number of elements. Let us then look at some examples with infinite sets.

Example I.34 (Equal cardinality for a set and its proper subsets).
Consider the set $\mathbb{N}$ of natural numbers, and the set $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ of nonnegative integers,

$$\begin{aligned} \mathbb{N} &= \{1, 2, 3, 4, \ldots\} \\ \mathbb{N}_0 &= \{0, 1, 2, 3, 4 \ldots\}. \end{aligned}$$

Because of the subset relation $\mathbb{N} \subset \mathbb{N}_0$, we have $\mathbb{N} \preceq \mathbb{N}_0$. On the other hand, we also have a surjective function $f \colon \mathbb{N} \to \mathbb{N}_0$ given by $f(n) = n - 1$ for $n \in \mathbb{N}$, so that $\mathbb{N} \succeq \mathbb{N}_0$ holds, too. Therefore the sets $\mathbb{N}$ and $\mathbb{N}_0$ have equal cardinalities.

In particular an infinite set can have a proper subset with equal cardinality: the set $\mathbb{N}_0$ has one extra element compared to $\mathbb{N}$ (namely zero), but this does not affect the "size" of the set; these two infinite sets have equal cardinalities.

Example I.35 (Another equal cardinality for a set and its proper subsets).
Consider the set $\mathbb{N}$ of natural numbers and the set $A = \{2, 4, 6, \ldots\}$ of even natural numbers. Because of the subset relation $A \subset \mathbb{N}$, we have $A \preceq \mathbb{N}$. On the other hand, we also have a surjective function $f \colon A \to \mathbb{N}$ given by $f(m) = \frac{1}{2}m$ for $m \in A$, so that $A \succeq \mathbb{N}$ holds, too. Therefore the sets $\mathbb{N}$ and A have equal cardinalities.

---

[19] The idea is the following. Suppose that A is the set of all students of this course, and B is the set of all exercise groups. Every student is assigned to exactly one exercise group, so that the assignment defines a function $A \to B$. The function is surjective if every exercise group has at least one student. The gist of this observation is that in such a case we can conclude that there are at least as many students as there are exercise classes (and conversely: an assignment that leaves no exercise group empty exists if there are at least as many students as there are exercise classes).

[20] Indeed, if $f \colon A \to B$ and $g \colon B \to C$ are surjective, then also the composition $g \circ f \colon A \to C$ is surjective.

[21] Indeed, choose some element $b_0 \in B$ of the nonempty set B, and define a function f from A to the subset $B \subset A$ by

$$f(a) = \begin{cases} a & \text{if } a \in B \subset A \\ b_0 & \text{if } a \in A \setminus B. \end{cases}$$

Then f is surjective.

The set $\mathbb{N}_0$ has infinitely many extra elements compared to A (namely the odd natural numbers), but this does not affect the "size" of the set; these two infinite sets have equal cardinalities.

Exercise I.11 (The sets of natural numbers and integers have equal cardinalities).
Prove that we have $\mathbb{N} \preceq \mathbb{Z}$ and $\mathbb{N} \succeq \mathbb{Z}$.
Hint: *One direction is obvious; for the other, construct some surjectve function* $\mathbb{N} \to \mathbb{Z}$.

These examples show that many common infinite sets have the same cardinality as the set of natural numbers. This motivates the following definition.

Definition I.36 (Countable and uncountable infinite sets).
An infinite set A is **countably infinite** if $A \preceq \mathbb{N}$, i.e., if there exists a surjective function $f \colon \mathbb{N} \to A$; otherwise A is **uncountably infinite**.

Remark I.37 (Enumeration of elements of a countably infinite set).
If A is a countably infinite set, then it is possible to list or enumerate all elements of A as follows. If $f \colon \mathbb{N} \to A$ is a surjective function, then we may form a sequence

$$(a_1, a_2, a_3, \ldots)$$

with $a_n = f(n)$ for $n \in \mathbb{N}$. By surjectivity, each element of A appears at least once in this sequence.

It is not possible to enumerate elements of uncountably infinite sets in this way!

In Examples I.34 and I.35 and Exercise I.11 we have seen some examples of countably infinite sets. The most important example of an uncountably infinite set is the set of real numbers.

Theorem I.38 (The set of real numbers is uncountably infinite).
The set $\mathbb{R}$ of real numbers is uncountably infinite.

We postpone the proof to Lecture II, where real numbers are discussed in detail. The most important point to take note of is that you *cannot* enumerate all real numbers in a list!

By contrast, and perhaps surprisingly, the set of rational numbers is in fact countable. In this sense the set $\mathbb{Q}$ of rational numbers is much smaller than the set $\mathbb{R}$ of real numbers.

Theorem I.39 (The set of rational numbers is countably infinite).
The set $\mathbb{Q}$ of rational numbers is countably infinite.

The key to the proof of Theorem I.39 is the following frequently useful lemma.

Lemma I.40 (The Cartesian product of two countable sets is countable).
Suppose that A and B are two countably infinite sets. Then their Cartesian product $A \times B$ is also countably infinite.

It is still clearest to prove this lemma by first explicitly addressing the following special case.

Lemma I.41 (The set of pairs of natural numbers is countable)**.**
The set $\mathbb{N} \times \mathbb{N} = \{(n, m) \mid n \in \mathbb{N},\ m \in \mathbb{N}\}$ *of pairs of natural numbers is countably infinite.*

Proof. We must show that there exists a surjective function $t\colon \mathbb{N} \to \mathbb{N} \times \mathbb{N}$. This can be done in various ways; the enumeration illustrated in Figure I.1 is easy to visualize, but let us choose a different construction. Namely, the function $t\colon \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ given by

$$t(n) = \begin{cases} (k, \ell) & \text{if } n = 2^{k-1}\,3^{\ell-1} \text{ for } k, \ell \in \mathbb{N} \\ (1,1) & \text{if } n \text{ contains other prime factors besides 2 and 3} \end{cases}$$

is well-defined because of the unique prime factorization of natural numbers, and it is surjective, since any $(k, \ell) \in \mathbb{N} \times \mathbb{N}$ is obtained as a value at $n = 2^{k-1}\,3^{\ell-1} \in \mathbb{N}$. □

Proof of Lemma I.40. Assume that A and B are countable. To prove countability of $A \times B$, we must exhibit a surjective function $\mathbb{N} \to A \times B$.

By countability of A and B, there exists surjective functions

$$g_1\colon\ \mathbb{N} \to A \qquad\qquad\qquad g_2\colon\ \mathbb{N} \to B.$$

We may combine them into a function $g\colon \mathbb{N} \times \mathbb{N} \to A \times B$ defined by

$$g(n, m) = \big(g_1(n),\, g_2(m)\big),$$

which is clearly surjective.[22] From Lemma I.41 we get the existence of a surjective function $t\colon \mathbb{N} \to \mathbb{N} \times \mathbb{N}$. Now the composition $g \circ t\colon \mathbb{N} \to A \times B$,

$$\begin{tikzcd} \mathbb{N} \arrow[r, "t"] \arrow[rr, bend right, "g\circ t"'] & \mathbb{N} \times \mathbb{N} \arrow[r, "g"] & A \times B, \end{tikzcd}$$

is surjective. Countability of $A \times B$ follows. □

Proof of Theorem I.39. To show countability of the set $\mathbb{Q}$ of rational numbers, we must exhibit a surjective function $\mathbb{N} \to \mathbb{Q}$.

By definition, rational numbers $q \in \mathbb{Q}$ are of the form $q = \frac{m}{n}$, with $m \in \mathbb{Z}$ and $n \in \mathbb{N}$. In other words, the function

$$r\colon \mathbb{Z} \times \mathbb{N} \to \mathbb{Q} \qquad \text{given by} \qquad r(m, n) = \frac{m}{n} \qquad \text{for } (m, n) \in \mathbb{Z} \times \mathbb{N}$$

is surjective.

The set $\mathbb{Z}$ of integers is countable by Exercise I.11, and the set $\mathbb{N}$ of natural numbers is obviously countable. By Lemma I.40 the Cartesian product $\mathbb{Z} \times \mathbb{N}$ is therefore countable, so there exists a surjective function

$$s\colon \mathbb{N} \to \mathbb{Z} \times \mathbb{N}.$$

Now the composition $r \circ s\colon \mathbb{N} \to \mathbb{Q}$,

$$\begin{tikzcd} \mathbb{N} \arrow[r, "s"] \arrow[rr, bend right, "h\circ s"'] & \mathbb{Z} \times \mathbb{N} \arrow[r, "r"] & \mathbb{Q}, \end{tikzcd}$$

is surjective. Countability of $\mathbb{Q}$ follows. □

[22]Indeed, suppose that $(a, b) \in A \times B$. By surjectivity of $g_1\colon \mathbb{N} \to A$ there exists an $n \in \mathbb{N}$ such that $g_1(n) = a$ and by surjectivity of $g_2\colon \mathbb{N} \to B$ there exists an $m \in \mathbb{N}$ such that $g_1(m) = b$. Therefore we have $g(n, m) = \big(g_1(n),\, g_2(m)\big) = (a, b)$, showing surjectivity of g.

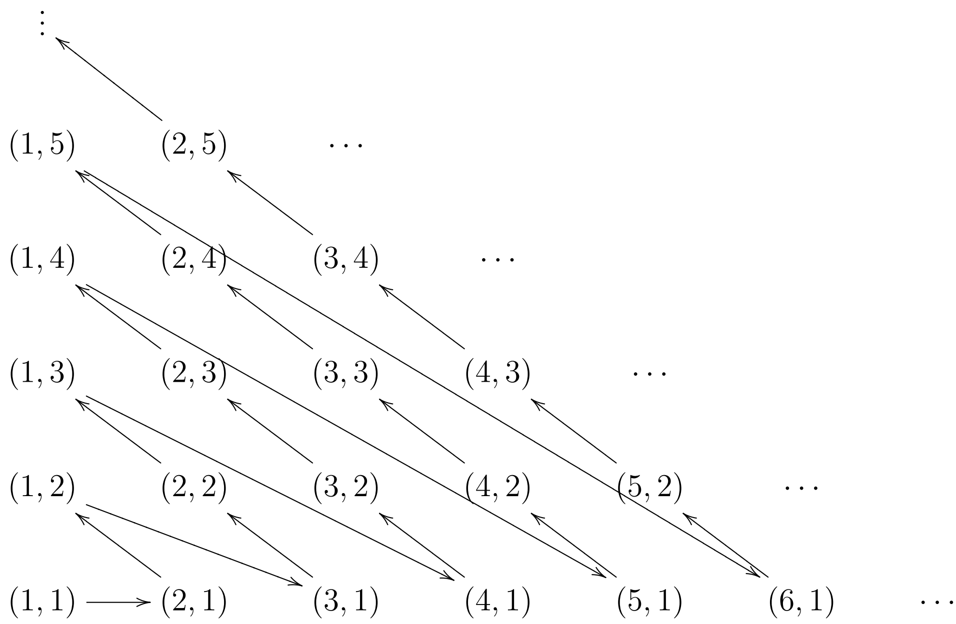


FIGURE I.1. The arrows indicate an enumeration of $\mathbb{N} \times \mathbb{N}$:

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

The idea is to observe that for pairs $(n, m) \in \mathbb{N} \times \mathbb{N}$, the sum $n + m$ takes values $2, 3, 4, \ldots$, and for each value of the sum it is straightforward to list the are finitely many pairs (n, m) in increasing order of the corresponding m (for example). This enumeration gives another construction of a surjective function $\mathbb{N} \to \mathbb{N} \times \mathbb{N}$.

Lecture II

Real numbers

Real numbers should surely feel familiar already! But examining their precise definition may reveal, however, that they are not as innocent as they may seem.

We start by discussing mostly familiar issues chosen to concretely exemplify, in the special case of the real line $\mathbb{R}$, certain topological notions that we encounter later on in this course. Namely, we consider the absolute value of real numbers and the related notion of distance on the real line, which is our first example of a metric, although the general definition will have to wait until Lecture V. This is the notion that underlies for example the convergence of sequences of real numbers, and we look at basic properties of of convergence as an instructive example of the more general theory to be developed later in Lecture VIII. We also show that both rational and irrational numbers are dense on the real line — which in particular gives the topological property of *separability* of the real line.[1]

We then turn to foundational questions again: what exactly are real numbers? We give the axiomatic properties, most of which are entirely unsurprising. Only the *completeness axiom* of the real numbers is likely to have any feeling of novelty to it. It involves the notions of *supremum* (least upper bound) and *infimum* (greatest lower bound), which are indispensable in advanced mathematics, and which we treat in some detail, and for which we give a few example applications.

Finally, we provide the proof of the statement that the set of real numbers is uncountably infinite, which was omitted in the previous lecture.

II.1. Absolute value, distances between numbers, and triangle inequality

At the core of topological considerations of numbers is the concept of distances on the real line. They are based on the absolute value — which itself is the distance of a given number to the origin.

The **absolute value** of a real number $x \in \mathbb{R}$ is defined by

$$|x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0. \end{cases} \tag{II.1}$$

It is easy to see that another expression for it is

$$|x| = \sqrt{x^2}, \tag{II.2}$$

[1] Separability is not in fact discussed generally in this short course, but is crucial in many areas of mathematics, including functional analysis, probability theory, etc.

where $\sqrt{\cdot}$ denotes the non-negative square root of a non-negative number. Clearly the absolute value of satisfies

$$|x| = \big|-x\big|, \qquad |x| \geq 0, \qquad x \leq |x|, \qquad -x \leq |x|$$

for any $x \in \mathbb{R}$, and

$$\big|xy\big| \; = \; |x|\,|y|$$

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

The **distance**[2] between two real numbers $x, y \in \mathbb{R}$ is the absolute value of their difference,

$$|x - y|. \tag{II.3}$$

The following **triangle inequality** is a simple and intuitive result, but as we will see later from Lecture V on, a rather straightforward generalization of it serves as the basis of a very fruitful general theory of metric spaces.

Proposition II.1 (Triangle inequality on $\mathbb{R}$)**.**
For all $x, y \in \mathbb{R}$ we have

$$\Big||x| - |y|\Big| \;\leq\; \big|x + y\big| \;\leq\; |x| + |y|. \tag{II.4}$$

Proof. Let $x, y \in \mathbb{R}$. There are two inequalities to prove in (II.4).

