Mathematics for Computer science

Frank-Olaf Schreyer

January 27, 2020

Contents

Introduction

1 Sets, logic and proofs

1.1 Sets

Definition 1.1. A **set** is a collection of well-defined **elements**.
Examples:

- $A = \{a, b, c, \ldots, z\}$ the set of letters of the alphabet,
- $B = \{\text{students in this lecture}\}$,
- $C = \{2, 3, 5, 7\} = \{p \mid p \text{ is a prime number } \leq 10\}$.

Sets can be specified by listing its elements or by a characterising property of its elements. Some special sets:

- $\emptyset = \{\ \}$ the empty set,
- $\mathbb{N} = \{1, 2, \ldots\}$ the set natural numbers,
- $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ the set natural numbers including 0,
- $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$ integral numbers,
- $\mathbb{Q} = \{\frac{a}{b} \mid a, b \in \mathbb{Z}, b \neq 0\}$ rational numbers,
- $\mathbb{R} = \{\text{real numbers}\} = \{\text{not necessarily periodic decimal numbers}\}$.

If x is an element of a set M then we write

$$x \in M.$$

If every element of a set N is also an element of a set M then we write

$$N \subset M.$$

$N \subseteq M$ is also in use. We use the notation $N \subsetneqq M$ for $N \subset M$ and $N \neq M$. $N \not\subset M$ means N is not a subset of M.

Definition 1.2. We denote by $|M|$ the number of elements of a set M.

- $|\emptyset| = 0, \quad |\{0\}| = 1$,
- $|\{a, b, c, \ldots, z\}| = 26$.

If M has infinitely many elements then we write

$$|M| = \infty.$$

Alternative notation: $\#M := |M|$.

- $\#\{\text{students in this lecture older than 40}\} = 0\,$?

Definition 1.3 (Intersections, unions, complements)**.** Let A, B denote sets. Then

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

denotes the **union**, and

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

denotes the **intersection**.

A and B are called **disjoint** if $A \cap B = \emptyset$. If A and B are disjoint, then

$$|A \cup B| = |A| + |B|.$$

In general,

$$|A \cup B| + |A \cap B| = |A| + |B|$$

holds because in the sum $|A| + |B|$ the elements of $A \cap B$ are counted twice.

Distribution laws. Let A, B, C denote sets. Then

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

and

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

hold.

If we consider only subset of a fixed set M, then

$$\overline{A} = \{x \in M \mid x \notin A\}$$

denotes the **complement** of A in M

De Morgan's laws. $A, B \subset M$. Then

- $\overline{A \cap B} = \overline{A} \cup \overline{B}$,

- $\overline{A \cup B} = \overline{A} \cap \overline{B}$,
- $\overline{\overline{A}} = A$

hold.

More general, $A \setminus B = \{a \in A \mid a \notin B\}$ denotes the difference of B in A. Notice:

$$A \setminus B = A \setminus (B \cap A).$$

Definition 1.4 (Cartesian product, power set)**.** If A, B are sets, then

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

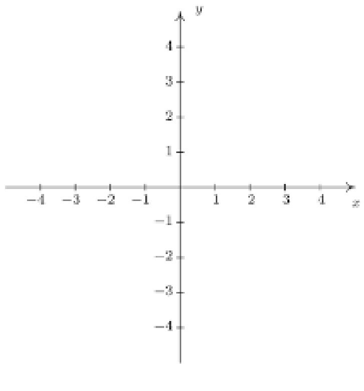
denotes the set of pairs (a, b), the **Cartesian product** of A and B.
Examples.

$$\{a, \ldots, h\} \times \{1, \ldots, 8\} = \{(a, 1), \ldots, (h, 8)\}$$

is used in chess. We can describe with

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x, y) \mid x \in \mathbb{R}, y \in \mathbb{R}\}$$

the set of points in the plane.



Note

$$|A \times B| = |A| \cdot |B|.$$

For a fixed set M the set of all subsets

$$2^M = \{A \mid A \subset M\}$$

is called the **power set** of M.

Theorem 1.5. *If M is a finite set, then*

$$|2^M| = 2^{|M|}.$$

In other words: a set with n elements has 2^n subsets. Examples:

- $2^{\emptyset} = \{\emptyset\}$, $2^{\{1\}} = \{\emptyset, \{1\}\}$
- $2^{\{1,2\}} = \{\emptyset, \{1\}, \{2\}, \{1,2\}\}$

How to prove such a theorem? In this section we will introduce two general methods for proofs:

1. Proof by contradiction
2. Proof by induction

Theorem 1.5 is proved by induction. We postpone the proof of Theorem 1.5 and start the discussion with a famous proof by contradiction.

1.2 Proof by contradiction

Let A be a statement which can be either true or false. We want to prove that A is true. For this we deduce from the assumption A *is false* a contradiction. Then the statement A is not true is false, hence A is true.

Recall a prime number p is a natural number $p \geq 2$ which has precisely two factors, namely 1 and p.

Theorem 1.6. *There exist infinitely many prime numbers.*

Proof. (Euclid) Suppose there are only finitely many prime numbers which we could denote by $p_1, p_2, \ldots, p_n$. Consider the product

$$m = p_1 \cdot p_2 \cdot \ldots \cdot p_n.$$

The integer $m + 1$ has none of the numbers $p_1, \ldots, p_n$ as a factor, since division by each p_i leaves the remainder 1. On the other hand, $m + 1$ has a prime factor q. Then $q \notin \{p_1, \ldots, p_n\}$ and $p_1, \ldots, p_n$ is not the complete list of prime numbers. The assumption that there are only finitely many prime numbers is false. □

1.3 Propositional logic

Notation 1.7. It is useful to have a short logical notation in complicated proofs. Let A and B be statements. Then

- $A \wedge B$ denotes the statement "A and B are true",
- $A \vee B$ denotes "A or B (or both) are true",
- $A \Rightarrow B$ denotes "if A is true then B is true",
  or in other words "A implies B",
- $A \Leftrightarrow B$ denotes "A is true if and only if B is true",
  or in other words "A is equivalent to B'",
- $\neg A$ denotes the statement "A is not true".

If $A \Rightarrow B$ and $B \Rightarrow C$ hold, then also $A \Rightarrow C$ holds. Differently formulated, the statement

$$((A \Rightarrow B) \wedge (B \Rightarrow C)) \Rightarrow (A \Rightarrow C)$$

is always true. Clearly, this fact is used in many proofs to divide the argument into smaller pieces.

The statement $A \Rightarrow B$ should not be confused with $B \Rightarrow A$. For example, consider the statements

$A_1 =$ The gate is closed and $B_1 =$ Some train crosses.

Then

$A_1 \Rightarrow B_1$ is false because the gate is closed earlier,

and

$B_1 \Rightarrow A_1$ is (hopefully) true.

Theorem 1.8 (Laws of de Morgan)**.** *Let A and B be logical statements. Then*

$$\neg(A \wedge B) = (\neg A) \vee (\neg B)$$
$$\neg(A \vee B) = (\neg A) \wedge (\neg B)$$

Proof. Think of A and B as logical variables which can take the binary values

true $= 1$ or false $= 0$.

Then there are four possibilities for the values of A and B, and depending on these, we obtain for the expressions occurring successively in the formula above the following values:

A	B	$A \wedge B$	$\neg(A \wedge B)$	$\neg A$	$\neg B$	$(\neg A) \vee (\neg B)$
1	1	1	0	0	0	0
1	0	0	1	0	1	1
0	1	0	1	1	0	1
0	0	0	1	1	1	1

The first formula holds because the fourth and seventh column have the same values. The proof of the second formula is similar. □

Proposition 1.9. *Let A and B be logical statements. Then*

$$A \Rightarrow B = (\neg A) \vee B.$$

Proof. The proof is similar.

A	B	$A \Rightarrow B$	$\neg A$	$(\neg A) \vee B$
1	1	1	0	1
1	0	0	0	0
0	1	1	1	1
0	0	1	1	1

□

Further identities

Proposition 1.10. *Let A, B, C be logical statements . Then*

- *associative*

$$A \vee (B \vee C) = (A \vee B) \vee C,$$
$$A \wedge (B \wedge C) = (A \wedge B) \wedge C$$

- *commutative*

$$A \vee B = B \vee A,$$
$$A \wedge B = B \wedge A$$

- *distributative*

$$(A \vee B) \wedge C = (A \wedge C) \vee (B \wedge C),$$
$$(A \wedge B) \vee C = (A \vee C) \wedge (B \vee C)$$

- *idempotent*
$$\begin{aligned} A \vee A &= A, \\ A \wedge A &= A \end{aligned}$$

- *tertium non datur*
$$\begin{aligned} A \vee \neg A &= 1, \\ A \wedge \neg A &= 0 \end{aligned}$$

- *double negation*
$$\neg(\neg A) = A$$

- *units*
$$\begin{aligned} A \wedge 1 &= A, \quad A \wedge 0 = 0, \\ A \vee 1 &= 1, \quad A \vee 0 = A \end{aligned}$$

Due to the associative law, we can drop brackets and write $A \vee B \vee C$ for $(A \vee B) \vee C = (A \vee B) \vee C$ and similarly for more factors or $\wedge$. Similar to the convention $a + b \cdot c = a + (b \cdot c)$ for addition and multiplication of numbers, we may drop some brackets by giving $\neg$ the strongest binding strength $\wedge$ and $\vee$ middle binding strength and $\Rightarrow$, $\Leftrightarrow$ the weakest strength.

That the statement
$$T = (A \Rightarrow B) \wedge (B \Rightarrow C) \Rightarrow (A \Rightarrow C)$$
is always true can be proved by the rules above:
$$\begin{aligned}
T &= (\neg A \vee B) \wedge (\neg B \vee C) \Rightarrow (\neg A \vee C) && \text{1.9} \\
&= \neg((\neg A \vee B) \wedge (\neg B \vee C)) \vee (\neg A \vee C) && \text{1.9} \\
&= (\neg(\neg A \vee B) \vee \neg(\neg B \vee C) \vee (\neg A \vee C) && \text{de Morgan} \\
&= (A \wedge \neg B) \vee (B \wedge \neg C) \vee \neg A \vee C && \text{de Morgan} \\
&= (A \wedge \neg B) \vee \neg A \vee (B \wedge \neg C) \vee C && \text{commutative, associative} \\
&= ((A \vee \neg A) \wedge (\neg B \vee \neg A)) \vee ((B \vee C) \wedge (\neg C \vee C)) && \text{distributive} \\
&= ((1 \wedge (\neg B \vee \neg A)) \vee ((B \vee C) \wedge 1)) && \text{tertium non datur} \\
&= (\neg B \vee \neg A) \vee (B \vee C) && \text{units} \\
&= \neg B \vee B \vee \neg A \vee C && \text{commutative, associative} \\
&= 1 \vee \neg A \vee C && \text{tertium non datur} \\
&= 1 && \text{units}
\end{aligned}$$

Using Proposition 1.9 one can remove $\Rightarrow$ and $\Leftrightarrow$ from any logical expression. Using de Morgan and the further rules any logical expression S in finitely many Boolean variables $A_1, A_2, \ldots, A_r$ can be brought into two normal forms:

- conjunctive normal form

  $$(x_{11} \vee x_{12} \vee \ldots \vee x_{1n_1}) \wedge (x_{21} \vee x_{22} \vee \ldots \vee x_{2n_2}) \wedge \ldots \wedge (x_{m1} \vee x_{m2} \vee \ldots \vee x_{2n_m})$$

- disjunctive normal form

  $$(y_{11} \wedge y_{12} \wedge \ldots \wedge y_{1n_1}) \vee (y_{21} \wedge y_{22} \wedge \ldots \wedge y_{2n_2}) \vee \ldots \vee (y_{m1} \wedge y_{m2} \wedge \ldots \wedge y_{2n_m})$$

where the $x_{ij}, y_{ij} \in \{A_1, \neg A_1, \ldots, A_r, \neg A_r\}$.

In the disjunctive normal form it is easy to check whether the formula S is satisfiable, meaning that there exist values 0 or 1 for each A_i which gives S the value 1. Logical statements in programs which assure the correctness of the program most frequently are in the conjunctive normal form. No fast general algorithm is known which answers the question whether an expression in the conjunctive normal form is satisfiable, and it is believed that a fast algorithm does not exist. This is at the heart of the famous $N \neq NP$ problem of complexity theory.

1.4 Principal of induction

1.11. Suppose we have a statement $A(n)$ or each $n \in \mathbb{N}$. If

1. (base of the induction) $A(1)$ holds, and
2. (induction step) $A(n) \Rightarrow A(n+1)$ holds for all $n \geq 1$,

then $A(n)$ holds for all $n \geq 1$.

Indeed $A(1) \Rightarrow A(2) \Rightarrow A(3) \Rightarrow \ldots \Rightarrow A(n) \Rightarrow \ldots$. The principal of induction is not a theorem but rather an axiom which specifies our intuition about natural numbers.

Example 1.12. *Proof of Theorem 1.5.* We want to prove the statement
$A(n) =$ "A set M with n elements has 2^n subsets'.'
Base of the induction. If M has one element, say $M = \{1\}$ then $2^M = \{\emptyset, \{1\}\}$, so has $2 = 2^1$ elements.

Induction step. Suppose that the statement holds for sets with n elements. Let $M = \{1, \ldots, n+1\}$ be the set with $n+1$ elements. Then

$$\begin{aligned} 2^M &= \{A \subset M\} = \{A \subset M \mid n+1 \notin A\} \cup \{A \subset M \mid n+1 \in M\} \\ &= \{A \subset \{1, \ldots, n\}\} \cup \{A \cup \{n+1\} \mid A \subset \{1, \ldots, n\}\} \\ . \end{aligned}$$

Since this is a disjoint union, we obtain

$$\begin{aligned} |2^M| &= 2^n + 2^n \\ &= 2 \cdot 2^n = 2^{n+1} = 2^{|M|} \end{aligned}$$

from the induction hypothesis. □

Remark 1.13. Sometimes one takes $A(0)$ as the base of the induction. In the case above also $A(0)$ makes sense and is true.

Induction is frequently used in the proof of sum or product formulas.

Definition 1.14. Suppose $a_1, \ldots, a_n \in \mathbb{R}$ are real numbers. Then

$$\sum_{k=1}^{n} a_k = a_1 + \ldots + a_n$$

denotes their sum, and

$$\prod_{k=1}^{n} a_k = a_1 \cdot \ldots \cdot a_n$$

denotes their product. By convention

$$\sum_{k=1}^{0} a_k = 0 \text{ and } \prod_{k=1}^{0} a_k = 1$$

We make this convention since then the recursive formulas

$$\sum_{k=1}^{n} a_k = \sum_{k=1}^{n-1} a_k + a_n$$

and

$$\prod_{k=1}^{n} a_k = (\prod_{k=1}^{n-1} a_k) \cdot a_n$$

make sense for all $n \geq 1$. In an implementation on a computer we would use the recursive formulas in a **for**-loop.

Example 1.15.

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}.$$

Proof by induction on n. For $n = 1$ we have

$$\sum_{k=1}^{1} k = 1 = \frac{1 \cdot 2}{2}.$$

Induction step $n \to n+1$:

$$\begin{aligned} \sum_{k=1}^{n+1} k &= \sum_{k=1}^{n} k + (n+1) && \text{definition} \\ &= \frac{n(n+1)}{2} + (n+1) && \text{induction hypothesis} \\ &= (n+1)(\frac{n}{2}+1) = \frac{(n+1)(n+2)}{2} \end{aligned}$$

as desired. □

For a different proof:

	1	+ 2	+ ...	+ n	
	+n	+ (n-1)	+ ...	+ 1	
=	(n+1)	+ (n+1)	+ ...	+ (n+1)	= n(n+1).

An anecdote says that Gauß discovered this proof as an 3-rd grader.

Example 1.16.

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

Definition 1.17. The number of k-element subsets A of $\{1, \ldots, n\}$ is denoted by

$$\binom{n}{k} = |\{A \subset \{1, \ldots, n\} \mid |A| = k\}|.$$

For any $m \in \mathbb{N}$ the integer m-factorial is

$$m! = \prod_{k=1}^{m} = 1 \cdot 2 \cdot \ldots \cdot m.$$

Thus $0! = 1$ by convention.

Theorem 1.18. *Let k, n be integers with $0 \leq k \leq n$. Then*

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

Proof by induction on n. For $n = 0$ the assertion is true: The empty set has precisely one 0-element subset, namely $\emptyset$. Thus

$$\binom{0}{0} = 1 = \frac{0!}{0! \cdot 0!}$$

holds. More generally, for any n-element set the empty set is the only 0-element subset. Thus $\binom{n}{0} = 1 = \frac{n!}{0! \cdot n!}$ holds for all n.
Induction step $n \to n+1$. By the preceding we may assume $k \geq 1$. Then

$$\begin{aligned}\binom{n+1}{k} &= ||\{A \subset \{1, \dots, n+1\} \mid |A| = k\}| \\ &= |\{A \subset \{1, \dots, n\} \mid |A| = k\}| + |\{A \cup \{n+1\} \mid A \subset \{1, \dots, n\} \text{ and } |A| = k-1\}| \\ &= \binom{n}{k} + \binom{n}{k-1} \\ &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k-1)!(n-k+1)!}\end{aligned}$$

by the induction hypothesis. Thus

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

as desired. □

Corollary 1.19 (of the proof)**.**

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$$

□

Theorem 1.20. *Let $a, b \in \mathbb{R}$ be real numbers and let $n \geq 1$ be an integer. Then*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k.$$

Proof. The formula is true for $n = 1$:

$$a + b = \binom{1}{0}a + \binom{1}{1}b = \sum_{k=0}^{1} \binom{1}{k} a^{1-k}b^k$$

Induction step.

$$\begin{aligned}
(a+b)^{n+1} &= (a+b)^n(a+b) \\
&= \big(\sum_{k=0}^{n} \binom{n}{k} a^{n-k}b^k\big)(a+b) \\
&= \sum_{k=0}^{n} \binom{n}{k} a^{n-k+1}b^k + \sum_{k=0}^{n} \binom{n}{k} a^{n-k}b^{k+1} \\
&= \sum_{k=0}^{n} \binom{n}{k} a^{n-k+1}b^k + \sum_{k=1}^{n+1} \binom{n}{k-1} a^{n-k+1}b^k \\
&= \binom{n}{0}a^{n+1} + \sum_{k=1}^{n}(\binom{n}{k} + \binom{n}{k-1})a^{n-k+1}b^k + \binom{n}{n}b^{n+1} \\
&= \sum_{k=0}^{n+1} \binom{n+1}{k} a^{n+1-k}b^k.
\end{aligned}$$

□

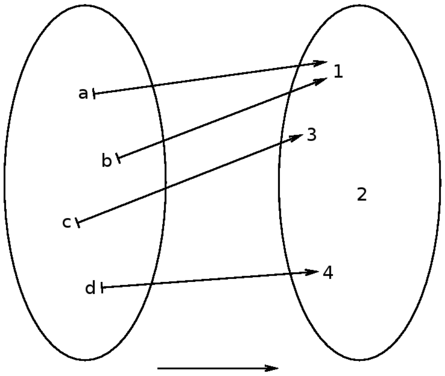
A quick way to compute all binomial coefficients $\binom{n}{k}$ for small n is to use Pascal's triangle:

$$\begin{array}{ccccccccccccccc}
&&&&&&&1&&&&&&& \\
&&&&&&1&&1&&&&&& \\
&&&&&1&&2&&1&&&&& \\
&&&&1&&3&&3&&1&&&& \\
&&&1&&4&&6&&4&&1&&& \\
&&1&&5&&10&&10&&5&&1&& \\
&1&&6&&15&&20&&15&&6&&1& \\
1&&7&&21&&35&&35&&21&&7&&1
\end{array}$$

2 Maps and counting

2.1 Maps and graphs

Definition 2.1. Let M and N be two sets. A **map** $f\colon M \to N$ is a rule which associates for each $x \in M$ an element $f(x) \in N$. We write $x \mapsto f(x)$.
Example:



$$f\colon M \to N \text{ with } M = \{a,b,c,d\} \text{ and } N = \{1,2,3,4\}$$
$$a \mapsto 1, b \mapsto 1, c \mapsto 3, d \mapsto 4$$

If $A \subset M$ is a subset, then $f(A) = \{f(x) \mid x \in A\} \subset N$ is called the **image** of A. For $B \subset N$ the set $f^{-1}(B) = \{x \in M \mid f(x) \in B\}$ is called the **preimage** of B under f.

In the example we have

$$f(\{a,b\}) = \{1\} \text{ and } f^{-1}(\{2\}) = \emptyset.$$

For one-element subsets $B = \{y\}$ we abbreviate $f^{-1}(y) = f^{-1}(\{y\})$. In general, the preimage of an element can have more than one element.

In the example: $f^{-1}(1) = \{a,b\}$.
If $f\colon M \to N$ is a map and $A \subset M$ a subset, then the **restriction** of f to A is defined by

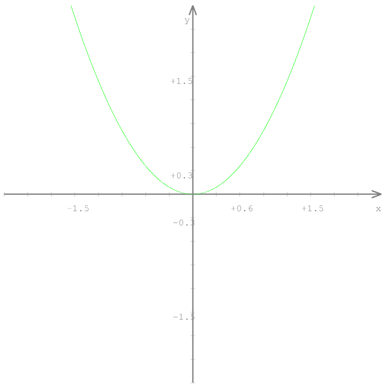
$$f|_A : A \to N, a \mapsto f(a).$$

2.2. The set

$$\Gamma_f = \{(x,y) \in M \times N \mid y = f(x)\}$$

is called the **graph** of f.

Example: $f\colon \mathbb{R} \to \mathbb{R}, x \mapsto f(x) = x^2$



One can recover the map from its graph: If $x_0 \in M$ is mapped to $y_0 = f(x_0) \in N$ then

$$\Gamma_f \cap (\{x_0\} \times N) = \{(x_0, y_0)\} \subset M \times N.$$

2.3. The most common maps are **real-valued functions**

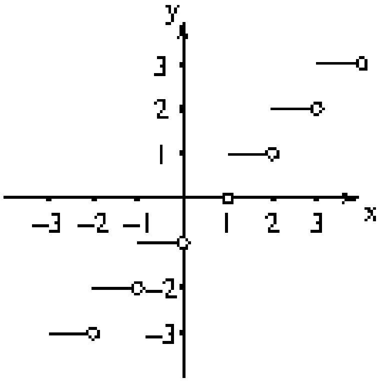
$$f \colon D \to \mathbb{R}$$

where $D \subset \mathbb{R}$. D is called the domain of definition of f

Examples

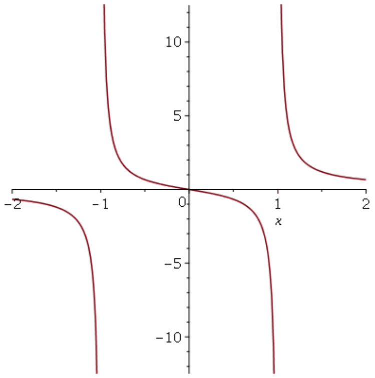
1. $y = x$

2. $$entier : \mathbb{R} \to \mathbb{R}, entier(x) := \lfloor x \rfloor = \max\{n \in \mathbb{Z} \mid n \leq x\}$$



3. $y = x^2$

4. $y = \frac{x}{x^2-1}$.



In this case the natural domain of definition is $D = \mathbb{R} \setminus \{\pm 1\}$, since the formula does not define a value for $x = \pm 1$.

Two properties of maps deserve a special name.

Definition 2.4. A map $f\colon M \to N$ is called **injective** if

$$x_1, x_2 \in M, x_1 \neq x_2 \implies f(x_1) \neq f(x_2)$$

holds. $f\colon M \to N$ is called **surjective** if for all $y \in N$ there exists an $x \in M$ with $f(x) = y$.

In the example of 2.1 f is neither injective nor surjective.
A map $f\colon M \to N$ is called **bijective** if it is both injective and surjective. In this case there is the **inverse map**

$$f^{-1} : N \to M$$

defined by $f^{-1}(y) = x$ for the unique $x \in M$ with $f(x) = y$.

Note that the symbol f^{-1} is overloaded. For bijective f it usually refers to the inverse map, while for arbitrary maps $f\colon M \to N$ it refers to the map

$$f^{-1} : 2^N \to 2^M, B \mapsto f^{-1}(B)$$

defined by taking the preimage.

2.2 Counting

If $f\colon M \to N$ is bijective then $|M| = |N|$. Conversely, we have

Theorem 2.5. *Let $f\colon M \to N$ be a map between finite sets with $|M| = |N|$. The following are equivalent (TFAE)*

a) f is injective,

b) f is surjective,

c) f is bijective.

Proof. Since $(\, a) \wedge b)\,) \Leftrightarrow c)$ holds by definition, it is enough to prove $a) \Leftrightarrow b)$. For $b) \Rightarrow a)$ suppose f is surjective. Then

$$\sum_{n\in N} |f^{-1}(n)| \geq \sum_{n\in N} 1 = |N|,$$

since each set $f^{-1}(n)$ is non-empty. On the other hand the sets $f^{-1}(n)$ are disjoint for differrent n's. Hence

$$\sum_{n\in N} |f^{-1}(n)| = |\bigcup_{n\in N} f^{-1}(n)| = |M|.$$

So $|M| \geq |N|$. The assumption $|M| = |N|$ implies that we have equality in each step of the chain

$$|M| = \sum_{n\in N} |f^{-1}(n)| \geq \sum_{n\in N} 1 = |N|.$$

Thus $|f^{-1}(n)| = 1$ for each $n \in N$, i.e., f is injective.

Conversely, for $a) \Rightarrow b)$ assume that f is injective. Then

$$|M| = \sum_{n\in N} |f^{-1}(n)| \leq \sum_{n\in N} 1 = |N|,$$

since each set $f^{-1}(n)$ contains at most one element. So $|M| = |N|$ implies $|f^{-1}(n)| = 1$ for each $n \in N$. Hence f is surjective. □

Corollary 2.6 (of the proof)**.** *If $f\colon M \to N$ is injective then $|M| \leq |N|$. If $f\colon M \to N$ is surjective then $|M| \geq |N|$.*

Corollary 2.7 (Pigeonhole principal)**.** *A map $f : M \to N$ with $|M| > |N|$ is not injective.*

2.8. Example. For arbitrary n^2+1 points $p_i = (x_i, y_i)$ in a square

$$Q = \{(x,y) \in \mathbb{R}^2 \mid 0 \leq x < n, 0 \leq y < n\}$$

of area n^2 there exist two points $p_i, p_j, i \neq j$ with distance $dist(p_i, p_j) \leq \sqrt{2}$.

Proof. By Pythagoras, two points in a square of area 1 have distance $\leq \sqrt{2}$. We decompose Q into n^2 disjoint squares

$$Q_{(k,l)} = \{(x,y) \mid k-1 \leq x < k, l-1 \leq y < l\}$$

and define a map

$$f\colon \{1, \dots, n^2+1\} \to \{(k,l) \in \mathbb{N} \times \mathbb{N} \mid 1 \leq k < n, 1 \leq l < n\}$$

by $f(i) = (k,l)$ if $p_i \in Q_{(k,l)}$.

It is enough to prove that f is not injective. This is clear by the pigeonhole principle. □

Example 2.9.

$$\sum_{i=1}^{n} i \binom{n}{i} = n2^{n-1}$$

This can be proved by induction. A much nicer proof uses for $M = \{1, \dots, n\}$ the correspondence

$$C = \{(x, A) \in M \times 2^M \mid x \in A\}.$$

We count the elements of C in two ways. Consider the projection $\pi_2 : C \to 2^M$ onto the second factor. The fact that $|\pi_2^{-1}(A)| = |A|$ depends only on the number of elements of A gives the left hand side:

$$|C| = \sum_{A \in 2^M} |\pi_2^{-1}(A)| = \sum_{i=1}^{n} i \binom{n}{i}.$$

Projection to the first component $\pi_1 : C \to M$ gives the right hand side because

$$|\pi_1^{-1}(k)| = |\{\{k\} \cup A \mid k \notin A\}| = 2^{n-1}$$

is independent of $k \in M$.

Notation 2.10. Given sets M and N we denote by

$$N^M = \{f\colon M \to N\}$$

the set of all maps from M to N. This is compatible with the notion 2^M of the power set: For each subset $A \subset M$ we define its characteristic function

$$\chi_A : M \to \{0,1\}$$

by

$$\chi_A(m) = \begin{cases} 0 & \text{if } m \notin A \\ 1 & \text{if } m \in A \end{cases}.$$

Then

$$\chi : 2^M \to \{0,1\}^M, A \mapsto \chi_A$$

is a bijection. Note

$$|N^M| = |N|^{|M|}.$$

Definition 2.11. If $f\colon M \to N$ and $g\colon N \to K$ are maps, then the composition

$$g \circ f\colon M \to K$$

of f and g is defined by

$$(g \circ f)(m) = g(f(m)).$$

Composition of maps is associative:

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

holds any triple of maps

$$f\colon M \to N, g\colon N \to K, h\colon K \to L.$$

$$\begin{tikzcd}
M \arrow[r, "f"] \arrow[rr, bend left, "g\circ f"] \arrow[rrr, bend left=50, "h\circ(g\circ f)"] \arrow[rrr, bend right=50, "(h\circ g)\circ f"'] & N \arrow[r, "g"] \arrow[rr, bend right, "h\circ g"'] & K \arrow[r, "h"] & L.
\end{tikzcd}$$

Indeed,

$$\begin{aligned}((h \circ g) \circ f)(m) &= (h \circ g)(f(m)) \\ &= h(g(f(m))) \\ &= h((g \circ f)(m)) \\ &= (h \circ (g \circ f))(m)\end{aligned}$$

holds for all $m \in M$.

2.3 Existence and all quantifiers

2.12. The phrases **for all** and **there exists** are used very often in mathematics. One abbreviates

$$\forall = \text{ for all} \qquad \text{and} \qquad \exists = \text{ there exists.}$$

Example.

$$\begin{aligned}&f\colon M \to N \text{ is surjective} \\ &\iff \text{ for all } n \in N \text{ there exists an } m \in M \text{ such that } f(m) = n \\ &\iff \forall n \in N \, \exists m \in M : f(m) = n.\end{aligned}$$

Here we replace ":" by "such that". Under negation the quantifiers $\forall$ and $\exists$ are interchanges in a de Morgan style rule.

$$\begin{aligned}&f\colon M \to N \text{ is not surjective} \\ &\iff \neg(\forall n \in N \, \exists m \in M : f(m) = n) \\ &\iff \exists n \in N : \neg(\exists m \in M : f(m) = n) \\ &\iff \exists n \in N : \forall m \in M \, \neg(f(m) = n) \\ &\iff \exists n \in N : \forall m \in M \; f(m) \neq n \\ &\iff \text{ there exists an } n \in N \text{ such that for all } m \in M,\ f(m) \text{ is not equal } n.\end{aligned}$$

translates correctly under de Morgan's rule.

2.13. Arbitrary unions and intersections. Given an arbitrary family $(A_i)_{i \in I}$ of subsets of a set M, i.e., a map

$$I \to 2^M,\ i \mapsto A_i$$

we define the union and the intersection by

$$\bigcup_{i\in I} A_i = \{x \in M \mid \exists i \in I : x \in A_i\}$$

and

$$\bigcap_{i\in I} A_i = \{x \in M \mid \forall i \in I\, x \in A_i\}.$$

De Morgan's rule still holds:

$$\overline{\bigcup_{i\in I} A_i} = \bigcap_{i\in I} \overline{A_i} \quad \text{and} \quad \overline{\bigcap_{i\in I} A_i} = \bigcup_{i\in I} \overline{A_i}.$$

Indeed,

$$\begin{aligned}
\overline{\bigcup_{i\in I} A_i} &= \{x \in M \mid \neg(\exists i \in I : x \in A_i)\} \\
&= \{x \in M \mid \nexists i \in I : x \in A_i\} \\
&= \{x \in M \mid \forall i \in I \neg(x \in A_i)\} \\
&= \{x \in M \mid \forall i \in I\; x \notin A_i\} \\
&= \{x \in M \mid \forall i \in I\; x \in \overline{A_i}\} = \bigcap_{i\in I} \overline{A_i}.
\end{aligned}$$

3 Equivalence relations and congruences

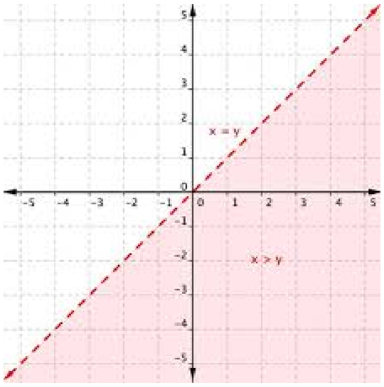
In mathematics and computer science one frequently considers relations.
Example. $\geq$ (greater or equal) is a relation on $\mathbb{R}$: For any two real numbers x, y the relation

$$x \geq y \text{ is either true or false.}$$

Formally, we define a relation as follows.

Definition 3.1. Let M be a set. A relation R is a subset $R \subset M \times M$. For $x, y \in M$ the relation R is satisfied, if $(x, y) \in R$.
Examples. For $\geq$ we have

$$R_{\geq} = \{(x, y) \in \mathbb{R}^2 | x \geq y\}$$



$R_{=}$ is the diagonal in $\mathbb{R}^2$.

3.1 Equivalence relations

In this section we will study equivalence relations. Our goal is to weaken the notion of **equal** to a notion of **equivalent** or **similar**.

Example 3.2. 1) Let $f\colon M \to N$ be a map. We say $a, b \in M$ are equivalent, in symbols $a \sim b$, if $f(a) = f(b)$.

2) Let $M = \mathbb{Z}$ and let n be a positive integer. Two integers a, b are called congruent modulo n, in symbols

$$a \equiv b \mod n$$

if $a - b$ is divisible by n.

What are the desired properties on an equivalence relation?

Definition 3.3. A subset $R \subset M \times M$ is called an **equivalence relation** (we write $a \sim b$, if $(a, b) \in R$) if the following properties hold:

i) (reflexive) $a \sim a \quad \forall a \in M$

ii) (symmetry) $a \sim b \Rightarrow b \sim a \quad \forall a, b \in M$

iii) (transitivity) $a \sim b$ and $b \sim c \Rightarrow a \sim c \quad \forall a, b, c \in M$

Examples.

0) Equality is an equivalence relation on any set M.

1) If $f\colon M \to N$ is a map, then $a \sim b \iff f(a) = f(b)$ is a equivalence relation because $=$ is a equivalence relation on N.

2) $a \equiv b \mod n$ is an equivalence relation:

   i) $a - a = 0 \cdot n$

   ii) $a - b = \alpha n \Rightarrow b - a = (-\alpha)n$

   iii) $a - b = \alpha n$ and $b - c = \beta n \Rightarrow a - c = (\alpha + \beta)n$

3) The relation $\geq$ on $\mathbb{R}$ is not an equivalence relation. i) and iii) are satisfied since

$$x \geq x \; \forall x \in \mathbb{R}$$

and

$$x \geq y \text{ and } y \geq z \Rightarrow x \geq z \; \forall x, y, z \in \mathbb{R}$$

hold. But ii) does not hold: $x \geq y \not\Rightarrow y \geq x$.

Definition-Theorem 3.4. Let $\sim$ be an equivalence relation on M. For $a \in M$ we call

$$[a] = \{b \in M \mid b \sim a\}$$

the **equivalence class** of a. Any element $b \in [a]$ is called a **representative** of the equivalence class $[a]$.

Any two equivalence classes $[a]$ and $[b]$ are either equal or disjoint.

Example. The equivalence classes of $\equiv \mod 3$ are
$[0] = \{0, \pm 3, \pm 6, \ldots\}$
$[1] = \{\ldots, -5, -2, 1, 4, 7, \ldots\}$
$[2] = \{\ldots, -4, -1, 2, 5, 8, \ldots\}$.

Proof. Suppose $[a] \cap [b] \neq \emptyset$, say $c \in [a] \cap [b]$. We have to show $[a] = [b]$.

If $d \in [a]$ then $d \sim a \sim c \sim b$ by ii). So iii) implies $d \sim b$. Thus $d \in [b]$. This proves $[a] \subset [b]$. The converse inclusion $[b] \subset [a]$ follows by the same argument. □

Definition 3.5. Let $\sim$ be an equivalence relation on M. Then

$$M/\sim := \{[a] \mid a \in M\} \subset 2^M,$$

"M modulo $\sim$", denotes the set of equivalence classes.

$$\pi : M \to M/\sim, a \mapsto [a]$$

is called the canonical map or projection to the quotient $M/\sim$

Example. For $\equiv \mod 3$ the quotient map is

$$\begin{aligned} \pi : \mathbb{Z} &\to \{[0], [1], [2]\} \\ n &\mapsto [\text{remainder of } n \text{ divided by } 3]. \end{aligned}$$

Remark. Apparently, we have

$$\pi(a) = \pi(b) \iff [a] = [b] \iff a \sim b$$

and

$$\pi^{-1}([a]) = [a].$$

Thus we can recover the equivalence relation from π, and in principal every equivalence relation is as Example 3.2 1). However, the main purpose of equivalence relations is to use suitable M and $\sim$ to construct new interesting sets $M/\sim$.

Example 3.6. (The construction of $\mathbb{Q}$ from $\mathbb{Z}$). We assume that $\mathbb{Z}$ together with the operations $+ : \mathbb{Z} \times \mathbb{Z}, (a, b) \mapsto a + b$ and $\cdot : \mathbb{Z} \times \mathbb{Z}, (a, b) \mapsto a \cdot b$ are given. We want to construct the field of rational numbers $\mathbb{Q}$.

For this we consider

$$M = \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$$

and the equivalence relation on M defined by

$$(p_1, q_1) \sim (p_2, q_2) \iff p_1q_2 = p_2q_1.$$

This is indeed a equivalence relation. i) and ii) are clear. For the transitivity, suppose

$$(p_1, q_1) \sim (p_2, q_2) \sim (p_3, q_3)$$

hence

$$p_1q_2 = p_2q_1 \text{ and } p_2q_3 = p_3q_2.$$

Then

$$p_1q_2q_3 = p_2q_1q_3 = p_2q_3q_1 = p_3q_2q_1$$

which implies

$$q_2(p_1q_3 - p_3q_1) = 0$$

Since $q_2 \neq 0$ this implies $p_1q_3 - p_3q_1 = 0 \in \mathbb{Z}$. So $(p_1, q_1) \sim (p_3, q_3)$.

As usual we denote the equivalence class $[(p, q)]$ by $\frac{p}{q}$. We now define $\mathbb{Q}$ as a set by

$$\mathbb{Q} = M/\sim = \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})/\sim .$$

Addition and multiplication on $\mathbb{Q}$ are defined via representatives:

$$\frac{p}{q} + \frac{r}{s} := \frac{ps + qr}{qs}$$

and

$$\frac{p}{q}.\frac{r}{s} := \frac{pr}{qs}$$

To verify that this defines a well-defined map

$$+ : (M/\sim) \times (M/\sim) \to (M/\sim)$$

we have to show that the result $[(ps + qr, qs)]$ does not depend on the choice of the representatives. So we have to show that

$$(p_1, q_1) \in [(p, q)] \text{ and } (r_1, s_1) \in [(r, s)] \implies (p_1s_1 + q_1r_1, q_1s_1) \in [(ps + qr, qs)].$$

We change the first representative and keep the second.

$$\begin{aligned}(p_1, q_1) \sim (p, q) &\implies p_1q = pq_1\\ &\implies p_1sqs = psq_1s\\ &\implies (p_1s + q_1r)qs = (ps + qr)q_1s\\ &\implies (p_1s + q_1r_1, q_1s) \sim (ps + qr, qs)\end{aligned}$$

shows the independence from the choice of a representative of the first summand. The argument for the second summand is similar. That the multiplication is well-defined follows by a similar but easier computation.

Finally, we can embed $\mathbb{Z}$ into $\mathbb{Q}$ by

$$\mathbb{Z} \hookrightarrow \mathbb{Q}, n \mapsto \frac{n}{1}.$$

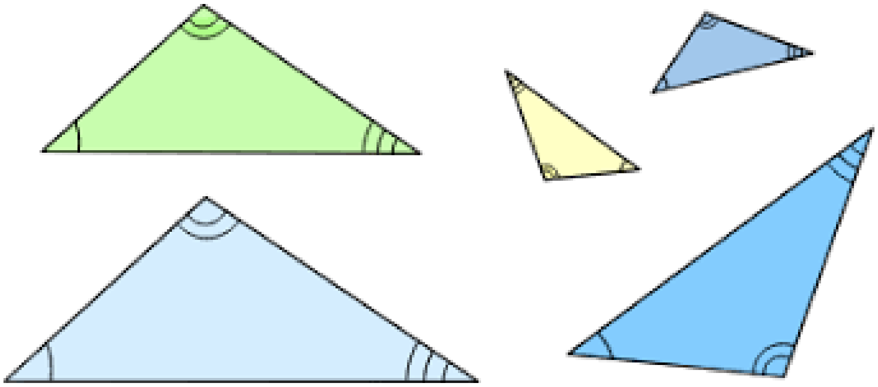
(The symbol $\hookrightarrow$ is used for injective maps which one would like to regard as an inclusion.)

Each element of $\mathbb{Q}$ has a distinguished representative, namely

$$(p, q) \in \frac{p}{q} \text{ with } p, q \text{ coprime and } q \geq 1.$$

Remark. In general there are no distinguished representatives for equivalence classes.

Example (Similarity of triangles). Two triangles with angles (α, β, γ) and $(\alpha', \beta', \gamma')$ are similar if the angles coincide up to their order. Similarity is an equivalence relation on the set of plane triangles.



How would you define a distinguished representative?

3.2 Congruences

3.7. In the following we will study the equivalence relation $\equiv \mod n$ in detail. We abbreviate $\mathbb{Z}/(\equiv \mod n)$ by $\mathbb{Z}/n$. Every element of $\mathbb{Z}/n$ has a distinguished representative $i \in \{0, 1, \ldots, n-1\}$ given by the remainder of the division by n. The residue class of i is

$$[i] = \{i + kn | k \in \mathbb{Z}\} \subset \mathbb{Z}.$$

Frequently, the notation

$$\bar{i} = [i]$$

is used. The set

$$\mathbb{Z}/n = \{\bar{0}, \ldots, \overline{n-1}\}$$

has precisely n elements. The elements of $\mathbb{Z}/n$ can be added and multiplied:

$$\bar{i} + \bar{j} = \overline{i+j}, \bar{i} \cdot \bar{j} = \overline{i \cdot j}.$$

These $+$ and $\cdot$ on $\mathbb{Z}/n$ satisfy the usual laws of addition and multiplication:

- (assocative) $(\bar{i} + \bar{j}) + \bar{k} = \bar{i} + (\bar{j} + \bar{k}), \quad (\bar{i} \cdot \bar{j}) \cdot \bar{k} = \bar{i} \cdot (\bar{j} \cdot \bar{k})$,
- (distributive) $(\bar{i} + \bar{j}) \cdot \bar{k} = \bar{i} \cdot \bar{k} + \bar{j} \cdot \bar{k}$,
- (commutative) $\bar{i} + \bar{j} = \bar{j} + \bar{i}, \quad \bar{i} \cdot \bar{j} = \bar{j} \cdot \bar{i}$.

Example. $n = 6$. Then

$$(\bar{2} + \bar{2}) + \bar{5} = \bar{4} + \bar{5} = \bar{9} = \bar{3} \in \mathbb{Z}/6$$

coincides with

$$\bar{2} + (\bar{2} + \bar{5}) = \bar{2} + \bar{7} = \bar{2} + \bar{1} = \bar{3}$$

and

$$(\bar{2} \cdot \bar{2}) \cdot \bar{5} = \bar{4} \cdot \bar{5} = \overline{20} = \bar{2} \in \mathbb{Z}/6$$

coincides with

$$\bar{2} \cdot (\bar{2} \cdot \bar{5}) = \bar{2} \cdot \overline{10} = \bar{2} \cdot \bar{4} = \bar{8} = \bar{2}.$$

To see these laws in general it is best to prove that in the definition

$$\bar{i} + \bar{j} = \overline{i+j}, \bar{i} \cdot \bar{j} = \overline{i \cdot j}$$

the result does not depend on the choice of the representative $i \in \bar{i}$ and $j \in \bar{j}$. So we have to show

$$\begin{aligned} & i_1 \equiv i_2, j_1 \equiv j_2 \mod n \\ \Longrightarrow\ & i_1 + i_2 \equiv j_1 + j_2 \text{ and } i_1 i_2 \equiv j_1 j_2 \mod n \\ & . \end{aligned}$$

Indeed

$$\begin{aligned} & i_1 - i_2 = kn \text{ and } j_1 - j_2 = ln \\ \implies & (i_1 + j_1) - (i_2 + j_2) = (k + l)n \\ & \text{ and } i_1 j_1 - i_2 j_2 = (i_1 l + k j_1 + kln)n. \end{aligned}$$

The calculation rules in $\mathbb{Z}/n$ follow now from the corresponding rules in $\mathbb{Z}$.

Remark. In $\mathbb{Z}/n$ it is possible that

$$\overline{a} \cdot \overline{b} = \overline{0} \text{ for some } \overline{a} \neq \overline{0} \text{ and } \overline{b} \neq \overline{0}.$$

Example. $\overline{2} \cdot \overline{3} = \overline{0} \in \mathbb{Z}/6$

Hence in general, there is no sensible ‘division’ in $\mathbb{Z}/n$. An exception is the case if $n = p$ is a prime number. (For $a, b \in \mathbb{Z}$ we write $a|b$ for the phrase a **divides** b).

$$\overline{a} \cdot \overline{b} = \overline{0} \implies a \cdot b \equiv 0 \mod p \implies p|ab \implies p|a \text{ or } p|b.$$

Theorem-Definition 3.8. Let p be a prime number and let $\overline{a} \in \mathbb{Z}/p$ be an element $\overline{a} \neq \overline{0}$. Then multiplication by $\overline{a}$ defines a bijective map

$$\mathbb{Z}/p \to \mathbb{Z}/p, \quad \overline{b} \mapsto \overline{b} \cdot \overline{a}.$$

The **inverse** of $\overline{a} \in \mathbb{Z}/p$ is the preimage of $\overline{1}$ which we denote

$$(\overline{a})^{-1}.$$

It is represented by a $u \in \mathbb{Z}$ such that $ua \equiv 1 \mod p$. u is called an **inverse** of a $\mod p$.

Proof. $\overline{b_1}\overline{a} = \overline{b_2}\overline{a} \implies (\overline{b_1} - \overline{b_2})\overline{a} = \overline{0}, \implies p|(b_1 - b_2)|a$ and $p \nmid a \implies p|(b_1 - b_2) \implies \overline{b_1} = \overline{b_2}$. Thus the map is injective and hence also surjective, since $\mathbb{Z}/p$ is finite. So $\exists \overline{u}$ with $\overline{u} \cdot \overline{a} = \overline{1} \implies \exists u \in \mathbb{Z}$ with $ua \equiv 1 \mod p$ □

3.9. Simultaneous solutions of congruences. Given two integers $m, n > 1$ we have a map

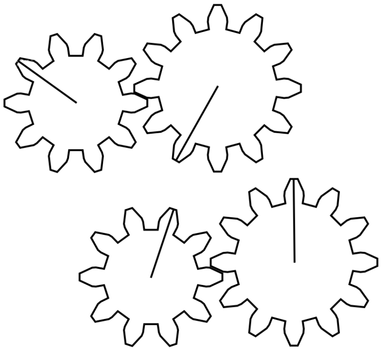
$$\mathbb{Z} \to \mathbb{Z}/m \times \mathbb{Z}/n, i \mapsto (i \bmod m, i \bmod n).$$

We ask whether a given pair $(\overline{a}, \overline{b}) \in \mathbb{Z}/m \times \mathbb{Z}/n$ lies in the image. In other words we ask whether the congruences

$$\begin{aligned} x &\equiv a \mod m \\ x &\equiv b \mod n \end{aligned}$$

can be solved simultaneously by an $x \in \mathbb{Z}$.

Example.(Cog wheels)



Is it possible to turn the wheels from the first position to the second position? The answer is no. We would have to solve the congruences

$$\begin{aligned} x &\equiv 2 \mod 10 \\ x &\equiv 7 \mod 12. \end{aligned}$$

But $2 \not\equiv 7 \mod 2$.

A necessary condition for the solvability of

$$\begin{aligned} x &\equiv a \mod n \\ x &\equiv b \mod m \end{aligned}$$

is $a \equiv b \mod \gcd(n, m)$.

How to compute the greatest common divisor? A method applied in high school uses factoring. Given $n, m \in \mathbb{N}$ factor

$$\begin{aligned} n &= p_1^{e_1} \cdots, p_r^{e_r} = \prod_{i=1}^{r} p_i^{e_i} \\ m &= p_1^{f_1} \cdots, p_r^{f_r} = \prod_{i=1}^{r} p_i^{f_i} \end{aligned}$$

into primes, where we allow also $e_i = 0$ or $f_i = 0$. Then

$$\gcd(m, n) = \prod_{i=1}^{r} p_i^{\min(e_i, f_i)}.$$

But factoring is hard. So hard that it is actually the basis of one of the first public-key cryptosystems.

3.10. RSA (Rivest-Shamir and Adleman, 1978) Bob wants to send Alice a message through an open channel such that Eve who might listen to the channel cannot decipher the message. In classical cryptosystems Alice and Bob might share a common secret on which the encryption, decryption is built. For this however Bob and Alice should have met or should have interchanged the secret through a trust worthy third party. In the internet this is not feasible. The idea of public-key encryption overcomes this difficulty.

RSA relies on the difficulty of factoring and of Fermat's little Theorem.
Definition. The Euler φ-function

$$\varphi : \mathbb{N} \to \mathbb{N}$$

is defined by

$$\varphi(n) = |\{a \mid 1 \leq a < n \text{ with } \gcd(a, n) = 1\}|.$$

Clearly, $\varphi(p) = p - 1$ for a prime number p. One of the remarkable properties of φ is $\varphi(mn) = \varphi(m)\varphi(n)$ if $\gcd(m, n) = 1$.

Theorem 3.11 (Fermat)**.** *Let $x, n \in \mathbb{Z}$ be integers with $\gcd(x, n) = 1$. Then*

$$x^{\varphi(n)} \equiv 1 \mod n.$$

Now, Alice chooses two large prime numbers p_A, q_A computes $n_A = p_A q_A$ and $\varphi(n_A) = (p_A - 1)(q_A - 1)$. In practise p_A and q_A will have at least 100 digits. In addition Alice chooses d_A, e_A with $d_A e_A \equiv 1 \mod \varphi(n_A)$. Then

$$n_A \text{ and } d_A$$

will be public while

$$p_A, q_A, \varphi(n_A) \text{ and } e_A$$

remain secret.

Bob wants to send a message

$$x \in \{0, \dots, n_A - 1\}$$

to Alice. With nearly 100% propabiltity we will have $\gcd(x, n_A) = 1$, since

$$\frac{\varphi(n_A)}{n_A} = \frac{p_A q_A - p_a - q_A + 1}{p_A q_A} \approx 1.$$

Bob computes

$$y \equiv x^{d_A} \mod n_A, y \in \{0, \dots, n_A - 1\}$$

and sends y through the public channel to Alice. Alice computes

$$z \equiv y^{e_A} \mod n_A, z \in \{0, \dots, n_A - 1\}$$

Since

$$\begin{aligned} z &\equiv (x^{d_A})^{e_A} \equiv x^{1+k\varphi(n_A)} \mod n_A \\ &\equiv x \cdot (x^{\varphi(n_A)})^k \equiv x \cdot 1 \mod n_A \end{aligned}$$

by Fermat's little theorem, Alice can recover x. Eve, who knows y, n_A and d_A, needs for the decryption the secret e_A. Knowing e_A allows with not too much effort to factor $n_A = p_A q_A$. However no fast factoring algorithm is known. So it is plausible that Eve cannot find e_A in a limited amount of time.

3.3 Euclidean algorithm and Chinese remainder theorem

Algorithm 3.12. (Extended Euclidean Algorithm)
Input: Integers $a > b > 1$.
Output: $d = gcd(a, b)$ and integers $u, v \in \mathbb{Z}$ satisfying $d = ua + vb$.

```
1. (Initialize)
   $x_1 = a,$   $u_1 = 1$, $v_1 = 0$
   $x_2 = b,$   $u_2 = 0$, $v_2 = 1$
   $i = 2$

2. while $x_i > 0$ do (
   (Division with remainder) Compute $q_i \in \mathbb{Z}$ with
   $x_{i+1} = x_{i-1} - q_i x_i$ and $0 \leq x_{i+1} \leq x_i$;
   Set $u_{i+1} = u_{i-1} - q_i u_i$, $v_{i+1} = v_{i-1} - v_i q_i$;
   $i = i + 1$)

3. Set $n = i - 1$ and return $d = x_n$ and $u_n, v_n$.
```

Proof of correctness. The algorithm terminates with $i = n + 1$ and $x_{n+1} = 0$ because the non-negative integers $x_i > x_{i+1}$ get smaller in the loop. We prove that

$$x_i = u_i a + v_i b$$

holds for each $i = 1, \dots, n$ by induction on i. This formula holds for $i = 1, 2$ by the choice of the initial values. Now suppose that $i > 2$ and that the formula holds for i and $i - 1$. Then

$$\begin{aligned} x_{i+1} &= x_{i-1} - q_i x_i \\ &= u_{i-1}a + v_{i-1}b - q_i(u_i a + v_i b) \\ &= (u_{i-1} - q_i u_i)a + (v_{i-1} - q_i v_i)b \\ &= u_{i+1}a + v_{i+1}b \end{aligned}$$

gives the induction step.

From $x_n = u_n a + v_n b$ we see that every common divisor of a and b divides x_n. It remains to prove that x_n itself is a divisor of a and b. For this we prove $x_n | x_{n-j}$ for all j with $0 \leq j \leq n - 1$ by induction on j

The cases $j = 0$ and $j = 1$ are clear since $x_{n+1} = 0$. The induction step follows from the identity $x_{n-(j+1)} = q_{n-j} x_{n-j} + x_{n-(j-1)}$.

In particular x_n divides $x_1 = a$ and $x_2 = b$. Hence $d = x_n$ is the greatest common divisor of a and b. □

Corollary 3.13. *Let p be a prime number and $a, b \in \mathbb{Z}$. Then*

$$p|ab \implies p|a \textit{ or } p|b.$$

Proof. Suppose $p \nmid a$. Then $gcd(a, p) = 1$ since 1 and p are the only divisors of p. By the Eulcidean algorithm there exist $u, v \in \mathbb{Z}$ with $1 = ua + vp$. Hence $p|(uab + vbp) = b$ since p divides both summands. □

Theorem 3.14 (Chinese remainder theorem)**.** *Let $m, n \in \mathbb{Z}$ be positive integers with $\gcd(n, m) = d$ and let $a, b \in \mathbb{Z}$. The simultaneous congruence*

$$\begin{aligned} x &\equiv a \mod m \\ x &\equiv b \mod n \end{aligned}$$

has a solution if and only if $a \equiv b \mod d$. If existent, then the solution is unique modulo the leat common multiple of m and n. If x is a solution, then

$$\{x + k \cdot \operatorname{lcm}(n, m) \mid k \in \mathbb{Z}\}$$

is the set of all solutions.

Proof. We already saw that the condition is necessary. Sufficient: Suppose $a = b + kd$. Write $d = un + vm$. Then $x_0 = vm$ solve the congruence

$$\begin{aligned} x_0 &\equiv 0 \mod m \\ x_0 &\equiv d \mod n \end{aligned}$$

and $x = a - kvm$ solves the desired congruence. If x' is a further solution then $x' - x \equiv 0 \mod \operatorname{lcm}(n, m) = \frac{nm}{d}$. □

3.4 Factorisation into prime numbers

Theorem 3.15 (Fundamental theorem of arithmetic)**.** *Every integer $n \neq 0$ can be factored*

$$n = \epsilon \prod_{i=1}^{r} p_i$$

with $p_1, \dots, p_r$ prime numbers and $\epsilon = \pm 1$. The factorisation is unique up to the order of the factors.

Proof. We may assume that $n > 0$ in which case $\epsilon = 1$.
Existence: We prove this by induction on n. If $n = 1$, then $r = 0$. If $n > 1$ and $n = p$ is a prime number, then $r = 1$ and $p_1 = p$. Otherwise we can factor $n = ab$ with $1 < a, b < n$. By the induction hypothesis a and b have factorisations into primes $a = p_1 \cdot \ldots \cdot p_r,$ and $b = q_1 \cdot \ldots \cdot q_s$ and

$$n = p_1 \cdot \ldots \cdot p_r \cdot q_1 \cdot \ldots \cdot q_s$$

is one for n.
Uniqueness: We proof this by induction on r. Suppose we have two factorizations

$$p_1 \cdot \ldots \cdot p_r = q_1 \cdot \ldots \cdot q_s.$$

Then since $p_r \mid ab \implies p_r \mid a$ or $p_r \mid b$ the prime p_r has to divide one of the factors q_j. Since q_j is prime we obtain $p_r = q_j$. After re-numbering $q_1 \ldots, q_s$ we may assume $j = s$. Then division by p_r gives $p_1 \cdot \ldots \cdot p_{r-1} = q_1 \cdot \ldots \cdot q_{s-1}$. The induction hypothesis implies $s - 1 = r - 1$ (so $p_r = q_r$) and after re-ordering $p_i = q_i$ for $i = 1, \dots, r - 1$.

□

4 The real numbers

In this section we will begin to collect the essential properties of $\mathbb{R}$ in the form of axioms. All our results in real analysis will be deduced from these axioms. The axiomatic method was introduced in Euclid's elements. Hilbert in 20-th century reintroduced this approach into all parts mathematics.

4.1 The axioms of a field

The first collection of axioms says that $\mathbb{R}$ together with the operations

$$+ : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, (a,b) \mapsto a+b$$

and

$$\cdot : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, (a,b) \mapsto a \cdot b$$

is a field in the sense of the following definition.

Definition 4.1. A triple $(K,+,\cdot)$ of a set K together with two maps $+ : K \times K \to K, (a,b) \mapsto a+b$ and $\cdot : K \times K \to K, (a,b) \mapsto a \cdot b$ is a **field** if the following axioms are satisfied.

K1 (Axioms for addition)

K1.1 (associativity of addition)

$$(a+b)+c = a+(b+c) \quad \forall\, a,b,c \in K$$

K1.2 (commutativity of addition)

$$a+b = a+b \quad \forall\, a,b \in K$$

K1.3 (existence of the zero element)

$$\exists\, 0 \in K \text{ such that } 0+a = a \quad \forall\, a \in K$$

K1.4 (existence of negative elements)

$$\forall\ a \in K\ \exists\, a' \in K \text{ such that } a'+a = 0$$

We call a' the negative of a denoted by $-a$.

K2 (Axioms for multiplication)

K2.1 (associativity of mutiplication)

$$(a \cdot b) \cdot c = a \cdot (b \cdot c) \quad \forall\, a, b, c \in K$$

K2.2 (commutativity of multiplictaion)

$$a \cdot b = a \cdot b \quad \forall\, a, b \in K$$

K2.3 (existence of the 1-element)

$$\exists\, 1 \in K \setminus \{0\} \text{ such that } 1 \cdot a = a \quad \forall\, a \in K$$

K2.4 (existence of inverse elements)

$$\forall\; a \in K \setminus \{0\} \;\exists\, a' \in K \text{ such that } a' \cdot a = 1$$

We call a' the inverse of a denoted by a^{-1} or $\frac{1}{a}$.

K3 (distributivity)

$$(a + b) \cdot c = a \cdot c + b \cdot c \quad \forall\, a, b, c \in K$$

$$a \cdot (b + c) = a \cdot b + a \cdot c \quad \forall\, a, b, c \in K$$

Example 4.2. 1. $\mathbb{Q}$ and $\mathbb{R}$ are fields.

2. $\mathbb{F}_p := (\mathbb{Z}/p, +, \cdot)$ for p a prime number is a field. In particular $\mathbb{F}_2$ is a field. Addition and multiplication can alternatively be defined by the Cayley tables:

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

3. $(\mathbb{Z}, +, \cdot)$ is not a field, since $2 \in \mathbb{Z}$ has no inverse in $\mathbb{Z}$.

Remark 4.3. A triple $(R, +, \cdot)$ of a set together with with two operations, which satisfy the field axioms except $K2.4$ is called a commutative ring with 1. So $(\mathbb{Z}, +, \cdot)$ and $(\mathbb{Z}/n, +, \cdot)$ are examples of commutative rings with 1.

A (general) ring is a triple where one requires only the axioms K1, K2.1 and K3.

In a field we have $ab = 0 \implies a = 0$ or $b = 0$. Indeed $a \neq 0 \implies 0 = a^{-1}0 = a^{-1}(ab) = ((a^{-1}a)b = 1 \cdot b = b$ by the property 3 of Proposition 4.4. In rings this is not necessarily true: $\overline{2} \cdot \overline{3} = 0 \in \mathbb{Z}/6$.

Proposition 4.4 (Properties of fields)**.** *Let $(K, +, \cdot)$ be a field. Then*

1. *0 and 1 are uniquely determined.*

2. *The negative of $a \in K$ and the inverse a^{-1} of $a \in K^* = K \setminus \{0\}$ are uniquely determined. We write $a - b$ for $a + (-b)$ and $\frac{a}{b}$ for $a \cdot b^{-1}$.*

3. *$(-1) \cdot (-1) = 1$ and $0 \cdot a = 0 \ \forall a \in K$.*

4. *In sums and products the result does not depend on how we set the brackets. The result does not depend on the order the summands or factors respectively.*

Proof.

1. Suppose $0' \in K$ is a further zero element. Then $0' + a = a \ \forall a \in K$, in particular

$$0 = 0' + 0 \overset{\text{K1.2}}{=} 0 + 0' \overset{\text{K1.3}}{=} 0'.$$

Similarly,

$$1 = 1' \cdot 1 = 1 \cdot 1' = 1'.$$

2. Let a' be a further negative element of a. Then $a' + a = 0$ and hence

$$\begin{aligned} -a &= 0 + (-a) = (a' + a) + (-a) \\ &= a' + (a + (-a)) = a' + ((-a) + a) \\ &= a' + 0 = 0 + a' = a'. \end{aligned}$$

For the inverse a^{-1} we argue similarly.

3. We have

$$0 \cdot a = (0 + 0) \cdot a = 0 \cdot a + 0 \cdot a.$$

Hence

$$\begin{aligned} 0 &= -(0\cdot a)+0\cdot a = -0\cdot a+(0\cdot a+0\cdot a) \\ &= (-(0\cdot a)+0\cdot a)+0\cdot a = 0+0\cdot a \\ &= 0\cdot a. \end{aligned}$$

We argue similarly for the other statement:

$$\begin{aligned} 0 &= 0\cdot(-1) = ((-1)+1)\cdot(-1) \\ &= (-1)\cdot(-1)+1\cdot(-1) \\ &= (-1)\cdot(-1)+(-1). \end{aligned}$$

This shows $(-1)\cdot(-1)=1$ by the uniqueness of the negative:

$$1+(-1)=0 \implies 1=-(-1).$$

4. We consider addition and argue with induction on the number n of summands. For $n=2$ there is nothing to prove. The case $n=3$, i.e.,

$$(a_1+a_2)+a_3 = a_1+(a_2+a_3)$$

is the associativity law K1.1.

For the induction step $n-1\to n$ we assume that $n\geq 4$ and that the assertion has been proved for smaller n. Let

$$(a_1+\cdots+a_k)+(a_{k+1}+\cdots+a_n)$$

be the outermost brackets.. We show

$$(a_1+\cdots+a_k)+(a_{k+1}+\cdots+a_n) = a_1+(a_2+(\cdots+(a_{n-1}+a_n)\cdots)).$$

In case $k=1$ this is clear by the induction hypothesis. So suppose $k\geq 2$. Then the induction hypothesis gives

$$a_1+\cdots+a_k = a_1+(a_2+\cdots+a_k),$$

and

$$\begin{aligned} (a_1+\cdots+a_k)+(a_{k+1}+\cdots+a_n) &= (a_1+(a_2+\cdots+a_k))+(a_{k+1}+\cdots+a_n) \\ &\overset{\text{K1.1}}{=} a_1+((a_2+\cdots+a_k)+(a_{k+1}+\cdots+a_n)) \\ &\overset{I.H.}{=} a_1+(a_2+(a_3+\cdots+a_n)). \end{aligned}$$

The argument for multiplication is the same. For the order of the summands we note that by the commutativity law K1.2 and by the result proved so far we have

$$a_1 + a_2 + \cdots + a_n = a_2 + a_1 + a_3 + \cdots + a_n.$$

The result follows since every permutation of the summands can be achieved by repeatedly interchanging neighbouring summands.