Let us start by proving the second inequality. Calculate the square of the absolute value of the sum, and estimate each term from above by its absolute value to get

$$\begin{aligned} \big|x+y\big|^2 = \big(x+y\big)^2 &= x^2 + 2xy + y^2 \\ &\leq |x|^2 + 2|x||y| + |y|^2 \\ &= \Big(|x| + |y|\Big)^2. \end{aligned}$$

Taking square roots ($\sqrt{\cdot}$ is an increasing function and thus respects the inequality), we obtain that

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

To prove the first inequality, we use the second one already proved above. Applying the second inequality to the numbers $x + y$ and $-y$, on the other hand, gives

$$|x| \;=\; \big|(x+y) + (-y)\big| \;\leq\; |x+y| + |-y| \;=\; |x+y| + |y|,$$

and by subtracting $|y|$ from both sides, we get

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

A similar application of the second inequality to the numbers $x + y$ and $-x$ (or just interchanging the roles of x and y above) yields

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

Combining these gives the first asserted inequality

$$\Big||x| - |y|\Big| \;\leq\; |x + y|,$$

and the proof is complete. □

[2]Later, in Lecture V, we will see that this notion of a distance (II.3) is a special case of a *metric*, and it makes the real line a *metric space*. For now, however, we will work exclusively with the concrete case of real numbers, and we will not refer to the general notion.

Another common form of the triangle inequality, which features the distance $|x-y|$ in exactly this form is

$$|x-y| \leq |x| + |y|. \tag{II.5}$$

It is obtained from the second inequality in (II.4) by plugging in x and $-y$, and noting $|-y| = |y|$.

Theere is a straightforward generalization of the triangle inequality to sums with finitely many many terms.

Proposition II.2 (Triangle inequality for finite sums)**.**
Let $x_1, \ldots, x_n \in \mathbb{R}$. Then we have

$$\big|x_1 + \cdots + x_n\big| \;\leq\; |x_1| + \cdots + |x_n|. \tag{II.6}$$

Exercise II.1 (Proof of Proposition II.2)**.**
Prove Proposition II.2 by induction over $n \in \mathbb{N}$, using Proposition II.1.

Remark II.3 (Triangle inequality for infinite series)**.**
Let $x_1, x_2, x_3, \ldots \in \mathbb{R}$. The natural generalization of (II.6) to infinite series would be

$$\Big|\sum_{j=1}^{\infty} x_j\Big| \overset{?}{\leq} \sum_{j=1}^{\infty} |x_j|.$$

With suitable interpretations, this generalization would indeed be valid. However, it immediately requires consideration of the convergence of the infinite series on both sides, and/or suitable interpretations in the case that one or both of the series diverge. For now, we will not try to use these types of generalizations, but we invite the reader to think about what can and cannot be said about it.

II.2. Sequences on the real line

Sequences of real numbers

A **sequence** (of real numbers) is an "infinite list"

$$(a_1, a_2, a_3, \ldots)$$

of real numbers $a_1, a_2, a_3, \ldots \in \mathbb{R}$. A precise definition of a sequence is a function

$$a \colon \mathbb{N} \to \mathbb{R},$$

and the "list" consists of the values of this function: $a_n = a(n)$ for $n \in \mathbb{N}$. We usually use notations such as

$$(a_1, a_2, a_3, \ldots) \;=\; (a_n)_{n=1}^{\infty} = (a_n)_{n\in\mathbb{N}}$$

for a sequence, and it is even common to use the lazy notation (a_n), although this fails to explicitly indicate that the indexing of the members of the sequence uses natural numbers.

Occasionally we also let the indexing start from something other than $n = 1$. When we do so, this should be quite clear from the context. The notation is then modified in obvious ways: e.g., $(a_0, a_1, a_2, \ldots) = (a_n)_{n=0}^{\infty}$ or $(a_{25}, a_{26}, a_{27}, \ldots) = (a_n)_{n=25}^{\infty}$; the precise interpretation is as functions $\{0, 1, 2, \ldots\} \to \mathbb{R}$ or $\{25, 26, 27, \ldots\} \to \mathbb{R}$. The definitions in the following subsections must then be adapted in obvious ways. Some properties of interest (particularly limits) in fact do not even depend on any finitely many initial members of the sequence, in which case the choice of the starting index is irrelevant.

Monotonicity properties of number sequences

The real axis has an order. Sequences which respect that order have some particularly nice properties.

Definition II.4 (Monotonicity properties of sequences of real numbers)**.**
A real number sequence $(a_n)_{n\in\mathbb{N}}$ is

- **increasing**, if we have[3] $a_{n+1} \geq a_n$ for all $n \in \mathbb{N}$
- **strictly increasing**, if we have[4] $a_{n+1} > a_n$ for all $n \in \mathbb{N}$
- **decreasing**, if we have[5] $a_{n+1} \leq a_n$ for all $n \in \mathbb{N}$
- **strictly decreasing**, if we have[6] $a_{n+1} < a_n$ for all $n \in \mathbb{N}$
- **monotone**, if it is either increasing or decreasing.

Example II.5 (An increasing sequence)**.**
The sequence $(1, 2, 4, 8, 16, 32, \ldots)$ is (strictly) increasing: its n:th member is given by the formula $a_n = 2^{n-1}$, and we have $a_{n+1} = 2^n > 2^{n-1} = a_n$ for all $n \in \mathbb{N}$.

Example II.6 (Another increasing sequence)**.**
Let $(a_n)_{n\in\mathbb{N}}$ be defined by the formula $a_n = \frac{n}{n+1}$ for $n \in \mathbb{N}$. We claim that this sequence is increasing.

One way to verify this is to calculate the difference of consequtive members,

$$a_{n+1} - a_n \;=\; \frac{n+1}{n+2} - \frac{n}{n+1} \;=\; \frac{(n+1)^2 - n(n+2)}{(n+2)(n+1)} \;=\; \frac{1}{(n+2)(n+1)} \;>\; 0,$$

which gives $a_{n+1} > a_n$ (by adding a_n to both sides in the inequality).

Another (perhaps easier?) way is to calculate the ratio of consequtive terms

$$\frac{a_n}{a_{n+1}} \;=\; \frac{n/(n+1)}{(n+1)/(n+2)} \;=\; \frac{n(n+2)}{(n+1)^2} \;=\; \frac{n^2+2n}{n^2+2n+1} \;<\; 1,$$

which gives $a_n < a_{n+1}$ (by multiplying both sides by the positive number a_{n+1}).

Example II.7 (A decreasing sequence)**.**
The sequence $(1, \frac{1}{4}, \frac{1}{9}, \frac{1}{16}, \frac{1}{25}, \ldots)$ is (strictly) decreasing: its n:th member is given by the formula $a_n = \frac{1}{n^2}$, and we have $a_{n+1} = \frac{1}{(n+1)^2} < \frac{1}{n^2} = a_n$ for all $n \in \mathbb{N}$.

Boundedness properties of number sequences

Definition II.8 (Boundedness properties of sequences of real numbers)**.**
A real number sequence $(a_n)_{n\in\mathbb{N}}$ is

- **bounded from above**, if there exists some $B \in \mathbb{R}$ (called an **upper bound** for the sequence) such that $a_n \leq B$ for all $n \in \mathbb{N}$;

[3] In other words, a sequence is increasing if $a_1 \leq a_2 \leq a_3 \leq \cdots$.
[4] In other words, a sequence is strictly increasing if $a_1 < a_2 < a_3 < \cdots$.
[5] In other words, a sequence is decreasing if $a_1 \geq a_2 \geq a_3 \geq \cdots$.
[6] In other words, a sequence is strictly decreasing if $a_1 > a_2 > a_3 > \cdots$.

- **bounded from below**, if there exists some $A \in \mathbb{R}$ (called a **lower bound** for the sequence) such that $a_n \geq A$ for all $n \in \mathbb{N}$;
- **bounded**, if there exists some $M \in \mathbb{R}$ such that $|a_n| \leq M$ for all $n \in \mathbb{N}$.

Exercise II.2 (Boundedness from both above and below).
Prove that a real number sequence $(a_n)_{n\in\mathbb{N}}$ is bounded if and only if it is both bounded from above and bounded from below.

Example II.9 (A sequence bounded from below).
The sequence $(1, 2, 4, 8, 16, 32, \ldots)$ is bounded from below by, for example, the constant $A = 1$.[7] To see this, note that the n:th member of the sequence is given by the formula $a_n = 2^{n-1}$, and we have $a_n = 2^{n-1} \geq 1 = A$ for all $n \in \mathbb{N}$.

This sequence is not bounded from above. Indeed for any $B \in \mathbb{R}$, we can take a natural number $n > \log_2(B) + 1$, and then $a_n = 2^{n-1} > 2^{\log_2(B)} = B$, so B does not work as an upper bound. Since $B \in \mathbb{R}$ was arbitrary, no upper bound exists for this sequence.

Example II.10 (A bounded sequence).
Let $(a_n)_{n\in\mathbb{N}}$ be defined by the formula $a_n = \frac{n}{n+1}$ for $n \in \mathbb{N}$. We claim that this sequence is bounded by, for example, the constant $M = 1$.[8] To see this, note that for any $n \in \mathbb{N}$ we have

$$|a_n| = \left|\frac{n}{n+1}\right| = \frac{n}{n+1} \leq 1 = M.$$

II.3. Limits of sequences on the real line

The definition of the limit of a sequence

Definition II.11 (Limit of a sequence of numbers).
A sequence $(a_n)_{n\in\mathbb{N}}$ of real numbers **converges** to a **limit** $\alpha \in \mathbb{R}$ if for every $\varepsilon > 0$ there exists an index $n_\varepsilon \in \mathbb{N}$ such that $|a_n - \alpha| < \varepsilon$ whenever $n \geq n_\varepsilon$.

Remark II.12 (Meaning of convergence as a logical statement).
Definition II.11 is written in plain English (a natural language), as usual. It nevertheless has a precise logical meaning, and it is instructive to unravel the definition using logical symbols (a formal language). The meaning of the statement $a_n \to \alpha$ as $n \to \infty$ is:

$$\forall \varepsilon > 0 : \quad \exists n_\varepsilon \in \mathbb{N} : \qquad n \geq n_\varepsilon \implies |a_n - \alpha| < \varepsilon. \tag{II.7}$$

A few advantages of the formal statement (II.7) are that it is succinct, unambiguous, and understandable to mathematicians irrespective of whether their mother tongue is English, Finnish, Swedish, or some other natural language (being familiar with the logical symbols, you can easily "read it" in your native language).

The logical definition is crucial, because it gives the precise meaning to the word *limit* (and *converge*). But of course it is still also important to have an intuitive idea about the notion as well! So how should you think about Definition II.11? In Figure II.1 and its caption we attempt to illustrate and describe this.

---

[7] As a lower bound here we could equally well use $A = 0$ or $A = -7$ or indeed any number $A \leq 1$.
[8] As a bound here we could equally well use $M = 42$ or $M = 10^{23}$ or indeed any number $M \geq 1$.

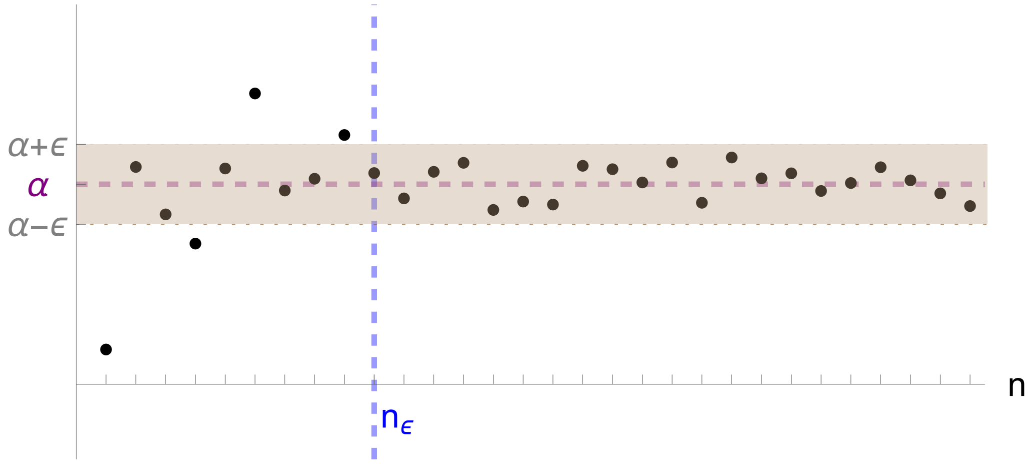


FIGURE II.1. An illustration of the limit of a number sequence. The limit of the sequence $(a_n)_{n\in\mathbb{N}}$ is $\alpha \in \mathbb{R}$ if for however small error $\varepsilon > 0$ we are willing to tolerate, the entire tail of the sequence starting from some index $n_\varepsilon \in \mathbb{N}$ lies within that error range from the value α; i.e., for all $n \geq n_\varepsilon$ we have $\alpha - \varepsilon < a_n < \alpha + \varepsilon$.

Basic properties of limits of sequences

Having given the precise definition of limit, we are ready to state and prove some basic properties of limits.

The first one says that the property of convergence implies that of boundedness.

Proposition II.13 (A convergent sequence of number is bounded)**.**
If a sequence $(a_n)_{n\in\mathbb{N}}$ of real numbers converges, then it is bounded.

Exercise II.3 (Proof of Proposition II.13)**.**
Prove Proposition II.13.

Exercise II.4 (Boundedness does not imply convergence)**.**
Show that there exists a bounded sequence which is not convergent. Conclude that the implication in the converse direction compared to Proposition II.13 does not hold.

Lemma II.14 (Preservation of inequalities)**.**
Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be two convergent sequences of real numbers, with respective limits $\alpha = \lim_{n\to\infty} a_n$ and $\beta = \lim_{n\to\infty} b_n$. If we have

$$a_n \leq b_n \qquad \textit{for all } n \in \mathbb{N},$$

then we also have the corresponding inequality

$$\alpha = \lim_{n\to\infty} a_n \leq \lim_{n\to\infty} b_n = \beta$$

for the limits.

Proof. UNDER CONSTRUCTION! □

Exercise II.5 (An improvement to Lemma II.14)**.**
Show that in Lemma II.14 it suffices to assume that the inequality holds for all sufficiently large n; more presicely that there exists $N \in \mathbb{N}$ such that $a_n \leq b_n$ for all $n \geq N$.

For sequences bounded from above or below, if a limit exists, it must also lie within the same bounds.

Corollary II.15 (Preservation of bounds)**.**
Let $(a_n)_{n\in\mathbb{N}}$ be a convergent sequences of real numbers, with limit $a = \lim_{n\to\infty} a_n$. If for some $c \in \mathbb{R}$ we have

$$a_n \leq B \qquad \textit{for all } n \in \mathbb{N},$$

then we also have the corresponding inequality for the limit,

$$\alpha \;=\; \lim_{n\to\infty} a_n \;\leq\; B.$$

Similarly, if for some $A \in \mathbb{R}$ we have $a_n \geq A$ for all $n \in \mathbb{N}$, then we also have $\lim_{n\to\infty} a_n \geq d$.