□

Finite fields play an important role in computer science in particular for error-correcting codes coding and in cryptography. We give a further example.

Example 4.5. The Cayley tables

$+$	0	1	-1
0	0	1	-1
1	1	-1	0
-1	-1	0	1

$\cdot$	0	1	-1
0	0	0	0
1	0	1	-1
-1	0	-1	1

give $K = \{0, 1, -1\}$ the structure of a field. With the map $K \to \mathbb{Z}/3$ induced by $0 \mapsto \overline{0}, 1 \mapsto \overline{1}$ and $-1 \mapsto \overline{2}$, we see that K can be identified with $\mathbb{F}_3$.

In the exercise we will answer the question whether there exists a field with precisely 4 elements.

4.2 The order axioms

$\mathbb{Q}$ and $\mathbb{R}$ are ordered fields in the sense of the following definition.

Definition 4.6. An **ordered field** is a field K together with a subset of positive elements

$$\{x \in K \mid x > 0\}$$

such that the following axioms hold:

A1: Each element $x \in K$ satisfies precisely one of the properties $x > 0$, $x = 0$ or $-x > 0$.

A2: If $x > 0$ and $y > 0$, then also $x + y > 0$.

A3: If $x > 0$ and $y > 0$, then also $x \cdot y > 0$.

In an ordered field we say $x > y$ if $x - y > 0$. We define $x \geq y :\iff x > y$ or $x = y$. Note that $\geq$ is a transitive relation on K.

If K is ordered, then we define the **absolute value** of $x \in K$ by

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

Notice:

1. $|x| \geq 0$, and $|x| = 0$ if and only if $x = 0$,
2. $|x \cdot y| = |x| \cdot |y| \; \forall x, y \in K$,
3. (Δ-inequality) $|x + y| \leq |x| + |y| \; \forall x, y \in K$.

The last statement is proved by considering all possible cases $x > 0, x = 0, x < 0, y > 0, y = 0, y < 0$.

For example, suppose $x > 0$ and $y < 0$. If $x + y \geq 0$, then $|x + y| = x + y < x - y = |x| + |y|$. If instead $x + y < 0$ then $|x + y| = -x - y < x - y = |x| + |y|$.

Remark 4.7. If K is an ordered field than $\mathbb{Q}$ embeds into K as follows. We write 1_K to distinguish $1 \in \mathbb{Q}$ from the 1-element in K. Since $1_K \neq 0$, we have $-1_K > 0$ or $1_K > 0$ by A1. Since $1_K = 1_K \cdot 1_K = (-1_K) \cdot (-1_K)$ we have in fact $1_K > 0$ by A3. We define a map

$$\mathbb{N} \hookrightarrow K \text{ by } n \mapsto \sum_{i=0}^{n} 1_K = n \cdot 1_K.$$

By A2 the image consists of strictly positive elements of K. So the map is indeed injective because if $n \cdot 1_K = m \cdot 1_K$ for $n > m$ then $(n - m)1_K = 0_K$ would not be strictly positive in K. Next we extend this map to a map

$$\mathbb{Z} \hookrightarrow K \text{ by } 0 \mapsto 0_K \text{ and } -n \mapsto -(n \cdot 1_K).$$

Finally, we define

$$\iota : \mathbb{Q} \hookrightarrow K$$

by

$$\frac{n}{m} \mapsto (n1_K) \cdot (m1_K)^{-1}.$$

It is easy to see that ι respects the field structures: $\iota(a+b) = \iota(a) + \iota(b)$ and $\iota(a \cdot b) = \iota(a) \cdot \iota(b)$ holds for all $a, b \in \mathbb{Q}$. Thus we may regard $\mathbb{Q}$ as a subfield of K.

A finite field $\mathbb{F}$ cannot be ordered: There is no injective map $\mathbb{N} \to \mathbb{F}$ by the pigeonhole principal.

If $\mathbb{N} \to K, n \mapsto n1_K$ is not injective then

$$p = \mathrm{char}(K) = \min\{n \in \mathbb{N} \mid n1_K = 0_K\}$$

is called the **characteristic** of K. The integer p is a prime number.

Indeed, if $p = ab$ for $0 < a, b < p$, then $a1_K, b1_K \neq 0$ by the definition of p and $ab1_K = a1_K \cdot b1_K \neq 0$ gives a contradiction.

We say K has characteristic zero, $\mathrm{char}(K) = 0$, if $n1_K \neq 0 \, \forall n \in \mathbb{N}$. $\mathbb{Q}$ embeds in every field of characteristic zero by the argument above.

Definition 4.8. An ordered field K is called **archimedean** if the following axiom holds

A4: For every $x \in K$ there exits an $n \in \mathbb{N}$ such that $n1_K > x$.

$\mathbb{Q}$ and $\mathbb{R}$ are archimedean ordered fields. There exists non-archimedean ordered fields, but right now we do not have techniques to describe an example. In a certain sense non-archimedean fields contain infinitely large elements: elements which are larger then any $n \in \mathbb{N} \subset \mathbb{Q} \subset K$.

4.3 Irrational numbers

Before we discuss the final axiom needed for $\mathbb{R}$, we recall why rational numbers are no sufficient. By Pythagoras the diagonal in a unit square has length c satisfying $c^2 = 1^2 + 1^2 = 2$. So $c = \sqrt{2}$. We claim that

$$\sqrt{2} \notin \mathbb{Q}.$$

Proof. Suppose $\sqrt{2} = \frac{p}{q}$. We may take p and q coprime. Then $2q^2 = p^2$. Hence p is not an odd number because squares of odd numbers are odd. So $2 \mid p$. If we write $p = 2p'$, then we obtain $q^2 = 2p'^2$ and q is even as well. Hence p and q are not coprime, a contradiction. □

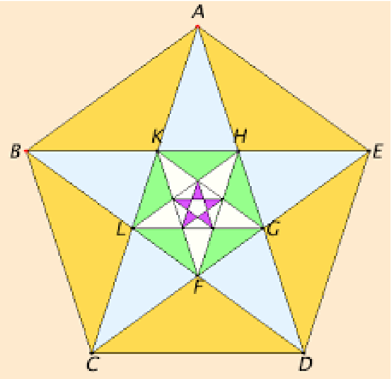
So rational numbers alone are not sufficient to compare the length of line segments.

Two line segments a and b are commensurable if there exist a line segment d such that $a = md$ and $b = md$.

In that case the ratio of the length $a : b = m : n$ is a rational number. If no such subsegment d exists, then a and b are called incommensurable. So the diagonal and the side of square are incommensurable.

If a and b are commensurable, then following Euclid's algorithm we can find the common line segment d by successively marking off the smaller segment from larger. After finitely many steps the remaining line sequences are equal and the algorithm stops with d.

The diagonal and the sided of a regular pentagon are visibly incommensurable:



We observe that ABL is an isosceles triangle. Indeed, $\measuredangle ABL = \measuredangle ABD = \measuredangle CAE$ holds by the rotational symmetry, and $\measuredangle CAE = \measuredangle CLD = \measuredangle BLA$ holds since $BEDB$ is a parallelogram. Hence the length s of the side $\overline{AB}$ coincides with the length of $\overline{BL}$. Hence $\overline{LC} = \overline{BL} = \overline{KB}$ has the length $d - s$ where d denotes the length of the diagonal $\overline{AC}$. $B, K, F,$ and C build for of five corners of another regular pentagon, with side length $d - s$ and diagonal length s. Finally, $\overline{LK}$ of length $s - (d - s)$ is a the side of the new pentagon $KHGFL$ in the center with diagonals $\overline{KF}$ of length $d - s$. The process does not stop.

Actually, $x = \frac{d}{s}$ is the golden ratio: $\frac{d}{s} = \frac{s}{d-s}$ implies that $x = \frac{1}{x-1}$. Hence x is a root of the equation $x^2 - x - 1 = 0$ which has solutions $\frac{1 \pm \sqrt{5}}{2}$. Since $x > 0$, we get

$$\frac{d}{s} = \frac{1 + \sqrt{5}}{2}.$$

5 Convergence and the completeness axiom

Convergence is the central idea of analysis. We define when a sequence of real numbers converges. This idea goes back to the ancient Greek mathematicians Eudoxus (390 - 337 BC) and Archimedes (287 - 212 BC). Isaac Newton (1643-1726) introduced this concept into modern science.

5.1 Sequences

Definition 5.1. A **sequence** of real numbers $(a_n) = (a_n)_{n \in \mathbb{N}}$ is a map

$$\mathbb{N} \to \mathbb{R},\ n \mapsto a_n$$

where we usually do not give the map a name, but instead use the index notation a_n for the n**-th term** or **element** of the sequence.

Example 5.2. 1. $(a_n) = (\frac{1}{n})$ i.e., $a_n = \frac{1}{n}$ has terms

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

2. $(b_n) = (n^2)$,

$$1, 4, 9, 16, \dots$$

is the sequence of square numbers,

3. $(c_n) = (2^n)$,

$$2, 4, 16, 64, \dots$$

is the sequence of powers of 2.

4. Sequences are used frequently in intelligence test where the task is to get the next term from an initial part of a sequence. For example,

$$2, 4, 3, 6, 5, 10, 9, 18, \dots$$

has the recursive rule

$$a_{n+1} = \begin{cases} 2a_n & \text{if } n \text{ is odd} \\ a_n - 1 & \text{if } n \text{ is odd} \end{cases}.$$

The sequence

$$(a_n) = (1, 2, 4, 6, 10, 12, 16, 18, 22, \ldots)$$

has the rule $a_n = n$-th prime number $- 1$.

5. Recursively defined sequences occur frequently.

   $$(f_n) = (0, 1, 1, 2, 3, 5, 8, 13, \ldots)$$

   defined by $f_0 = 0, f_1 = 1,$ and $f_{n+1} = f_n + f_{n-1}$ for $n \geq 2$ is called the sequence of Fibonacci numbers. In the exercise we will prove:

   $$f_n = \frac{1}{\sqrt{5}}\big((\frac{1+\sqrt{5}}{2})^n - (\frac{1-\sqrt{5}}{2})^n)\big).$$

   A general method to get closed formulas for sequences defined by linear recursions will be a topic of the lecture MfI 2.

6. In analysis sequences are used to get arbitrary good approximations of a real number. For example,

   $$(3, 3.1, 3.14, 3.141, 3.1415, \ldots)$$

   are better and better approximations of Archimedes constant

   $$\pi = 3.14592653589793238462643 3\ldots$$

   The letter π for (half) the perimeter is in use since the 18-th century.

5.2 Convergence

We give the concept of arbitrary good approximations a precise meaning:

Definition 5.3. Let (a_n) be a sequence of real numbers and a a further real number. We say (a_n) converges to a, in sympols

$$\lim_{n\to\infty} a_n = a,$$

if $\forall\,\varepsilon > 0\,\exists n_0 \in \mathbb{N}$ such that $|a_n - a| < \varepsilon\ \forall n \geq n_0$. We call a is the **limit** of the sequence (a_n).

Example 5.4.

1. The limit of $(a_n) = (\frac{1}{n})$ is $\lim_{n\to\infty} \frac{1}{n} = 0$. Indeed given $\varepsilon > 0$, there exits an $n_0 \in \mathbb{N}$ with $n_0 > \frac{1}{\varepsilon}$ by the archimedean axiom. Then
$$|\frac{1}{n} - 0| = \frac{1}{n} \leq \frac{1}{n_0} < \varepsilon\,\forall\, n \geq n_0.$$
2. We have
$$\lim_{n\to\infty} \frac{n+1}{n} = 1.$$
Indeed,
$$|\frac{n+1}{n} - 1| = \frac{1}{n} < \varepsilon\,\forall\, n \geq n_0$$
if we choose $n_0 = \lceil\frac{1}{\varepsilon}\rceil$.
3. The **constant sequence** (a_n) with $a_n = a$ for all n converges to a.
4. The sequence $((-1)^n)$ does not converges. Indeed suppose $\lim_{n\to\infty} = a$ for some a, then for $\varepsilon = 1$ there exists an $n_0 \in \mathbb{N}$ with $|(-1)^n - a| < 1 \forall n \geq n_0$. In particular,
$$\begin{aligned} 2 =& |(-1)^{n_0+1} - a + a - (-1)^{n_0}| \\ \overset{\Delta\text{-ineq}}{\leq}& |(-1)^{n_0+1} - a| + |a - (-1)^{n_0}| < 1 + 1 = 2, \end{aligned}$$
a contradiction. Sequences which do not converge are called **divergent**.

Remark 5.5. If (a_n) converges to a, then for arbitrary small $\varepsilon > 0$ all but finitely many terms a_n lie in the interval

$$]-\varepsilon + a, a + \varepsilon[.$$

Some times $]-\varepsilon + a, a + \varepsilon[$ is called an ε-neighborhood of a.

Definition 5.6. Let $a < b$ be two real numbers. Then intervals with boundary point a, b are defined by

$$\begin{array}{rcll} [a,b] & := & \{x \in \mathbb{R} \mid a \le x \le b\} & (\textbf{closed interval}), \\ ]a,b[& := & \{x \in \mathbb{R} \mid a < x < b\} & (\textbf{open interval}), \\ [a,b[& := & \{x \in \mathbb{R} \mid a \le x < b\} & (\textbf{half-open interval}), \\ ]a,b] & := & \{x \in \mathbb{R} \mid a < x \le b\} & (\textbf{half-open interval}). \end{array}$$

The notation $(a, b) =]a, b[, (a, b] =]a, b]$ etc. are also in use.

Remark 5.7. The limit $\lim a_n$ of a convergent sequence is uniquely determined.

Proof. Let a and a' be limits of (a_n). Suppose $a \neq a'$. Then for $\varepsilon = \frac{|a-a'|}{2} > 0$ there exist integers n_1, n_2 such that

$$|a - a_n| < \varepsilon \,\forall\, n \geq n_1 \text{ and } |a' - a_n| < \varepsilon \,\forall\, n \geq n_2.$$

Then for $n \geq \max(n_1, n_2)$ we obtain

$$\begin{array}{rcl} |a - a'| & = & |a - a_n + a_n - a| \\ & \overset{\triangle\text{-ineq}}{\leq} & |a - a_n| + |a_n - a'| < \varepsilon + \varepsilon = |a - a'|, \end{array}$$

a contradiction. □

Proposition 5.8 (Calculation rules for limits)**.** *Let (a_n) and (b_n) be convergent sequences with limits $a = \lim a_n$ and $b = \lim b_n$. Then the following holds:*

1. *The sequence $(a_n + b_n)$ is also convergent with limit $a + b$. In other words:*

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

   if the right hand side exists.

2. *The sequence $(a_n \cdot b_n)$ is also convergent with limit $a \cdot b$. In other words:*

   $$\lim_{n\to\infty} (a_n \cdot b_n) = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n,$$

   if the right hand side exists.

3. If $b_n \neq 0$ *and* $b \neq 0$*, then the sequence* $(\frac{a_n}{b_n})$ *is also convergent and*

$$\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{\lim a_n}{\lim b_n} = \frac{a}{b}.$$

Proof. 1. Let $\varepsilon > 0$ be given. By assumption $\exists n_1, n_2 \in \mathbb{N}$ such that

$$|a_n - a| < \frac{\varepsilon}{2}\ \forall n \geq n_1 \text{ and } |b_n - b| < \frac{\varepsilon}{2}\ \forall n \geq n_2.$$

Then

$$\begin{array}{rcl} |a_n + b_n - (a+b)| & = & |a_n - a + b_n - b| \\ & \overset{\Delta-\text{ineq}}{\leq} & |a_n - a| + |b_n - b| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{array}$$

holds for all $n \geq n_0 = \max(n_1, n_2)$.

2. We will use the Δ-inequality in the form

$$\begin{array}{rcl} |a_n b_n - ab| & = & |a_n b_n - a_n b + a_n b - ab| \\ & \leq & |a_n b_n - a_n b| + |a_n b - ab| = |a_n| \cdot |b_n - b| + |a_n - a| \cdot |b| \end{array}$$

By assumption there exists for $\varepsilon = 1$ an n_1 such that

$$|a_n - a| < 1\ \forall n \geq n_1 \implies |a_n| \leq |a| + 1\ \forall \geq n_1.$$

Now let $\varepsilon > 0$ be given. Then by the convergence condition for (b_n) applied to $\frac{\varepsilon}{2(|a|+1)}$ there exists an n_2 such that

$$|b_n - b| < \frac{\varepsilon}{2(|a|+1)}\ \forall n \geq n_2.$$

Similarly there exists an n_3 such that

$$|a_n - a| < \frac{\varepsilon}{2(|b|+1)}\ \forall n \geq n_3.$$

Then for $n \geq n_0 = \max(n_1, n_2, n_3)$ we obtain

$$\begin{array}{rcl} |a_n b_n - ab| & \leq & |a_n| \cdot |b_n - b| + |a_n - a| \cdot |b| \\ & \leq & (|a|+1) \cdot |b_n - b| + (|b|+1) \cdot |a_n - a| \\ & < & (|a|+1) \cdot \frac{\varepsilon}{2(|a|+1)} + (|b|+1) \cdot \frac{\varepsilon}{2(|b|+1)} \\ & = & \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{array}$$

3. This time the essential estimate is

$$\begin{aligned}
|\frac{a_n}{b_n} - \frac{a}{b}| &= |\frac{a_n b - a b_n}{b_n b}| \\
&= \frac{|a_n b - ab + ab - ab_n|}{|b_n b|} \\
&\leq \frac{|a_n b - ab| + |ab - ab_n|}{|b_n| \cdot |b|} \\
&\leq \frac{|a_n - a| \cdot |b|}{|b_n| \cdot |b|} + \frac{|a| \cdot |b - b_n|}{|b_n| \cdot |b|}.
\end{aligned}$$

For $\varepsilon = \frac{|b|}{2} > 0 \; \exists n_1$, so that $|b_n - b| < \frac{|b|}{2} \; \forall n \geq n_1$, i.e.

$$|b_n| > |b| - \frac{|b|}{2} = \frac{|b|}{2} \; \forall n \geq n_1$$

since (b_n) converges to $b \neq 0$. Hence for $n \geq n_1$ we obtain

$$|\frac{a_n}{b_n} - \frac{a}{b}| \leq \frac{|a_n - a|}{|b_n|} + |a| \cdot \frac{|b_n - b|}{|b_n| \cdot |b|} \leq \frac{|a_n - a|}{|b|/2} + \frac{|a| \cdot |b_n - b|}{|b|^2/2}.$$

Let $\varepsilon > 0$ be given. Then $\exists n_2$ such that $|a_n - a| < \frac{\varepsilon \cdot |b|}{4} \; \forall n \geq n_2$, and $\exists n_3$ such that $|b_n - b| < \frac{1}{(|a|+1)} \cdot \frac{|b|^2}{4} \cdot \varepsilon \; \forall n \geq n_3$. Hence we obtain

$$\begin{aligned}
|\frac{a_n}{b_n} - \frac{a}{b}| &< \frac{\varepsilon \cdot |b|}{4} \cdot \frac{2}{|b|} + (|a| + 1) \cdot \frac{|b|^2}{4 \cdot (|a| + 1)} \cdot \varepsilon \cdot \frac{2}{|b|^2} \\
&= \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

for all $n \geq n_0 = \max(n_1, n_2, n_3)$. □

5.3 Examples of sequences in Computer science

Let A be an algorithm which can have input of variable length n.

Example 5.9 (Addition of of integers)**.** Let a_n be the maximal run time for the addition for two n-digit numbers.

Given two integers

$$d = \sum_{i=0}^{n-1} d_i 10^i, \quad e = \sum_{i=0}^{n-1} e_i 10^i$$

with digits d_i and e_i we add them with the scheme

$$\begin{array}{cccccc}
 & d_{n-1} & \cdots & d_2 & d_1 & d_0 \\
 & e_{n-1} & \cdots & e_2 & e_1 & e_0 \\
c_n & c_{n-1} & \cdots & c_2 & c_1 & \\
\hline
f_n & f_{n-1} & \cdots & f_2 & f_1 & f_0.
\end{array}$$

For the addition of two (respectively three) one digit numbers one can use a look-up table. If the look-up takes t machine clock cycles and each cycle takes s seconds, then the run time of the algorithm is

$$a_n = t \cdot s \cdot n.$$

Definition 5.10 (Landau symbols)**.** Let (a_n) be a sequence of positive numbers and (b_n) a further sequence. We say

$$(b_n) \in O(a_n), \ (b_n) \text{ growth at most as fast as } (a_n),$$

if there exist a constant $c \in \mathbb{R}_{>0}$ and an integer $n_0 \in \mathbb{N}$ such that

$$|b_n| \leq c a_n \ \forall n \geq n_0.$$

We write

$$(b_n) \in o(a_n),$$

if $\lim_{n\to\infty} \frac{b_n}{a_n} = 0$.

Hence

$$o(a_n) = \{(b_n) \in \mathbb{R}^{\mathbb{N}} \mid \lim_{n\to\infty} \frac{b_n}{a_n} = 0\}.$$

There are further Landau symbols, whose definitions we do not give here.

Example. $t \cdot s \cdot n \in O(n)$. This statement that addition of two n digit numbers has run time in $O(n)$ is in many aspect better than the precise formula because it does not depend on the hardware or implementation details, or the question whether we use binary or decimal expansion.

Example 5.11. The multiplication scheme for two n digits numbers requires n^2 memory tasks. Hence the naive multiplication scheme has run time $O(n^2)$. It can be done faster:

Algorithm 5.12 (Karatsuba,1962)**.** Input: Two integers a, b with $n = 2^k$ binary digits.
Ouput: The product $a \cdot b$.

1. Write $a = a_0 + a_1 2^{k-1},\ b = b_0 + b_1 2^{k-1}$ where a_0, a_1, b_0, b_1 have only 2^{k-1} binary digits.

2. Call the algorithm recursively to compute

$$a_0 b_0, (a_0 + a_1)(b_0 + b_1), a_1 b_1.$$

3. Return

$$a_0 b_0 + [(a_0 + a_1)(b_0 + b_1) - a_0 b_1 - a_1 b_1] \cdot 2^{k-1} + a_1 b_1 2^k$$

The third step involves the addition of several integers with 2^{k-1} binary digits, hence the additions altogether are in $O(n) = O(2^k)$ since $\sum_{i=0}^{k-1} 2^i = 2^k - 1$. The crucial point is that we use in the second step only three instead of four multiplication. Hence we obtain an algorithm for the multiplication of two n digits numbers of cost

$$O(n^{log_2 3}) \subset O(n^{1.59}).$$

That is much better than $O(n^2)$. An even faster algorithm exists:

Theorem 5.13 (Schönhage-Strassen, 1981)**.** *Two n digit numbers can be multiplied with cost*

$$O(n \log n \log\log n).$$

5.4 The completeness axiom

We will formulate the completeness axiom fo $\mathbb{R}$. Roughly speaking it says, every sequence which looks like an convergent sequence does converge.

Definition 5.14. A sequence (a_n) of real numbers is **bounded from above**, **bounded from below** or **bounded** if there exists a bound $M \in \mathbb{R}$ such that

$$a_n \leq M \ \forall n,$$

$a_n \geq M \; \forall n$, or

$|a_n| \leq M \; \forall n$ respectively.

The sequence (a_n) is called **increasing** or **decreasing** if $a_{n+1} \geq a_n \; \forall n$ or $a_{n+1} \leq a_n \; \forall n$ respectively. (a_n) is called **monotone** if it is increasing or decreasing. We speak of **strictly increasing** or **strictly decreasing** sequences if the inequalities are strict.

If (a_n) is a sequence and (n_k) a strictly increasing sequence of integers, then we call $(a_{n_k})_{k \in \mathbb{N}}$ a **subsequence** of (a_n).

Remark 5.15. Convergent sequences are bounded.

Proof. Suppose $\lim a_n = a$. Then for $\varepsilon = 1$ there exists an n_0 such that $|a_n - a| < 1 \; \forall n \geq n_0$. Hence

$$|a_n| \leq C = \max(|a| + 1, |a_1|, |a_2|, \ldots, |a_{n_0 - 1}|) \; \forall n.$$

□

Our first version of the completeness axiom is

Theorem 5.16 (Completness axiom, first version)**.** *Every bounded monotone sequence (a_n) of real numbers converges.*

Sketch of the proof. We take $\mathbb{R} = \{\text{decimal numbers}\}$ as a definition and show that every decreasing sequence of positive numbers converges by describing the decimal expansion of the limit. Let Let $k = \lfloor a_1 \rfloor$ be the integral part of a_1. We divided the interval

$$[0, k+1[= [0, 1[\cup [1, 2[\cup \ldots \cup [k, k+1[$$

into disjoint subintervals. Since the sequence is decreasing and bounded from below by 0, there is precisely one integer i such that $[i, i+1[$ contains all but finitely many elements of the sequence. i will give us the integral part of the limit. Next we divide

$$[i, i+1[= [i, i + \frac{1}{10}[\cup \ldots \cup [i + \frac{8}{10}, i + \frac{9}{10}[\cup [i + \frac{9}{10}, i + \frac{10}{10}[$$

into ten subintervals. There exists precisely one $i_1 \in \{0, \ldots, 9\}$ such that all but finitely many elements of the sequence are contained in the interval $[i + \frac{i_1}{10}, i + \frac{i_1 + 1}{10})[$. Dividing this interval again in ten disjoint equal length subintervals will give the second decimal i_2 place after the decimal point. Continuing this way we obtain decimal expansion $i.i_1 i_2 i_3 \ldots$ of the limit. □

Our second version uses the concept of Cauchy-sequences.

Definition 5.17. A sequence (a_n) of real numbers is a **Cauchy sequence** if

$$\forall \varepsilon > 0 \; \exists n_0 \in \mathbb{N} \text{ such that } |a_n - a_m| < \varepsilon \; \forall n, m \geq n_0.$$

Remark 5.18.

1. A convergent sequence is a Cauchy sequence.
2. Cauchy sequences are bounded.

Proof. 1. Suppose $\lim a_n = a$. Then $\exists n_0$ such that $|a_n - a| < \varepsilon/2 \; \forall n \geq n_0$. Hence

$$\begin{aligned} |a_n - a_m| &= |a_n - a + a - a_m| \\ &\overset{\triangle-\text{ineq}}{\leq} |a_n - a| + |a - a_m| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \quad \forall n, m \geq n_0 \end{aligned}$$

2. Let (a_n) be a Cauchy sequence. For $\varepsilon = 1$ there exists an $n_0 \in \mathbb{N}$ with $|a_n - a_m| < \varepsilon \; \forall n, m \geq n_0$. Then

$$|a_n| \leq \max(|a_{n_0}| + 1, |a_1|, \ldots, |a_{n_0 - 1}|) \; \forall n \geq n_0.$$

□

Our second version of the completeness axiom says that the converse is true:

Theorem 5.19 (Completeness axiom, second version; Cauchy criterion)**.** *Every Cauchy sequence of real numbers converges.*

Proof of the logical equivalence of the two versions. For the proof that the Theorem 5.16 on bounded monotone sequences implies the Cauchy criterion Theorem 5.19 we need a Lemma.

Lemma 5.20. *Every sequence contains a monotone subsequence.*

Proof. Let (a_n) be a sequence of real numbers. An element a_n is called a "scenic point", if $a_n > a_m \forall m > n$.

There are two possibilities. Either (a_n) has infinitely may scenic points or (a_n) has only finitely many scenic points. In the first case the subsequence of scenic points is strictly decreasing. In the second case, let a_{n_0} be the last scenic point. Then for each $n > n_0$ we can find an integer $m > n$ such that $a_m \geq a_n$. In this case we start with any $n_1 > n_0$ and choose recursively $n_{k+1} > n_k$ such that $a_{n_k} \leq a_{n_{k+1}}$. Then $(a_{n_k})_{k\in\mathbb{N}}$ is an increasing subsequence. □

Let (a_n) be a Cauchy sequence and let (a_{n_k}) be a monotone subsequence. Both sequences are bounded. By Theorem 5.16 there exists a limit

$$\lim_{k\to\infty} a_{n_k} = a.$$

Using the Δ-inequality we see that $\lim_{n\to\infty} a_n = a$ holds as well: For given $\varepsilon > 0\, \exists n_1, k_1 \in \mathbb{N}$ such that

$$|a_n - a_m| < \frac{\varepsilon}{2}\ \forall n, m \geq n_1 \text{ and } |a_{n_k} - a| < \frac{\varepsilon}{2}\ \forall k \geq k_1.$$

Taking $n_0 = \max(n_1, n_{k_1})$ we find

$$|a_n - a| \leq |a_n - a_{n_{k_1}}| + |a_{n_{k_1}} - a| < \varepsilon.$$

For the other direction we have to prove that bounded monotone sequences are Cauchy sequences. Let $|a_n| \leq M\ \forall n$. For given $\varepsilon > 0$ we choose an integer $N > \lceil\frac{1}{\varepsilon}\rceil$. We divided the interval $[-M, M]$ into $2N$ subintervals

$$[-M, M] = \bigcup_{k=-N}^{N-1} [\frac{k}{N}, \frac{k+1}{N}].$$

Then by the monotonicity there exists at least one subinterval $[\frac{k}{N}, \frac{k+1}{N}]$ containing all but finitely many a_n. (Actually there is only one unless the sequence is eventually constant.) Let n_0 be the index of the last a_n not contained in the this interval. Then

$$|a_n - a_m| < \varepsilon\ \forall n > n_0.$$

This completes the proof of the equivalence.

□

A variation of the idea above leads to a proof of the following very useful theorem.

Theorem 5.21 (Bolzano-Weierstrass)**.** *Every bounded sequence (a_n) has a convergent subsequence (a_{n_k}).*

Proof. Suppose $|a_n| \leq M \; \forall M$. We start with $n_1 = 1, N_1 = -M$ and $M_1 = M$ and use recursion. Suppose we defined $n_k, N_k < M_k$ such that a_{n_k} and infinitely many terms of the sequence (a_n) lie in $[N_k, M_k]$. Then at least one of the subintervals of

$$[N_k, M - k] = [N_k, \frac{N_k + M_k}{2}] \cup [\frac{N_k + M_k}{2}, M_k]$$

contains infinitely many terms of (a_n). We choose

$$N_{k+1} = \begin{cases} \frac{N_k+M_k}{2} & \text{if } [\frac{N_k+M_k}{2}, M_k] \text{ contains infinitely } a_n \\ N_k & \text{otherwise} \end{cases},$$

$$M_{k+1} = \begin{cases} M_k & \text{if } [\frac{N_k+M_k}{2}, M_k] \text{ contains infinitely } a_n \\ \frac{N_k+M_k}{2} & \text{otherwise} \end{cases},$$

and $n_{k+1} > n_k$ such that $a_{n_{k+1}} \in [N_{k+1}, M_{k+1}]$. Then $(N_k), (M_k)$ are increasing respectively decreasing sequences with

$$-M \leq N_k \leq a_{n_k} \leq M_k \leq M \; \forall k.$$

Thus $(N_k), (M_k)$ converge and since $\lim(M_k - N_k) = \lim \frac{M}{2^{n-1}} = 0$ the subsequence (a_{n_k}) converges as well to common limit of the sequences (N_k) and (M_k). □

Definition 5.22. An **upper bound** of a subset $M \subset \mathbb{R}$ is a real number b such that $a \leq b \; \forall a \in M$. A subset $M \subset \mathbb{R}$ is bounded from above if an upper bound of M exists. The **supremum**

$$\sup M$$

is a smallest upper bound b', i.e., an upper bound b' such that $b' \leq b$ for all other upper bounds of M.

Theorem 5.23 (Existence of the supremum)**.** *Every nonempty subset $M \subset \mathbb{R}$ which is bounded from above, has a supremum.*

Proof. We start with an $a_1 \in M$ and an upper bound b_1 of M and use recursion. Given $a_k \in M$ and an upper bound b_k of M we consider $m = \frac{a_k+b_k}{2}$ and define

$$b_{k+1} = \begin{cases} m & \text{if } m \text{ is an upper bound of } M \\ b_k & \text{otherwise} \end{cases}.$$

If m is not an upper bound of M, then we choose $a_{k+1} \in M$ with $a_{k+1} > m$. Otherwise we take $a_{k+1} = a_k$.

So (a_k) and (b_k) are monotone bounded sequences with the same limit b'. We claim that b' is the supremum. Indeed, b' is an upper bound because, if there exists an $a \in M$ with $a > b'$, then there would exist a b_k with $a > b_k$ since $\lim b_k = b'$. But this contradicts the fact that b_k is an upper bound. There no smaller upper bounds b because, if $b < b'$, then there exist an a_k with $b < a_k$ since $\lim a_k = b'$. □

Remark 5.24.

1. Theorem 5.23 is another property of $\mathbb{R}$ which is equivalent to the completeness axiom. In some courses this is used for the completeness axiom.

2. We sometimes write $\sup M = +\infty$ for subsets which have no upper bound and $\sup \emptyset = -\infty$ for the empty set.

3. The notion bounded from below and the **infimum** $\inf M$ for the largest lower bound are defined similarly. Nearly the same argument as above proves the existence of an infimum for non-empty subset which are bounded from below. We set $\inf M = -\infty$ if M has no lower bound, and $\inf \emptyset = +\infty$.

5.5 Square roots

Further axioms are not needed to characterise $\mathbb{R}$. Let us prove the existence of square roots of real positive numbers using our axioms.

Theorem 5.25. *Let $b \in \mathbb{R}_{>0}$. Let $a_0 \in \mathbb{R}_{>0}$ any starting value and consider the recursively defined sequence*

$$a_{n+1} = \frac{1}{2}(a_n + \frac{b}{a_n}).$$

Then (a_n) is well-defined and converges to a limit $a \in \mathbb{R}_{>0}$ satisfying $a^2 = b$. We write $a = \sqrt{b}$.

Proof. We argue in several steps.

1. The recursion formula makes only sense if $a_n \neq 0$. We prove $a_n > 0$ for all n by induction. $a_0 > 0$ holds by assumption. For the induction step we note $a_n > 0 \implies \frac{b}{a_n} > 0 \implies (a_n + \frac{b}{a_n}) > 0 \implies a_{n+1} = \frac{1}{2}(a_n + \frac{b}{a_n}) > 0$

2. We establish $a_n^2 \geq b \; \forall n \geq 1$. Indeed,

$$\begin{aligned} a_n^2 - b &= \frac{1}{4}(a_{n-1} + \frac{b}{a_{n-1}})^2 - b \\ &= \frac{1}{4}(a_{n-1}^2 + 2b + \frac{b^2}{a_{n-1}^2}) - b \\ &= \frac{1}{4}(a_{n-1}^2 - 2b + \frac{b^2}{a_{n-1}^2}) \\ &= \frac{1}{4}(a_{n-1} - \frac{b}{a_{n-1}})^2 \\ &\geq 0. \end{aligned}$$

3. We show $a_{n+1} \leq a_n \; \forall n \geq 1$. Indeed,

$$\begin{aligned} a_n - a_{n+1} &= a_n - \frac{1}{2}(a_n + \frac{b}{a_n}) \\ &= \frac{1}{2}(a_n - \frac{b}{a_n}) = \frac{1}{2a_n}(a_n^2 - b) \\ &\geq 0 \end{aligned}$$

   by the previous steps.

4. Hence $(a_n)_{n\geq 1}$ is a decreasing sequence of positive numbers which by the first version of the completeness axiom has a limit $a = \lim_{n\to\infty} a_n \in \mathbb{R}$.

5. We prove $a^2 = b$. With (a_n) also (a_{n+1}) converges with the same limit. Hence

$$a^2 = \lim_{n\to\infty}(a_n \cdot a_{n+1}) = \lim_{n\to\infty} \frac{1}{2}(a_n^2 + b) = \frac{1}{2}(a^2 + b)$$

$$\Longrightarrow \frac{1}{2}a^2 = \frac{1}{2}b \Longrightarrow a^2 = b.$$

□

Example 5.26. We apply the algorithm of Theorem 5.25 to compute square roots approximately.

1. $b = 4$, $a_0 = 1$. The correct value is $\sqrt{b} = 2$. The algorithm gives

n	a_n	$\frac{b}{a_n}$
0	1	4
1	2.5	1.6
2	2.05	$1.\overline{95121}$
3	$2.0006097\ldots$	

2. For $b = 2$ und $a_0 = 1$ gives $\sqrt{2}$:

n	a_n	$\frac{b}{a_n}$
0	1	2
1	1.5	$1.3333\ldots$
2	$1.41666\ldots$	$1.411764\ldots$
3	$1.414215\ldots$	$1.414211\ldots$
4	$1.41421356237\ldots$	

Already the fourth value has 12 correct digits!

Remark 5.27. Let $b > 0$. The sequence

$$a_{n+1} = \frac{1}{2}\big(a_n + \frac{b}{a_n}\big)$$

converges to $a = \sqrt{b}$ remarkable fast. If we define the **relative error** f_n of a_n by the formula

$$a_n = a \cdot (1 + f_n),$$

then $f_n \geq 0$ for $n \geq 1$. Substituting into $a_{n+1} = \frac{1}{2}(a_n + \frac{b}{a_n})$ yields

$$a(1 + f_{n+1}) = \frac{1}{2}(a(1 + f_n) + \frac{a^2}{a(1 + f_n)},$$

hence

$$1 + f_{n+1} = \frac{1}{2}\big((1 + f_n) + \frac{1}{1 + f_n}\big) = \frac{1}{2} \cdot \frac{2 + 2f_n + f_n^2}{1 + f_n}.$$

We conclude

$$f_{n+1} = \frac{1}{2} \cdot \frac{f_n^2}{1 + f_n} \leq \frac{1}{2} \cdot \min(f_n, f_n^2).$$

If the relative error $f_n \geq 1$, then it will be halved in the next step. Once we reach $f_n < 1$ then $f_{n+1} = \frac{1}{2} \cdot f_n^2$. In this case, the number of correct decimal digits will double with each iteration step. One speaks of **quadratic convergence**.

5.6 The existence of real numbers

After discussing the axioms it might (and should) become questionable whether a triple $(\mathbb{R}, +, \cdot)$ satisfying the axioms K1-K3,A1-A4 and the completeness axiom does exist.

Certainly defining $\mathbb{R} = \{\text{decimal numbers}\}$ is not convincing since the definition of an infinite decimal number already uses the concepts of convergence. Certainly it will not explain properly why

$$0.999\ldots = 0.\overline{9} = 1$$

holds. Also aliens might find it rather bizarre that we use ten digits.

We will give two constructions starting from the rational numbers. But first let us establish that these axioms characterise $\mathbb{R}$ completely.

Theorem 5.28. *Let R and R' be two archimedean ordered fields satisfying the completeness axiom. Then there exists a unique bijection*

$$\varphi : R \to R'$$

compatible with all structures, for example $\varphi(a+b) = \varphi(a) + \varphi(b), \varphi(a \cdot b) = \varphi(a) \cdot \varphi(b)$ and $a > b \Rightarrow \varphi(a) > \varphi(b)$.

Proof. Since R and R' are ordered, we already know that we can regard $\mathbb{Q}$ as a subset both of R and R'. We will construct $\varphi : R \to R'$ by extending the identity map $\mathrm{id}_{\mathbb{Q}}$.

$$\begin{tikzcd}
R \arrow[r, "\exists\varphi"] & R' \\
\mathbb{Q} \arrow[u, hook] \arrow[r, "\mathrm{id}_{\mathbb{Q}}"] & \mathbb{Q} \arrow[u, hook]
\end{tikzcd}$$

Given $a \in R$ we choose a sequence (a_n) of rational numbers converging to a. Such a sequence exists because each interval $]-\frac{1}{n}+a, a+\frac{1}{n}[\subset R$ contains a rational number a_n of the form $\frac{m}{n}$. Then $(\varphi(a_n))$ is a Cauchy sequence R' since $\varphi|_{\mathbb{Q}} = \mathrm{id}_{\mathbb{Q}}$, and by the archimedean axiom it suffices to consider all ε of the form $\varepsilon_N = \frac{1}{N}$ in the definition of the Cauchy sequence. We define

$$\varphi(a) = \lim_{n\to\infty} \varphi(a_n).$$

Then $\varphi : R \to R'$ is well-defined: It is independent of the choice of the sequence (a_n) of rational numbers converging to a because the difference of two such sequences will be a sequence which converges to 0.

φ is bijective since we can define a map $\psi : R' \to R$ by the same method and $\psi \circ \varphi = \mathrm{id}_R$ since the restrictions to $\mathbb{Q}$ are all $\mathrm{id}|_{\mathbb{Q}}$. Then $\varphi(a+b) = \varphi(a) + \varphi(b), \varphi(a \cdot b) = \varphi(a) \cdot \varphi(b)$ follows from calculation rules for convergent sequences. Finally, R is mapped to positve elements of R', since we can recover the set of positive elements

$$\{x \in R \mid x > 0\} = \{x \in R \mid \exists y \in R, y \neq 0 \text{ with } x = y^2\}$$

as the set of non-zero squares by Theorem 5.25. □

By the theorem above it does not matter how we construct $\mathbb{R}$. We sketch two constructions as $\{$Cauchy sequences in $\mathbb{Q}\}/\{$zero sequences$\}$ or as Dedekind cuts.

Construction 5.29 ($\mathbb{R}$ as Cauchy sequences modulo zero sequences)**.** A **zero sequence** is a sequence which converges to 0. Consider

$$M = \{(a_n) \in \mathbb{Q}^{\mathbb{N}} \mid (a_n) \text{ is a Cauchy sequence}\}.$$

Elementwise addition and multiplication gives M the structure of a commutative ring with 1, where 1 corresponds to the constant sequence (1). We define on M an equivalence relation by

$$(a_n) \sim (b_n) :\iff (a_n - b_n) \text{ is a zero sequence},$$

and define

$$\mathbb{R} := M/\sim$$

as a set. Then

$$[(a_n)] + [(b_n)] := [(a_n + b_n)]$$

is well-defined since the sum of two zero-sequences is a zero sequence, and

$$[(a_n)] \cdot [(b_n)] := [(a_n \cdot b_n)]$$

is well-defined because the product of a bounded sequence with a zero sequence is a zero sequence.

To verify that this gives $M/\sim$ the structure of a field, we have to prove the existence of inverse elements: If a Cauchy sequence (a_n) is not a zero sequence, then only finitely many a_n are zero. Take n_1 to be the smallest integer such that $a_n \neq 0$ for all $n \geq n_1$. Then the sequence (b_n) with

$$b_n = \begin{cases} 1 & \text{if } n < n_1 \\ 1/a_n & \text{if } n \geq n_1 \end{cases}$$

is Cauchy sequence which represents a class with $[(b_n)] = [(a_n)]^{-1}$.

Finally, we verify the completeness axiom: If $(x_k)_{k\in\mathbb{N}} = ([(a_{kn})_{n\in\mathbb{N}}])_{k\in\mathbb{N}}$ is a Cauchy sequence in $M/\sim$, then the diagonal sequence

$$(a_{kk})$$

is a Cauchy sequence which represents $\lim_{k\to\infty} x_k \in M/\sim$.

With a little bit of experience you will be able to fill in all missing details of this construction. □

The second approach has as its basic idea that a real number a is uniquely determined by the set

$$U = \{x \in \mathbb{Q} \mid x < a\}.$$

Definition 5.30. A **Dedekind cut** (U, V) is a pair of non-empty subsets of $\mathbb{Q}$ satisfying

$$U \cup V = \mathbb{Q} \text{ and } u < v \; \forall u \in U \; \forall v \in V.$$

If $r \in \mathbb{Q}$, then $(U(r), V(r)) = (\{u \in \mathbb{Q} \mid u < r\}, \{v \in \mathbb{Q} \mid v \geq r\}$ is called a **well-chosen rational** Dedekind cut, and $(U'(r), V'(r)) = (\{u \in \mathbb{Q} \mid u \leq r\}, \{v \in \mathbb{Q} \mid v > r\}$ is a badly chosen rational Dedekind cut. All other Dedekind cuts are called **irrational**. A Dedekind cut is called **good** if it is a well-chosen rational or an irrational Dedekind cut.

Construction 5.31 ($\mathbb{R}$ **as a set of Dedekind cuts).** We define $\mathbb{R}$ as a set

$$\mathbb{R} := \{(U, V) \in 2^{\mathbb{Q}} \times 2^{\mathbb{Q}} \mid (U, V) \text{ is a good Dedekind cut}\}.$$

To define the field structure we take as addition

$$(U_1, V_1) + (U_2, V_2) = (U_3, V_3)$$

by $U_3 = \{u_1 + u_2 \mid u_1 \in U_1, u_2 \in U_2\}$ and $V_3 = \mathbb{Q} \setminus U_3$. It is not difficult but a bit tedious to define all the structures. For example, the negative of (U, V) is not always $(-V, -U) = (\{-v \mid v \in V\}, \{-u \mid u \in U\})$ since this might be a badly chosen rational cut. For details see E. Landau, xxx

The completeness axiom is best verified for bounded increasing sequences. If $((U_n, V_n))_{n\in N}$ is an increasing sequence bounded from above by (U_M, V_M), i.e.,

$$U_n \subset U_{n+1} \subset U_M \; \forall n$$

then (U, V) with

$$U = \bigcup_{n\geq 1} U_n \subset U_M$$

and $V = \mathbb{Q} \setminus U$ is the limit. □

5.7 Complex numbers

We complete the construction of number systems by constructing the field of complex numbers $\mathbb{C}$ from $\mathbb{R}$. This is by far the easiest step in the chain

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{R} \subset \mathbb{C}$$

of constructions.

Definition 5.32. As a set we define

$$\mathbb{C} = \mathbb{R} \times \mathbb{R}.$$

Addition and multiplication on $\mathbb{C}$ are defined by

$$(a_1, b_1) + (a_2, b_2) := (a_1 + a_2, b_1 + b_2)$$

and

$$(a_1, b_1) \cdot (a_2, b_2) = (a_1 a_2 - b_1 b_2, a_1 b_2 + b_1 a_2)$$

The associativity, commutativity and distributivity laws are verified in a straight forward way using these laws in $\mathbb{R}$.

The element $0 = (0, 0) \in \mathbb{C}$ is the zero element, and $1 = (1, 0) \in \mathbb{C}$ is the one element. The map

$$\mathbb{R} \hookrightarrow \mathbb{C}, a \mapsto (a, 0)$$

is an embedding which respects addition and multiplication. The imaginary unit $i = (0, 1) \in \mathbb{C}$ is an element with $i^2 = -1 = (-1, 0) \in \mathbb{C}$. We usually write $c = a + ib$ instead of (a, b). To compute $(a_1 + ib_1)(a_2 + ib_2) = (a_1 a_2 - b_1 b_2) + i(a_1 b_2 + a_2 b_1)$ one only uses the usual laws and memorise $i^2 = -1$.

The inverse of $c = a - ib$ is given by the formula

$$c^{-1} = \frac{a - ib}{a^2 + b^2}.$$

Hence $(\mathbb{C}, +, \cdot)$ is a field. $\mathbb{C}$ cannot be ordered because $i \neq 0$ would imply $-1 = i^2 > 0$, a contradiction.

For $c = a + ib$ we call $\mathrm{Re}\ c = a, \mathrm{Im}\ c = b$ the real and imaginary part of c. The distance of c from 0

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

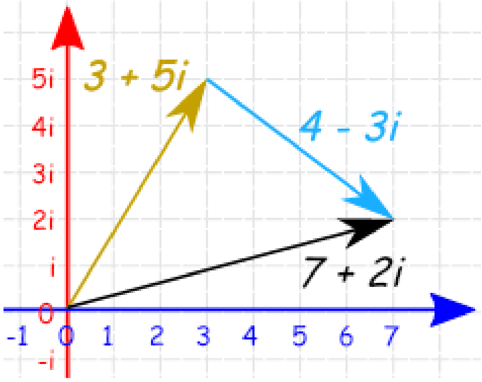
is called the absolute value of c. The number

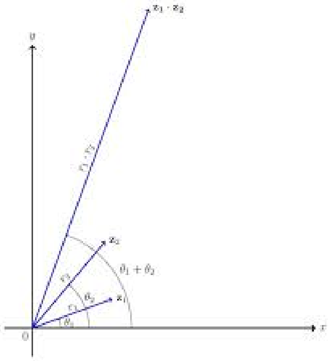
$$\bar{c} = a - ib$$

is called the complex conjugate of c. The formula for the inverse can be also memorised by

$$c^{-1} = \frac{1}{c} = \frac{\bar{c}}{c\bar{c}} = \frac{\bar{c}}{|c|^2}.$$

Geometric interpretation of addition and multiplication. Addition is simply vector addition.





Multiplication is best understood in polar coordinates: If we write

$$c = r(\cos\alpha + i\sin\alpha),$$

then $r = |c|$, and α is called the **argument** of c

$$c_1c_2 \quad = \quad r_1(\cos\alpha_1 + i\sin\alpha_1) \cdot r_2(\cos\alpha_2 + i\sin\alpha_2)$$

$$= \quad r_1 r_2(\cos(\alpha_1 + \alpha_2) + i \sin(\alpha_1 + \alpha_2))$$

holds by the addition laws 7.11 of $\sin$ and $\cos$. Hence in multiplication the absolute values multiply and the arguments are added.

For $c \in \mathbb{C}^* = \mathbb{C} \setminus 0$ and $d \in \mathbb{C}$ the map

$$\mathbb{C} \to \mathbb{C}, z \mapsto cz + d$$

is a rotation by the argument α of c combined with a stretching by the factor $|c|$ followed by a translation by d.

The absolute value satisfies

1. $|z| \geq 0$, and $|z| = 0$ if and only if $z = 0$,
2. $|z \cdot w| = |z| \cdot |w| \; \forall z, w \in \mathbb{C}$,
3. (Δ-inequality) $|z + w| \leq |z| + |w| \; \forall z, w \in \mathbb{C}$.

The name triangle inequality is clear now: The third side of a triangle in $\mathbb{C}$ is at most as long as the sum of the length of the two other sides.

By definition, a sequence (z_n) of complex numbers converges to $z \in \mathbb{C}$, if

$$\forall \varepsilon > 0 \; \exists n_0 \text{ such that } |z_n - z| < \varepsilon \; \forall n \geq n_0,$$

equivalently if the sequences of real and imaginary parts $(\mathrm{Re}\; z_n)$ and $(\mathrm{Im}\; z_n)$ converge to $\mathrm{Re}\; z$ and $\mathrm{Im}\; z$ respectively.

Complex numbers first showed up in the work of Cardano (1501 –1576), who used them to give a formula for the three roots of a cubic polynomial with real coefficients. The name imaginary is used since first i was an imagined solution of the equation $x^2 + 1 = 0$. There is nothing mysterious about imaginary numbers, except that they do not lie on the real number line, but in the complex number plane.

Complex numbers play an important role in mathematics because of the following theorem.

Theorem 5.33 (Fundamental theorem of algebra)**.** *Let $p(z) = c_n z^n + c_{n-1} z^{n-1} + \ldots + c_1 z + c_0$ be a polynomial of degree n with complex coefficients c_k. Then p has a complex root, equivalently, there exists $z_1, \ldots, z_n \in \mathbb{C}$ such that*

$$p = c_n \prod_{k=1}^{n} (z - z_k)$$

factors completely in linear factors.

Thus adding the imaginary root $i = \sqrt{-1}$ to our number system, all polynomials have roots. We do not give a proof of this theorem in this course.

For polynomials $p(x)$ with real coefficients, we have

$$p(w) = 0 \implies p(\overline{w}) = 0.$$

Thus the roots of of real polynomials consist of a certain number n_1 of real roots (counted with multiplicity) and n_2 pairs of complex conjugate roots (counted with multiplicity), where

$$n_1 + 2n_2 = \deg p(x).$$

add up to the degree of $p(x)$. Since

$$q(x) = (x - w)(x - \overline{w}) = x^2 - 2\text{Re } w + |w|^2$$

real polynomials factor in linear forms and quadric factors over $\mathbb{R}$.

In Cardano's formula, complex numbers are needed to describe the roots of a cubic real polynomial with three real roots. In case there is only one real root, complex numbers can be avoided to give a formula for the single real root.

In physics complex numbers are used in quantum mechanics. In the course complex numbers will show up in the description of the range of convergence of power series. They are also crucial for the Fourier transform, which is used in sound and image compression in Computer science.

5.8 Countable sets

Definition 5.34. A set M is called countable if there exists a surjective map

$$\varphi : \mathbb{N} \to M.$$

We call M uncountable if M is not countable.

Example 5.35. 1. Every finite set is countable.

2. $\mathbb{Z}$ is countable: $\mathbb{Z} = \{0, 1, -1, 2, -2, \dots\}$, more precisely

$$\varphi \colon \mathbb{N} \to \mathbb{Z},\ \varphi(n) = \begin{cases} 0, & n = 1, \\ \frac{1}{2}n, & n \text{ even}, \\ -\frac{1}{2}(n-1), & n \text{ odd } \geq 3. \end{cases}$$

3. $\mathbb{N} \times \mathbb{N}$ is countable. We define $\varphi \colon \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ by

$$\begin{tikzcd}
(1,1) \arrow[r] & (1,2) \arrow[dl] & (1,3) \arrow[dl] & (1,4) \arrow[dl] & \dots \\
(2,1) \arrow[urr] & (2,2) \arrow[dl] & (2,3) \arrow[dl] & \dots & \\
(3,1) \arrow[uurrr] & (3,2) \arrow[dl] & \dots & & \\
(4,1) & \dots & & &
\end{tikzcd}$$

Remark 5.36. If M is countably infinite, then there exists also a bijective map $\psi : \mathbb{N} \to M$.

Proof. Let $\varphi : \mathbb{N} \to M$ be surjective. We set $n_1 = 1$,

$$\psi(1) = \varphi(1),$$

and recursively, if $\psi(1), \dots, \psi(k)$ are already defined, we consider

$$n_{k+1} = \min\{n \in \mathbb{N} \mid \varphi(n) \notin \{\psi(1), \dots, \psi(k)\}\,\}$$

and set

$$\psi(k+1) = \varphi(n_{k+1}).$$

□

Theorem 5.37. *A countable union $M = \bigcup_{k=1}^{\infty} M_k$ of countable sets M_k is countable.*

Proof. Let

$$\psi : \mathbb{N} \to \mathbb{N} \times \mathbb{N}, n \mapsto \psi(n) = (\psi_1(n), \psi_2(n))$$

be an enumeration of $\mathbb{N} \times \mathbb{N}$ and let $\varphi_k : \mathbb{N} \to M_k$ be enumerations of M_k. Then $\Phi : \mathbb{N} \to M$ with

$$\Phi(n) = \varphi_{\psi_1(n)}(\psi_2(n))$$

is an enumeration of M. □

Corollary 5.38. $\mathbb{Q}$ *is countable.*

Proof.

$$\mathbb{Q} = \bigcup_{b=1}^{\infty} \{\frac{a}{b} \mid a \in \mathbb{Z}\}.$$

□

Theorem 5.39 (Cantor's second diagonal argument, 1877)**.** $\mathbb{R}$ *is uncountable.*

Proof. It is enough to prove that $[0,1[\subset \mathbb{R}$ is uncountable. Suppose

$$\mathbb{N} \to [0,1[, n \mapsto a_n$$

is an enumeration. Consider the decimal expansion of the a_n:

$$a_n = 0.a_{n1}a_{n2}\dots a_{nk}\dots$$

with $a_{nk} \in \{0,\dots,9\}$ the k-th decimal digit after the point. We consider the number $c = 0.c_1c_2\dots c_k\dots$ with digits

$$c_k = \begin{cases} 1 & \text{if } a_{nn} \neq 1 \\ 0 & \text{if } a_{nn} = 1 \end{cases}.$$

Then $c \neq a_n$, since $c_n \neq a_{nn}$. Hence

$$\mathbb{N} \to [0,1[, n \mapsto a_n$$

is not surjective, a contradiction. □

The set theory of Cantor extends counting from finite sets to infinite set.

Definition 5.40 (Cantor)**.** Two sets M and N have the same cardinality, in notation

$$\mathrm{card}(M) = \mathrm{card}(N),$$

if there exists a bijection $M \to N$. We say that N has at least the cardinality of M, $\mathrm{card}(M) \leq \mathrm{card}(N)$ if there exists an injective map $M \to N$.

Using the axiom of choice of set theory one can prove that $\mathrm{card}(M) \leq \mathrm{card}(N)$ and $\mathrm{card}(N) \leq \mathrm{card}(M)$ implies $\mathrm{card}(M) = \mathrm{card}(N)$.

Axiom of choice. *Let* $(M_i)_{i\in I}$ *be a family of non-empty sets* M_i*. Then there exists a map*

$$a : I \to \bigcup_{i\in I} M_i$$

such that $a(i) \in M_i$ *for all* $i \in I$. In other words, it is possible to select elements $a(i) \in M_i$ simultaneously for all $i \in I$.

One can prove that for any set M, the power set 2^M has always strictly larger cardinality than M.

6 Infinite series

Definition 6.1. Let $(a_k)_{k\in\mathbb{N}}$ a sequence of real (or complex) numbers. The sequence (s_n) of partial sums

$$s_n = \sum_{k=1}^{n} a_n$$

is called a series, which we denote by

$$\sum_{k=1}^{\infty} a_k.$$

If the sequence (s_n) converges, then

$$\sum_{k=1}^{\infty} a_k = \lim_{n\to\infty} s_n$$

also denotes the limit. Thus $\sum_{k=1}^{\infty} a_k$ is an overloaded notation which can mean two things

1. the sequence of partial sums or

2. the limit $\lim_{k\to\infty} s_n$.

Example 6.2. 1. $\sum_{k=1}^{\infty} \frac{1}{k}$

2. $\sum_{k=1}^{\infty} \frac{1}{2^k}$

3. (Decimal expansion): $d_k \in \{0, \dots, 9\}$ digits and

$$\sum_{k=1}^{\infty} d_k \cdot 10^{-k}.$$

We will see that series of this kind always converge.

4. Let (c_n) be a sequence. Then $a_1 = c_1$ and $a_k = c_k - c_{k-1}$ for $k \geq 2$ defines a series

$$\sum_{k=1}^{n} a_k$$

whose partial sums are $s_n = c_n$. Hence infinite sums is not really a new concept. Sometimes this can be used to compute the limit.

Example 6.3. (Telescope series) We show

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)} = 1.$$

Indeed,

$$\begin{aligned} s_n = \sum_{k=1}^{n} \frac{1}{k(k+1)} &= \sum_{k=1}^{n} (\frac{1}{k} - \frac{1}{k+1}) \\ &= 1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \ldots + \frac{1}{n} - \frac{1}{n+1} \\ &= 1 - \frac{1}{n+1}. \end{aligned}$$

Hence $\lim_{n\to\infty} s_n = 1$, as claimed.

Theorem 6.4 (Cauchy criterion for series)**.** *A series $\sum_{k=1}^{\infty} a_n$ converges if and only if*

$$\forall \varepsilon > 0 \; \exists n_0 \textit{ such that } |\sum_{k=n}^{m} a_k| \le \varepsilon \; \forall m, n \ge n_0.$$

In particular, the summands (a_k) of a convergent series $\sum_{k=1}^{\infty} a_k$ form a zero sequence. □

6.1 Convergence criteria for infinite series

To prove the convergence of a series is often easy and possible without computing the limit.

Definition 6.5. An **alternating series** is a series of the form

$$\sum_{k=0}^{\infty} (-1)^k a_k$$

where all $a_k \ge 0$.

Theorem 6.6. *If (a_k) is a monotone decreasing zero sequence, then*

$$\sum_{k=0}^{\infty} (-1)^k a_k$$

converges.

Proof. We consider the subsequence (s_{2n}) and (s_{2n+1}) of even and odd partial sums. Then

$$\begin{aligned} s_{2n+2} &= s_{2n} - a_{2n+1} + a_{2n+2} \leq s_{2n} \quad \text{because } a_{2n+1} \geq a_{2n+2}, \\ s_{2n+1} &= s_{2n-1} - a_{2n} + a_{2n+1} \geq s_{2n-1} \quad \text{and} \end{aligned}$$

$$s_{2n+1} \leq s_{2n}$$

hold. Hence

$$s_0 \geq s_2 \geq \ldots \geq s_{2n} \geq \ldots \geq s_{2n+1} \geq \ldots \geq s_3 \geq s_1$$

holds. Hence (s_{2n}) and (s_{2n+1}) converges since they are bounded monotone sequences. Their limit coincides because

$$\lim_{n\to\infty} s_{2n} - \lim_{n\to\infty} s_{2n+1} = \lim_{n\to\infty}(s_{2n} - s_{2n+1}) = \lim_{n\to\infty} a_{2n+1} = 0.$$

□

Example 6.7. The series

$$\sum_{k=1}^{\infty}(-1)^{k+1}\frac{1}{k} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots$$

and

$$\sum_{k=0}^{\infty}(-1)^{k}\frac{1}{2k+1} = 1 - \frac{1}{3} + \frac{1}{5} - \ldots$$

converge. To compute the limits is much more difficult. In the end of the course we will be able to prove

$$\sum_{k=1}^{\infty}(-1)^{k+1}\frac{1}{k} = \ln 2$$

and

$$\sum_{k=0}^{\infty}(-1)^{k}\frac{1}{2k+1} = \frac{\pi}{4}.$$

Theorem 6.8. *Let $q \in \mathbb{R}$. The **geometric series** $\sum_{k=0}^{\infty} q^n$ converges if and only if $|q| < 1$. In this case*

$$\sum_{k=0}^{\infty} q^n = \frac{1}{1-q}.$$

Example 6.9.

$$\sum_{n=0}^{\infty} \frac{1}{2^n} = \sum_{n=0}^{\infty} (\frac{1}{2})^n = \frac{1}{1-\frac{1}{2}} = 2.$$

Proof. $|q| < 1$ is necessary for convergence since the terms of a convergent series form a zero sequence. To prove that this suffices, we first prove

$$(1-q)\sum_{k=0}^{n} q^k = 1 - q^{n+1}$$

by induction. Hence

$$s_n = \sum_{k=0}^{n} q^k = \frac{1-q^{n+1}}{1-q} \underset{n\to\infty}{\longrightarrow} \frac{1}{1-q}$$

if $|q| < 1$. □

Example 6.10. As known from high school

$$0.999\ldots = 0.\overline{9} = \sum_{n=1}^{\infty} 9 \cdot 10^{-n} = \frac{9}{10}\sum_{k=0}^{\infty} 10^{-k} = \frac{9}{10} \cdot \frac{1}{1-\frac{1}{10}} = 1.$$

Example 6.11. The series

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots$$

is called the harmonic series. It does not converge. Indeed,

$$1 + \frac{1}{2} + \underbrace{(\frac{1}{3} + \frac{1}{4})}_{\geq \frac{1}{4}+\frac{1}{4}=\frac{1}{2}} + \underbrace{(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8})}_{\geq \frac{4}{8}=\frac{1}{2}} + \underbrace{(\frac{1}{9} + \cdots + \frac{1}{16})}_{\geq \frac{8}{16}=\frac{1}{2}} + \cdots,$$

hence

$$s_{2^k} = \sum_{n=1}^{2^k} \frac{1}{n} \geq 1 + \underbrace{\frac{1}{2} + \cdots + \frac{1}{2}}_{k \text{ Sumands}} \geq 1 + \frac{k}{2} = \frac{k+2}{2}.$$

So the sequence of partial sums grows unbounded.

Remark 6.12. If (a_k) is a sequence of non-negative real numbers, then we write sometimes

$$\sum_{k=1}^{\infty} a_k < \infty$$

to indicate that this series converges. This is justified by our first version of the completeness axiom: Since $a_n \geq 0$, the sequence of partial sums (s_n) is monotonously increasing. Hence it is convergent if and only if it stays bounded.