Exercise II.6 (Proof of Corollary II.15)**.**
Prove the first part of Corollary II.15, by using a constant sequence $(b_n)_{n\in\mathbb{N}}$ in Lemma II.14.

Prove the second part by a similar idea.

Exercise II.7 (An improvement to Corollary II.15)**.**
Show that in Corollary II.15 it suffices to assume that the inequality holds for all sufficiently large n.

Note that from Definition II.11 it is *not immediately obvious* that there could not exist several different numbers that satisfy what is required of a limit. In other words, if the condition (II.7) holds for $\alpha \in \mathbb{R}$, then could it also hold for some other real number $\alpha' \in \mathbb{R}$ in place of α?[9] Fortunately the following result lifts any concerns about the possibility of such ambiguity.[10]

Corollary II.16 (Uniqueness of limits)**.**
Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of real numbers. If both $\alpha \in \mathbb{R}$ and $\alpha' \in \mathbb{R}$ are limits of this sequence, then $\alpha = \alpha'$.

Proof. UNDER CONSTRUCTION! □

We next state a very practical result, which is in some sense an improved version of the preservation of inequalities — in it we do not even need to assume the existence of the limit, but the existence is a part of the conclusion! It states that if a sequence is "squeezed" in between two convergent sequences with same limit, then the sequence itself has to converge to this limit, too. This is often called *the squeeze theorem* (though we regard it more as a lemma). It also has various other affectionate nicknames. Some call it *the sandwich principle* — since it talks about a sequence "sandwiched" in between two others. Also *the lemma of two policemen* is descriptive: the idea being that one sequence "guards" the sequence of interest from above,

[9] The notation $\lim_{n\to\infty} a_n$ *suggests* that the limit is uniquely determined, but we should actually *prove* that this is so, in order to be sure that the notation is unambiguously defined!

[10] It would not have been difficult to prove the uniqueness of limits more directly, immediately after the definition. In fact, in Chapter VIII we treat sequences and limits in a more general setup, and we give a more direct (as well as more general) proof of the uniqueness. Here, however, we think it instructive to give a proof based on the preservation of inequalities.

preventing its escape to the upwards direction, and another sequence "guards" the sequence of interest from below, preventing its escape to the downwards direction.[11]

Lemma II.17 (Squeeze theorem)**.**
Let $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$, $(c_n)_{n\in\mathbb{N}}$ be three sequences of real numbers. Suppose that for all $n \in \mathbb{N}$ we have

$$a_n \;\leq\; b_n \;\leq\; c_n.$$

Suppose also that the sequences $(a_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ are convergent and have the same limit

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n = \beta \in \mathbb{R}.$$

Then the sequence $(b_n)_{n\in\mathbb{N}}$ is also convergent and its limit is the same,

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

Exercise II.8 (Proof of Lemma II.17)**.**
Prove the squeeze theorem (Lemma II.17).

Example II.18 (An example of the squeeze theorem)**.**
Consider the sequence $(b_n)_{n\in\mathbb{N}}$ given by

$$b_n = 3 + 4^{-n}\,\sin(5n) \qquad \text{for } n \in \mathbb{N}.$$

Since $-1 \leq \sin(\theta) \leq +1$ for all $\theta \in \mathbb{R}$, we get

$$3 - 4^{-n} \;\leq\; b_n \;\leq\; 3 + 4^{-n} \qquad \text{for all } n \in \mathbb{N}.$$

We have $\lim_{n\to\infty}(3-4^{-n}) = 3$ and $\lim_{n\to\infty}(3+4^{-n}) = 3$, so we can use the sequences $(a_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ defined by $a_n = 3-4^{-n}$ and $c_n = 3+4^{-n}$ in the squeeze theorem (Lemma II.17) to conclude that

$$\lim_{n\to\infty} b_n = 3.$$

Rules of calculation with limits

Given two real number sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$, we often consider new sequences formed from these for example by[12]

$$\big(a_n + b_n\big)_{n\in\mathbb{N}}\,, \qquad \big(a_n\, b_n\big)_{n\in\mathbb{N}}\,, \qquad \big(a_n/b_n\big)_{n\in\mathbb{N}}\,.$$

Theorem II.19 (Rules of calculation with limits)**.**
Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be two convergent sequences with respective limits $\alpha = \lim_{n\to\infty} a_n$ and $\beta = \lim_{n\to\infty} b_n$. Then we have

$$\lim_{n\to\infty} (a_n + b_n) \;=\; \alpha + \beta, \tag{II.8}$$

$$\lim_{n\to\infty} (a_n\, b_n) \;=\; \alpha\beta, \tag{II.9}$$

[11] Clearly the result must be important, given how imaginative nomenclature it has inspired!

[12] The last of these is only well defined if $b_n \neq 0$ for all $n \in \mathbb{N}$ — otherwise we encounter a division by zero in a_n/b_n. But in fact, since we will be concerned with calculating limits as $n \to \infty$, we do not necessarily need to care if such a problem occurs just finitely many times along the sequence. If there are only finitely many indices $n \in \mathbb{N}$ for which $b_n = 0$, then we can still consider the sequence $(a_n/b_n)_{n=n_0}^{\infty}$ starting from a sufficiently large index n_0 such that $b_n \neq 0$ for all $n \geq n_0$. It is meaningful to consider the limit of this sequence, which we still write as $\lim_{n\to\infty} a_n/b_n$.

and if moreover $\beta \neq 0$, then

$$\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{\alpha}{\beta}. \tag{II.10}$$

Before addressing the proof, we note the following. From (II.8) and (II.9) above, we get the following special cases by choosing $(b_n)_{n\in\mathbb{N}}$ to be a constant sequence $b_n = c$ for all $n \in \mathbb{N}$.

Corollary II.20 (Additive and multiplicative constants in limits)**.**
Let $(a_n)_{n\in\mathbb{N}}$ be a convergent sequence with limit $\alpha = \lim_{n\to\infty} a_n$, and let $c \in \mathbb{R}$ be a constant. Then we have

$$\lim_{n\to\infty} (a_n + c) = \alpha + c, \qquad\qquad \lim_{n\to\infty} (c\, a_n) = c\alpha.$$

Proof of Theorem II.19. The proofs of all cases are quite similar, so we only do (II.9) and leave the other two as exercises.

So suppose that $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ are two convergent sequences with limits $\alpha = \lim_{n\to\infty} a_n$ and $\beta = \lim_{n\to\infty} b_n$. We must prove that the sequence $(a_n\, b_n)_{n\in\mathbb{N}}$ converges to $\alpha\beta$, and we will do this directly using the Definition II.11 of limits. For this purpose, we first estimate

$$\begin{aligned}
&\big|a_n\, b_n - \alpha\beta\big| \\
=\;& \Big|a_n\, b_n - \alpha\, b_n + \alpha b_n - \alpha\beta\Big| && \text{(added and subtracted the same term)} \\
=\;& \Big|(a_n - \alpha)\, b_n + \alpha\,(b_n - \beta)\Big| && \text{(rearranged terms)} \\
\leq\;& \big|(a_n - \alpha)\, b_n\big| \;+\; \big|\alpha\,(b_n - \beta)\big| && \text{(triangle inequality)} \\
=\;& |b_n|\,\big|a_n - \alpha\big| \;+\; |\alpha|\,\big|b_n - \beta\big|.
\end{aligned}$$

By Proposition II.13, the convergent sequence $(b_n)_{n\in\mathbb{N}}$ is bounded, so there exists a constant $M > 0$ such that $|b_n| \leq M$ for all $n \in \mathbb{N}$. Plugging this in the above estimate, we obtain[13]

$$\big|a_n\, b_n - \alpha\beta\big| \;\leq\; M\,\big|a_n - \alpha\big| \;+\; |\alpha|\,\big|b_n - \beta\big|. \tag{II.11}$$

Now let $\varepsilon > 0$. Because $\lim_{n\to\infty} a_n = \alpha$ and $\frac{\varepsilon}{3M}$ is a positive number, there exists an $n' \in \mathbb{N}$ such that we have

$$|a_n - \alpha| \;<\; \frac{\varepsilon}{3M} \qquad \text{for } n \geq n'.$$

Similarly, because $\lim_{n\to\infty} b_n = \beta$ and $\frac{\varepsilon}{3\,(|\alpha|+1)}$ is a positive number, there exists an $n'' \in \mathbb{N}$ such that we have

$$|b_n - \beta| \;<\; \frac{\varepsilon}{3\,(|\alpha|+1)} \qquad \text{for } n \geq n''.$$

Then if $n \geq \max\{n', n''\}$, we have both $n \geq n'$ and $n \geq n''$, so both of the above estimates hold, and from (II.11) we get

$$\begin{aligned}
\big|a_n\, b_n - \alpha\beta\big| \;&\leq\; M\,\underbrace{\big|a_n - \alpha\big|}_{\leq \varepsilon/3M} + |\alpha|\;\underbrace{\big|b_n - \beta\big|}_{\leq \varepsilon/3(|\alpha|+1)} \\
&\leq\; \underbrace{\frac{M}{3M}}_{\leq 1/3}\,\varepsilon \;+\; \underbrace{\frac{|\alpha|}{3\,(|\alpha|+1)}}_{\leq 1/3}\,\varepsilon \\
&\leq\; \frac{2}{3}\,\varepsilon \;<\; \varepsilon.
\end{aligned}$$

[13]The idea now is that assumed convergence of $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ implies that both $|a_n - \alpha|$ and $|b_n - \beta|$ become small for n large. This should allow us to show that the expression (II.11) becomes small for large n. The rest of the proof is about making this idea precise.

If we set $n_\varepsilon = \max\{n', n''\}$, then from the above we got $\big|a_n\, b_n - \alpha\beta\big| < \varepsilon$ for all $n \geq n_\varepsilon$. Since $\varepsilon > 0$ was arbitrary, this by definition shows that $\lim_{n\to\infty}(a_n\, b_n) = \alpha\beta$, proving (II.9).

The proofs of (II.8) and (II.10) are left as exercises. □

Exercise II.9 (Completing the proof of Theorem II.19)**.**
Prove (II.8) and (II.10).

II.4. Density of rational and irrational numbers

Density of rational numbers

Theorem II.21 (Density of rational numbers)**.**
UNDER CONSTRUCTION!

Density of irrational numbers

Theorem II.22 (Density of irrational numbers)**.**
UNDER CONSTRUCTION!

II.5. Axioms of the real numbers

So far we have talked about real numbers, but we have not in fact precisely said what they are! This turns out to be not an entirely trivial philosophical matter.[14]

As with many other mathematical structures, the standard rigorous approach is to take certain properties of real numbers as axioms (statements that are simply accepted as true), and require that any other statement made about real numbers

[14] Issues related to the meaning of real numbers have also been a source of many historical controversies. A few famous ones, informally described, are the following.

- The Pythagoreans believed that only rational numbers exist — but they were aware of the irrationality of the length $\sqrt{2}$ of the diagonal of a unit square, which must have been a source of some significant cognitive dissonance!
- Cantor's diagonal argument showed that the set of all real numbers has a strictly bigger cardinality than the more familiar infinite set of all natural numbers. This conclusion was initially met with ridicule even by very reputable mathematicians.
- The Banach-Tarski paradox states that one cannot extend the notion of length on the real axis from intervals (an interval $[a, b] \subset \mathbb{R}$ is declared to have length $b - a$) to all subsets of $\mathbb{R}$ in such a way that the length satisfies reasonable additivity properties (it is a *measure*) and shifted versions of the same set have the same length. So not everything that we can intuitively think of (with the real numbers in particular) is logically possible!
- Gödel's (second) incompleteness theorem implies that it is not possible to prove that the axioms of real numbers do not lead to logical contradictions. This is in fact the case already in apparently more benign setups, such as the natural numbers — but should this be viewed as bad news for our approach to real numbers based on axioms?

Unsettling? No doubt questions about the foundations of mathematics quickly run into non-trivial philosophical dilemmas. But we shall be brave, like many mathematicans before us: we summon the courage to work with real numbers in spite of how terrifying they are!

is logically deduced from these axioms.[15] We want the axioms to be as modest as possible, so that the starting point is uncontroversial. And since everything else is deduced from the starting point by logical reasoning, also everything else will be uncontroversial — which of course is the point of mathematics!

The set of real numbers is denoted by $\mathbb{R}$, and it is equipped with two binary operations:

$$\begin{aligned} \text{addition “}+\text{”:} \qquad & \mathbb{R} \times \mathbb{R} \to \mathbb{R} \\ & (x, y) \mapsto x + y, \\ \text{multiplication “}\cdot\text{”:} \qquad & \mathbb{R} \times \mathbb{R} \to \mathbb{R} \\ & (x, y) \mapsto x \cdot y \\ & \text{(omitted in usual notation, so } x \cdot y = xy\text{)}, \end{aligned}$$

and a binary relation

$$\text{less than “}<\text{”.}$$

The axioms of real numbers fall into three types: the *field axioms* concern only the operations of addition and multiplication, the *order axioms* concern the order relation $<$ and how it behaves under addition and multiplication, and the *completeness axiom*, which is the most consequential one in terms of topology. Rational numbers, for example, obey exactly the same field axioms and order axioms, so the “only” difference between real numbers and rational numbers stems from the completeness axiom.

♡ Field axioms

The field axioms concern two operations of calculation with numbers: addition and multiplication, and we of course use the standard notational convention that multiplication takes precedence over addition.

[15] In addition to giving the axioms, it would be desirable to provide a (set theoretic) construction of real numbers, for which the axioms can be shown to be true. Suffice to say that there are such constructions, but we will not discuss them in detail here. Any such construction of course will admit something else as a starting point. An apparently benign starting point is to assume “uncontroversial” axioms of natural numbers (but in view of Gödel’s incompleteness theorem this is in fact not without its own issues!).

From the natural numbers, it is not difficult to construct the rational numbers. From the rational numbers, the real numbers can be constructed for example using what is known as Dedekind cuts. There are many other constructions as well. Importantly, a sort of uniqueness property of real numbers holds: any two constructions will yield results that are in all essential ways the same (isomorphic). So we do not need to care about the chosen construction. The axioms of real numbers themselves are indeed a good starting point.