We will work out the idea to compare a series with a simpler one next.

Definition 6.13. Let $\sum_{n=1}^{\infty} b_n$, $\sum_{n=1}^{\infty} a_n$ be two series. Then $\sum a_n$ **majorizes** $\sum b_n$ if

$$|b_n| \leq a_n \forall n.$$

Theorem 6.14 (Comparison theorem)**.** *Suppose $\sum a_n$ majorizes $\sum b_n$. Then the following holds*

1. *If $\sum_{n=1}^{\infty} a_n$ converges, then $\sum_{n=1}^{\infty} b_n$ converges.*
2. *If $\sum_{n=1}^{\infty} b_n$ diverges, then $\sum_{n=1}^{\infty} a_n$ diverges.*

Proof. We have

$$|\sum_{k=n}^{m} b_k| \leq \sum_{k=n}^{m} |b_k| \leq \sum_{k=n}^{m} a_n.$$

Hence if the Cauchy criterium is satisfied for $\sum a_n$, it is also satisfied for $\sum b_n$. This proves 1.) The second statement 2.) is logically equivalent to 1.) □

Example 6.15. The series $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges. To prove this, it suffices to prove that $\sum_{n=1}^{\infty} \frac{1}{(n+1)^2}$ converges. The telescope series $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$ is a convergent majorizing series. To compute the limit needs substantially more techniques. With the help of Fourier series one can prove

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

We deduce further convergence criteria from the comparison theorem.

Theorem 6.16 (Quotient criterion)**.** *Let $\sum_{n=0}^{\infty} a_n$ be a series with $a_n \neq 0$ $\forall n$. Suppose that there exists a q with $0 < q < 1$ such that*

$$|\frac{a_{n+1}}{a_n}| \leq q \; \forall n$$

Then $\sum a_n$ converges.

Proof. $\sum |a_0| q^n$ is a convergent majorizing series. □

It is enough to require the bound $|\frac{a_{n+1}}{a_n}| \leq q$ for all but finitely many n. This yields:

Corollary 6.17 (Quotient test)**.** *Let $\sum_{n=0}^{\infty} a_n$ be a series with $a_n \neq 0$ $\forall n$. Suppose that the limit*

$$q = \lim_{n\to\infty} |\frac{a_{n+1}}{a_n}|$$

exists.

1. *If $q < 1$, then $\sum_{n=1}^{\infty} a_n$ converges.*
2. *if $q > 1$, then $\sum_{n=1}^{\infty} a_n$ diverges.*
3. *if $q = 1$, then this test yields no information.*

Example 6.18.

$$\lim_{n\to\infty} \frac{\frac{1}{n+1}}{\frac{1}{n}} = \lim_{n\to\infty} \frac{n}{n+1} = 1 = \lim_{n\to\infty} \frac{n^2}{(n+1)^2},$$

$\sum_n \frac{1}{n}$ diverges, while $\sum_n \frac{1}{n^2}$ converges.

Theorem 6.19 (Root test)**.** *Let (a_n) be a sequence and suppose that the limit*

$$r = \lim_{n\to\infty} \sqrt[n]{|a_n|}$$

exists.

1. *If $r < 1$, then $\sum_{n=1}^{\infty} a_n$ converges.*
2. *if $r > 1$, then $\sum_{n=1}^{\infty} a_n$ diverges.*
3. *if $r = 1$, then this test yields no information.*

Proof. Suppose $r < 1$. Then there exists an n_0 such that $\sqrt[n]{|a_n|} \leq q = \frac{r+1}{2} < 1$ for all $n \geq n_0$. Hence $\sum_{n=n_0}^{\infty} q^n$ is a convergent majorizer of $\sum_{n=n_0}^{\infty} a_n$. If $r > 1$, then there exists an n_0 such that $\sqrt[n]{|a_n|} \geq q = \frac{r+1}{2} > 1$ for all $n \geq n_0$. Then $|a_n| \geq q^n$ is not a zero sequence. □

6.2 Rearrangement of series

Consider the alternating harmonic series

$$\sum_{k=1}^{\infty}(-1)^k\frac{1}{k} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \ldots.$$

By Leibniz' criterion this sum has a limit s satisfying

$$1 = s_1 \geq s \geq s_2 = \frac{1}{2}.$$

Now consider the following rearrangements of the summands

$$1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{6} - \frac{1}{8} + \frac{1}{5} - \frac{1}{10} - \frac{1}{12} + \cdots,$$

i.e., we consider

$$\sum_{k=1}^{\infty}(\frac{1}{2k-1} - \frac{1}{4k-2} - \frac{1}{4k}).$$

In this series each fraction $\frac{1}{n}$ occurs precisely once with the correct sign. Since $\frac{1}{2k-1} - \frac{1}{4k-2} = \frac{1}{4k-2}$, we obtain

$$\begin{aligned}\sum_{k=1}^{\infty}(\frac{1}{2k-1} - \frac{1}{4k-2} - \frac{1}{4k}) &= \sum_{k=1}^{\infty}(\frac{1}{4k-2} - \frac{1}{4k}) \\ &= \frac{1}{2}\sum_{k=1}^{\infty}(\frac{1}{2k-1} - \frac{1}{2k}) = \frac{1}{2}s \neq s,\end{aligned}$$

since $s \neq 0$. Thus in general an infinite sum has not the same limit if we rearrange its terms. Our next results, says that under an additional assumption, arbitrary rearrangements of a series have the same limit.

Definition 6.20. A series $\sum_{n=1}^{\infty} a_n$ of real or complex numbers a_n is **absolutely convergent** if the series

$$\sum_{n=1}^{\infty}|a_n|$$

of the absolute values converges.

The triangle inequality and the Cauchy criterion give:

$$\text{absolute convergence} \implies \text{convergence.}$$

Moreover in case of absolute convergence, we have a generalised triangle inequality

$$|\sum_{n=1}^{\infty} a_n| \leq \sum_{n=1}^{\infty} |a_n|$$

since for each finite sum $|\sum_{n=1}^{N} a_n| \leq \sum_{n=1}^{N} |a_n| \leq \sum_{n=1}^{\infty} |a_n| < \infty$ holds.

Theorem 6.21. *Let $\sum_{n=1}^{\infty} a_n$ be an absolutely convergent series. Then the following holds:*

1. *(Little rearrangement theorem.) Let $\sigma : \mathbb{N} \to \mathbb{N}$ be a bijection. Then also $\sum_{n=1}^{\infty} a_{\sigma(n)}$ is absolutely convergent and*

$$\sum_{n=1}^{\infty} a_{\sigma(n)} = \sum_{n=1}^{\infty} a_n.$$

2. *(Big rearrangement theorem.) Let $(I_k)_{k\in\mathbb{N}}$ be a family of finite or infinite disjoint subsets $I_k \subset \mathbb{N}$ such that $\bigcup_{k=1}^{\infty} I_k = \mathbb{N}$. Then the sums $s_k = \sum_{n\in I_k} a_n$ and $\sum_{k=1}^{\infty} s_k$ are absolutely convergent, and*

$$\sum_{n=1}^{\infty} a_n = \sum_{k=1}^{\infty} s_k.$$

Proof. The first case is a special case of the second with $I_k = \{a_{\sigma(k)}\}$. However we use 1.) in the statement of 2.). The notation $s_k = \sum_{n\in I_k} a_n$ only makes sense since it does not matter how we enumerate I_k by 1.). Indeed, in case that I_k is infinite, we obtain for any bijection $\tau : \mathbb{N} \to I_k$ the same limit $\sum_{n=1}^{\infty} a_{\tau(n)}$ which we denote by $s_k = \sum_{n\in I_k} a_n$.

We first prove the absolute convergence of each s_k. For finite subsets $I' \subset I'' \subset I_k$, we have

$$\sum_{j=I'} |a_j| \leq \sum_{j\in I''} |a_j|$$

Hence is suffices to prove that these sums stay bounded. This holds, since

$$\sum_{j\in I'} |a_j| \leq \sum_{n=0}^{N} |a_n| \leq \sum_{n=0}^{\infty} |a_n| < \infty,$$

for $N = \max\{j \mid j \in I'\}$. For the absolute convergence of $\sum_{k=1}^{\infty} s_k$ we argue similarly:

$$\begin{aligned}
\sum_{k=1}^{l} |s_k| &= \sum_{k=1}^{l} \lim_{N\to\infty} \Big| \sum_{j\in I_k, j\le N} a_j \Big| \\
&\le \lim_{N\to\infty} \sum_{k=1}^{l} \sum_{j\in I_k, j\le N} |a_j| \\
&= \lim_{N\to\infty} \sum_{j\in\bigcup_{k=1}^{l} I_k,\, j\le N} |a_j| \\
&\le \lim_{N\to\infty} \sum_{j=1}^{N} |a_j| = \sum_{n=1}^{\infty} |a_n| < \infty.
\end{aligned}$$

To prove the equality, we consider $s = \sum_{n=1}^{\infty} a_n$ and for $\varepsilon > 0$ an n_0 such that

$$\sum_{n=n_0}^{\infty} |a_n| < \varepsilon \text{ and } \Big| s - \sum_{n=1}^{n_0-1} a_n \Big| < \varepsilon$$

holds. Let

$$k_0 := \max\big\{ k \mid \{1,\dots,n_0-1\} \cap I_k \neq \emptyset \big\}.$$

Then for $k_1 \ge k_0$ we obtain

$$\begin{aligned}
\Big| s - \sum_{k=1}^{k_1} s_k \Big| &\le \Big| s - \sum_{n=1}^{n_0-1} a_n \Big| + \sum_{k=1}^{k_1} \sum_{n\in I_k, n\ge n_0} |a_n| \\
&\le \Big| s - \sum_{n=1}^{n_0-1} a_n \Big| + \sum_{n\ge n_0} |a_n| < 2\varepsilon.
\end{aligned}$$

□

Theorem 6.22 (Cauchy–Product of two series)**.** *Let $\sum_{i=0}^{\infty} a_i$ and $\sum_{j=0}^{\infty} b_j$ be two absolutely convergent series and let (d_k) be defined by*

$$d_k = \sum_{i=0}^{k} a_i b_{k-i}.$$

Then also the series $\sum_{k=0}^{\infty} d_k$ is absolutely convergent with limit

$$\sum_{k=0}^{\infty} d_k = \big(\sum_{i=0}^{\infty} a_i\big) \cdot \big(\sum_{j=0}^{\infty} b_j\big).$$

Proof. We consider a bijective enumeration

$$\varphi\colon \mathbb{N}_0 \to \mathbb{N}_0 \times \mathbb{N}_0,\ n \mapsto \varphi(n) = (\alpha(n), \beta(n))$$

and the sequence $\sum_{n=0}^{\infty} a_{\alpha(n)} b_{\beta(n)}$. We first show that this series is absolutely convergent as well. It is enough to prove

$$\sum_{n=0}^{N} |a_{\alpha(n)} b_{\beta(n)}| \leq \big(\sum_{i=0}^{\infty} |a_i|\big) \cdot \big(\sum_{j=0}^{\infty} |b_j|\big) < \infty$$

for arbitrary N. Let N be given. We consider

$$\begin{aligned} i_0 &= \max\{\alpha(0), \ldots, \alpha(N)\}, \\ j_0 &= \max\{\beta(0), \ldots, \beta(N)\}. \end{aligned}$$

Then

$$\begin{aligned} \sum_{n=0}^{N} |a_{\alpha(n)} b_{\beta(n)}| &\leq \sum_{i=0}^{i_0} |a_i| \cdot \sum_{j=0}^{j_0} |b_j| \\ &\leq \big(\sum_{i=0}^{\infty} |a_i|\big) \cdot \big(\sum_{j=0}^{\infty} |b_j|\big) < \infty \end{aligned}$$

holds by assumption. The sequence $\sum_{k=0}^{\infty} d_k = \sum_{k=0}^{\infty} \sum_{i=0}^{k} a_i b_{k-i}$ is a rearrangement of $\sum_{n=1}^{\infty} a_{\alpha(n)} b_{\beta(n)}$. The product $\sum_{i=0}^{\infty} a_i \cdot \sum_{j=0}^{\infty} b_j = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_i b_j$ is another rearrangement. By the big rearrangement theorem both series are absolutely convergent with the same limit $\sum_{k=0}^{\infty} d_k = (\sum_{i=0}^{\infty} a_i)(\sum_{j=0}^{\infty} b_j)$. □

Definition 6.23. Let (q_n) be a sequence of numbers. The **infinite product** $\prod_{k=1}^{\infty} q_k$ is convergent if the limit $\lim_{n\to\infty} \prod_{k=1}^{n} q_k$ of the partial products exists. In that case $q = \prod_{k=1}^{\infty} q_k$ denotes the limit as well.

Theorem 6.24 (Euler product)**.** *Let p_k denote the k-th prime number and let $s > 1$ denote an integer. Then*

$$\prod_{k=1}^{\infty} \frac{1}{1 - p_k^{-s}} = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

For $s = 1$ the product is divergent. In particular there are infinitely many primes.

Proof. We prove this in several steps.

1. For $s \geq 2$ we have $\frac{1}{n^s} \leq \frac{1}{n^2}$. Hence the series $\sum_{n=1}^{\infty} \frac{1}{n^s}$ converges absolutely by the comparison theorem: The series $\sum_{n=1}^{\infty} \frac{1}{n^2}$ is a convergent majoriser.

2. The series
$$\frac{1}{1-\frac{1}{p^s}} = \sum_{l=0}^{\infty} (\frac{1}{p^s})^l$$
converges for every prime p, since it is a geoemetric series and $(\frac{1}{p})^s < \frac{1}{p} < 1$ holds.

3. The identity between the finite product on the left hand side,
$$\prod_{k=1}^{r} \sum_{e=0}^{N} (\frac{1}{p_k^e})^s \quad = \quad \sum_{\substack{n \in \mathbb{N},\ n \text{ has} \\ \text{prime factors } p_1, \ldots, p_r \\ \text{with multiplicity } \leq N}} \quad \frac{1}{n^s}$$
and the finite sum on the right hand side holds because of the unique factorisation of integers into primes. We deduce for $N \to \infty$:
$$\prod_{k=1}^{r} \frac{1}{1-p_k^{-s}} = \sum_{\substack{n \in \mathbb{N},\ n \text{ has} \\ \text{prime factors } p_1, \ldots, p_r}} \quad \frac{1}{n^s}$$
from the rearrangement theorem.

4. Finally,
$$\prod_{k=1}^{\infty} \frac{1}{1-p_k^{-s}} = \sum_{n=1}^{\infty} \frac{1}{n^s},$$
follows once more from the rearrangement theorem.

□

Corollary 6.25. *Let N be a large integer and let ω_N be the probability that two randomly chosen integers $a, b \in \mathbb{N}$ with $0 < a \leq N$, $0 < b \leq N$ have no common factor. Then*
$$\lim_{N\to\infty} \omega_N = \frac{6}{\pi^2} = 0.60792\cdots \approx 60\%.$$

Sketch of the proof. As common prime factor is a $p \leq N$. A prime $p \ll N$ does not divide a randomly chosen $a \in \{1, \dots, N\}$ with probability about $\frac{p-1}{p} = 1 - \frac{1}{p}$. The probability that both a and b do not have p as a common divsor is approximately $\frac{p^2-1}{p^2} = 1 - \frac{1}{p^2}$.

By the Chinese remainder theorem the conditions are independent for different primes. Thus

$$\omega_N \approx \prod_{p \leq N} (1 - \frac{1}{p^2})$$

and $\prod_{p \leq N} \frac{1}{1-\frac{1}{p^2}} \approx \omega_N^{-1}$. Hence

$$\lim_{N \to \infty} \frac{1}{\omega_N} = \prod_{p \text{ a prime}} \frac{1}{1 - \frac{1}{p^2}} \overset{\text{Euler}}{=} \sum_{n=1}^{\infty} \frac{1}{n^2} \overset{\text{Fourier series}}{=} \frac{\pi^2}{6}$$

and

$$\lim_{N \to \infty} \omega_N = \frac{6}{\pi^2} = 0.60792\dots$$

□

7 Power series

Many important functions are defined via power series.

Definition 7.1. Let (a_n) be a sequence of numbers, and $x \in \mathbb{R}$. Then the series

$$\sum_{n=0}^{\infty} a_n x^n.$$

is called a power series.

Example 7.2.

1. $\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$ for $|x| < 1$.

2. The exponential function

   $$\exp : \mathbb{R} \to \mathbb{R}$$

   is defined by

   $$\exp(x) := \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

   This power series converges for all $x \in \mathbb{R}$ by the quotient test:

   $$\left| \frac{\frac{x^{n+1}}{(n+1)!}}{\frac{x^n}{n!}} \right| = \frac{|x|}{n+1} \underset{n\to\infty}{\longrightarrow} 0$$

   The value $e = \exp(1) = 2.71828\cdots$ is called Euler's number. The notation

   $$e^x := \exp(x)$$

   is also in use.

3. Likewise we introduce $\sin$ and $\cos$ as power series

   $$\begin{aligned} \sin(x) &:= \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots \\ \cos(x) &:= \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots \end{aligned}$$

   The usual properties will be deduced from these formulas later.

The convergence behaviour of power series is best explained if we allow complex numbers.

Proposition 7.3. *Let $\sum_{n=0}^{\infty} a_n z^n$ be a power series. If the power series converges for $z_0 \in \mathbb{C}$, then it converges absolutely for every element z of the disc*

$$\{z \in \mathbb{C} \mid |z| < |z_0|\}.$$

Proof. Since $\sum_{n=0}^{\infty} a_n z_0^n$ converges, its terms form a zero sequence. In particular the terms are bounded, say $|a_n z_0^n| \leq M$ for all n. Then

$$\sum_{n=0}^{\infty} M \left| \frac{z}{z_0} \right|^n$$

is a majoriser of $\sum_{n=0}^{\infty} |a_n z^n|$, which converges if $|z| < |z_0|$. □

Definition 7.4. Let $\sum_{n=0}^{\infty} a_n z^n$ be a power series. Then

$$R := \sup \Big\{ |z_0| \mid \sum_{n=0}^{\infty} a_n z_0^n \text{ converges} \Big\} \in [0, \infty]$$

is called the **radius of convergence** of the power series. The power series converges for all z in the disc $\{z \in \mathbb{C} \mid |z| < R\}$ and diverges for all z with $|z| > R$ by Proposition 7.3.

Theorem 7.5. *Let $(a_n)_{n \in \mathbb{N}_0}$ be a sequence of complex numbers with $a_n \neq 0 \ \forall n$. If the limit $q = \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right|$ exists, then the power series $\sum_{n=0}^{\infty} a_n z^n$ has the radius of convergence*

$$R = \begin{cases} \frac{1}{q}, & \text{if } q > 0, \\ \infty, & \text{if } q = 0. \end{cases}$$

Proof. Quotient test □

The formula does not always apply, e.g., for sin. A formula which always applies is the following:

Definition 7.6. Let (b_n) be a sequence real numbers. Then

$$\limsup_{n\to\infty}(b_n) := \lim_{n\to\infty} \sup \{ b_k \mid k \geq n \}$$

is called the **limit superior** of (b_n). If (b_n) is not bounded from above, then we set

$$\limsup b_n = +\infty.$$

Similarly,

$$\liminf_{n\to\infty} b_n := \lim_{n\to\infty} \inf\{b_k \mid k \geq n\},$$

is called the **limit inferior** of (b_n).

Theorem 7.7. *Let (a_n) be a sequence of complex numbers and*

$$q = \limsup_{n\to\infty}(\sqrt[n]{|a_n|}).$$

Then the power series

$$\sum_{n=0}^{\infty} a_n z^n$$

has radius of convergence

$$R = \begin{cases} 0, & \textit{falls } q = \infty, \\ \frac{1}{q}, & \textit{falls } 0 < q < \infty, \\ \infty, & \textit{falls } q = 0. \end{cases}$$

Proof. We use a more precise version of the root test. □

7.1 The complex exponential function

Definition 7.8. The complex exponential function

$$\exp : \mathbb{C} \to \mathbb{C}$$

is defined by

$$\exp(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

Theorem 7.9 (Funktional equation of the exponential function)**.**

$$\exp(z + w) = \exp(z) \cdot \exp(w)$$

holds for all $z, w \in \mathbb{C}$.

Proof. We consider the Cauchy product of

$$\sum_{k=0}^{\infty} \frac{z^k}{k!} \text{ and } \sum_{n=0}^{\infty} \frac{w^n}{n!}.$$

The binomial formula gives

$$d_n = \sum_{k=0}^{n} \frac{z^k}{k!} \cdot \frac{w^{n-k}}{(n-k)!} = \sum_{k=0}^{n} \binom{n}{k} \frac{z^k \cdot w^{n-k}}{n!} = \frac{(z+w)^n}{n!}.$$

Hence the rearrangement theorem implies

$$\begin{aligned} \exp(z+w) &= \sum_{n=0}^{\infty} \frac{(z+w)^n}{n!} \\ &= (\sum_{k=0}^{\infty} \frac{z^k}{k!}) \cdot (\sum_{l=0}^{\infty} \frac{w^l}{l!}) \\ &= \exp(z) \cdot \exp(w). \end{aligned}$$

□

Corollary 7.10. *The following holds*

1. $\exp(0) = 1$,
2. $\exp(-z) = \frac{1}{\exp(z)}$. *In particular* $\exp(z) \in \mathbb{C}^* \; \forall z \in \mathbb{C}$, *where* $\mathbb{C}^* := \mathbb{C} \backslash \{0\}$.

Proof. $1 = \exp(0) = \exp(z + (-z)) = \exp(z) \cdot \exp(-z)$. □

Evaluating the exponential function at purely imaginary numbers $z = iy$ with $y \in \mathbb{R}$ gives

$$\begin{aligned} \exp(iy) &= \sum_{n=0}^{\infty} \frac{(iy)^n}{n!} \\ &= \sum_{k=0}^{\infty} (-1)^k \frac{y^{2k}}{2k!} \; + \; i \cdot \sum_{k=0}^{\infty} (-1)^k \frac{y^{2k+1}}{(2k+1)!} \\ &= \cos(y) + i \sin(y), \end{aligned}$$

i.e., a connection between the complex exponential function and sine and cosine. The functional equation of the complex exponential function implies the addition laws for sine and cosine.

Corollary 7.11 (Addition laws for sine and cosine)**.** *Let* $\alpha, \beta \in \mathbb{R}$. *Then*

$$\begin{aligned}\cos(\alpha+\beta) &= \cos\alpha\cdot\cos\beta - \sin\alpha\cdot\sin\beta,\\ \sin(\alpha+\beta) &= \sin\alpha\cdot\cos\beta + \cos\alpha\cdot\sin\beta.\end{aligned}$$

In particular (with the notation $\sin^k\alpha := (\sin\alpha)^k$ *and similarly for* $\cos$*):*

$$1 = \sin^2\alpha + \cos^2\alpha.$$

Proof. We have

$$\begin{aligned}\cos(\alpha+\beta) + i\sin(\alpha+\beta) &= \exp(i(\alpha+\beta))\\ &= \exp(i\alpha)\cdot\exp(i\beta)\\ &= (\cos\alpha + i\sin\alpha)\cdot(\cos\beta + i\sin\beta)\\ &= (\cos\alpha\cos\beta - \sin\alpha\sin\beta)\\ &\qquad +i(\cos\alpha\sin\beta + \sin\alpha\cos\beta),\end{aligned}$$

where we used multiplication in $\mathbb{C}$ for the last equation. Taking real and imaginary parts gives the assertion.

The last formula follows from

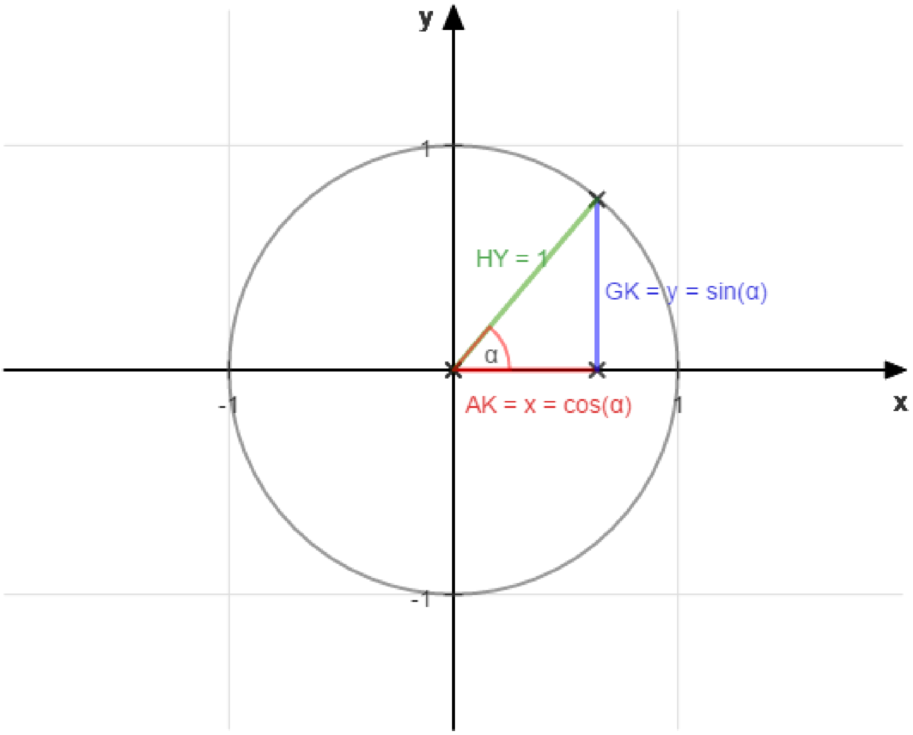
$$\begin{aligned}1 = \cos(0) &= \cos\bigl(\alpha + (-\alpha)\bigr)\\ &= \cos(\alpha)\cos(-\alpha) - \sin(\alpha)\sin(-\alpha)\\ &= \sin^2\alpha + \cos^2\alpha\end{aligned}$$

because $\cos(-\alpha) = \cos(\alpha)$ and $\sin(-\alpha) = -\sin(\alpha)$ hold, since the power series of $\cos$ has only even terms and the one for $\sin$ has only odd terms. □

The last formula says that

$$t \mapsto (\cos t, \sin t)$$

is a parametrization of the unit circle. Here t is measured in radian, as we will see later.


y
1
HY = 1
GK = y = sin(α)
α
-1
AK = x = cos(α)
1
x
-1

8 Continuity

Let $D \subset \mathbb{R}$ and $f\colon D \to \mathbb{R}$ be a real valued function. D is called the domain of definition of f, and typically D is a union of intervals.

The heaviside function $h : \mathbb{R} \to \mathbb{R}$ with

$$h(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

has a jump at $x = 0$. Roughly speaking, continuous functions are functions which do not have any jumps. We give a precise definition.

Definition 8.1. Let $f\colon D \to \mathbb{R}$ be a function on a domain $D \subset \mathbb{R}$. f is called **continuous** at a point $x_0 \in D$ if

$$\forall \varepsilon > 0 \, \exists \, \delta > 0 : \ |f(x) - f(x_0)| < \varepsilon \, \forall x \in D \text{ with } |x - x_0| < \delta.$$

f is called continuous on D if f is continuous at x_0 for all $x_0 \in D$.

Hence f is not continuous in x_0 if

$$\exists \varepsilon > 0 \text{ such that } \forall \delta > 0 \, \exists x \in D \text{ with } |x - x_0| < \delta \text{ and } |f(x) - f(x_0)| \geq \varepsilon.$$

Example 8.2.

1. $f(x) = x$ is continuous in all points of $\mathbb{R}$. For $\varepsilon > 0$ given one can take $\delta = \varepsilon$.

2. $f(x) = x^2$ is continuous on $\mathbb{R}$: We have the estimate

   $$\begin{aligned} |x^2 - x_0^2| &= |x + x_0| \cdot |x - x_0| \\ &\leq |2x_0 + 1| \cdot |x - x_0| \quad \forall x \text{ with } |x - x_0| < 1. \end{aligned}$$

   Hence $|x^2 - x_0^2| < \varepsilon \, \forall x$ with $|x - x_0| < \delta$, where we choose

   $$\delta = \min\left\{ 1, \ \frac{\varepsilon}{2|x_0| + 1} \right\}.$$

   In this case δ depends both on ε and x_0.

3. The heaviside function h from above is not continuous in $x_0 = 0$: For any $\varepsilon \leq 1$ and arbitrary small $\delta > 0$ there exist negative numbers $x \in] - \delta, \delta[$. For these x we have

   $$|h(x) - h(x_0)| = |0 - 1| = 1 \not< \varepsilon.$$

Theorem 8.3 (Limit criterion for continuity)**.** *Let $f\colon D \to \mathbb{R}$ be a function and $x_0 \in D \subset \mathbb{R}$ a point. f is continuous in x_0 if and only if for all sequences $(x_n)_{n\in\mathbb{N}}$ with $x_n \in D$ and $\lim_{n\to\infty} x_n = x_0$ the identity*

$$f(x_0) = \lim_{n\to\infty} f(x_n)$$

holds

Proof. Suppose f is continuous in x_0. Let (x_n) be a sequence in D which converges to x_0. Let $\varepsilon > 0$ be given. Then there exists a $\delta > 0$ such that

$$|f(x) - f(x_n)| < \varepsilon\, \forall x \in D \text{ with } |x - x_0| < \delta.$$

Since $\lim x_n = x_0$ there exists an n_0 such $|x_n - x_0| < \delta\; \forall n \geq n_0$. Hence

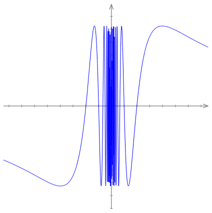
$$|f(x_n) - f(x_0)| < \varepsilon\, \forall n \geq n_0.$$

This proves one direction.

For the other direction we assume that f is not continuous in x_0. Then there exists $\varepsilon > 0$ such that for all $\delta > 0$ there exists an $x \in D$ with $|x - x_0| < \delta$ and $|f(x) - f(x_0| > \varepsilon$. If we apply this statement for $\delta = \frac{1}{n}$, we get $x_n \in D$ with $|f(x_n) - f(x_0)| > \varepsilon$. The sequence (x_n) to converges x_0, but $f(x_n)$ does not converge to $f(x_0)$. □

Example 8.4. Consider the function $f\colon \mathbb{R} \to \mathbb{R}$ with

$$f(x) = \begin{cases} 0 & \text{for } x = 0 \\ \sin(\frac{1}{x}) & \text{else} \end{cases}$$



As known from high school (we will reprove this in Section 10) one has

$$1 = \sin(\frac{1}{2}\pi) = \sin(\frac{1}{2}\pi + 2\pi n)$$

for $n \in \mathbb{Z}$. Thus $x_n = \frac{1}{\frac{1}{2}\pi + 2\pi n}$ is a sequence with $x_n \underset{n\to\infty}{\longrightarrow} 0$ and $f(x_n) = 1$. This shows that f is not continues in $x_0 = 0$. Similarly, $x'_n = \frac{1}{\frac{3}{2}\pi + 2\pi n}$ is a sequence with $x'_n \underset{n\to\infty}{\longrightarrow} 0$ and $f(x'_n) = -1$. Hence there is no way to give $f(0)$ a value, which would make the function continuous.

Theorem 8.5. *Let $f, g\colon D \to \mathbb{R}$ be functions.*

1. *If f, g are continuous in x_0, then $f + g$ and $f \cdot g$ are also continuous in x_0*
2. *If f, g are continuous in x_0 and $g(x_0) \neq 0$, then*

$$\frac{f}{g}\colon D' \to \mathbb{R}$$

mit $D' = \{x \in D \mid g(x) \neq 0\} \subset D$ is also continuous in $x_0 \in D'$

Proof. Similar to the corresponding rules for limits. □

Corollary 8.6. 1. *Polynomials*

$$f(x) = a_n x^n + \ldots + a_1 x + a_0$$

with coefficients $a_k \in \mathbb{R}$ define continuous functions $f\colon \mathbb{R} \to \mathbb{R}$.