The field axioms of real numbers state that the set $\mathbb{R}$ equipped with the operations $+$ and $\cdot$ is a field:

$$\begin{aligned}
&\text{Commutativity of addition:} && \\
&\forall x, y \in \mathbb{R}: && x + y = y + x \\
&\text{Associativity of addition:} && \\
&\forall x, y, z \in \mathbb{R}: && x + (y + z) = (x + y) + z \\
&\text{Neutral element for addition:} && \\
&\exists 0 \in \mathbb{R}: \ \forall x \in \mathbb{R}: && x + 0 = x \\
&\text{Opposite elements:} && \\
&\forall x \in \mathbb{R}: \ \exists (-x) \in \mathbb{R}: && x + (-x) = 0 \\
&\text{Commutativity of multiplication:} && \\
&\forall x, y \in \mathbb{R}: && x \cdot y = y \cdot x \\
&\text{Associativity of multiplication:} && \\
&\forall x, y, z \in \mathbb{R}: && x \cdot (y \cdot z) = (x \cdot y) \cdot z \\
&\text{Neutral element of multiplication:} && \\
&\exists 1 \in \mathbb{R} \setminus \{0\}: \ \forall x \in \mathbb{R}: && x \cdot 1 = x \\
&\text{Inverse elements:} && \\
&\forall x \in \mathbb{R} \setminus \{0\}: \ \exists x^{-1} \in \mathbb{R}: && x \cdot x^{-1} = 1 \\
&\text{Distributivity of addition over multiplication:} && \\
&\forall x, y, z \in \mathbb{R}: && x \cdot (y + z) = x \cdot y + x \cdot z
\end{aligned}$$

Hopefully the reader finds these uncontroversial enough to be admitted as axioms.

Note that the axioms explicitly state the existence of two real numbers, 0 ("zero", the neutral element for addition) and 1 ("one", the neutral element for multiplication), and they require that these two are different, $1 \neq 0$ (since $1 \in \mathbb{R} \setminus \{0\}$). Moreover, zero and one are uniquely determined by the properties required of them in the axioms.

Example II.23 (Uniqueness of zero).

Claim: If both $0 \in \mathbb{R}$ and $0' \in \mathbb{R}$ satisfy the property of being neutral elements for addition, then we have $0 = 0'$.

Proof: Suppose that $x + 0 = x$ and $x + 0' = x$ for all $x \in \mathbb{R}$. Then we get

$$0 \ = \ 0 + 0' \ = \ 0' + 0 \ = \ 0',$$

where we first used the neutral element property of $0'$, then commutativity of addition, and finally the neutral element property of 0. □

Exercise II.10 (Uniqueness of one).

Prove from the field axioms of real numbers that if both $1 \in \mathbb{R}$ and $1' \in \mathbb{R}$ satisfy the property of being neutral elements for multiplication, then we have $1 = 1'$.

Exercise II.11 (Uniqueness of opposite elements and inverse elemens).

Prove that for any $x \in \mathbb{R}$, the opposite element $(-x)$ is unique, and that for any $x \in \mathbb{R} \setminus \{0\}$, the inverse element x^{-1} is unique.

The following example indicates how yet a few other familiar facts about the real numbers are proved starting from the field axioms.

Example II.24 (Natural numbers from the field axioms).
Since we know that there exists a real number $1 \in \mathbb{R}$, we can use addition to define a new number $2 := 1+1 \in \mathbb{R}$. Similarly we define $3 := 2+1$, $4 := 3+1$, $5 := 4+1$, $6 := 5+1$, etc. These satisfy the usual properties. As an example, let us verify the following.

Claim: We have $2+2 = 4$.

Proof: Calculate, using the definition of 2, the associativity of addition, the definition of 3, and the definition of 4:

$$2+2 \;=\; 2+(1+1) \;=\; (2+1)+1 \;=\; 3+1 = 4.$$

This proves the claim. □

Exercise II.12 (Other exciting properties of natural numbers).
Prove some other exciting properties of these numbers, for example $2+3 = 5$ and $2 \cdot 3 = 6$.

Example II.25 (Yet another consequence of the field axioms).
We next verify a familiar property of real numbers which involves this number $2 := 1+1$ and an arbitrary real number x.

Claim: For any $x \in \mathbb{R}$ we have $x + x = 2x$.

Proof: Using the neutral element for multiplication (twice), distributivity, and definition of 2, we get the following expression for $x + x$:

$$x+x \;=\; 1\cdot x \,+\, 1\cdot x \;=\; (1+1)\cdot x \;=\; 2\cdot x.$$

The claim follows. □

Exercise II.13 (Multiplication by zero).
Prove that for any $x \in \mathbb{R}$ we have $0 \cdot x = 0$.

Ok, you get the point! The familiar facts about the real numbers that you knew since kindergarten can be deduced from the axioms. In order that we get somewhere in this course, we will not actually require you to provide detailed proofs of totally commonplace statements such as $0\cdot x = 0$ or $2\cdot 2 = 4$ or $(-1)^2 = 1$ or $(x-y)^2 = x^2 - 2xy + y^2$ — as long as you realize that *in principle* they should be obtained as logical consequences of the axioms (and familiar notational conventions and definitions).[16]

Other fields

There are many other fields besides $\mathbb{R}$, and they all satisfy the field axioms. In this course we do not use other fields than $\mathbb{R}$ and $\mathbb{Q}$, but for perspective we mention a few prominent examples of fields:

[16] Nowadays there are actually formally verified proofs of a vast amount of basic statements about real numbers and even metric space topology; certainly including the ones here. This means that the statements have been phrased in an entirely formal language, proofs have also been given in the formal language, and a computer has verified that the proofs adhere to the rules of logic and actually demonstrate the statements. I dare to doubt, however, that all of that makes for much more pleasant reading than these lecture notes... Moreover, producing the fully formal proofs is so tedious that already in this undergraduate course we actually get deeper into mathematics than where hundreds of mathematicians and computer scientists have gotten in the past decades with formally verified mathematics.

- the field of rational numbers $\mathbb{Q}$;
- the field $\mathbb{Q}(\sqrt{5})$ of rational numbers adjoined with a square root of 5;
- the field of complex numbers $\mathbb{C}$;
- the finite field $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ of p elements, where p is a prime number;
- the finite field $\mathbb{F}_{p^k}$ of p^k elements, where p is a prime number and $k \in \mathbb{N}$;
- the field $\mathbb{Q}_p$ of p-adic numbers, where p is a prime number;
- the field $\mathbb{K}(q)$ of rational functions in a single variable q over another field $\mathbb{K}$;
- $\cdots$

In particular all (familiar) consequences of the field axioms hold for these as well: for example if $z, w \in \mathbb{C}$ are complex numbers, then $(z + w) = z^2 + 2zw + w^2$, where $2 := 1 + 1 \in \mathbb{C}$. Note, however, that for instance in the two element field $\mathbb{F}_2$, we have $2 = 0$. So you should not try to derive the property $2 \neq 0$ of real numbers from the field axioms (you would fail, it is not a logical consequence of them!).

♡ Order axioms

Since there are other fields besides just the real numbers, the real numbers must have some further specific properties that distinguish them. Having an order relation $<$ turns out to be quite special already. To work with the order relation $<$, we adopt the usual additional notations such as: $y > x$ means $x < y$, $x \leq y$ means that either $x < y$ or $x = y$, etc.

The order axioms of the real number field are:

$$\begin{aligned}
&\text{Alternatives:} && \\
&\forall x, y \in \mathbb{R} : && \text{exactly one of the relations} \\
& && x < y,\ x = y,\ x > y \text{ holds} \\
&\text{Transitivity:} && \\
&\forall x, y, z \in \mathbb{R} : && x < y \text{ and } y < z \quad \Longrightarrow \quad x < z \\
&\text{Compatibility with addition:} && \\
&\forall x, y, z \in \mathbb{R} : && x < y \quad \Longrightarrow \quad x + z < y + z \\
&\text{Compatibility with multiplication:} && \\
&\forall x, y \in \mathbb{R} : && x > 0 \text{ and } y > 0 \quad \Longrightarrow \quad x \cdot y > 0
\end{aligned}$$

From these and the field axioms together, one can derive consequences.

Example II.26 (Some consequences of the order axioms)**.**

Claim: We have $0 < 1$.

Proof: Since we know that $0 \neq 1$, there are two mutually exclusive alternatives left: $0 < 1$ or $0 > 1$. Let us prove that the second one is impossible. So assume that $0 > 1$. By adding -1 and using compatibility with addition, we get (after simplifying by neutral element of addition and opposite element properties) that $-1 > 0$. Then by compatibility with multiplication we get $(-1)(-1) > 0$. But from the field axioms one can show that $(-1)(-1) = 1$, so this simplifies to $1 > 0$. In this case we have both $1 > 0$ and $0 > 1$, which is not allowed by the alternatives. So we have to discard the possibility that $0 > 1$, and we conclude that $0 < 1$ holds. □

As the example of the field of two elements shows, the property $2 \neq 0$ of real numbers cannot be derived from the field axioms alone. Having also the order axioms, we can now prove it.

Claim: We have $2 \neq 0$ in $\mathbb{R}$.

Proof: We already know that $1 > 0$. Adding 1 and using compatibility with addition, we find $2 > 1$. Transitivity then implies $2 > 0$. This rules out the possibility that $2 = 0$ by the axiom that exactly one of the alternatives $2 > 0$, $2 = 0$, or $2 < 0$ holds. □

Exercise II.14 (Product of two negative numbers).
Prove that the product of any two negative real numbers is positive.

Ok, you get the point again. All the familiar properties you already knew in kindergarten are logical consequences of the axioms.

Note that the rational number field $\mathbb{Q}$ also satisfies the order axioms. Therefore if $\mathbb{R} \neq \mathbb{Q}$ there must be yet something else that distinguishes the real numbers.

Completeness axiom

After the field axioms and order axioms, which probably look like a totally paranoid and hopeless approach to math[17], we now come to the one really subtle axiom of the real numbers: the completeness axiom. But before stating it, we first need to introduce a few notions that will make an appearance here.

Supremum and infimum

Let us start from notions of boundedness for subsets of the real line. If you compare the following with Definition II.8, a common theme becomes evident.

Definition II.27 (Upper and lower bounds).
Let $A \subset \mathbb{R}$ be a subset.

A real number $t \in \mathbb{R}$ is an **upper bound** for A if for all $a \in A$ we have $a \leq t$.

A real number $s \in \mathbb{R}$ is a **lower bound** for A if for all $a \in A$ we have $a \geq s$.

Example II.28 (Examples of upper and lower bounds).

- Consider the open interval $(-\pi, +\pi) \subset \mathbb{R}$. The number π is an upper bound for the set $(-\pi, +\pi)$. But also any number $t > \pi$ is an upper bound for $(-\pi, +\pi)$. Likewise, any number $s \leq -\pi$ is a lower bound for the set $(-\pi, +\pi)$.
- Consider the set $\mathbb{N} \subset \mathbb{R}$ of natural numbers. The number 1 is a lower bound for $\mathbb{N}$, and so is any number $s < 1$. The set $\mathbb{N}$ does not have any upper bounds: for any $t \in \mathbb{R}$ there exists a natural number $n \in \mathbb{N}$ such that $n > t$.
- The sets $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ do not have any upper or lower bounds.
- Consider the empty set $\emptyset \subset \mathbb{R}$. Any real number $t \in \mathbb{R}$ is an upper bound for $\emptyset$.[18] Likewise, any real number $s \in \mathbb{R}$ is a lower bound for $\emptyset$.

From these examples we learn that upper and lower bounds for arbitrary sets do not need to exist. Moreover, when they exist, they are not unique.

[17] Still to be fair, they may not be the quickest approach to fancy calculations, but they have value. They bring clarity of the philosophical position to mathematics, and they actually feature in mathematically profound generalizations (for example the p-adic number fields).

[18] Indeed, there are no elements $a \in \emptyset$ for which we would get a nontrivial condition that t needs to satisfy to qualify as an upper bound for the empty set.

Definition II.29 (Sets bounded from above and below).

A subset $A \subset \mathbb{R}$ is **bounded from above** if it has an upper bound, and **bounded from below** if it has a lower bound.

The following definition introduces the key notion in the completeness axiom: the least upper bound or supremum. The supremum and its counterpart infimum also feature in a number of contexts in real analysis — we will mention some applications in Section II.6.

Definition II.30 (Supremum and infimum).

Let $A \subset \mathbb{R}$ be a nonempty subset ($A \neq \emptyset$) which is bounded from above. A number $t_0 \in \mathbb{R}$ is called the **least upper bound** or **supremum** of A, if t_0 is an upper bound for A (i.e. $\forall a \in A : \ a \leq t_0$) and for all upper bounds t of A we have $t_0 \leq t$. We then denote $t_0 = \sup A$.

Let $B \subset \mathbb{R}$ be a nonempty subset ($B \neq \emptyset$) which is bounded from below. A number $s_0 \in \mathbb{R}$ is called the **greatest lower bound** or **infimum** of B, if s_0 is a lower bound for B (i.e. $\forall b \in B : \ b \geq s_0$) and for all lower bounds s of B we have $s_0 \geq s$. We then denote $s_0 = \inf B$.

Implicit in the notations $\sup A$ and $\inf A$ (and in our use of the definite article above) is that the least upper bound and the greatest lower bound are uniquely determined by the above properties. This is indeed not difficult to check, and we leave it as an exercise to the reader.

The notions of supremum and infimum are crucial in this course as well as mathematics more generally, so let us give a more practical characterization.

Lemma II.31 (Characterization of supremum).

A number $t_0 \in \mathbb{R}$ is the supremum of a nonempty subset $A \subset \mathbb{R}$ which is bounded from above, if and only if the following two conditions hold:

- *For all $a \in A$ we have $a \leq t_0$;*[19]
- *For all $\varepsilon > 0$ there exists some $a \in A$ such that $a > t_0 - \varepsilon$.*[20]

Exercise II.15 (Proof of Lemma II.31).

Prove Lemma II.31.

Hint: *Compare the defining properties of supremum in Definition II.30 with the interpretations of the two conditions in the lemma, given in the footnotes.*

Exercise II.16 (Characterization of infimum).

Formulate and prove a the counterpart of Lemma II.31 for infimum, i.e., two conditions characterizing the infimum of a non-empty set $B \subset \mathbb{R}$ which is bounded from below.

Supremum and infimum are generalizations of the notions of maximum (i.e., the largest element) and minimum (i.e., the smallest element). For comparison, let us look at these more familiar notions.

[19]This condition simply says that t_0 is an upper bound for A.

[20]This condition says that any smaller number $t_0 - \varepsilon$ can not be an upper bound for A.

Definition II.32 (Maximum and minimum).
Let $A \subset \mathbb{R}$ be a subset.

An element $a' \in A$ is the **maximum** (i.e. the **largest element**) of A, if for all $a \in A$ we have $a \leq a'$. We then denote $a' = \max A$.

An element $a'' \in A$ is the **minimum** (i.e. the **smallest element**) of A, if for all $a \in A$ we have $a \geq a''$. We then denote $a'' = \min A$.

The crucial difference between the maximum and the supremum (or an upper bound more generally) is that the maximum of A is required to be an element of the set A itself. A similar remark applies to the difference between the minimum and the infimum.

Note, however, that not every subset $A \subset \mathbb{R}$ has a maximum (resp. minimum) — even if we assume A to be non-empty and bounded from above (resp. from below).

Example II.33 (Open intervals have no maximum and minimum).
The open interval $(-\pi, \pi)$ has no largest element and no smallest element ($\nexists \max(-\pi,\pi)$, $\nexists \min(-\pi,\pi)$). In fact, open intervals never have a maximum or minimum.

Example II.34 (A bounded nonempty set).
The set $A = \left\{ \frac{1}{n} \;\middle|\; n \in \mathbb{N} \right\}$ is bounded and non-empty. The largest element of this set is $\max A = 1$. The set A has no smallest element ($\nexists \min A$). The least upper bound of this set is $\sup A = 1$ and the greatest lower bound of it is $\inf A = 0$. Note that in this case the infimum is not an element of the set, $0 \notin A$.

Exercise II.17 (The relationship between supremum and maximum).
Let $A \subset \mathbb{R}$.

(a) Prove that if $\max A$ exists, then we have $\sup A = \max A$.
(b) Prove that if $A \subset \mathbb{R}$ and $\sup A \in A$ (so in particular the supremum exists), then the maximum exists and we have $\max A = \sup A$.

Exercise II.18 (The relationship between infimum and minimum).
Formulate and prove an analogous result for the infimum and minimum.

Formulations of the completeness axiom

There are various equivalent ways of formulating the completeness axiom of $\mathbb{R}$. We first state three formulations, and later discuss their equivalence.

The **completeness axiom** of $\mathbb{R}$ refers to any of the following (equivalent) statements:

(C1): Every non-empty subset $A \subset \mathbb{R}$ which is bounded from above has a least upper bound $\sup A \in \mathbb{R}$.
(C2): Every increasing real number sequence $(a_n)_{n \in \mathbb{N}}$ which is bounded from above has a limit $\lim_{n\to\infty} a_n \in \mathbb{R}$.
(C3): Every collection $(I_n)_{n \in \mathbb{N}}$ of closed intervals $I_n \subset \mathbb{R}$, which is nested in the sense that $I_{n+1} \subset I_n$ for every $n \in \mathbb{N}$, has a nonempty intersection

$$\bigcap_{n \in \mathbb{N}} I_n \neq \emptyset.$$

Furthermore, it is easy to see that the first two formulations above are equivalent with the following two, respectively:

- **(C1'):** Every non-empty subset $A \subset \mathbb{R}$ which is bounded from below has a greatest lower bound $\inf A \in \mathbb{R}$.
- **(C2'):** Every decreasing real number sequence $(a_n)_{n\in\mathbb{N}}$ which is bounded from below has a limit $\lim_{n\to\infty} a_n \in \mathbb{R}$.

Exercise II.19 (Equivalence of (C1) and (C1')).
Prove that (C1) and (C1') are equivalent.

Exercise II.20 (Equivalence of (C2) and (C2')).
Prove that (C2) and (C2') are equivalent.

As an example application of the formulation (C2) of the completeness axiom, we discuss the existence of real numbers given by decimal expansions.

Example II.35 (Decimal expansions).
Let $(d_1, d_2, d_3, \ldots) = (d_k)_{k\in\mathbb{N}}$ be a sequence of numbers $d_k \in \{0, 1, 2, \ldots, 8, 9\}$. The precise definition of the **decimal number**

$$0.d_1 d_2 d_3 \ldots$$

is the sum of the series

$$\sum_{k=1}^{\infty} d_k\, 10^{-k}. \tag{II.12}$$

In particular, the existence of this decimal number requires the series to be convergent.

The partial sums of the series

$$a_n = \sum_{k=1}^{n} d_k\, 10^{-k}$$

clearly form an increasing sequence $(a_n)_{n\in\mathbb{N}}$, because

$$a_{n+1} - a_n \;=\; \sum_{k=1}^{n+1} d_k\, 10^{-k} - \sum_{k=1}^{n} d_k\, 10^{-k} \;=\; d_{n+1}\, 10^{-n-1} \;\geq\; 0.$$

The sequence of the partial sums is also bounded from above, because

$$\begin{aligned}
a_n \;&=\; \sum_{k=1}^{n} d_k\, 10^{-k} \\
&\leq\; \sum_{k=1}^{n} 9\cdot 10^{-k} && \text{(since } d_k \leq 9\text{)} \\
&=\; \frac{9}{10} \sum_{\ell=0}^{n-1} 10^{-\ell} && \text{(change summation index to } \ell = k-1\text{)} \\
&=\; \frac{9}{10}\, \frac{1-10^{-n}}{1-\frac{1}{10}} && \text{(finite geometric sum)} \\
&\leq\; \frac{9}{10}\, \frac{1}{9/10} \;=\; 1.
\end{aligned}$$

According to the formulation (C2) of the completeness axiom, the sequence $(a_n)_{n\in\mathbb{N}}$ of partial sums converges, because it is increasing and bounded from above. This means that the series (II.12) converges.

We conclude that every decimal expansion indeed represents some real number.

♡ *Equivalence of the formulations of the completeness axiom*

Let us now prove the equivalence of (C1), (C2), and (C3). We prove separately the implications

$$(C1) \implies (C2), \qquad (C2) \implies (C3), \qquad (C3) \implies (C1).$$

The equivalence of all three follows by combining these implications.

Proof of (C1) ⇒ (C2). Assume (C1). Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of real numbers which is increasing and bounded from above. We must prove that this sequence converges.

Since the sequence $(a_n)_{n\in\mathbb{N}}$ is bounded from above, the set

$$A = \big\{a_n \;\big|\; n \in \mathbb{N}\big\}$$

of its values is a subset in $\mathbb{R}$ which is bounded from above. By assumption (C1), then, this subset has a least upper bound $t_0 := \sup\big\{a_n \;\big|\; n \in \mathbb{N}\big\} \in \mathbb{R}$. We will prove that the sequence $(a_n)_{n\in\mathbb{N}}$ converges to t_0.

Let $\varepsilon > 0$. Since $t_0 - \varepsilon$ is not an upper bound for $\big\{a_n \;\big|\; n \in \mathbb{N}\big\}$, there exists some $n_\varepsilon \in \mathbb{N}$ such that $a_{n_\varepsilon} > t_0 - \varepsilon$. Since the sequence is increasing, for all $n \geq n_\varepsilon$ we must then also have $a_n > t_0 - \varepsilon$. On the other hand, since t_0 is an upper bound for $\big\{a_n \;\big|\; n \in \mathbb{N}\big\}$, we have $a_n \leq t_0$ for all $n \in \mathbb{N}$. For $n \geq n_\varepsilon$ we have thus obtained

$$t_0 - \varepsilon \;<\; a_n \;\leq\; t_0 \;<\; t_0 + \varepsilon,$$

which implies $|a_n - t_0| < \varepsilon$. Since such an $n_\varepsilon \in \mathbb{N}$ was found for an arbitrary $\varepsilon > 0$, we have by definition of limits shown that $\lim_{n\to\infty} a_n = t_0$.

Property (C2) is thus established. □

Proof of (C2) ⇒ (C3). Assume (C2). Let $(I_n)_{n\in\mathbb{N}}$ be a collection of closed intervals $I_n = [a_n, b_n] \subset \mathbb{R}$ such that $I_{n+1} \subset I_n$ for every $n \in \mathbb{N}$. We must prove that $\bigcap_{n\in\mathbb{N}} I_n \neq \emptyset$.

The nesting condition

$$[a_{n+1}, b_{n+1}] \;=\; I_{n+1} \;\subset\; I_n \;=\; [a_n, b_n]$$

simply amounts to the following inequalities among the endpoints of the intervals:

$$a_n \;\leq\; a_{n+1} \;\leq\; b_{n+1} \;\leq\; b_n.$$

Therefore the sequence $(a_n)_{n\in\mathbb{N}}$ of the left endpoints of the intervals is increasing. This sequence is also bounded from above, because for any $n \in \mathbb{N}$ we have $a_n \leq b_n \leq b_1$ (the second inequality follows from the fact that the sequence of right endpoints is decreasing). By assumption (C2), the sequence $(a_n)_{n\in\mathbb{N}}$ therefore has a limit $\alpha = \lim_{n\to\infty} a_n$. We will show that $\alpha \in \bigcap_{n\in\mathbb{N}} I_n$, and it will follow that the intersection is nonempty.

To show that $\alpha \in \bigcap_{m\in\mathbb{N}} I_m$, by definition of intersection we must show that $\alpha \in I_m$ for all $m \in \mathbb{N}$. Since $I_m = [a_m, b_m]$, this amounts to proving that

$$a_m \;\overset{?}{\leq}\; \alpha \;\overset{?}{\leq}\; b_m.$$

The first inequality above is clear: for $n \geq m$ we have $a_n \geq a_m$ (the left endopints form an increasing sequence), so by preservation of constant inequalities (Corollary II.15) we have $\alpha = \lim_{n\to\infty} a_n \geq a_m$. To prove the other inequality, we argue by contradiction. Suppose that it is not true, i.e., that $\alpha > b_m$. By definition of the limit $\alpha = \lim_{n\to\infty} a_n$, for the positive number $\varepsilon = \alpha - b_m > 0$, there exists an $n_\varepsilon \in \mathbb{N}$ so that for $n \geq n_\varepsilon$ we have $|a_n - \alpha| < \varepsilon = \alpha - b_m$, which implies

$$a_n > \alpha - \varepsilon = \alpha - (\alpha - b_m) = b_m.$$

In particular for $n \geq \max\{n_\varepsilon, m\}$, since the sequence of right endpoints is decreasing, we would get $a_n > b_m \geq b_n$. This, however, is a contradiction with the ordering $a_n \leq b_n$ of the left and right endpoints of $I_n = [a_n, b_n]$. With this contradiction, we have also established the other claimed inequality. We have thus concluded $\alpha \in I_m$ for all $m \in \mathbb{N}$, which shows that the intersection $\bigcap_{m\in\mathbb{N}} I_m$ is not empty, establishing (C3). □

Proof of (C3) ⇒ (C1). Assume (C3). Let $A \subset \mathbb{R}$ be a nonempty subset which is bounded from above. We must prove that A has a least upper bound.

Since A is bounded from above, it has an upper bound, and since A is moreover nonempty, it is possible to choose an upper bound $b_0 \in \mathbb{R}$ so that the number $a_0 = b_0 - 1$ is not an upper bound for A.

Consider the midpoint $c_0 = \frac{a_0+b_0}{2}$. If c_0 is an upper bound for A, we set $a_1 = a_0$ and $b_1 = c_0$. Otherwise we set $a_1 = c_0$ and $b_1 = b_0$. With such choices, we are guaranteed that b_1 is an upper bound for A, while a_1 is not.

Continue inductively. When b_n is an upper bound for A and a_n is not, consider the midpoint $c_n = \frac{a_n+b_n}{2}$. Depending on whether c_n is an upper bound for A or not, set $a_{n+1} = a_n$ and $b_{n+1} = c_n$, or $a_{n+1} = c_n$ and $b_{n+1} = b_n$. By construction we get a decreasing sequence $(b_n)_{n\in\mathbb{N}}$ of upper bounds for A, and an increasing sequence $(a_n)_{n\in\mathbb{N}}$ of numbers that are not upper bounds for A, and moreover $a_n < b_n$ for all $n \in \mathbb{N}$. The distances are halved at every step, so $b_n - a_n = 2^{-n}$.

From the above properties we get that the closed intervals $I_n = [a_n, b_n]$ for $n \in \mathbb{N}$ are nested, $I_{n+1} \subset I_n$ for all $n \in \mathbb{N}$. By the assumed property (C3), the intersection $\bigcap_{n\in\mathbb{N}} I_n$ is then nonempty. On the other hand, the intersection cannot contain two distinct points: if $x, y \in \bigcap_{n\in\mathbb{N}} I_n$, then for all $n \in \mathbb{N}$ we have $x, y \in I_n = [a_n, b_n]$, which implies that $|x - y| \leq b_n - a_n = 2^{-n} \to 0$, so $|x - y| = 0$ and $x = y$. We conclude that the intersection contains exactly one point ξ, i.e., it is a singleton $\bigcap_{n\in\mathbb{N}} I_n = \{\xi\}$. From the same estimate we also get that

$$\xi - 2^{-n} \;\leq\; a_n \;\leq\; \xi \;\leq\; b_n \;\leq\; \xi + 2^{-n}.$$

By the squeeze theorem (Lemma II.17), this implies $\lim_{n\to\infty} a_n = \xi$ and $\lim_{n\to\infty} b_n = \xi$.

Our goal was to show that the set A has a least upper bound. We will show that ξ is it.

Let us first verify that ξ is an upper bound for A. So let $x \in A$. By construction each b_n is an upper bound for A, so $b_n \geq x$ for all $n \in \mathbb{N}$. Since we have $\xi = \lim_{n\to\infty} b_n$, from the preservation of bounds (Corollary II.15) we get that $\xi \geq x$. This was true for an arbitrary element $x \in A$, so indeed ξ is an upper bound for A.

Let us then show that no smaller number $\xi' < \xi$ is an upper bound for A. Since $\lim_{n\to\infty} a_n = \xi$, for any $\xi' < \xi$ we have $a_n > \xi'$ for large enough n (take $\varepsilon = \xi - \xi' > 0$ in the definition of limit), and since a_n was not an upper bound for A, the smaller number $\xi' < a_n$ can not be either.

This proves that $\sup A = \xi$ exists, establishing (C1). □

Conventions to extend the notion of supremum and infimum

In Definition II.30 we considered the supremum only for subsets A which are nonempty and bounded from above. If A fails one of these properties, we will use the following conventions:

- If a set $A \subset \mathbb{R}$ is not bounded from above (and thus has no upper bounds), we interpret the least upper bound as the symbol $+\infty$ (not a number),

$$\sup A \;=\; +\infty.$$

- For the empty set $\emptyset \subset \mathbb{R}$ (for which any real number is an upper bound), we interpret the least upper bound as the symbol $-\infty$ (not a number),

$$\sup \emptyset \;=\; -\infty.$$

Symmetrically for the infimum we use the conventions:

- If a set $A \subset \mathbb{R}$ is not bounded from below (and thus has no lower bounds), we interpret the greatest lower bound as the symbol $-\infty$ (not a number),
$$\inf A \; = \; -\infty.$$
- For the empty set $\emptyset \subset \mathbb{R}$ (for which any real number is a lower bound), we interpret the greatest lower bound as the symbol $+\infty$ (not a number),
$$\inf \emptyset \; = \; +\infty.$$

II.6. ♡ Applications of supremum and infimum

The notions of supremum and infimum may seem a bit abstract. Indirectly, via a different formulation of the completeness axiom, we saw that supremum is important for instance for the existence of real numbers with given decimal expansions (Example II.35). But it is good to realize that supremum is also quite directly used in many common constructions in analysis. We describe a few examples here, but the intention is not to elaborate on the details in full.