2. *A rational functions , i.e., maps $\frac{f}{g}\colon D \to \mathbb{R}$ with f, g defined by polynomials, are continues on the domain $D = \{x \in \mathbb{R} \mid g(x) \neq 0\}$.*

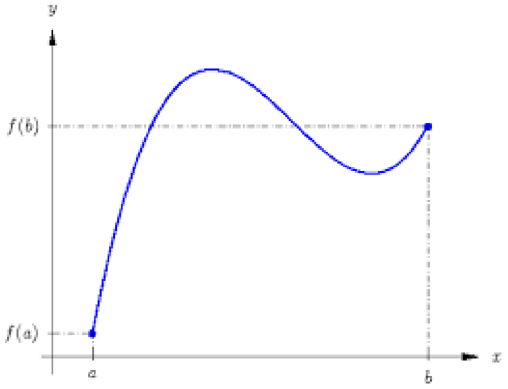
8.1 Intermediate value theorem and applications

An important property of continuous functions is the following:

Theorem 8.7 (Intermediate value theorem)**.** *Let $f\colon [a, b] \to \mathbb{R}$ be a continuous function and let c be a value between $f(a)$ und $f(b)$, i.e., $f(a) \leq c \leq f(b)$ or $f(a) \geq c \geq f(b)$. Then there exists a $\xi \in [a, b]$ such that*

$$f(\xi) = c.$$

In particular, every continuous function f with $f(a) < 0$ and $f(b) > 0$ has a zero in $[a, b]$.



Proof. Passing to the function ±$(f-c)$, we may assume that $f(a)<0, f(b)>0$ and $c=0$. We will construct sequences (x_n), (y_n) which converge to a zero ξ by the bisection algorithm. We start with $x_0=a$ and $y_0=b$. If x_n, y_n are already constructed, then we consider $\overline{x}=\frac{x_n+y_n}{2}$ and $f(\overline{x})$ and define

$$x_{n+1}=\begin{cases}\overline{x}, & \text{if } f(\overline{x})<0,\\ x_n, & \text{else},\end{cases} \quad \text{and} \quad y_{n+1}=\begin{cases}y_n, & \text{if } f(\overline{x})<0,\\ \overline{x}, & \text{else}.\end{cases}$$

We have

1. $f(x_n)<0\ \forall n$ and $f(y_n)\geq 0\ \forall n$.
2. $|y_n-x_n|=2^{-n}(b-a)$.
3. (x_n) is monotonously increasing, and (y_n) is monotonously decreasing.

Hence both sequences converge, and since

$$\lim_{n\to\infty}(y_n-x_n)=\lim_{n\to\infty}2^{-n}(b-a)=0$$

the limits $\xi=\lim_{n\to\infty}x_n=\lim_{n\to\infty}y_n$ coincide. The continuity of f gives:

$$f(\xi)=\lim_{n\to\infty}f(x_n)\leq 0$$

since $f(x_n)<0\ \forall n$ and

$$f(\xi)=\lim_{n\to\infty}f(y_n)\geq 0$$

since $f(y_n)\geq 0\ \forall n$. Hence $f(\xi)=0$. □

Theorem 8.8 (Existence of a maximum and a minimum)**.** *Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function on a closed bounded intervall. Then there exist $x_{\max}, x_{\min} \in [a,b]$ with*

$$\begin{aligned} f(x_{\max}) &= \sup\big\{ f(x) \mid x \in [a,b] \big\}, \\ f(x_{\min}) &= \inf\big\{ f(x) \mid x \in [a,b] \big\}. \end{aligned}$$

In particular, f is bounded.

We say: f takes its **maximum** in $x_{\max}$ and write

$$\max_{x\in[a,b]} f(x) := f(x_{\max}).$$

Similarly for the **minimum**: $\min_{x\in[a,b]} f(x) := f(x_{\min})$.

Remark 8.9. The condition that $[a,b]$ is a bounded closed interval is essential. The function $f\colon]0,\infty[\to \mathbb{R}$ with $f(x) = \frac{1}{x}$ has neither a maximum nor a minimum.

Proof. Set

$$M = \sup\{f(x) \mid x \in [a,b]\} \in \mathbb{R} \cup \{\infty\}.$$

By the definition of the supremum there exists a sequence (x_n) in $[a,b]$ such that $\lim_{n\to\infty} f(x_n) = M$ (or that $f(x_n)$ grows unboundedly in case $M = \infty$). By Bolzano-Weierstrass there exists a convergent subsequence (x_{n_k}). Let $x_{\max} = \lim_{k\to\infty} x_{n_k}$. Then $x_{\max} \in [a,b]$ and

$$f(x_{\max}) = \lim_{k\to\infty} f(x_{n_k}) = M$$

by the continuity of f. In particular $M < \infty$. □

Definition 8.10. A function : $I \to \mathbb{R}$ is defined on an interval I **grows monotonously** (or is **monotone increasing**) if points $x_1 < x_2$ in I the values satisfy $f(x_1) \leq f(x_2)$. It is **strictly** monotone increasing if $f(x_1) < f(x_2)$. **Monotone decreasing** and **strictly monotone decreasing** are defined similarly. f is (strictly) monotone if it is (strictly) monotone increasing or (strictly) monotone decreasing.

Proposition 8.11.

1. *Let $f : I \to \mathbb{R}$ be a continuous function on an interval I. The $J = f(I) \subset \mathbb{R}$ is a interval as well.*

2. *If in addition f is strictly monotone, then $f\colon I \to J$ is bijective, and the inverse map $f^{-1} : J \to I$ is also continuous.*

Proof. 1.) $J \subset \mathbb{R}$ is an interval if with two points $y_1 < y_2$ of J the whole interval $[y_1, y_2]$ is contained in J. This holds by the intermediate value theorem.

2.) If f is strictly monotone, then f is injective. Hence $f\colon I \to J$ is injective and surjective, i.e., f is bijective. Let (y_n) be a sequence in J, which converges to $y_0 \in J$. Then $x_n = f^{-1}(y_n)$ defines a sequence in I. Suppose (x_n) does not converge to $x_0 = f^{-1}(y_0)$. Then there exists a monotone subsequence (x_{n_k}) which does not converge to x_0: Infinitely many x_n would lie outside a ε-neighborhood of x_0, and from these elements we can choose a monotone subsequence by Lemma 5.20. The corresponding sequence (y_{n_k}) is monotone as well. Hence (y_{n_k}) lies in the interval with boundary points y_0 and y_{n_1}. This implies that (x_{n_k}) is a monotone sequence contained in the interval with boundary points x_0 and x_{n_1}. We conclude (x_{n_k}) is convergent. By our assumption the limit $x_0' = \lim_{k\to\infty} x_{n_k}$ is a point not equal to x_0. The continuity of f implies $f(x_0') = \lim f(x_{n_k}) = \lim y_{n_k} = y_0 = f(x_0)$. This contradicts the fact that f is strictly monotone.

□

Definition 8.12. Let $f\colon D \to \mathbb{R}$ be a function and let $a \in \mathbb{R} \setminus D$ be a point such that there exists a sequence (x_n) in D with $\lim x_n = a$. If for each sequence (x_n) in D with $\lim x_n = a$, there exists the limit $b = \lim_{n\to\infty} f(x_n)$, and if these limits are equal for all possible sequences, then the common limit

$$\lim_{x\to a} f(x) = b$$

is called the limit of $f(x)$ when x approaches a. The notation

$$\lim_{x\nearrow a} f(x)$$

is used if we consider only the sequence (x_n) with $\lim x_n = a$ and $x_n < a$. The limit from above $\lim_{x\searrow a} f(x)$ is defined similarly.

Example 8.13. The rational function $f(x) = \frac{x^2-1}{x-1}$ is only defined on $D = \mathbb{R}\setminus\{1\}$. However, since

$$\lim_{x\to 1} \frac{x^2-1}{x-1} = \lim_{x\to 1}(x+1) = 2,$$

we can extend this function to the continuous function $\tilde{f}\colon \mathbb{R} \to \mathbb{R}$ defined by $\tilde{f}(x) = x + 1$.

9 Differentiation

Definition 9.1. Let $f\colon I \to \mathbb{R}$ be a function on an interval I and let $x_0 \in I$ be a point. We call f **differentiable** at x_0 if the limit

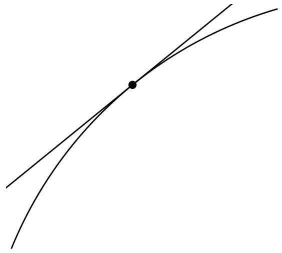
$$\lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0}$$

exists.

Geometrically, we can interpret the difference quotient

$$\frac{f(x)-f(x_0)}{x-x_0}$$

as the slope of the secant line through the points $(x, f(x))$ and $(x_0, f(x_0))$ on the graph G_f. Hence the limit may be interpreted as the slope of the tangent line to G_f at $(x_0, f(x_0))$. Thus f is differentiable at x_0 if we can associate in a sensible way a tangent line to G_f at $(x_0, f(x_0)$.



Definition 9.2. A function $f\colon I \to \mathbb{R}$ is called differentiable on I if f is differentiable at every point of I. In that case the function

$$f'\colon I \to \mathbb{R}$$

defined by

$$f'(x) := \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$$

is called the **derivative** of f. Thus $f'(x_0)$ is the slope of the tangent to G_f at $(x_0, f(x_0))$. In case f' is continuous, we call f **continuously differentiable**.

Remark 9.3.
1. In Physics the derivative is basic to even define the concept of speed. If $f : I \to \mathbb{R}, t \mapsto f(t)$ describes the motion of a point on a line, then $f'(t)$ is the speed at time t.

2. The notation $f'(x)$ (and $\dot{f}(t)$ for derivates after time) goes back to Newton. Leibniz used the notation

$$\frac{df}{dx}(x).$$

Proposition 9.4. *Let $f : I \to \mathbb{R}$ be a function on an interval and let $x_0 \in I$. Then*

$$f \text{ differentiable in } x_0 \quad \Longrightarrow \quad f \text{ continuous in } x_0.$$

Proof.

$$\lim_{x \to x_0} (f(x) - f(x_0)) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \cdot \lim_{x \to x_0} (x - x_0) = 0$$

holds because the right hand side limits exist. □

Example 9.5. 1. $f(x) = x^2$ is differentiable since

$$\lim_{x \to x_0} \frac{x^2 - x_0^2}{x - x_0} = \lim_{x \to x_0} (x + x_0) = 2x_0.$$

Hence $f'(x) = 2x$.

2. Constant functions $f(x) = c \; \forall x$ have derivative 0.

3. Linear functions $f(x) = mx + c$ have derivative $f'(x) = m$, a constant function.

Theorem 9.6 (Calculation rules for derivatives)**.** *Let $f, g : I \to \mathbb{R}$ be functions on an interval which are differentiable at $x_0 \in I$. Then*

1. $f + g : I \to \mathbb{R}$ is differentiable in x_0 with

$$(f + g)'(x_0) = f'(x_0) + g'(x_0).$$

*2. **Product rule.** $fg : I \to \mathbb{R}$ is differentiable in x_0 with*

$$(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0).$$

*3. **Quotient rule.** If $g(x_0) \neq 0$ then*

$$\frac{f}{g} : \{x \in I \mid g(x) \neq 0\} \to \mathbb{R}$$

is differentiable in x_0 with

$$(\frac{f}{g})'(x_0) = \frac{f'g - fg'}{g^2}(x_0).$$

Proof. The first assertion is clear. For the second we write

$$\begin{aligned}\frac{f(x)g(x)-f(x_0)g(x_0)}{x-x_0} &= \frac{f(x)g(x)-f(x_0)g(x)+f(x_0)g(x)-f(x_0)g(x_0)}{x-x_0}\\ &= \frac{f(x)-f(x_0)}{x-x_0}g(x)+f(x_0)\frac{g(x)-g(x_0)}{x-x_0}.\end{aligned}$$

Hence

$$\lim_{x\to x_0}\frac{f(x)g(x)-f(x_0)g(x_0)}{x-x_0} = f'(x_0)g(x_0)+f(x_0)g'(x_0).$$

3.) We treat the special case $\frac{1}{g}$ first.

$$\begin{aligned}\lim_{x\to x_0}\frac{\frac{1}{g(x)}-\frac{1}{g(x_0}}{x-x_0} &= \lim_{x\to x_0}\frac{1}{g(x)g(x_0)}\frac{g(x_0)-g(x)}{x-x_0}\\ &= -\frac{g'}{g^2}(x_0).\end{aligned}$$

The general case follows now with the product rule:

$$(f\frac{1}{g})(x_0) = (f\frac{(-g')}{g^2}+f'\frac{1}{g})(x_0) = \frac{f'g-fg'}{g^2}(x_0).$$

□

Example 9.7. The function $f(x) = x^n$ is for arbitrary $n \in \mathbb{Z}$ on its domain of definition differentiable with the exponent rule

$$f'(x) = nx^{n-1}.$$

The case $n \geq 0$ follows by induction on n. The induction hypothesis was established above.

$$(x^n)' = (x\cdot x^{n-1})' = 1\cdot x^{n-1}+x\cdot(n-1)x^{n-2} = nx^{n-1}$$

shows the induction step using the product rule. For negative n, say $n = -k$, the quotient rule gives

$$(x^n)' = (\frac{1}{x^k})' = \frac{-kx^{k-1}}{x^{2k}} = -kx^{-k-1}.$$

Rational functions $r = \frac{f}{g}$ are differentiable on their domain of definition. If f and g have no common factor, then a zero of g are called a **pole** of r.

Theorem 9.8 (Chain rule)**.** *Suppose $f\colon I \to \mathbb{R}$ and $g : J \to \mathbb{R}$ are functions with $f(I) \subset J$. If f is differentiable at x_0 and g is differentiable at $f(x_0)$, then the composition $g \circ f\colon I \to \mathbb{R}$ is differentiable at x_0 and*

$$(g \circ f)(x_0) = g'(f(x_0))f'(x_0).$$

The factor $f'(x_0)$ in this formula is called the ***inner derivative****.*

Proof. We have:

$$\frac{(g \circ f)(x_0 + h) - (g \circ f)(x_0)}{h} = \frac{g(f(x_0 + h)) - g(f(x_0))}{f(x_0 + h) - f(x_0)} \cdot \frac{f(x_0 + h) - f(x_0)}{h}.$$

With $h \to 0$ also $f(x_0 + h) \to f(x_0)$ holds because f is continuous at x_0.

$$\frac{g(f(x_0 + h)) - g(f(x_0))}{f(x_0 + h) - f(x_0)} \underset{h\to 0}{\longrightarrow} g'(f(x_0))$$

and

$$\frac{f(x_0 + h) - f(x_0)}{h} \to f'(x_0).$$

This argument is valid if $f(x_0+h)-f(x_0) \neq 0$. On the other hand, if $f(x_0+h_n)-f(x_0) = 0$ for a zero sequence (h_n), then $f'(x_0) = 0$ and $\frac{g(f(x_0+h_n))-g(f(x_0))}{h_n} = 0$. Hence also in the case we have

$$\lim_{n\to\infty} \frac{g(f(x_0 + h_n)) - g(f(x_0))}{h_n} = 0 = g'(f(x_0)) \cdot f'(x_0).$$

□

Example 9.9. We consider $f(x) = x^2 + 1$ and $g(x) = x^3$. The composition is $(g \circ f)(x) = (x^2 + 1)^3$. Using the chain rule we obtain

$$(g \circ f)'(x) = 3(x^2 + 1)^2 2x.$$

First expanding and then taking the derivative

$$(x^6 + 3x^4 + 3x^2 + 1)' = 6x^5 + 12x^3 + 6x,$$

gives the same result.

Theorem 9.10 (Derivative of the inverse function)**.** *Let $f\colon I \to \mathbb{R}$ be a strictly monotone function, and denote by $J = f(I)$. If f is differentiable in x_0 with $f'(x_0) \neq 0$, then the inverse function $f^{-1} : J \to I \subset \mathbb{R}$ is differentiable in $y_0 = f(x_0)$ and*

$$(f^{-1})'(y_0) = \frac{1}{f'(f^{-1}(y_0))} = \frac{1}{f'(x_0)}.$$

Proof. Since by assumption $f'(x_0) \neq 0$, we have $\frac{f(x)-f(x_0)}{x-x_0} \neq 0$ for x nearby x_0. Since $x \to x_0$ implies $y = f(x) \to y_0 = f(x_0)$, we obtain:

$$\frac{f^{-1}(y) - f^{-1}(y_0)}{y - y_0} = \frac{x - x_0}{f(x) - f(x_0)} \underset{y \to y_0}{\longrightarrow} \frac{1}{f'(x_0)}.$$

□

Remark 9.11. A way to memorise the formula is to use the chain rule: $f^{-1} \circ f = \mathrm{id}_I$ gives

$$1 = (f^{-1})'(f(x_0))f'(x_0) = (f^{-1})'(y_0)f'(x_0).$$

Example 9.12. The **k-th root** $g(x) = \sqrt[k]{x} = x^{1/k}$, $k \in \mathbb{N}$, is defined as the inverse function of the strictly montonous function $f\colon \mathbb{R}_{\geq 0} \to \mathbb{R},\ f(x) = x^k$. We have $f'(x) = kx^{k-1} \neq 0$ for $x > 0$. Hence $g\colon \mathbb{R}_{\geq 0} \to \mathbb{R}$ is differentiable on $\mathbb{R}_{>0}$ with

$$g'(x) = \frac{1}{k(\sqrt[k]{x})^{k-1}} = \frac{1}{k}x^{\frac{1-k}{k}} = \frac{1}{k}x^{\frac{1}{k}-1}.$$

Again the exponent rule is valid.

9.1 Local extrema and the mean value theorem

The derivative is often used to find local maxima or minima.

Definition 9.13. Let $f\colon I \to \mathbb{R}$ on an interval and $x_0 \in I$. We say, f has in x_0 a **local maximum**, if $\exists\ h > 0$ such that $]x_0 - h, x_0 + h[\subset I$ and

$$f(x_0) \geq f(x) \forall x \in\ ]x_0 - h, x_0 + h[.$$

holds. **Local minima** are defined similarly. A **local extremum** is a local maximum or a local minimum.
If

$$f(x_0) > f(x)\ \forall x \in]x_0 - h, x_0 + h[,\ x \neq x_0,$$

holds, then we speak of an **isolated maximum**. Isolated minima and isolated extrema are defined accordingly.

An **absolute maximum** or **global maximum** is a point $x_0 \in I$ such that

$$f(x_0) \geq f(x) \; \forall x \in I$$

holds. **Absolute Minima** and **absolute extrema** are defined similarly.

Theorem 9.14. *Let $f\colon]a,b[\to \mathbb{R}$ be a differentiable function. If $x_0 \in]a,b[$ is a local extremum of f, then*

$$f'(x_0) = 0.$$

Proof. We consider the case of a local maximum. We have

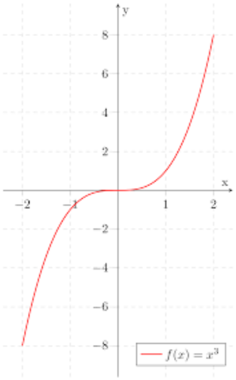
$$0 \leq \frac{f(x) - f(x_0)}{x - x_0} \text{ for } x < x_0 \text{ and } \frac{f(x) - f(x_0)}{x - x_0} \leq 0 \text{ for } x > x_0.$$

Hence

$$0 \leq \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0) \leq 0$$

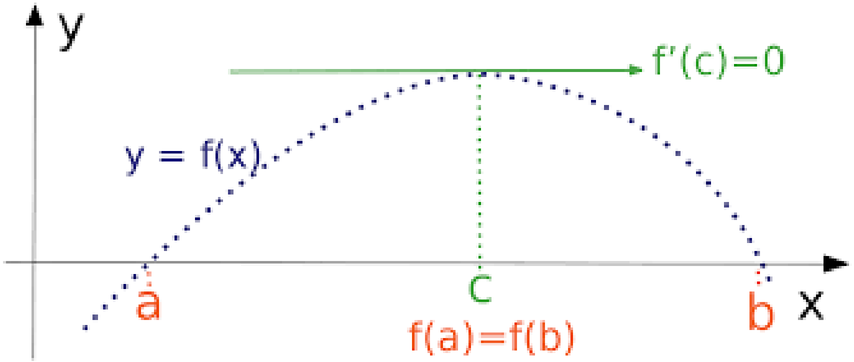
as claimed. □

Remark 9.15. $f'(x) = 0$ is a necessary but not sufficient condition for a local extremum for a differentiable function. For example, the function $f(x) = x^3$ satisfies $f'(0) = 0$, but $x_0 = 0$ is not a local extremum.



Theorem 9.16 (Rolle's theorem)**.** *Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function with $f(a) = f(b)$, which is differentiable on $]a,b[$. Then there exists a $\xi \in]a,b[$ with*

$$f'(\xi) = 0.$$

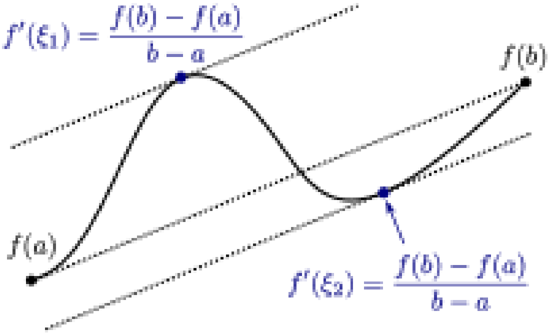


Proof. f takes its global maximum and global minimum on $[a, b]$ by Theorem 8.8. If both of these points lie on the boundary, then f is constant and all $\xi \in]a, b[$ satisfy f′$(\xi) = 0$. Otherwise at least one extrema lies in the interior $]a, b[$. This point ξ is also a local extremum, hence $f'(\xi) = 0$. □

Theorem 9.17 (Mean value theorem)**.** *Let* $f : [a, b] \to \mathbb{R}$ *be a continues function, which is differentiable on* $]a, b[$*. Then there exists a* $\xi \in]a, b[$ *such that*

$$\frac{f(b) - f(a)}{b - a} = f'(\xi)$$

.


$f'(\xi_1) = \frac{f(b)-f(a)}{b-a}$
$f(b)$
$f(a)$
$f'(\xi_2) = \frac{f(b)-f(a)}{b-a}$

Proof. We apply Rolle's theorem to the function

$$F(x) := f(x) - \frac{f(b) - f(a)}{b - a}(x - a)$$

with $F(a) = f(a) = F(b)$. Thus there exists a point $\xi \in]a, b[$ with

$$0 = F'(\xi) = f'(\xi) - \frac{f(b) - f(a)}{b - a}.$$

□

Corollary 9.18. *Let $f\colon [a, b] \to \mathbb{R}$ be continuous and in $]a, b[$ differentiable. Suppose we have constants $m, M \in \mathbb{R}$ such that $m \leq f'(x) \leq M$ for all $x \in]a, b[$. Then*

$$m(x_2 - x_1) \leq f(x_2) - f(x_1) \leq M(x_2 - x_1)$$

holds for all $x_1 < x_2$ in $[a, b]$.

Proof. If not, then the mean value theorem gives a ξ with $f'(\xi) = \frac{f(x_2)-f(x_1)}{x_2-x_1} \notin [m, M]$. □

Corollary 9.19. *Let $f\colon [a, b] \to \mathbb{R}$ be continuous and in $]a, b[$ differentiable. Suppose $\forall x \in]a, b[$ we have*

1. *If $f'(x) \geq 0 \; \forall x \in]a, b[$, then f is monotone increasing.*
2. *If $f'(x) > 0 \; \forall x \in]a, b[$, then f is strictly monotone increasing.*
3. *If $f'(x) \leq 0 \; \forall x \in]a, b[$, then f is monotone decreasing.*
4. *If $f'(x) < 0 \; \forall x \in]a, b[$, then f is strictly monotone decreasing.*
5. *If $f'(x) = 0 \; \forall x \in]a, b[$, then f is constant.*

Proof. 1.) Suppose f is not monotone increasing. Then $\exists x_1 < x_2 \in [a, b]$ with $f(x_1) > f(x_2)$ and the mean value theorem gives a $\xi \in]x_1, x_2[\subset]a, b[$ with $f'(\xi) < 0$, contradicting our assumption. The other assertions follow similarly. □

9.2 Higher derivatives

Definition 9.20. Let $f\colon I \to \mathbb{R}$ be differentiable. f is twice differentiable, f' is differentiable again. In that case $f'' := f^{(2)} := (f')'$ denotes the second derivative.

Recursively, we define f is n-times differentiable if $f^{(n-1)}$ is differentiable, and then

$$f^{(n)} = (f^{(n-1)})'$$

denotes the n-th derivative.

Theorem 9.21 (Sufficient criterion for local extrema)**.** *Let $f\colon\,]a,b[\to \mathbb{R}$ be two-times differentiable, and let $x_0 \in]a,b[$. If $f'(x_0) = 0$ and $f''(x_0) \neq 0$, then f has an isolated local extremum at x_0.*

It is a local maximum if $f''(x_0) < 0$ and a local minimum, if $f''(x_0) > 0$.

Proof. We assume that $f'(x_0) = 0$ and $f''(x_0) < 0$. Since

$$\lim_{x \to x_0} \frac{f'(x) - f'(x_0)}{x - x_0} < 0,$$

we conclude that there exists an $\varepsilon > 0$ such that

$$f'(x) > 0 \text{ for } x \in\,]x_0 - \varepsilon, x_0[\text{ and} f'(x) < 0 \text{ for } x \in\,]x_0, x_0 + \varepsilon[.$$

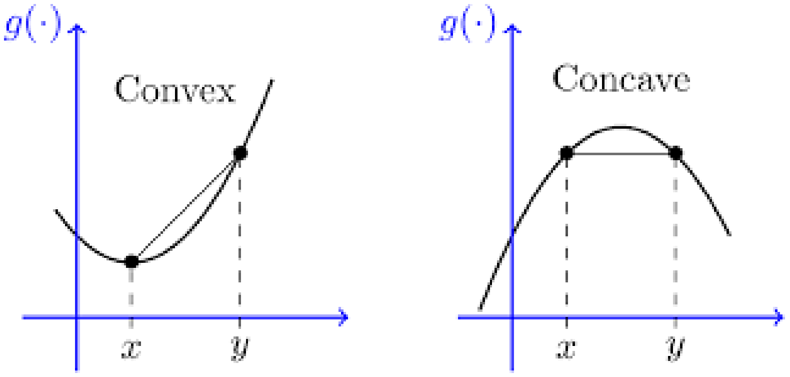
So f is strictly increasing on $]x_0 - \varepsilon, x_0[$ and strictly decreasing on $]x_0, x_0 + \varepsilon[$. □

Example 9.22. Let $f(x) = x^2$. Then $f'(x) = 2x$ and $f''(x) = 2 > 0$. The function f has a local minimum at $x_0 = 0$.

Definition 9.23. Let $I \subset \mathbb{R}$ be an interval. A function $f\colon I \to \mathbb{R}$ is called **convex** if $\forall x_1, x_2 \in I$ and all λ with $0 \leq \lambda \leq 1$ we have

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2).$$

f is called **concave** if $-f$ is convex.



Theorem 9.24. *Let $f\colon I \to \mathbb{R}$ be a twice differentiable function on an interval. f is convex if and only if $f''(x) \geq 0$ $\forall x \in I$.*

Proof. Suppose $f''(x) \geq 0 \; \forall x \in I$. Then the derivative f' is monotone increasing. We assume $x_1 < x_2$ and consider $\overline{x} = \lambda x_1 + (1-\lambda)x_2$ for a fixed $\lambda \in]0,1[$. By the mean value theorem there exist ξ_1, ξ_2 with $x_1 < \xi_1 < \overline{x} < \xi_2 < x_2$ and

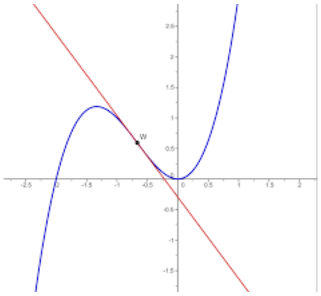
$$\frac{f(\overline{x}) - f(x_1)}{\overline{x} - x_1} = f'(\xi_1) \leq f'(\xi_2) = \frac{f(x_2) - f(\overline{x})}{x_2 - \overline{x}}.$$

Since $\overline{x} - x_1 = (1-\lambda)(x_2 - x_1)$ and $x_2 - \overline{x} = \lambda(x_2 - x_1)$ holds, the inequality above is equivalent to

$$\frac{f(\overline{x}) - f(x_1)}{1-\lambda} \leq \frac{f(x_2) - f(\overline{x})}{\lambda} \iff f(\overline{x}) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

as desired. We leave the other direction as an exercise. □

Remark 9.25. If $f\colon I \to \mathbb{R}$ is tree times differentiable, then points $x_0 \in I$ with $f'''(x_0) = 0$ are called a **flex** of f. Typically but not always flexes are points where f'' changes sign. If this is the case, then the function turns from convex to concave or from concave to convex at x_0.



9.3 Newton method

The result above allows to make curve discussions for sufficiently often differentiable functions f. One issue, which we have not discussed, is how to find zeroes of the functions f and f'. The Newton method gives an answer.

Consider a point x_k with a non horizontal tangent. Then

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

is the intersection point of a tangent line with the x-axis.

Theorem 9.26 (Newton method)**.** *Let $f\colon [a,b] \to \mathbb{R}$ be a two times differentiable convex function with $f(a) < 0$ and $f(b) > 0$. Then the following holds*

1. *There is a unique zero $\xi \in [a,b]$ of f.*
2. *If $x_0 \in [a,b]$ is an arbitrary start value with $f(x_0) \geq 0$, then the sequence (x_n) with $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ converges monotonously decreasing to ξ.*
3. *If $f'(x) \geq c > 0$ and $f''(x) < k$ $\forall x \in [\xi, b]$, then we have the following estimate*
$$|x_{n+1} - x_n| \leq |\xi - x_n| \leq \frac{k}{2c}|x_n - x_{n-1}|^2.$$
*Hence the Newton method converges **quadratically**.*

Remark 9.27. For $f(x) = x^2 - a$ the Newton method recovers our algorithm to approximate the square root of a:

$$x_{k+1} = x_k - \frac{x_k^2 - a}{2x_k} = \frac{1}{2}(x_k + \frac{a}{x_k}).$$

Proof. 1.) Suppose f has two zeroes $\xi_1 < \xi_2$ then by the Rolle's theorem f' would have a zero $\xi_3 \in]\xi_1, \xi_2[$. Since f' is monotonously increasing, we get $f'(x) < 0$ for $x \leq \xi_3$ and $f(x) \geq 0$ for $x \leq \xi_1$. This contradicts $f(a) < 0$.

2.) We prove by induction on n that $\xi \leq x_{n+1} \leq x_n$ holds. First $x_n \geq \xi$ implies $f'(x_n) \geq f'(\xi)$ and $f(x_n) \geq 0$. Thus $\frac{f(x_n)}{f'(x_n)} \geq 0$ and $x_{n+1} \leq x_n$. For the second inequality we note that $f'(x) \leq f'(x_n)$ for all $x \in [\xi, x_n]$. Hence

$$f(x_n) = f(x_n) - f(\xi) \leq f'(x_n)(x_n - \xi) \iff \xi \leq x_{n+1}$$

follows from Corollary 9.18.

Hence (x_n) is a bounded decreasing sequence. Let $\overline{x}$ be the limit. Then

$$\overline{x} = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty}(x_n - \frac{f(x_n)}{f'(x_n)}) = \overline{x} - \frac{f(\overline{x})}{f'(\overline{x})}$$

since f and f' are continuous. Thus $f(\overline{x}) = 0$ and $\overline{x} = \xi$ by 1.)