♡ *Riemann integration*

Let
$$f \colon [a,b] \to \mathbb{R}$$
be a real valued function defined on a closed interval $[a,b]$. Assume also that f is bounded in the sense that there exists some $M \in \mathbb{R}$ such that $|f(x)| \leq M$ for all $x \in [a,b]$.

Consider a finite subdivision of the interval $[a,b]$, consisting of points
$$a = x_0 < x_1 < x_2 < \cdots < x_n = b.$$
On the subdivision interval $[x_{j-1}, x_j]$ for $j \in \{1, \ldots, n\}$, the values of the function have the greatest lower bound and least upper bound
$$s_j = \inf \Big\{ f(x) \;\Big|\; x_{j-1} \leq x \leq x_j \Big\} \qquad \text{and} \qquad t_j = \sup \Big\{ f(x) \;\Big|\; x_{j-1} \leq x \leq x_j \Big\}.$$
Using these, we define the lower and upper Riemann sums
$$S = \sum_{j=1}^{n} s_j \left(x_j - x_{j-1} \right) \qquad \text{and} \qquad T = \sum_{j=1}^{n} t_j \left(x_j - x_{j-1} \right)$$
associated with the subdivision $a = x_0 < x_1 < x_2 < \cdots < x_n = b$.

It is not particularly difficult to show that the lower and upper Riemann sums satisfy the following:

- If a subdivision $a = x'_0 < x'_1 < x'_2 < \cdots < x'_m = b$ is a refinement of the above subdivision (i.e. $\{x_0, x_1, \ldots, x_n\} \subset \{x'_0, x'_1, \ldots, x'_m\}$) and S' and T' are the associated lower and upper Riemann sums, then $S \leq S'$ and $T' \leq T$.
- If $\tilde{S}$ and $\tilde{T}$ are the lower and upper Riemann sums associated with any subdivision, then $\tilde{S} \leq T$ (and symmetrically $S \leq \tilde{T}$).

Exercise II.21 (♡ Proof of the properties of upper and lower Riemann sums)**.**
Prove the two claims above.

Since increasingly fine subdivisions lead to larger lower Riemann sums, it is natural to define the lower integral $I_-(f)$ as the least upper bound for the set of all lower Riemann sums for all subdivisions; or explicitly

$$I_-(f) := \sup \left\{ \sum_{j=1}^{n} \Big(\inf_{x \in [x_{j-1}, x_j]} f(x) \Big) \big(x_j - x_{j-1} \big) \;\middle|\; a = x_0 < x_1 < \ldots < x_n = b \right\}.$$

Similarly, since increasingly fine subdivisions lead to smaller upper Riemann sums, it is natural to define the upper integral $I_+(f)$ as the greatest lower bound for the set of all upper Riemann sums for all subdivisions; or explicitly

$$I_+(f) := \inf \left\{ \sum_{j=1}^{n} \Big(\sup_{x \in [x_{j-1}, x_j]} f(x) \Big) \big(x_j - x_{j-1} \big) \;\middle|\; a = x_0 < x_1 < \ldots < x_n = b \right\}.$$

From the earlier observations one straightforwardly gets $I_-(f) \leq I_+(f)$.

Definition II.36 (Riemann integral)**.**
A function $f\colon [a,b] \to \mathbb{R}$ is said to be **(Riemann) integrable**, if $I_-(f) = I_+(f)$. In this case we define its **(Riemann) integral** as

$$\int_a^b f(x)\,\mathrm{d}x \;:=\; I_-(f) \;=\; I_+(f).$$

From characterizations of supremum and infimum, it is quite straightforweard to prove that

Theorem II.37 (Characterization of Riemann integrability)**.**
A function $f\colon [a,b] \to \mathbb{R}$ is Riemann integrable if and only if for every $\varepsilon > 0$ there exists a subdivision such that the associated upper and lower Riemann sums satisfy $T - S < \varepsilon$.

Exercise II.22 (♡ Proof of Theorem II.37)**.**
Prove the above theorem.

To properly define the familiar integral, therefore, we end up using the notions of supremum and infimum — in fact many times.

♡ *Infinite sums with non-negative terms*

If we have an arbitrary collection $(s_i)_{i \in I}$ of nonnegative terms $s_i \geq 0$, then there is a quick and rather simple way to give a meaning to the sum

$$\sum_{i \in I} s_i.$$

This is rather remarkable, since we can allow the index set I to be arbitrarily large, meaning that the number of terms in the sum can be finite, countably infinite, or even uncountably infinite.

Of course a partial sum that includes only finitely many terms $s_{i_1}, \ldots, s_{i_k}$, with $i_1, \ldots, i_k \in I$ distinct, is trivially defined as

$$s_{i_1} + \cdots + s_{i_k}.$$

The idea is that the other terms omitted are non-negative, so the "true sum" must be at least as large. In other words the "true sum" should be an upper bound for the set of all finite partial sums. The definition

$$\sum_{i \in I} s_i \;:=\; \sup \left\{ s_{i_1} + \cdots + s_{i_k} \;\middle|\; k \in \mathbb{N}_0,\ i_1, \ldots, i_k \in I \text{ distinct} \right\}$$

declares the "true sum" as the least upper bound for the set of all finite partial sums — i.e., it requires that the gap in between the finite partial sums and the "true sum" can be made arbitrarily small by including enough terms.

The set of all finite partial sums is nonempty (we always have the possibility of including $k = 0$ terms in the partial sum), but it may or may nor be bounded, so we obviously have to allow the possibility $\sum_{i \in I} s_i = +\infty$ (in keeping with the convention of supremum of unbounded sets) if arbitrarily large finite partial sums exist.

This definition also has the advantage that it is clear that the sum does not depend on the "order of elements" (in fact we did not even require the index set I to be ordered!).

But to be fair, this definition crucially relies on the terms being nonnegative. It does not admit easy generalizations to infinite sums with terms of both signs.[21]

Exercise II.23 (♡ Uncountably many strictly positive terms yields infinite sum)**.**
Let $(s_i)_{i \in I}$ be a collection of nonnegative terms $s_i \geq 0$. Consider the sum $\sum_{i \in I} s_i$.

(a) Suppose that $\sum_{i \in I} s_i < +\infty$. Show that for any $m \in \mathbb{N}$, there can only exist finitely many indices $i \in I$ such that $s_i \geq \frac{1}{m}$.

(b) Suppose again that $\sum_{i \in I} s_i < +\infty$. Show that there can only exist countably many indices $i \in I$ such that $s_i > 0$.
Hint: *Consider all different $m \in \mathbb{N}$ in (a). Recall that countable unions of finite sets are countable.*

(c) Now prove that if there are uncountably infinitely many indices i corresponding to a strictly positive term $s_i > 0$, then we have $\sum_{i \in I} s_i = +\infty$.
Hint: *Contrapositive of (b).*

II.7. ♡ Uncountability of real numbers

UNDER CONSTRUCTION!

Proof of Theorem I.38. UNDER CONSTRUCTION! □

♡ Cantor set

UNDER CONSTRUCTION!

[21] Somewhat more general sums (and integrals) are studied in measure theory.

Lecture III

Sequences and functions on the real line

The main theme of this section is continuous functions of a real variable — certainly to some extent a familiar topic already. We will in particular verify that polynomials and rational functions are continuous. We will examine continuous functions on closed intervals, and show that they always have a maximum and a minimum — a crucial existence result for optimization tasks[1], for example — and especially that they are bounded. We also look into the intermediate value theorem (Bolzano's theorem), according to which continuous functions on intervals "can not skip values".

One of the main objectives is to precisely define the notion of continuity, and to rigorously prove the above (probably familiar) facts. Another goal is to start drawing attention to the underlying topological reasons behind such important results: later in the course it is possible to appreciate the role of compactness (of closed intervals) and connectedness (of intervals) in the results.

As a tool we will still use real number sequences, and in particular their judiciously chosen subsequences. A key role is played by the fundamental fact that from a bounded sequence it is always possible to pick some convergent subsequence (a fact that makes use of the completeness axiom of the real number field, and will itself in later chapters be recognized as a compactness statement).

III.1. Real number sequences

Subsequences

UNDER CONSTRUCTION!

Lemma III.1 (All subsequences of a convergent sequence have the same limit)**.**
Suppose that $(x_n)_{n\in\mathbb{N}}$ *is a real number sequence which converges to a limit* $\lim_{n\to\infty} x_n = x$*. Then any subsequence* $(x_{\varphi(n)})_{n\in\mathbb{N}}$ *also converges to the same limit,* $\lim_{n\to\infty} x_{\varphi(n)} = x$*.*

Exercise III.1 (Proof of Lemma III.1)**.**
Prove Lemma III.1.

Example III.2 (The alternating sign sequence)**.**
Consider the sequence $(a_n)_{n\in\mathbb{N}}$ where $a_n = (-1)^n$, i.e.,

$$(a_n)_{n\in\mathbb{N}} = (-1, +1, -1, +1, -1, \ldots).$$

[1] Mathematically, optimization is formulated as the task of finding the maximum (or minimum) of some objective function: e.g., "profit", "accuracy", "efficiency", ... (or "cost", "error", "risk", ...). If the maximum (or minimum) does not exist, the problem is not even well posed and it is meaningless to look for the optimum!

The subsequence of even members, obtained with $\varphi_{\text{even}}(n) = 2n$, is

$$(a_{2n})_{n\in\mathbb{N}} = (+1, +1, +1, +1, \ldots),$$

and it obviously converges to $+1$.

The subsequence of odd members, obtained with $\varphi_{\text{odd}}(n) = 2n - 1$, is

$$(a_{2n-1})_{n\in\mathbb{N}} = (-1, -1, -1, -1, \ldots),$$

and it obviously converges to -1.

The fact that these two subsequences have different limits implies (by Lemma III.1) that the original sequence $(a_n)_{n\in\mathbb{N}}$ does not converge. (Of course this could also be proven directly!)

Example III.3 (A sequence without convergent subsequences)**.**
Consider the sequence $(a_n)_{n\in\mathbb{N}}$ where $a_n = -n$, i.e.,

$$(a_n)_{n\in\mathbb{N}} = (-1, -2, -3, -4, \ldots).$$

It is easy to see that this sequence has no convergent subsequences (in fact it has no bounded subsequences, so this follows from Proposition II.13).

Existence of monotone subsequences

Proposition III.4 (Any real number sequence has a monotone subsequence)**.**
UNDER CONSTRUCTION!

Proof. UNDER CONSTRUCTION! □

Bounded sequences have convergent subsequences

Theorem III.5 (A bounded real number sequence has a convergent subsequence)**.**

UNDER CONSTRUCTION!

Proof. Let $(x_n)_{n\in\mathbb{N}}$ be a real number sequence which is bounded. By Proposition III.4, this sequence has a monotone subsequence $(x_{\varphi(n)})_{n\in\mathbb{N}}$. But this subsequence is also bounded (by the same constant as the original sequence). As a bounded monotone sequence of real numbers, $(x_{\varphi(n)})_{n\in\mathbb{N}}$ converges by the completeness axiom. This shows that $(x_n)_{n\in\mathbb{N}}$ has a convergent subsequence. □

III.2. Functions of real variable

We will now consider real-valued functions of a real variable. Being a function of a real variable means that we take the domain of the function to be a subset $A \subset \mathbb{R}$ of the real axis. Real-valuedness means that we (can) take the codomain of the functions to be $\mathbb{R}$. Therefore we are interested in functions

$$f : A \to \mathbb{R} \qquad \text{where } A \subset \mathbb{R}.$$

Important special cases of the choice of the domain include, e.g., closed intervals $A = [a, b] \subset \mathbb{R}$ and the whole real axis $A = \mathbb{R}$.

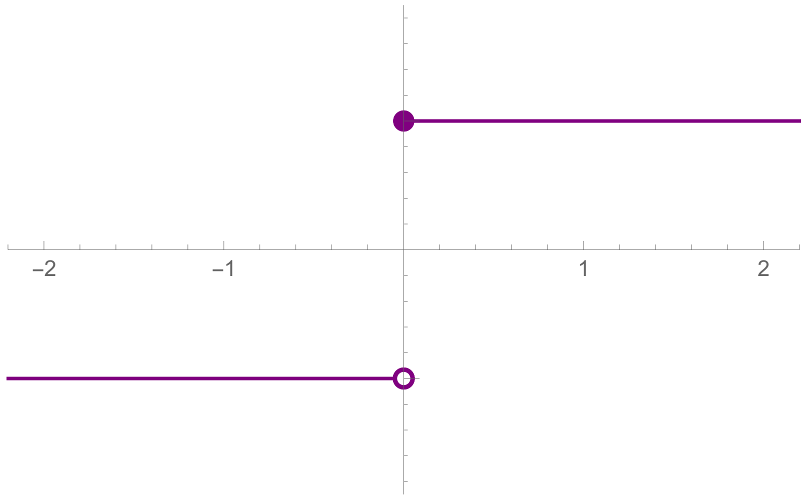


FIGURE III.1. The step function of Exercise III.2 is a basic example of a function that is not continuous.

III.2.1. Continuity of a function of real variable

The intuitive idea of continuous functions should be familiar from calculus courses. As a precise definition, we take the following.[2]

Definition III.6 (Continuity of a real-valued function of a real variable)**.**
Let $A \subset \mathbb{R}$, and let $f \colon A \to \mathbb{R}$ be a function. We say that f is **continuous** if the following holds: whenever $(a_n)_{n \in \mathbb{N}}$ is a sequence in A which is convergent and its limit is in the domain, $\lim_{n\to\infty} a_n \in A$, then we have

$$\lim_{n\to\infty} f(a_n) = f\Big(\lim_{n\to\infty} a_n\Big).$$

To elaborate — given a sequence $(a_n)_{n \in \mathbb{N}}$ in the domain A of the function $f \colon A \to \mathbb{R}$, we map the members $a_n \in A$ of the sequence to real numbers $f(a_n) \in \mathbb{R}$, and form a real number sequence $\big(f(a_n)\big)_{n \in \mathbb{N}}$ of these images under f, i.e., of the corresponding function values. Continuity of f (Definition III.6) is the requirement that if the original sequence is convergent and its limit is in the domain A, then also the image sequence is convergent and its limit is the image of the limit of the original sequence. In somewhat imprecise terms, "we are allowed to interchange the order of (i) taking the limit and (ii) applying the function".[3]

Certainly a function which abruptly jumps from one value to another should be the simplest example of a function that is not continuous; see Figure III.1. In the next exercise you check the discontinuity of such a step function directly from the definition.