3.) By step 2.) we already know $|x_{n+1} - x_n| \leq |\xi - x_n|$. Since $f'(x) \geq c > 0$ $\forall x \in [\xi, b]$, we obtain

$$x_n - \xi \leq \frac{f(x_n)}{c}$$

from the mean value theorem. To estimate $f(x_n)$ we consider

$$\psi(x) = f(x) - f(x_{n-1}) - f'(x_{n-1})(x - x_{n-1}) - \frac{k}{2}(x - x_{n-1})^2.$$

Then

$$\begin{aligned} \psi'(x) &= f'(x) - f'(x_{n-1}) - k(x - x_{n-1}) \\ \psi''(x) &= f''(x) - k \leq 0 \; \forall x \in]\xi, b[. \end{aligned}$$

So ψ' is monotone decreasing. Since $\psi'(x_{n-1}) = 0$, we conclude :

$$\psi'(x) \geq 0 \; \forall x \in]\xi, x_{n-1}[.$$

Next we use $\psi(x_{n-1}) = 0$ to deduce $\psi(x) \leq 0 \; \forall x \in]\xi, x_{n-1}[$. In particular $\psi(x_n) \leq 0$, i.e., $f(x_n) \leq \frac{k}{2}(x_n - x_{n-1})^2$ because $f(x_{n-1}) + f'(x_{n-1})(x_n - x_{n-1}) = 0$. This proves

$$|x_{n+1} - x_n| \leq |\xi - x_n| \leq \frac{k}{2c}(x_n - x_{n-1})^2.$$

□

10 Special functions

10.1 The exponential function

In section 7 we introduced the exponential function with a power series 7.2:

$$e^x = \exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

and established $e^{x_1+x_2} = e^{x_1}e^{x_2}$, $e^{-x} = \frac{1}{e^x}$, $e = \exp(1)$. What is the derivative of exp?

$$\frac{\exp(x)-1}{x-0} = \sum_{n=1}^{\infty} \frac{x^{n-1}}{n!} \quad \xrightarrow[x\to 0]{} \quad ?$$

Lemma 10.1 (Error term of the exponential series)**.** *For* $N \in \mathbb{N}$ *we define* $r_{N+1}(x)$ *by the identity*

$$\exp(x) = \sum_{n=0}^{N} \frac{x^n}{n!} + r_{N+1}(x).$$

Then the estimate

$$|r_{N+1}(x)| \leq 2\frac{|x|^{N+1}}{(N+1)!} \quad \textit{for} \ \ |x| \leq 1 + \frac{N}{2}$$

holds.

Proof. By definition we have $r_{N+1}(x) = \sum_{n=N+1}^{\infty} \frac{x^n}{n!}$. Hence

$$\begin{aligned}
|r_{N+1}(x)| &\leq \sum_{n=N+1}^{\infty} \frac{|x^n|}{n!} \\
&= \frac{|x|^{N+1}}{(N+1)!}\left(1 + \frac{|x|}{N+2} + \frac{|x|^2}{(N+2)(N+3)} + \cdots\right) \\
&\leq \frac{|x|^{N+1}}{(N+1)!} \cdot \sum_{k=0}^{\infty} \left(\frac{|x|}{N+2}\right)^k \\
&\leq \frac{|x|^{N+1}}{(N+1)!} \cdot \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k = \frac{|x|^{N+1}}{(N+1)!} \cdot \frac{1}{1-\frac{1}{2}} \\
&= 2 \cdot \frac{|x|^{N+1}}{(N+1)!},
\end{aligned}$$

□

Theorem 10.2. *The exponential function* $\exp : \mathbb{R} \to \mathbb{R}$ *is differentiable with* $\exp'(x) = \exp(x)$.

Proof. We start by showing $\exp'(0) = 1$ using the error term: Since

$$0 \leq |\frac{r_{N+1}(x)}{x}| \leq \frac{1}{2}\frac{|x|^N}{(N+1)!} \quad \xrightarrow[x\to 0]{} \quad 0$$

holds, we get

$$\exp'(0) = \lim_{x\to 0}\frac{\exp(x)-\exp(0)}{x} = \lim_{x\to 0}(\sum_{n=1}^{N}\frac{x^{n-1}}{n!}) + 0 = 1.$$

For general x_0 we use the addition theorem:

$$\frac{\exp(x_0+h)-\exp(x_0)}{h} = \exp(x_0)\cdot\frac{\exp(h)-1}{h} \quad \xrightarrow[h\to 0]{} \quad \exp(x_0)\cdot 1 = \exp(x_0).$$

□

Corollary 10.3. *The exponential function* $\exp$ *is strictly monotone increasing and convex.*

The reason for the frequent occurring of the exponential function in nature is that $y = e^{cx}$ is a solution of the differential equation $y' = cy$. More precisely:

Proposition 10.4. *Let* $f\colon I \to \mathbb{R}$ *be a differential function on an interval satisfying* $f' = cf$. *Then*

$$f(x) = f(x_0)e^{c(x-x_0)},$$

where $x_0 \in I$ *is an arbitrary fixed point.*

Proof. Consider the function $h(x) = f(x)\cdot e^{-cx}$. Then

$$h'(x) = f'(x)e^{-cx} + f(x)(-c)e^{-cx} = cf(x)e^{-cx} - cf(x)e^{-cx} = 0.$$

Thus h is a constant function and

$$h(x) = f(x)e^{-cx} = h(x_0) = f(x_0)e^{-cx_0},$$

yields $f(x) = f(x_0)e^{c(x-x_0)}$. □

The map $\exp\colon \mathbb{R} \to \mathbb{R}_{>0}$ is bijektiv and is a so called *isomorphism of groups* $(\mathbb{R},+) \to (\mathbb{R}_{>0},\cdot)$ because $\exp(x+y) = \exp(x)\cdot\exp(y)$ for all $x, y \in \mathbb{R}$. Groups will be a topic in the next term.

10.2 The Logarithm

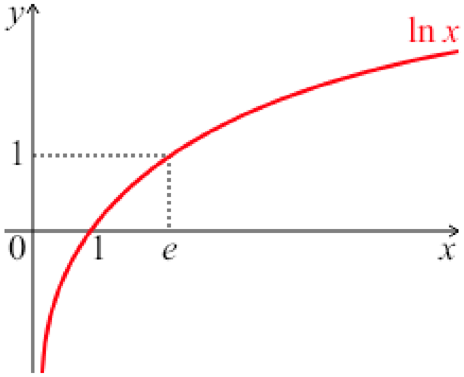
Definition 10.5. The inverse function

$$\ln\colon \mathbb{R}_{>0} \to \mathbb{R}$$

of $\exp$ is called the **natural logarithm**.

Proposition 10.6 (Properties of the logarithm)**.** *The following holds:*

1. $\ln(x_1 \cdot x_2) = \ln x_1 + \ln x_2$.
2. $\ln$ *ist differentiable with* $(\ln x)' = \frac{1}{x}$.
3. $\ln$ *is concave und monotone increasing*



Proof. 1. This follows from the addition theorem of $\exp$. The *slide rule* relies on this property.

2. The formula of the derivative of an inverse function gives

$$\ln'(x) = \frac{1}{\exp(\ln x)} = \frac{1}{x}.$$

3. Follows from 2.).

$\square$

Definition 10.7. Let $a \in \mathbb{R}_{>0}$ be a positive real number. We define the **exponentiation with base** a by

$$a^x := e^{x \cdot \ln a}.$$

Theorem 10.8. *The following holds:*

1. $a^{x_1+x_2} = a^{x_1} \cdot a^{x_2}$.

2. *If* $x = \frac{p}{q} \in \mathbb{Q}$ *is a rational number, then* $a^x = \sqrt[q]{a^p}$. *Hence* a^x *extends our previous notation* $a^{\frac{p}{q}} = \sqrt[q]{a^p}$.

3. *The function* $x \mapsto a^x$ *is differentiable with* $(a^x)' = \ln a \cdot a^x$.

Proof. The first and last assertion are clear. For 2.) we start considering integral exponents $n \in \mathbb{Z}_{>0}$. We have

$$a^n (= \underbrace{a \cdots a}_{n \text{ times}}) = (e^{\ln a})^n = e^{n \ln a}.$$

For $n \in \mathbb{Z}_{<0}$, hence $k = -n > 0$, we get

$$a^n = \frac{1}{a^k} = \frac{1}{e^{k \ln a}} = e^{-k \ln a} = e^{n \ln a}.$$

For the general case we have to show that $\sqrt[q]{a^p} = e^{\frac{p}{q} \ln a}$ holds. This is equivalent to $a^p \overset{!}{=} (e^{\frac{p}{q} \ln a})^q = e^{p \ln a}$, which we established above. $\square$

Definition 10.9. For $a \in \mathbb{R}_{>0}$ we denote by

$$\log_a \colon \mathbb{R}_{>0} \to \mathbb{R}$$

the inverse function of $x \mapsto a^x$.

Remark 10.10. $\log_a x$ is differentiable with

$$(\log_a)'(x) = \frac{1}{\ln a \cdot a^{\log_a x}} = \frac{1}{x \ln a}.$$

In computer science $\log_2 n$ is important: $\lfloor \log_2 n \rfloor + 1$ is the number of **binary digits** of an integer $n > 0$, i.e., the number of bits needed to specify n.

Functions like $x \mapsto x^x = e^{x \ln x}$ show up in complexity theory: One of the best algorithms known to factor an integer n with $x = \log_2 n$ binary digits has run time $O(e^{\frac{1}{2} x \ln x})$.

10.3 Trigometric functions

We have defined sine and cosine by power series in section 7.

$$\sin(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!} \text{ and } \cos(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}.$$

Moreover, we showed the addition laws and $\sin^2 x + \cos^2 x = 1 \ \forall x \in \mathbb{R}$ in Theorem 7.11

Theorem 10.11. *The functions* $\sin, \cos \colon \mathbb{R} \to \mathbb{R}$ *are differentiable with*

$$\sin' = \cos, \quad \cos' = -\sin .$$

Proof. We first prove $\sin'(0) = 1 = \cos(0)$ and $\cos'(0) = 0 = \sin 0$:

$$\begin{aligned} \sin'(0) &= \lim_{h \to 0} \frac{\sin h - 0}{h} = \lim_{h \to 0} \Big(\sum_{k=0}^{\infty} (-1)^k \frac{h^{2k}}{(2k+1)!} \Big) = 1, \\ \cos'(0) &= \lim_{h \to 0} \frac{\cos h - 1}{h} = \lim_{h \to 0} \Big(\sum_{k=1}^{\infty} (-1)^k \frac{h^{2k-1}}{(2k)!} \Big) = 0 \end{aligned}$$

by using an error estimates similar to the case of the exponential function. For general $x_0 \in \mathbb{R}$ we obtain from the addition law 7.11

$$\frac{\sin(x_0 + h) - \sin(x_0)}{h} = \sin x_0 \cdot \frac{\cos h - 1}{h} + \cos x_0 \cdot \frac{\sin h - 0}{h} \xrightarrow[h \to 0]{} \cos x_0.$$

Similarly for cosine

$$\frac{\cos(x_0 + h) - \cos(x_0)}{h} = \cos x_0 \cdot \frac{\cos h - 1}{h} - \sin x_0 \cdot \frac{\sin h - 1}{h} \xrightarrow[h \to 0]{} -\sin x_0.$$

□

We next define Archimedian's constant π. To define π as the ratio of the diameter to the circumference of a circle or to prove that the unit disc has area π, one needs integration. We will come back to this later.

Proposition 10.12. *The following holds*

1. $\cos(0) = 1$, $\cos(2) < 0$.
2. $\sin(x) > 0$ *for* $x \in]0, 2[$.
3. *The function* $\cos$ *has a unique zero in* $[0, 2]$.

Proof. 1. By the definition as a power series $\cos(0) = 1$. To estimate $\cos(2)$ we note that

$$\frac{2^{2k}}{2k!} > \frac{2^{2k+2}}{(2k+2)!} \iff (2k+1)(2k+2) > 2^2 \iff k \geq 1.$$

Hence

$$-1 = 1 - \frac{2^2}{2} \leq \cos 2 \leq 1 - \frac{2^2}{2} + \frac{2^4}{4!} = 1 - 2 + \frac{16}{24} = -1 + \frac{2}{3} = -\frac{1}{3}.$$

2. We have

$$\frac{x^{2k+1}}{(2k+1)!} > \frac{x^{2k+3}}{(2k+3)!} \iff (2k+2)(2k+3) > x^2.$$

For $k \geq 0$ and $0 \leq x \leq 2$ we have $2k+3 \geq 2k+2 \geq x \geq 0$, hence, $(2k+2)(2k+3) > x^2$. For $x \in]0, 2]$ we get

$$x \geq \sin x \geq x - \frac{x^3}{6} = x \cdot \big(1 - \frac{x^2}{6}\big) \geq x \cdot \big(1 - \frac{2^2}{6}\big) > 0,$$

as desired.

3. Since $\cos$ is continuous and $\cos(1) = 1 > 0$ and $\cos(2) < 0$, there is a zero of $\cos$ in $[0, 2]$. Since $\cos' = -\sin$ is negative on $]0, 2]$, the function $\cos$ is decreasing on $[0, 2]$. Hence this zero is unique.

□

We now define π as follows:

Definition 10.13. We define **Archimedean's constant** π as the unique real number such that $\frac{\pi}{2}$ is the zero of $\cos$ in $[0, 2]$.

Note that

$$\cos\left(\frac{\pi}{2}\right) = 0 \implies \sin\left(\frac{\pi}{2}\right) = 1.$$

Hence the addition laws give the following identities:

Theorem 10.14 (Phase shift of sine and cosine)**.**

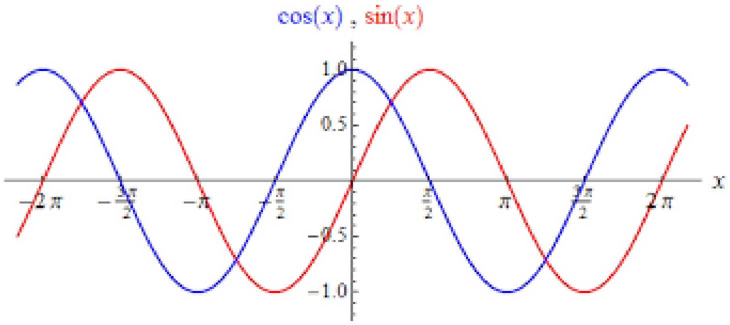
1. $\sin(x + \frac{\pi}{2}) = \cos x,\ \cos(x + \frac{\pi}{2}) = -\sin x.$

2. $\sin(x + \pi) = -\sin x,\ \cos(x + \pi) = -\cos x.$

3. $\sin(x + 2\pi) = \sin x,\ \cos(x + 2\pi) = \cos x.$

□

We say $\sin$ and $\cos$ are 2π-periodic functions. The value of π is $3.1415\ldots$.



Remark 10.15 (Importance of sine und cosine)**.**

1. $[0, 2\pi[\to \mathbb{R}^2,\ t \mapsto (\cos t, \sin t)$ is a parametrization of the unit circle.

2. If f is a solution of the differential equation

$$y'' = -w^2 y,$$

then $f(x) = a\cos(wx) + b\sin(wx)$. Hence sine und cosine occur in oscillations.

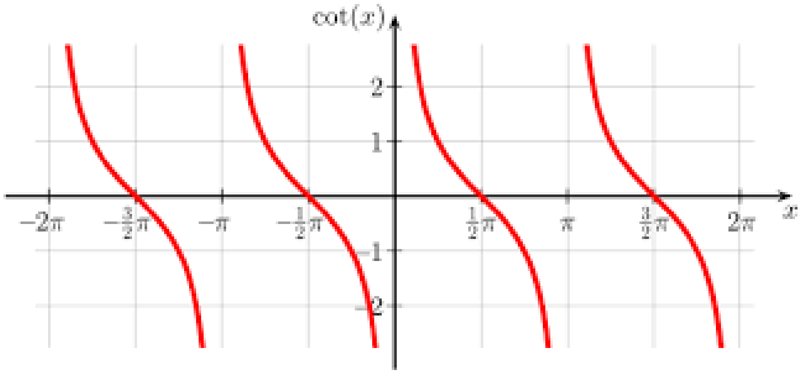
Definition 10.16. The function

$$\tan\colon \mathbb{R}\setminus\{\frac{\pi}{2}+\pi k \mid k\in\mathbb{Z}\}\to\mathbb{R},\ \ x\mapsto \tan x=\frac{\sin x}{\cos x}$$

is called **tangent function**. Its reciprocal

$$\cot\colon \mathbb{R}\setminus\{k\pi \mid k\in\mathbb{Z}\}\to\mathbb{R},\ \ x\mapsto \cot x=\frac{1}{\tan x}=\frac{\cos x}{\sin x}$$

is the **cotangent**.



Definition 10.17.

1. $\tan\colon]-\frac{\pi}{2},\frac{\pi}{2}[\to\mathbb{R}$ is strictly increasing. The inverse function of $\tan$

   $$\arctan\colon \mathbb{R}\ \to\ \left]-\frac{\pi}{2},\frac{\pi}{2}\right[\ \subset\ \mathbb{R}$$

   is called **arc tangent**.

2. The map

   $$\sin\colon \left[-\frac{\pi}{2},\frac{\pi}{2}\right]\ \to\ [-1,1]$$

   is strictly increasing. The inverse function

   $$\arcsin\colon [-1,1]\ \to\ \left[-\frac{\pi}{2},\frac{\pi}{2}\right]$$

   is called **arc sine**.

Theorem 10.18.

1. $\arcsin$ *is differentiable on* $]-\frac{\pi}{2},\frac{\pi}{2}[$ *with*

$$\arcsin'(x) = \frac{1}{\sqrt{1-x^2}}.$$

2. $\arctan$ *is differentiable with*

$$\arctan'(x) = \frac{1}{1+x^2}.$$

Proof.

1. We know $\sin' = \cos$. Hence

$$\begin{aligned}\arcsin'(x) &= \frac{1}{\cos(\arcsin(x))}\\ &= \frac{1}{\sqrt{1-\sin^2(\arcsin(x))}}\\ &= \frac{1}{\sqrt{1-x^2}}.\end{aligned}$$

2. We know $\tan'(x) = \frac{1}{\cos^2(x)}$. Hence

$$\begin{aligned}\arctan'(x) &= \cos^2(\arctan(x))\\ &= \frac{\cos^2(\arctan(x))}{\sin^2(\arctan(x))+\cos^2(\arctan(x))}\\ &= \frac{1}{1+\tan^2(\arctan(x))}\\ &= \frac{1}{1+x^2}.\end{aligned}$$

□

11 Asymptotic behavior and L' Hospital's rule

Theorem 11.1 (L'Hospital's rule)**.** *Let $f, g\colon [a,b] \to \mathbb{R}$ be continuous functions which are differentiable on $]a,b]$ with $g'(x) \neq 0\ \forall x \in\ ]a,b[$ and $f(a) = g(a) = 0$. If the limit $\lim_{x \searrow a} \frac{f'(x)}{g'(x)}$ exists, then the limit $\lim_{x \searrow a} \frac{f(x)}{g(x)}$ also exists and*

$$\lim_{x \searrow a} \frac{f(x)}{g(x)} = \lim_{x \searrow a} \frac{f'(x)}{g'(x)}.$$

holds

Example 11.2. $f(x) = \sin(x)$, $g(x) = e^x - 1$, $[a,b] = [0,1]$. The quotient $\frac{f(0)}{g(0)} = \frac{0}{0}$ makes no sense, but

$$\frac{f'(x)}{g'(x)} = \frac{\cos x}{e^x}$$

is continuous in $x = 0$ with

$$\lim_{x \searrow 0} \frac{f'(x)}{g'(x)} = \frac{\cos 0}{e^0} = \frac{1}{1} = 1.$$

Hence

$$\lim_{x \searrow 0} \frac{\sin x}{e^x - 1} = \lim_{x \searrow 0} \frac{\cos x}{e^x} = \frac{1}{1} = 1.$$

The proof of L'Hospital's rule uses the following variant of the meanvalue theorem.

Lemma 11.3. *Let $f, g\colon [a,b] \to \mathbb{R}$ be continuous functions which are differentiable on $]a,b[$ with $g'(x) \neq 0\ \forall x \in]a,b[$ and $g(a) \neq g(b)$. Then there exists a $\xi \in\ ]a,b[$ such that*

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(\xi)}{g'(\xi)}.$$

Proof. We consider the function

$$h(x) = f(x) - \frac{f(b) - f(a)}{g(b) - g(a)}\big(g(x) - g(a)\big).$$

Since $h(a) = f(a) = h(b)$ holds, Rolle's theorem gives us a $\xi \in\ ]a,b[$ such that:

$$0 = h'(\xi) = f'(\xi) - \frac{f(b) - f(a)}{g(b) - g(a)} g'(\xi).$$

Since $g'(\xi) \neq 0$ by assumption, we deduce

$$\frac{f'(\xi)}{g'(\xi)} = \frac{f(b)-f(a)}{g(b)-g(a)}.$$

□

of L'Hospital's rule. Since $g'(x) \neq 0 \; \forall x \in \,]a,b[$ and $g(a) = 0$, is $g(x) \neq 0 \; \forall x \in \,]a,b[$ by Rolle's theorem. The Lemma gives a $\xi \in \,]a,x[$ with

$$\frac{f(x)}{g(x)} = \frac{f(x)-f(a)}{g(x)-g(a)} = \frac{f'(\xi)}{g'(\xi)}.$$

Since with $x \searrow a$ also $\xi \searrow a$ holds, we obtain

$$\lim_{x \searrow a} \frac{f(x)}{g(x)} = \lim_{\xi \searrow a} \frac{f'(\xi)}{g'(\xi)} = \lim_{x \searrow a} \frac{f'(x)}{g'(x)}.$$

□

There are various variants of L'Hospital's rule. To formulate them we need some notation.

Definition 11.4 (Convergence to ∞)**.** Let $f\colon [a,\infty[\to \mathbb{R}$ be a function. We say, $f(x)$ **approaches** $c \in \mathbb{R}$ for x to ∞, in symbols

$$\lim_{x \to \infty} f(x) = c,$$

if

$$\forall \varepsilon > 0 \; \exists \, N: \; |f(x) - c| < \varepsilon \; \forall x \geq N.$$

We say: $\lim_{x \to \infty} f(x) = \infty$, if

$$\forall M > 0 \; \exists \, N > 0: \; f(x) > M \; \forall x > N.$$

For $f\colon]a,b] \to \mathbb{R}$ we write $\lim_{x \searrow a} f(x) = \infty$ if

$$\forall M > 0 \; \exists \, \delta > 0: \; f(x) > M \; \forall x > a \text{ with } |x - a| < \delta.$$

Similarly we define $\lim_{x \to -\infty} f(x) = c$ or for example $\lim_{x \nearrow b} f(x) = -\infty$.

Theorem 11.5 (Variants of L'Hospital's rule)**.**

1. *Let* $f, g\colon]a,b[\to \mathbb{R}$ *be differentiable functions with* $g'(x) \neq 0 \; \forall x \in]a,b[$ *and*
$$\lim_{x \searrow a} f(x) = \infty = \lim_{x \searrow a} g(x).$$
If the limit $\lim_{x \searrow a} \frac{f'(x)}{g'(x)}$ *exist, then* $\lim_{x \searrow a} \frac{f(x)}{g(x)}$ *exists and*
$$\lim_{x \searrow a} \frac{f(x)}{g(x)} = \lim_{x \searrow a} \frac{f'(x)}{g'(x)}.$$

2. *Let* $f, g\colon [a,\infty[\to \mathbb{R}$ *be differentiable functions with* $g'(x) \neq 0 \; \forall x \in [a,\infty[$ *and*
$$\lim_{x \to \infty} f(x) = 0 = \lim_{x \to \infty} g(x)$$
or
$$\lim_{x \to \infty} f(x) = \infty = \lim_{x \to \infty} g(x).$$
If the limit $\lim_{x \to \infty} \frac{f'(x)}{g'(x)}$ *exists, then* $\lim_{x \to \infty} \frac{f(x)}{g(x)}$ *exists and*
$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = \lim_{x \to \infty} \frac{f'(x)}{g'(x)}.$$

Example 11.6. We show for all $n \in \mathbb{N}$
$$\lim_{x \to \infty} \frac{x^n}{e^x} = 0.$$
Since $\lim_{x \to \infty} x^n = \lim_{x \to \infty} e^x = \infty$, we can try to apply L'Hospital's rule.
$$\lim_{x \to \infty} \frac{x^n}{e^x} = \lim_{x \to \infty} \frac{nx^{n-1}}{e^x} = \cdots = n! \lim_{x \to \infty} \frac{1}{e^x} = 0.$$
Hence the exponential function grows faster than any polynomial.

The O-notion will also be used for functions:

Definition 11.7 (O– and o– notation for functions)**.** Let $f, g\colon [a,\infty[\to \mathbb{R}$ be functions. We write
$$f \in O(g) \text{ for } x \to \infty,$$
if $\exists\, c > 0 \; \exists\, M$, such that $|f(x)| \leq c \cdot g(x) \; \forall x \geq M$, and say f **lies in big** O **of** g. We say $f \in o(g)$, f **lies in small** o **of** g, ifs $\lim_{x \to \infty} \frac{f(x)}{g(x)} = 0$.

Example 11.8.

1. $x^n \in o(e^x)$ for $x \to \infty$ for every $n \in \mathbb{N}$ as shown above.

2. Let $f(x) = a_n x^n + \cdots + a_0 \in \mathbb{R}[x]$ be a polynomial of degree n. Then $f(x) \in O(x^n)$ for $x \to \infty$. More precisely: For every $C = |a_n| + \varepsilon, \varepsilon > 0$, $\exists\, M > 0$, such that
$$|f(x)| \leq C \cdot x^n \; \forall x \geq M.$$

11.1 Asymptotic behaviour of rational functions

Let $h(x) = \frac{f(x)}{g(x)}$ be a rational function with polynomials

$$\begin{aligned} f(x) &= a_n x^n + \cdots + a_0, \\ g(x) &= b_m x^m + \cdots + b_0, \end{aligned}$$

of degree n and m respectively, i.e., $a_n, b_m \neq 0$. L'Hospital's rule gives

$$\begin{aligned} \lim_{x\to\infty} \frac{f(x)}{g(x)} &= 0, \quad \text{if } n < m, \\ \lim_{x\to\infty} \frac{f(x)}{g(x)} &= \frac{a_n}{b_m}, \quad \text{if } n = m, \\ \lim_{x\to\infty} \frac{f(x)}{g(x)} &= \begin{cases} +\infty, & \text{if } n > m \text{ and } \frac{a_n}{b_m} > 0, \\ -\infty, & \text{if } n > m \text{ and } \frac{a_n}{b_m} < 0. \end{cases} \end{aligned}$$

Then the last case can be refined. We denote by

$$\mathbb{R}[x] = \{f : \mathbb{R} \to \mathbb{R} \mid \exists \text{ coefficients } a_0 \ldots, a_n \in \mathbb{R} : f(x) = a_n x^n + \ldots + a_0\}$$

the ring of polynomial functions.

Theorem 11.9 (Division with remainder)**.** *Let $f, g \in \mathbb{R}[x]$ be polynomials in one variable x with real coefficients. Then there exist uniquely determined polynomials $q, r \in \mathbb{R}[x]$, such that*

$$f = q \cdot g + r \quad \textit{and} \quad \deg r < \deg g.$$

Proof. Existence: If $\deg f < \deg g$ then we can choose $q = 0$ and $r = f$. If $n = \deg f \geq \deg g = m$, say

$$f(x) = a_n x^n + \cdots + a_0, \quad g(x) = b_m x^m + \cdots + b_0$$

with $a_n \neq 0 \neq b_m$, we consider

$$q_0(x) := \frac{a_n}{b_m} x^{n-m} \text{ and } f_1 := f - q_0 \cdot g.$$

Then $\deg f_1 < \deg f$. Inductively, we may assume that an expression $f_1 = q_1 g + r_1$ exists. Then

$$f = f_1 + q_0 \cdot g = (q_0 + q_1) \cdot g + r_1.$$

is the desired expression for f.
Uniqueness: Suppose $f = q \cdot g + r$ and $f = \tilde{q} \cdot g + \tilde{r}$ with $\deg g > \deg r$, $\deg g > \deg \tilde{r}$, are two different expressions. Then we must have $q \neq \tilde{q}$ and

$$0 = (\underbrace{q - \tilde{q}}_{=:\bar{q}}) \cdot g + (\underbrace{r - \tilde{r}}_{=:\bar{r}}) \text{ with } \bar{q} \neq 0.$$

Then

$$\deg(\bar{q} \cdot g) = \deg(\bar{q}) + \deg g \geq 0 + \deg g > \deg(\bar{r}),$$

so $\bar{q} \cdot g + \bar{r} \neq 0$, a contradiction. □

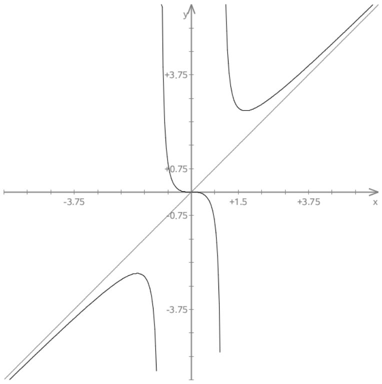
Example 11.10. Consider

$$h(x) = \frac{f(x)}{g(x)} = \frac{x^3}{x^2 - 1}.$$

Long division gives

$$\begin{array}{l} x^3 : (x^2 - 1) = x + \frac{x}{x^2-1}, \\ \underline{x^3 - x} \\ \qquad x \end{array}$$

i.e., $q(x) = x$, $r(x) = x$. Hence $h(x) \in x + o(1)$, i.e., asymptotically $h(x)$ has the same behaviour as x. Near 0 the function h is very different from $x \mapsto x$. For example, h has poles at $x = \pm 1$ and a saddle point at 0



In general we have

Corollary 11.11. *Let* $h(x) = \frac{f(x)}{g(x)}$ *be a rational function and*

$$f(x) = q(x) \cdot g(x) + r(x)$$

with $\deg r < \deg g$*. Then* h *behaves like* $q(x)$ *for* $x \to \pm\infty$ *, more precisely*

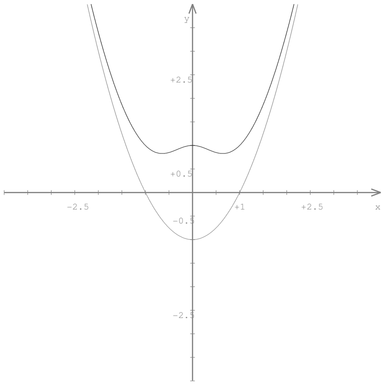
$$h(x) \in q(x) + o(1).$$

□

Example 11.12. We consider the rational function $h(x) = \frac{x^4+1}{x^2+1}$. Long division gives

$$\begin{array}{l} (x^4+1) : (x^2+1) = x^2 - 1 + \frac{2}{x^2+1}, \\ \underline{x^4 + x^2} \\ \quad - x^2 + 1 \\ \quad \underline{- x^2 - 1} \\ \qquad\qquad 2 \end{array}$$

i.e., $q = x^2 - 1$ and $r = 2$. Since $x^2 + 1$ has no zero in $\mathbb{R}$, the function $h(x)$ has no poles. To draw the graph, we compute the extrema of $h(x)$.



$$h'(x) = \frac{(x^2+1) \cdot 4 \cdot x^3 - 2 \cdot x \cdot (x^4+1)}{(x^2+1)^2} = \frac{2x^5+4x^3-2x}{(x^2+1)^2}.$$

Thus the extrema satisfy the equation $2x^5 + 4x^3 - 2x = 0$, i.e., $x = 0$ or $x^4 + 2x^2 - 1 = 0$. Substituting $z = x^2$ we get a quadratic equation $z^2 + 2z - 1 = 0$, which has the solutions

$$z_{1,2} = -1 \pm \sqrt{1+1}.$$

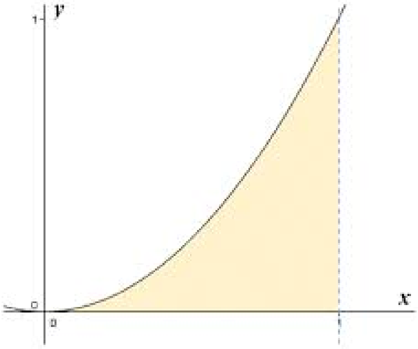
by the p, q–formula. Since $-1 - \sqrt{2} < 0$, only one of these solutions lead to solutions for $x^2 = z$. Hence $x = 0$ and

$$x_{1,2} = \pm\sqrt{-1+\sqrt{2}} \approx \pm 0.64.$$

are the extrema. One can check that $x = 0$ is a local maximum and that $x_{1,2}$ are local minima. Since $h(x)$ for $x \to \pm\infty$ behaves like $q(x) = x^2 - 1$, we obtain the graph above.