[2]Later, in Chapter VI, we define continuity in a more general context. The general definition we take will be shown to be equivalent to the definition used here, in the special case of real-valued functions of a real variable.

[3]Note, however, that we still have to assume convergence of the original sequence to a limit in the domain.

Exercise III.2 (Step function is discontinuous)**.**
Consider the function

$$f\colon\ \mathbb{R} \to \mathbb{R}, \qquad\qquad f(x) := \begin{cases} -1 & \text{for } x < 0 \\ +1 & \text{for } x \geq 0. \end{cases}$$

Show that it is not a continuous function $\mathbb{R} \to \mathbb{R}$.
Hint: *By Definition III.6 it suffices to find* some *convergent sequence* $(x_n)_{n\in\mathbb{N}}$ *on* $\mathbb{R}$ *such that* $\lim_n f(x_n) \neq f\big(\lim_n x_n\big)$.

Example III.7 (Another discontinuous function)**.**
Consider the function

$$f\colon\ [0, +\infty) \to \mathbb{R}, \qquad\qquad f(x) := \begin{cases} -1 & \text{for } x = 0 \\ \cos(\pi + 2\pi/x) & \text{for } x > 0. \end{cases}$$

Is this function continuous or not? (For a plot, see Figure III.2.)

Intuitively it seems that any potential problem should occur at $x = 0$, if at all. So let us try out some sequences $(x_n)_{n\in\mathbb{N}}$ which tend to 0.

First try the sequence with members $x_n = \frac{1}{n}$. This sequence has limit

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

The values of f at members of the sequence are

$$f(x_n) = \cos\Big(\pi + \frac{2\pi}{1/n}\Big) = \cos(\pi + 2\pi n) = -1 \qquad \text{for } n \in \mathbb{N}.$$

Therefore we find

$$\lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} (-1) = -1 = f(0) = f\big(\lim_{n\to\infty} x_n\big).$$

This is as required in the definition of continuity. (Note that $x_n = \frac{1}{n} \in [0, +\infty)$ for all $n \in \mathbb{N}$, so this is a sequence in the domain of definition of f, and also the limit $\lim_n x_n = 0$ remains in the domain, $0 \in [0, +\infty)$.)

If you were to try for example the sequence with members $x_n = 2^{-n}$, which also tends to zero, you would again find $\lim f(x_n) = f\big(\lim x_n\big)$.

But Definition III.6 requires a similar conclusion for *all* sequences (which are convergent in the domain of f). So a single counterexample will be sufficient to show discontinuity! For such a counterexample, take the sequence with members $x_n = \frac{4}{4n-1}$. One again notes that $x_n \in [0, +\infty)$ for all $n \in \mathbb{N}$, and $\lim x_n = 0$. Now the function values along this sequence are

$$f(x_n) = \cos\Big(\pi + \frac{2\pi}{4/(4n-1)}\Big) = \cos\Big(\pi + \frac{\pi}{2}(4n-1)\Big) = \cos\Big(\frac{\pi}{2} + 2\pi n\Big) = 0$$

for $n \in \mathbb{N}$. Therefore

$$\lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} 0 = 0 \neq -1 = f(0) = f\big(\lim_{n\to\infty} x_n\big).$$

This shows that f is not continuous.

Exercise III.3 (Indicator function of rational numbers)**.**
Consider the function

$$f\colon\ \mathbb{R} \to \mathbb{R}, \qquad\qquad f(x) := \begin{cases} 1 & \text{for } x \in \mathbb{Q} \\ 0 & \text{for } x \in \mathbb{R} \setminus \mathbb{Q}. \end{cases}$$

Is it continuous?

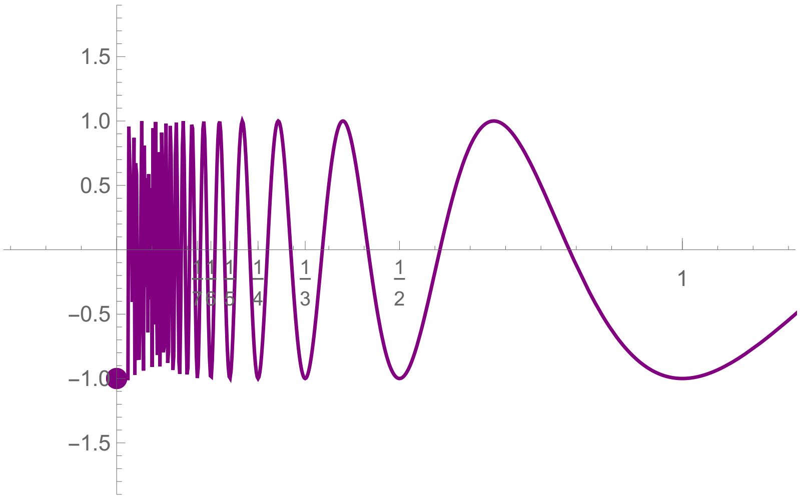


FIGURE III.2. The function of Example III.7.

Note that continuity depends on the domain of definition of the function: a discontinuous function restricted to a smaller set can become continuous, as you will see in the next example.

Exercise III.4 (The restriction of a discontinuous function can be continuous)**.**
Consider the step funtion $f\colon \mathbb{R} \to \mathbb{R}$ of Exercise III.2, and let $\widetilde{f} = f|_{\mathbb{R}\setminus\{0\}}$ be its restriction to the set $\mathbb{R} \setminus \{0\}$ non-zero real numbers. Show that $\widetilde{f}\colon \mathbb{R} \setminus \{0\} \to \mathbb{R}$ is continuous.

In the converse direction things are stable: a function never loses its continuity when restricting the domain.

Exercise III.5 (Restrictions of continuous functions are continuous)**.**
Suppose $f\colon A \to \mathbb{R}$ is a continuous function and $\widetilde{A} \subset A$ is a subset. Show that the restriction $\widetilde{f} = f|_{\widetilde{A}}$ is a continuous function $\widetilde{f}\colon \widetilde{A} \to \mathbb{R}$.

Operations on continuous functions

Because the definition of continuity relies on limits of real number sequences, our earlier results about limits easily yield some fundamental results about continuous functions. Particularly important is the fact that the following pointwise operations preserve continuity.

Definition III.8 (Pointwise operations on real-valued functions of real variable)**.**
Let $A \subset \mathbb{R}$, and let $f, g\colon A \to \mathbb{R}$ be two real-valued functions on A.

The **pointwise sum** of f and g is the function

$$f + g\colon\ A \to \mathbb{R} \qquad\qquad \big(f + g\big)(x) := f(x) + g(x) \quad \text{for } x \in A.$$

The **pointwise product** of f and g is the function

$$fg\colon\ A \to \mathbb{R} \qquad\qquad \big(fg\big)(x) := f(x)\, g(x) \quad \text{for } x \in A.$$

The **pointwise quotient** of f and g is the function

$$f/g\colon\ A' \to \mathbb{R} \qquad\qquad \big(f/g\big)(x) := \frac{f(x)}{g(x)} \quad \text{for } x \in A',$$

where $A' := \{x \in A \mid g(x) \neq 0\}$.

Note that the pointwise quotient f/g is only defined on the subset $A' \subset A$ of the domain, where the function g is non-vanishing — in order to avoid ill-defined division by zero.

Scalar multiplication of a function $f\colon A \to \mathbb{R}$ is a special case of the pointwise product: it is natural to interpret a real number $c \in \mathbb{R}$ also as the constant function $x \mapsto c$ on A, so $cf\colon A \to \mathbb{R}$ is defined by $(cf)(x) := c\,f(x)$ for $x \in A$.

Proposition III.9 (Continuity preserving operations)**.**
Let $A \subset \mathbb{R}$, and let $f, g\colon A \to \mathbb{R}$ be two continuous real-valued functions on A. Then also

(i) *the pointwise sum function $f + g\colon A \to \mathbb{R}$ is continuous,*
(ii) *the pointwise product function $fg\colon A \to \mathbb{R}$ is continuous,*
(iii) *the pointwise quotient function $f/g\colon A' \to \mathbb{R}$ is continuous on the subset $A' = \{x \in A \mid g(x) \neq 0\}$.*

Proof. The proofs of all cases are essentially similar, so let us provide the details only for (ii), and leave (i) and (ii) as exercises.

proof of (ii): Assume $f, g\colon A \to \mathbb{R}$ are continuous. We must prove that $fg\colon A \to \mathbb{R}$ is continuous, and we will do this directly using Definition III.6 of continuity. So suppose $(x_n)_{n\in\mathbb{N}}$ is a sequence such that $x_n \in A$ for all $n \in \mathbb{N}$, and $(x_n)_{n\in\mathbb{N}}$ converges to a limit in A. Let us denote the limit by $x = \lim_{n\to\infty} x_n \in A$. By definition of continuity, we must consider the sequence $\big((fg)(x_n)\big)_{n\in\mathbb{N}}$ of values of the pointwise product function fg and show that it converges to the limit $(fg)(x)$. By definition of pointwise products, the values are $(fg)(x_n) = f(x_n)\,g(x_n)$. By assumptions of continuity of f and g we know that $\lim_{n\to\infty} f(x_n) = f(x)$ and $\lim_{n\to\infty} g(x_n) = g(x)$, so by Lemma **??** we have

$$\lim_{n\to\infty} \Big(f(x_n)\,g(x_n)\Big) = \Big(\lim_{n\to\infty} f(x_n)\Big)\Big(\lim_{n\to\infty} g(x_n)\Big) = f(x)\,g(x).$$

By definition of the pointwise products the right hand side above is $(fg)(x)$, so we have shown

$$\lim_{n\to\infty} \Big((fg)(x_n)\Big) = (fg)(x),$$

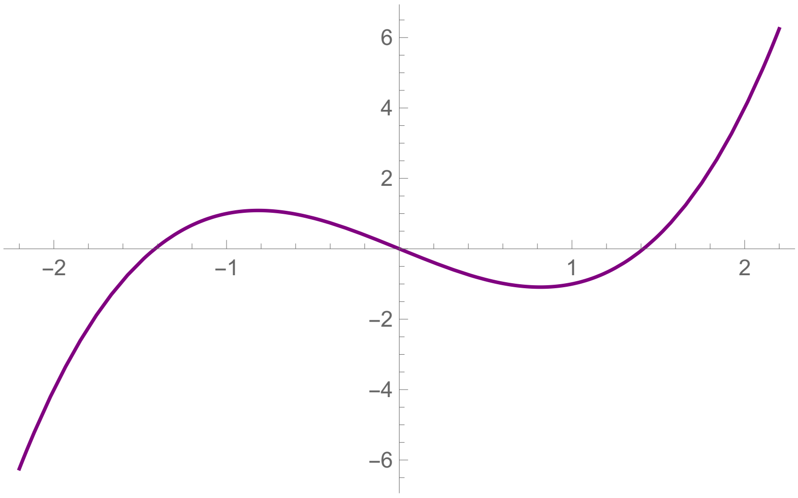
which is what was needed to prove continuity. □

Exercise III.6 (Proof of Proposition III.9 (i) and (iii))**.**
Prove parts (i) and (iii) of the above proposition.

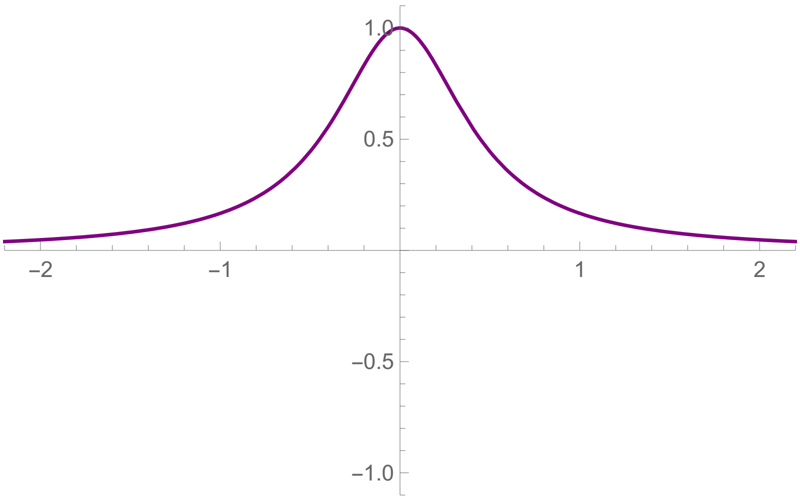
Exercise III.7 (Continuity of the tangent trigonometric function)**.**
Is the function

$$\tan\colon \mathbb{R} \setminus \Big\{\frac{\pi}{2} + n\pi \;\Big|\; n \in \mathbb{Z}\Big\} \to \mathbb{R}$$

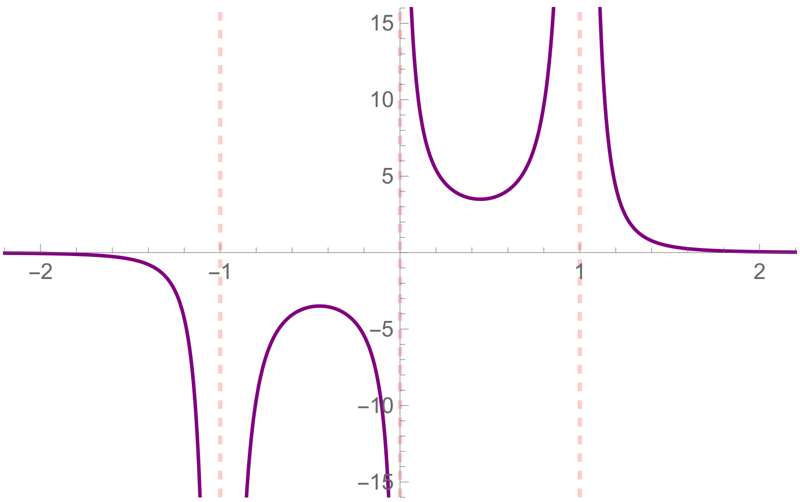
continuous?
Hint: *The continuity of the trigonometric functions* $\sin\colon \mathbb{R} \to \mathbb{R}$ *and* $\cos\colon \mathbb{R} \to \mathbb{R}$ *can be considered known here (a precise justification will be given in Chapter IX). Recall that* $\tan(x) = \frac{\sin(x)}{\cos(x)}$.



(a) The polynomial $x \mapsto x^3 - 2x$ defines a function $\mathbb{R} \to \mathbb{R}$.



(b) The function $x \mapsto \frac{1}{1+5x^2}$ is a rational function function $\mathbb{R} \to \mathbb{R}$.



(c) The rational function $x \mapsto \frac{1}{x(x^2-1)^2}$ has poles at $x = 0$ and $x = \pm 1$, and defines a (continuous) function $\mathbb{R} \setminus \{-1, 0, +1\} \to \mathbb{R}$.

FIGURE III.3. Polynomials and rational functions are continuous.

Corollary III.10 (Polynomials and rational functions are continuous). *Let $a_0, a_1, a_2, \ldots, a_n \in \mathbb{R}$ and consider the polynomial function*

$$P(x) := a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \qquad \text{for } x \in \mathbb{R}.$$

Then $P\colon \mathbb{R} \to \mathbb{R}$ is a continuous function. Let $a_0, a_1, a_2, \ldots, a_n \in \mathbb{R}$ and consider the polynomial function

Let also $b_0, b_1, b_2, \ldots, b_m \in \mathbb{R}$ and

$$Q(x) := b_0 + b_1 x + b_2 x^2 + \cdots + b_m x^m \qquad \text{for } x \in \mathbb{R}.$$

Consider the rational function

$$R(x) = \frac{P(x)}{Q(x)} \qquad \text{for } x \in A,$$

where $A = \{x \in \mathbb{R} \mid Q(x) \neq 0\}$. Then $R\colon A \to \mathbb{R}$ is a continuous function.

Proof. Let us first observe that the continuity of the identity function $x \mapsto x$ (as a function $\mathbb{R} \to \mathbb{R}$) is trivial from the definition. The function $x \mapsto x^2$ is a pointwise product of identity functions, and therefore also continuous by Proposition III.9(ii). Continuing to take pointwise products with the identity function, we find that the monomial functions $x \mapsto x^k$ are continuous for all $k \in \mathbb{N}$ (easy induction).

Then note that for any $c \in \mathbb{R}$ the continuity of the constant function $x \mapsto c$ (as a function $\mathbb{R} \to \mathbb{R}$) is also obvious from the definition. Taking $c = a_k \in \mathbb{R}$ and further taking pointwise product with a monomial function, we find that $x \mapsto a_k x^k$ is continuous, for any $k \in \mathbb{Z}_{\geq 0}$ (for the case $k = 0$ one does not even need a product with monomial function). The polynomial $P(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ is a finite sum of terms of this type, so continuous by Proposition III.9(i), inductively applied.

Since polynomial functions given by $P(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ and $Q(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_m x^m$ are continuous, we get from Proposition III.9(iii) that the rational function given by the pointwise quotient $R(x) = \frac{P(x)}{Q(x)}$ is continuous on the subset $A = \{x \in \mathbb{R} \mid Q(x) \neq 0\} \subset \mathbb{R}$ where the denominator is non-vanishing. □

Later in the course we will develop more powerful tools for verifying the continuity of functions. But knowing the continuity of polynomial and rational functions at least assures us that some nontrivial continuous functions exist; see Figure III.3 for a few concrete examples.

III.3. Continuous functions on a closed interval

Let us now address specifically the case when the domain of the function is closed interval on the real axis, $A = [a, b] \subset \mathbb{R}$. This case appears very frequently in applications. We address some key properties of continuous functions

$$f\colon [a, b] \to \mathbb{R}$$

on a closed interval; later in the course we will understand them from more general perspectives.

Let us recall a few key concepts: a function $f\colon A \to \mathbb{R}$ is **bounded** if there exists a $M \geq 0$ such that $|f(x)| \leq M$ for all $x \in A$. A function $f\colon A \to \mathbb{R}$ has a **maximum** if there exists a point $x_{\max} \in A$ such that $f(x) \leq f(x_{\max})$ for all $x \in A$, and it has a **minimum** if there exists a point $x_{\min} \in A$ such that $f(x) \geq f(x_{\min})$ for all $x \in A$.

Remark III.11 (Reinterpretation in terms of the range of the function).
Consider the image $f[A] = \{f(x) \mid x \in A\} \subset \mathbb{R}$ of the whole domain A (this is often called the range of f). It is straightforward to verify that the function f is bounded if and only if the set $f[A] \subset \mathbb{R}$ is bounded, and the function f has a maximum (resp. minimum) if and only if the set $f[A]$ has a maximum (resp. minimum).

The following important properties are related to compactness (see Chapter XI) of the closed interval.

Theorem III.12 (Extrema of continuous functions on a closed interval).
Let $f \colon [a,b] \to \mathbb{R}$ be a continuous function. Then,

(i) *f is bounded,*
(ii) *f has a maximum and a minimum.*

Before giving the proof, let us note that the conclusions of Theorem III.12 rely crucially on the assumption that the interval is closed. The examples in Figure III.4 illustrate what can go wrong on non-closed intervals.

Exercise III.8 (Continuous functions on non-closed intervals).

(a) Prove that the function of Figure III.4(a) is continuous but not bounded.
(b) Prove that the function of Figure III.4(b) is continuous but has no minimum.
(c) Find an example of a continuous function on an interval, which has no maximum.

Proof of Theorem III.12 Let us prove the two assertions separately.

proof of (i): Suppose, by contrapositive, that f is not bounded. Then for every $n \in \mathbb{N}$, it is possible to choose some $x_n \in [a,b]$ such that $|f(x_n)| \geq n$. Let us form a sequence $(x_n)_{n\in\mathbb{N}}$ using such choices. This sequence is bounded, since $a \leq x_n \leq b$ for all $n \in \mathbb{N}$. Therefore, by Theorem III.5, it has a convergent subsequence $(x_{\varphi(n)})_{n\in\mathbb{N}}$. If we denote the limit of such a convergent subsequence by $x = \lim_{n\to\infty} x_{\varphi(n)}$, then by the preservation of inequalities $a \leq x \leq b$ (see Lemma **??**). By continuity, then, $\big(f(x_{\varphi(n)})\big)_{n\in\mathbb{N}}$ converges to $f(x)$. But by the choice of x_n we have $|f(x_{\varphi(n)})| \geq \varphi(n) \geq n$ for all $n \in \mathbb{N}$, which means that the sequence $\big(f(x_{\varphi(n)})\big)_{n\in\mathbb{N}}$ is not bounded and so cannot be convergent (see Lemma **??**). This is a contradiction. We conclude that f had to be bounded.

proof of (ii): Let us only prove that f has a maximum — the existence of minimum can be concluded similarly (or by considering the maximum of the continuous function $-f$).

By (i) f is bounded, so the supremum of its values is finite,

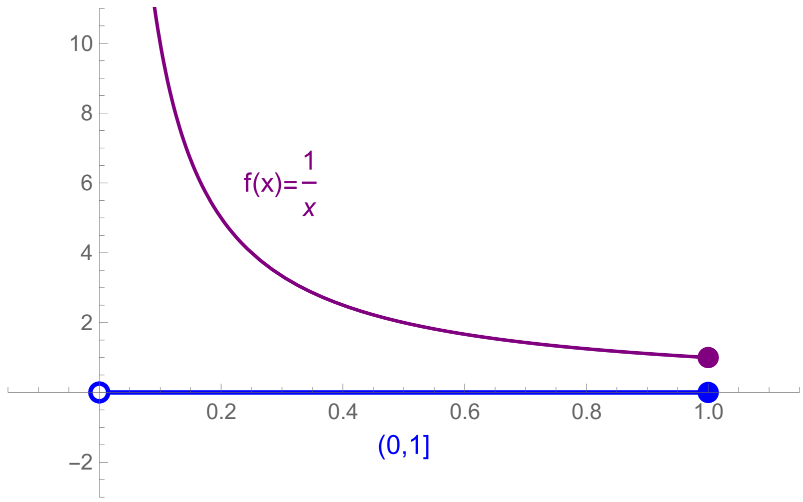
$$C \;:=\; \sup\big\{f(x) \;\big|\; x \in [a,b]\big\} \;\in\; \mathbb{R}.$$

For all $n \in \mathbb{N}$ there then exists some $x_n \in [a,b]$ such that $f(x_n) \geq C - \frac{1}{n}$. Let us form a sequence $(x_n)_{n\in\mathbb{N}}$ using such choices. As in part (i), this sequence is bounded and therefore has a convergent subsequence $(x_{\varphi(n)})_{n\in\mathbb{N}}$ whose limit $x = \lim_{n\to\infty} x_{\varphi(n)}$ is also on the interval $[a,b]$. By the choice of C (supremum of the values) and of x_n we have

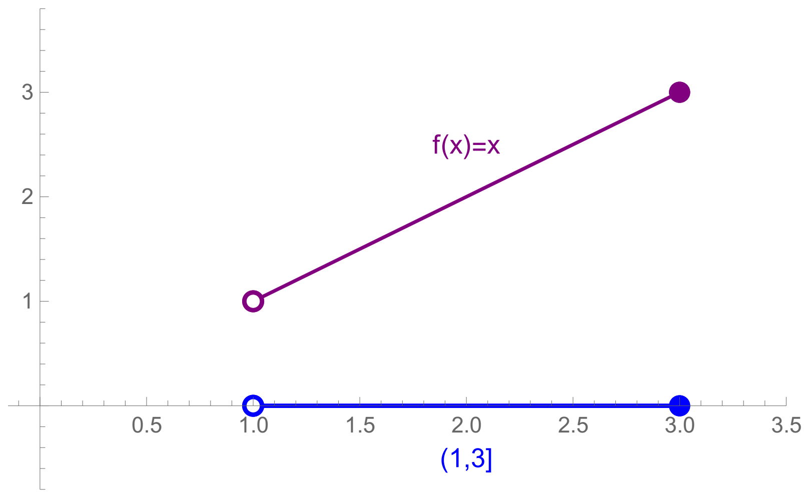
$$C \;\geq\; f(x_{\varphi(n)}) \;\geq\; C - \frac{1}{\varphi(n)} \;\geq C - \frac{1}{n}.$$

The squeeze theorem (Theorem **??**) thus gives $\lim_{n\to\infty} f(x_{\varphi(n)}) = C$. On the other hand, by continuity we have

$$f(x) = f\big(\lim_{n\to\infty} x_{\varphi(n)}\big) = \lim_{n\to\infty} f(x_{\varphi(n)}) = C.$$



(a) The function $f\colon (0,1] \to \mathbb{R}$ given by $f(x) = 1/x$ is continuous (a rational function) but not bounded.



(b) The function $f\colon (1,3] \to \mathbb{R}$ given by $f(x) = x$ is continuous (a polynomial function) but has no minimum.

FIGURE III.4. Continuous functions on general intervals do not need to be bounded and do not need to have minima and maxima.

We conclude that

$$f(x) = \sup\left\{ f(x) \;\middle|\; x \in [a,b] \right\},$$

which implies that the maximum of $f\colon [a,b] \to \mathbb{R}$ is attained at $x \in [a,b]$. $\square$

The following property is related to connectedness (see Chapter XII) of the closed interval.

Theorem III.13 (Mean value theorem)**.**
Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function. Assume that $f(a)f(b) < 0$.[4] *Then there exists a point $z \in (a,b)$ such that $f(z) = 0$.*

[4]This is just a concise way of saying that $f(a)$ and $f(b)$ are non-zero and have opposite signs: either $f(a) < 0$ and $f(b) > 0$, or $f(a) > 0$ and $f(b) < 0$.

Proof. We may assume that $f(a) < 0$ and $f(b) > 0$; the other case is similar (or is obtained from this one by considering the continuous function $-f$ with the same zeroes as f).

Consider the subset

$$N = \Big\{ x \in [a, b] \;\Big|\; f(x) < 0 \Big\}$$

of the interval $[a, b]$ where the function f is negative. Since $f(a) < 0$, we have at least $a \in N$, so this subset is non-empty, $N \neq \emptyset$. The subset is also bounded from above, because the right endpoint b of the interval is an upper bound for N. So let

$$z = \sup N.$$

The earlier observations yield $z \geq a$ and $z \leq b$, so we have $z \in [a, b]$. By a characterization of the supremum (Lemma II.31), there exists a sequence $(x_n^-)_{n\in\mathbb{N}}$ in the set N such that $\lim_{n\to\infty} x_n^- = z$. By continuity of f and by preservation of inequalities, we have

$$f(z) = f\big(\lim_{n\to\infty} x_n^-\big) = \lim_{n\to\infty} \underbrace{f(x_n^-)}_{<0} \leq 0.$$

In particular we have $z \neq b$, since $f(b) > 0$.

Then consider (for example) the sequence $(x_n^+)_{n\in\mathbb{N}}$ with $x_n^+ = z + \frac{1}{n}$, which also has the property $\lim_{n\to\infty} x_n^- = z$ and $f(x_n^+) \geq 0$ since $x_n^+ \notin N$ (note that for large enough n we have $x_n^+ = z + \frac{1}{n} \leq b$, so the members of the sequence are in the domain $[a, b]$ of f). By continuity of f and by preservation of inequalities, we have

$$f(z) = f\big(\lim_{n\to\infty} x_n^+\big) = \lim_{n\to\infty} \underbrace{f(x_n^+)}_{\geq 0} \geq 0.$$

In particular we have $z \neq a$, since $f(a) < 0$.

Therefore we have $z \in (a, b)$, and the above two inequalities combined yield $f(z) = 0$. □

Corollary III.14 (Continuous image of interval)**.**
If $I \subset \mathbb{R}$ is an interval and $f\colon I \to \mathbb{R}$ is continuous, then the image $f[I]$ is also an interval.

Example III.15 (The image of an interval under the sine function)**.**
Consider the sine function

$$f_1\colon\ (0, \pi) \to \mathbb{R} \qquad\qquad f_1(x) = \sin(x) \quad \text{for } x \in (0, \pi).$$

This function is continuous.[5] The domain of definition is the open interval $(0, \pi)$, and its image under f_1 is the half-open interval $f_1\big[(0, \pi)\big] = (0, 1]$.

If instead we considered the sine function on the longer interval

$$f_2\colon\ (0, 2\pi) \to \mathbb{R} \qquad\qquad f_2(x) = \sin(x) \quad \text{for } x \in (0, 2\pi),$$

then the image would be the closed interval $f_1\big[(0, \pi)\big] = [-1, 1]$.

This shows... UNDER CONSTRUCTION!

[5]The appropriate tools for the precise justification of the continuity of trigonometric functions are developed in Chapter IX; for now we just accept the continuity statement.

Index

References

Bibliography

[Hal74] Paul R. Halmos. *Naive set theory.* Undergraduate Texts in Mathematics. Springer, 1974.
[Väi99] Jussi Väisälä. *Topologia I.* Limes ry., 1999.