12 Integration

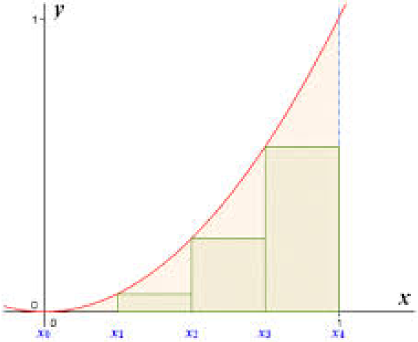
Let $f\colon [a,b] \to \mathbb{R}$ be a function on a closed interval. We want to compute the area under the graph of f.



The basic idea is to use an approximation by staircase functions.

Definition 12.1. A **staircase function** $\varphi\colon [a,b] \to \mathbb{R}$ is a function, such that there exists a subdivision $a = t_0 < t_1 < \cdots < t_n = b$ of the interval $[a,b]$, such that φ is constant on the open intervals $]t_{i-1}, t_i[$, i.e., for each $i \in \{1,\ldots,n\}$ there is a $c_i \in \mathbb{R}$, such that

$$\varphi|_{]t_{i-1},t_i[}\colon\ ]t_{i-1},t_i[\ \to \mathbb{R}, \quad \varphi|_{]t_{i-1},t_i[}(x) = \varphi(x) = c_i.$$



For $\varphi(t_i)$ we require nothing. The **integral of the staircase function** is

$$\int_a^b \varphi(x)\,dx := \sum_{i=1}^{n} c_i(t_i - t_{i-1}).$$

Each sum of this kind is called a **Riemann sum**. Using these, the integral of an arbitrary bounded function $f\colon [a,b] \to \mathbb{R}$ is defined as follows: The **upper integral** of f is

$$\int_a^{*b} f := \inf \Big\{ \int_a^b \psi \, dx \mid \psi \geq f,\ \psi \text{ a staircase function} \Big\},$$

and the **lower integral** is

$$\int_{*a}^{b} f := \sup \Big\{ \int_a^b \varphi \, dx \mid \varphi \leq f,\ \varphi \text{ a staircase function} \Big\}.$$

f is integrable (more precisely: **Riemann-integrable**), if

$$\int_a^{*b} f = \int_{*a}^{b} f$$

holds. If this is the case, then we define the **integral of the bounded function** $f\colon [a,b] \to \mathbb{R}$ by

$$\int_a^b f(x)\, dx := \int_a^{*b} f = \int_{*a}^{b} f.$$

Example 12.2. Staircase functions are integrable.

Theorem 12.3. *Monotone functions $f\colon [a,b] \to \mathbb{R}$ are integrable.*

Proof. Let $f\colon [a,b] \to \mathbb{R}$ be monotone increasing and $\varepsilon > 0$ given. We choose n so large that

$$\varepsilon > \frac{1}{n} \cdot (b-a) \cdot \big(f(b) - f(a)\big),$$

holds, take $h = \frac{b-a}{n}$ and consider the subdivision

$$t_i = a + i \cdot h,\ i = 0, \dots, n.$$

The staircase functions φ, ψ with

$$\varphi|_{[t_{i-1}, t_i[} = f(t_{i-1}), \quad \psi|_{[t_{i-1}, t_i[} = f(t_i)$$

and $\varphi(b) = \psi(b) = f(b)$ satisfy $\varphi \leq f \leq \psi$ because of the monotonicity. On the other hand:

$$\begin{aligned} \int \psi \, dx - \int \varphi \, dx &= \sum_{i=1}^{n} \big(f(t_i) - f(t_{i-1})\big) \cdot h \\ &= h \cdot \big(f(b) - f(a)\big) = \frac{b-a}{n} \cdot \big(f(b) - f(a)\big) < \varepsilon. \end{aligned}$$

Hence

$$0 \leq \int_a^{*b} f - \int_{*a}^b f < \varepsilon$$

for arbitrary small $\varepsilon > 0$. Thus upper and lower integral coincide. □

Example 12.4. Suppose $0 \leq b$. What is $\int_0^b x^2\,dx$?
$f(x) = x^2$ is monotone on $[0,b]$, hence the integral exists. We consider the **equidistant subdivision**

$$t_i = i \cdot \frac{b-0}{n} = ih, \quad i = 0, \dots, n,$$

of the interval and the staircase function ψ with $\psi|_{]t_{i-1},t_i[} = f(t_i)$. Then

$$\int_0^b \psi\,dx = \sum_{i=1}^n i^2 \cdot h^2 \cdot h = \frac{n(n+1)(2n+1)}{6} \cdot \frac{b^3}{n^3} \quad \underset{n\to\infty}{\longrightarrow} \quad \frac{b^3}{3}.$$

Hence $\int_0^b x^2\,dx = \frac{b^3}{3}$.

Example 12.5. The function

$$f\colon [0,1] \to \mathbb{R},\ f(x) = \begin{cases} 1, & \text{for } x \in \mathbb{Q}, \\ 0, & \text{for } x \notin \mathbb{Q}, \end{cases}$$

is not Riemann integrable. For each pair φ, ψ of staircase functions with $\varphi \leq f \leq \psi$ we have

$$\int_0^1 \varphi(x)\,dx \leq 0, \qquad \int_0^1 \psi(x)\,dx \geq 1,$$

since each interval $]t_{i-1}, t_i[$ contains rational and irrational points.

Theorem 12.6 (Integrability of continuous function)**.** *Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function on a bounded closed interval. Then f is integrable on $[a,b]$.*

For the proof we need more than pointwise continuity.

Definition 12.7. A function $f\colon I \to \mathbb{R}$ is **uniformly continuous** on I if $\forall \varepsilon > 0$ $\exists\, \delta > 0$, such that

$$|f(x_1) - f(x_0)| < \varepsilon\ \forall\ x_0, x_1 \in I \text{ with } |x_1 - x_0| < \delta.$$

The crucial difference to pointwise continuity is that δ does not depend on x_0.

Example 12.8. The functions

$$f\colon \mathbb{R}_{>0} \to \mathbb{R},\ f(x) = \frac{1}{x} \text{ and } g\colon \mathbb{R} \to \mathbb{R},\ g(x) = x^2$$

are continuous but not uniformly continuous.

Theorem 12.9. *Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function on a bounded closed interval. Then f is uniformly continuous.*

Proof. Suppose f is not uniformly continuous. Then there exists an $\varepsilon > 0$, a zero sequence (δ_n) and sequences (x_n), (y_n), such that

$$|f(x_n) - f(y_n)| \geq \varepsilon, \text{ although } |x_n - y_n| < \delta_n.$$

By Bolzano–Weierstrass the sequence (x_n) has a convergent subsequence (x_{n_k}). Let

$$x_0 = \lim_{k\to\infty} x_{n_k} \in [a,b]$$

be the limit. Since f is continuous in x_0, there exists a $\delta > 0$, such that

$$|f(x) - f(x_0)| < \frac{\varepsilon}{2} \text{ for } x \in [a,b] \text{ with } |x - x_0| < \delta.$$

Take k large enough such that $|x_{n_k} - x_0| < \frac{\delta}{2}$ and $\delta_{n_k} < \frac{\delta}{2}$. Then

$$|y_{n_k} - x_0| \leq |y_{n_k} - x_{n_k}| + |x_{n_k} - x_0| \leq \delta_{n_k} + \frac{\delta}{2} < \delta$$

and

$$|f(x_{n_k}) - f(y_{n_k})| \leq |f(x_{n_k}) - f(x_0)| + |f(y_{n_k}) - f(x_0)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

contradicting $|f(x_n) - f(y_n)| \geq \varepsilon\ \forall n$. □

Proof of Theorem 12.6. Since $f\colon [a,b] \to \mathbb{R}$ is continuous, it is uniformly continuous by Theorem 12.9. Let $\varepsilon > 0$ be given. We will construct staircase functions

$$\varphi \leq f \leq \psi \ \text{ with } \int_a^b (\psi - \varphi)\, dx < \varepsilon.$$

For $\frac{\varepsilon}{b-a} > 0$ we choose $\delta > 0$, such that

$$|f(x) - f(y)| < \frac{\varepsilon}{b-a} \,\forall\, x, y \in [a, b] \text{ with } |x - y| < \delta.$$

Choose n large enough such that $h = \frac{b-a}{n} < \delta$. Take $t_i = a + i \cdot h$ and staircase functions ψ, φ with $\varphi|_{[t_{i-1},t_i[} = \min\{f(x) \mid x \in [t_{i-1}, t_i]\}$ and $\psi|_{t_{i-1},t_i[} = \max\{f(x) \mid x \in [t_{i-1}, t_i]\}$. Since f has a maximum and minimum on $[t_{i-1}, t_i]$ and since $h < \delta$ holds, we get

$$0 \leq \psi(x) - \varphi(x) < \frac{\varepsilon}{b-a} \implies 0 \leq \int_a^b (\psi(x) - \varphi(x))\, dx < \frac{\varepsilon}{b-a} \cdot (b - a) = \varepsilon.$$

□

Theorem 12.10 (Properties of the integrals)**.** *Let $f, g\colon [a, b] \to \mathbb{R}$ be integrable functions, and let $c \in \mathbb{R}$ be a constant. Then the following holds:*

1. ***(Linearity of the integral)** The functions $c \cdot f$ and $f + g$ are integrable with*

$$\begin{aligned} \int_a^b c \cdot f(x)\, dx &= c \cdot \int_a^b f(x)\, dx, \\ \int_a^b \big(f(x) + g(x)\big)\, dx &= \int_a^b f(x)\, dx + \int_a^b g(x)\, dx. \end{aligned}$$

2. ***(Monotonicity of the integral)** $f \leq g \implies \int_a^b f(x)\, dx \leq \int_a^b g(x)\, dx$.*

3. *With f also the functions $f_+ := \max(f, 0)$, $f_- := \max(-f, 0)$ and $|f| = f_+ + f_-$ are integrable.*

4. *For $p \in \mathbb{R}_{>0}$ also $|f|^p$ is integrable. In particular the product*

$$f \cdot g = \frac{1}{4}\big(|f + g|^2 - |f - g|^2\big)$$

is integrable.

Proof. 1. If $\varphi \leq f$, $\psi \leq g$ are staircase functions, then $\varphi + \psi$ is a staircase function with $\varphi + \psi \leq f + g$. Hence

$$\int_{a*}^b (f + g) \geq \int_{a*}^b f + \int_{a*}^b g,$$

because on the right hand side we take the supremum only for staircase functions $\varphi + \psi$ as above. Similarly,

$$\int_a^{*b} (f+g) \leq \int_a^{*b} f \, + \int_a^{*b} g.$$

The chain of inequalities

$$\int_{a*}^{b} f \, + \int_{a*}^{b} g \leq \int_{a*}^{b} (f+g) \leq \int_a^{*b} (f+g) \, \leq \int_a^{*b} f \, + \int_a^{*b} g$$

are equalities since f and g are integrable

2. If $f \leq g$, then $(\varphi \leq f \implies \varphi \leq g)$. Hence

   $$\int_* f \, dx \leq \int_* g \, dx.$$

   Similarly $\int^* f \, dx \leq \int^* g \, dx$.

3. If φ is a staircase function so is $\varphi_+ = \max(\varphi, 0)$ and $\varphi \leq f \leq \psi \implies \varphi_+ \leq f_+ \leq \psi_+$. Moreover

   $$0 \leq \int_a^b (\psi_+ - \varphi_+) \, dx \leq \int_a^b (\psi - \varphi) \, dx < \varepsilon$$

   for suitable φ, ψ. Hence f_+ is integrable, and so are $f_- = -(f - f_+)$ and $|f| = f_+ + f_-$ because of the linearity the integral.

4. Since f is bounded, we may assume by 1.) that $0 \leq |f| \leq 1$ holds. For staircase functions $0 \leq \varphi \leq |f| \leq \psi \leq 1$ we obtain

   $$\varphi^p \leq |f|^p \leq \psi^p \text{ and } 0 \leq (\psi^p - \varphi^p) \leq p(\psi - \varphi)$$

   by the mean value theorem applied to the function $x \mapsto x^p$ on the interval $[0,1]$:

   $$\frac{\beta^p - \alpha^p}{\beta - \alpha} = p\xi^{p-1} \leq p \text{ for } [\alpha, \beta] \subset [0,1].$$

   Hence

   $$\int_a^b (\psi^p - \varphi^p) \leq p \int_a^b (\psi - \varphi) \, dx < \varepsilon$$

for suitable φ, ψ. Finally $f \cdot g$ is integrable follows from the results established so far because the formula

$$f \cdot g = \frac{1}{4}\big((f+g)^2 - (f-g)^2\big),$$

holds.

□

Theorem 12.11 (Mean value theorem of integral calculus)**.** *Let $f, g\colon [a,b] \to \mathbb{R}$ be functions, f continuous, g integrable and $g(x) \geq 0\ \forall x$. Then there exists $\xi \in [a,b]$, such that*

$$\int_a^b f(x)g(x)\,dx = f(\xi) \cdot \int_a^b g(x)\,dx.$$

In particular: $\exists\, \xi \in [a,b]$, such that

$$\int_a^b f(x)\,dx = f(\xi) \cdot (b-a).$$

Proof. Set

$$M := \max\{f(x) \mid x \in [a,b]\},\ \ m := \min\{f(x) \mid x \in [a,b]\}.$$

Then

$$mg(x) \leq f(x)g(x) \leq Mg(x),$$

holds because $g \geq 0$. Since the integral is monotone we obtain

$$m\int_a^b g(x)\,dx \ \leq \int_a^b f(x)g(x)\,dx \ \leq\ M\int_a^b g(x)\,dx.$$

If $\int_a^b g(x)\,dx = 0$, then there is nothing more to prove. Otherwise by the intermediate value theorem there exists an ξ with

$$m \leq \frac{\int_a^b f(x)g(x)\,dx}{\int_a^b g(x)\,dx} = f(\xi) \leq M.$$

The assertion follows. The special case is the case $g(x) = 1\ \forall x \in [a,b]$. □

12.1 Antiderivative

The key mathematical discovery of Newton (and Leibniz) was that for $f\colon [a,b] \to \mathbb{R}$ the integrals

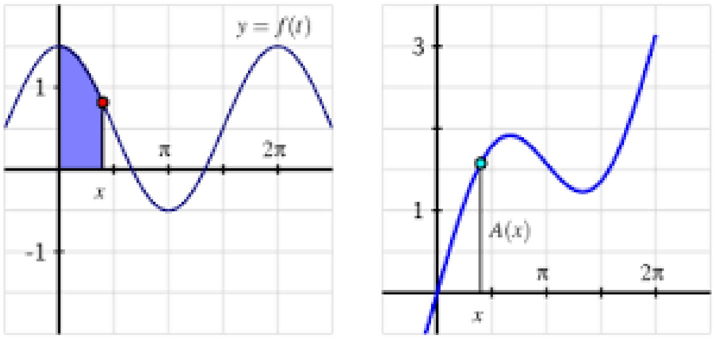
$$\int_a^t f(x)dx$$

for any $t \in [a,b]$ can be computed simultaneously using antiderivatives.

Proposition 12.12. *Let $f\colon [a,b] \to \mathbb{R}$ be a integrable function. Then f is also integrable over each closed subinterval of $[a,b]$ and*

$$\int_a^t f(x)\, dx + \int_t^b f(x)\, dx = \int_a^b f(x)\, dx$$

holds for all $t \in]a,b[$.



Proof. We can refine any subdivisions $a = t_0 < t_1 < \cdots < t_n = b$ of $[a,b]$ for staircase functions such that t is one of the t_i. The result follows. □

Definition 12.13. Let $f\colon [a,b] \to \mathbb{R}$ be an integrable function. We set

$$\int_b^a f(x)\, dx := -\int_a^b f(x)\, dx$$

for interchanged lower and upper boundary points.

Definition 12.14. Let $f\colon I \to \mathbb{R}$ be a continuous function on an interval. A differentiable function $F\colon I \to \mathbb{R}$ is called an antiderivative of f if $F' = f$.

Theorem 12.15 (Fundamental theorem of calculus)**.** *Let $f\colon I \to \mathbb{R}$ be a continuous function on an interval and let $a \in I$ be a point.*

1. *$F\colon I \to \mathbb{R}$ defined by $F(x) = \int_a^x f(t)\,dt$ is a antiderivative of f.*
2. *If $G\colon I \to \mathbb{R}$ is a antiderivative of f, then*

$$\int_a^b f(x)\,dx = G(b) - G(a).$$

for any subinterval $[a,b] \subset I$. Two short hand notations are in use

$$G(x)\,\Big|_a^b := \big[G(x)\big]_a^b := G(b) - G(a).$$

Proof. 1. By the proposition above and the mean value theorem of integral calculus we have

$$F(x) - F(x_0) = \int_{x_0}^x f(t)\,dt = f(\xi)(x - x_0)$$

for a value $\xi \in [x_0, x]$ (or $\xi \in [x, x_0]$ if $x < x_0$). Since with $x \to x_0$ also $\xi \to x_0$ holds, we deduce

$$F'(x_0) = \lim_{x \to x_0} \frac{F(x) - F(x_0)}{x - x_0} = \lim_{\xi \to x_0} f(\xi) = f(x_0).$$

since f is continuous.

2. Since G and F are both antiderivatives of f, they differ by a constant:

$$(G - F)' = f - f = 0 \implies G - F = c$$

for some $c \in \mathbb{R}$ since I is an interval. Hence $G = F + c$ and

$$G(b) - G(a) = F(b) - F(a) = \int_a^b f(x)\, dx$$

holds by the definition of F. □

Definition 12.16. The **indefinite integral** $\int f(x)\, dx$ denotes a antiderivative of f.

Example 12.17. Our computation of derivatives gives the following basic examples.

1. $\int x^\alpha\, dx = \frac{x^{\alpha+1}}{\alpha+1}, \quad \alpha \neq -1.$
2. $\int \frac{1}{x}\, dx = \ln|x|.$
3. $\int e^x\, dx = e^x.$
4. $\int \sin x\, dx = -\cos x.$
5. $\int \cos x\, dx = \sin x.$
6. $\int \frac{1}{1+x^2}\, dx = \arctan x.$
7. $\int \frac{1}{\sqrt{1-x^2}}\, dx = \arcsin x.$

Each differentiation rule gives a rule for computing integrals. The chain rule gives the following:

Theorem 12.18 (Substitution rule)**.** *Let $f\colon I \to \mathbb{R}$ be continuous, $\varphi\colon [a,b] \to I$ continuously differentiable and $\alpha = \varphi(a), \beta = \varphi(b)$. Then the following holds*

$$\int_\alpha^\beta f(x)\, dx = \int_a^b f(\varphi(t)) \cdot \varphi'(t)\, dt.$$

Proof. Let $F(x) = \int f(x)\, dx$ be an antiderivative of f. Then $F \circ \varphi$ is differentiable and

$$(F \circ \varphi)'(t) = F'(\varphi(t)) \cdot \varphi'(t) = f(\varphi(t)) \cdot \varphi'(t)$$

by the chain rule. Hence $F \circ \varphi$ is a antiderivative of $(f \circ \varphi) \cdot \varphi'$ and:

$$\int_a^b f(\varphi(t))\varphi'(t)\, dt = F(\varphi(b)) - F(\varphi(a)) = F(\beta) - F(\alpha) = \int_\alpha^\beta f(x)\, dx.$$

□

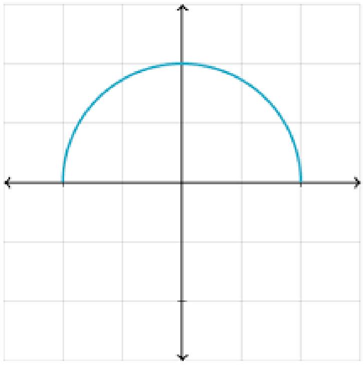
Example 12.19. A frequent application of the substitution rule is the following.

1. let $g\colon [a,b] \to \mathbb{R}$ continuously differentiable with $g(t) \neq 0 \; \forall t$. Then
$$\int \frac{g'(t)}{g(t)} \, dt = \ln |g(t)|.$$
With the notation of the substitution rule we have $f(x) = \frac{1}{x}$ and $\varphi = g$.
We check this directly: We may assume $g > 0$, i.e., $g(t) > 0 \; \forall t$ by the intermediate value theorem. If not, we could consider $-g$ with $(-g)' = -g'$. The chain rule gives
$$(\ln(g(t))' = \frac{1}{g(t)} \cdot g'(t).$$

2. For the tangent function this gives
$$\int \tan x \, dx = \int \frac{\sin x}{\cos x} \, dx = -\int \frac{-\sin x}{\cos x} \, dx = -\ln |\cos x|.$$

Remark 12.20. A way to memorise the substitution rule is to use Leibniz notation:
$$x = \varphi(t), \quad \frac{dx}{dt} = \varphi'(t),$$
hence "$dx = \varphi'(t) \, dt$". We cannot give a precise definition what "dx" mathematically is without more theory, but it gives the right result: If we replace x by $\varphi(t)$ and dx by $\varphi'(t) \, dt$ in $F(x) = \int f(x) \, dx$, then we obtain
$$F(\varphi(t)) = \int f(\varphi(t)) \cdot \varphi'(t) \, dt.$$

Example 12.21. We prove
$$\frac{\pi}{2} = \int_{-1}^{1} \sqrt{1 - x^2} \, dx$$
establishing that the circle with radius 1 has indeed area π.



Substituting $x = \sin t$, $dx = \cos(t)dt$ gives

$$\begin{aligned} \int_{-1}^{1} \sqrt{1-x^2}\, dx &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{1-\sin^2 t} \cdot \cos(t)dt \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2(t)dt \\ &= \frac{1}{2}(t + \sin t \cos t)|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = \frac{\pi}{2} \end{aligned}$$

since $\sqrt{1-\sin^2 t} = \cos^2 t$ on $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and

$$\int \cos^2(t)dt = \frac{1}{2}(t + \sin t \cos t).$$

From the product rule $(fg)' = f'g + fg'$ one deduces

Theorem 12.22 (Integration by parts)**.** *Let $f, g \colon I \to \mathbb{R}$ be continuously differentiable functions. Then*

$$\int f'(x)g(x)\, dx = f(x)g(x) - \int f(x)g'(x)\, dx$$

and

$$\int_a^b f'(x)g(x)\, dx = f(x)g(x)\,\big|_a^b - \int_a^b f(x)g'(x)\, dx$$

holds. □

Example 12.23.

1. $$\int_0^\pi x\sin x\,dx = (-x\cos x)\,\Big|_0^\pi + \sin x\,\Big|_0^\pi = -\pi\cdot(-1) = \pi,$$

   since in integration by parts we may choose $g(x) = x,\ g'(x) = 1$ and $f'(x) = \sin x$, i.e., $f(x) = -\cos x$.

2. To calculate $\int e^{-x}\sin x\,dx$ we use integration by parts twice. With $f'(x) = e^{-x}$, i.e., $f(x) = -e^{-x}$, $g(x) = \sin x$, $g'(x) = \cos x$, gives

   $$\int e^{-x}\sin x\,dx = -e^{-x}\sin x + \int e^{-x}\cos x\,dx.$$

   For the last term we apply integration by parts again

   $$\int e^{-x}\cos x\,dx = -e^{-x}\cos x - \int(-e^{-x})(-\sin x)\,dx.$$

   Hence

   $$2\int e^{-x}\sin x\,dx = -e^{-x}(\sin x + \cos x),$$

   and

   $$\int e^{-x}\sin x\,dx = -\frac{1}{2}e^{-x}(\sin x + \cos x).$$

   To verify that we did not make for example any sign mistake we check this formula:

   $$\big(-\frac{1}{2}e^{-x}(\sin x + \cos x)\big)' = \frac{1}{2}e^{-x}(\sin x + \cos x) - \frac{1}{2}e^{-x}(\cos x - \sin x),$$

   which equals $e^{-x}\sin x$ indeed.

3. We show that $\int_0^{2\pi}\sin^2 x\,dx = \pi$. Set $f(x) = \sin x$, $g' = \sin x$. Then $g = -\cos x$, $f' = -\cos x$ and we get:

   $$\int \sin^2 x\,dx = -\sin x\cos x + \int\cos^2 x\,dx.$$

   Since $\cos^2 x = 1 - \sin^2 x$ holds, we deduce

   $$\int\sin^2 x\,dx = \frac{1}{2}(x - \sin x\cos x).$$

   Evaluating at the boundary points 0 and 2π gives the claim.

4. Similarly, $\int_0^{2\pi} \sin x \cos x \; dx = \frac{1}{2} \sin^2 x \Big|_0^{2\pi} = 0$ because $f(x) = \sin x$ and $g'(x) = \cos x$ gives

$$\int \sin x \cos x \; dx = \sin^2 x - \int \cos x \sin x dx.$$

To give a closed formula for indefinite integrals can be difficult.

Definition 12.24. The set of **elementary functions** is the smallest set of functions satisfying

1. x^n, $\sin x$, $\tan x$, e^x and their inverse functions are elementary.
2. Sums, products and quotients of elementary functions are elementary.
3. Compositions of elementary functions are elementary.

Theorem 12.25 (Liouville's theorem)**.** *Not every elementary function has an elementary antiderivative.*

The proof of this theorem is far beyond the material of this course. Explicit examples are the functions e^{-x^2} and $\frac{1}{\sqrt{x^3-x}}$.

On the positive side one has

Theorem 12.26. *Rational functions have elementary antiderivatives.*

Example 12.27. We compute

$$\int \frac{1}{1-x^2} \; dx.$$

The function $\frac{1}{1-x^2}$ has poles at ± 1. Key idea is now to make a partial fraction decomposition:

$$\frac{1}{1-x^2} = \frac{A}{1-x} + \frac{B}{1+x}.$$

This **Ansatz** gives $A(1+x) + B(1-x) = 1$, i.e., $A = B = \frac{1}{2}$, hence

$$\frac{1}{1-x^2} = \frac{1}{2} \cdot \frac{1}{1-x} + \frac{1}{2} \cdot \frac{1}{1+x}.$$

Applied to our original problem this gives

$$\int \frac{1}{1-x^2} \; dx = \frac{1}{2}\big(-\ln|1-x| + \ln|1+x|\big) = \frac{1}{2} \ln \left|\frac{1+x}{1-x}\right|.$$

Proof of Theorem 12.26, sketch. There are three steps.

1. Recall

$$\begin{aligned}
\int \frac{1}{1+x^2}\,dx &= \arctan x,\\
\int \frac{1}{x^n}\,dx &= \begin{cases} \ln|x|, & \text{if } n=1,\\ \frac{1}{1-n}\cdot\frac{1}{x^{n-1}}, & \text{if } n>1,\end{cases}\\
\int \frac{x}{(1+x^2)^n}\,dx &= \begin{cases} \frac{1}{2}\ln(1+x^2), & \text{if } n=1,\\ \frac{1}{2(1-n)}\cdot\frac{1}{(1+x^2)^{n-1}}, & \text{if } n>1.\end{cases}
\end{aligned}$$

For $\int \frac{1}{(1+x^2)^n}\,dx$ in case $n>1$ we get a recursion formula. We have

$$\begin{aligned}
\int \frac{1}{(1+x^2)^n}\,dx &= \int \frac{dx}{(1+x^2)^{n-1}} - \int x\cdot\frac{x}{(1+x^2)^n}\,dx\\
&= \int \frac{dx}{(1+x^2)^{n-1}}\\
&\quad -\frac{x}{2(1-n)(1+x^2)^{n-1}} + \frac{1}{2(1-n)}\int\frac{dx}{(1+x^2)^{n-1}},
\end{aligned}$$

and the integrals on the right hand side are already known.

2. As in the example above we make a partial fraction decomposition. Starting from $f(x)=\frac{g(x)}{h(x)}$ division with remainder gives $q(x)$ and $r(x)$, such that

$$f(x)=\frac{g(x)}{h(x)}=q(x)+\frac{r(x)}{h(x)} \text{ with } \deg r<\deg h.$$

Next we factor the denominator $h(x)$ in linear factors $l_i(x)\in\mathbb{R}[x]$ and quadratic factors $q_j(x)\in\mathbb{R}[x]$ with no real zeroes using the fundamental theorem of algebra 5.33:

$$h(x)=\prod_{i=1}^{k} l_i(x)^{e_i}\cdot\prod_{j=1}^{l} q_j(x)^{f_j}.$$

It can be shown that there exist $a_{im}, b_{jn}, c_{jn}\in\mathbb{R}$ such that the following partial fractional decomposition holds

$$\frac{r(x)}{h(x)}=\sum_{i=1}^{k}\sum_{m=1}^{e_i}\frac{a_{im}}{l_i^m}+\sum_{j=1}^{l}\sum_{n=1}^{f_j}\frac{b_{jn}x+c_{jn}}{q_j^n}.$$

   To determine the constants a_{im}, b_{jn}, c_{jn} is a matter of linear algebra, a topic of the next semester.

3. The fraction $\frac{1}{l_i}$ can be transformed into $\frac{1}{x}$ by an affine substitution $x \mapsto ax + b$. Similarly $\frac{1}{q_j}$ can be transformed into $\frac{1}{1+x^2}$. Thus the formulas from step 1.) lead to elementary antiderivatives for each summand of the partial fractional decomposition.

□

12.2 Improper Integrals

Definition 12.28. Let $f\colon [a, \infty[\to \mathbb{R}$ be a continuous function. We define the **improper integral**

$$\int_a^\infty f(x)\, dx := \lim_{b\to\infty} \int_a^b f(x)\, dx,$$

in case the limit exists. In this case f is called integrable over $[a, \infty[$, and we say that $\int_a^\infty f(x)\, dx$ converges.

Example 12.29. For $f(x) = x^{-s}$ for $s \in \mathbb{R}$ the limit

$$\int_1^\infty x^{-s}\, dx = \lim_{b\to\infty} \big(\frac{1}{1-s} x^{1-s} \big|_1^b \big) = \lim_{b\to\infty} \frac{1}{1-s}(1 - b^{1-s})$$

exists if and only if $s > 1$. In this case

$$\int_1^\infty x^{-s}\, dx = \frac{1}{1-s}.$$

Theorem 12.30 (Integral criterion for convergence of series)**.** *Let $f\colon [0, \infty[\to \mathbb{R}$ be a monotone decreasing non-negative function. The sum $\sum_{n=0}^\infty f(n)$ converges if and only if the integral $\int_0^\infty f(x)\, dx$ converges.*

Proof. Since f is monotone decreasing, we obtain for every $k \in \mathbb{N}$ bounds

$$\sum_{k=0}^{k-1} f(n) \geq \int_0^k f(x)\,dx \geq \sum_{n=1}^{k} f(n).$$

The assertion follows. □

Corollary 12.31. *The limit*

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

exists for $s > 1$.

Proof. In example 12.29 we saw that $\int_1^\infty x^{-s}dx$ for $s > 1$ converges. □

$\zeta(s)$ is called the **Riemannian zeta function**. It satisfies the Euler product

$$\prod_{p \text{ prime}} \frac{1}{1-p^{-s}} = \sum_{n=1}^{\infty} \frac{1}{n^s},$$

compare 6.24.

Definition 12.32. Let $f\colon]a,b]$ be a continuous function on a semi-open interval. We define

$$\int_a^b f(x)\,dx := \lim_{t \searrow a} \int_t^b f(x)\,dx,$$

in case the limit exists.

Example 12.33.

$$\int_0^1 \frac{1}{\sqrt{x}}dx = \lim_{t \searrow 0} \big[\frac{1}{2}x^{\frac{1}{2}}\big]_t^1 = \frac{1}{2}.$$

More general, $\int_0^1 x^{-s}dx$ converges if $s < 1$ because $\int x^{-s}dx = \frac{1}{1-s}x^{1-s}$. In contrast $\int_0^1 \frac{1}{x}\,dx$ does not converge because $\ln x \xrightarrow[x\to 0]{} -\infty$.

Definition 12.34. A function $f\colon \mathbb{R} \to \mathbb{R}$ is integrable over $[-\infty, \infty]$ if the limits

$$\lim_{b\to\infty} \int_0^b f(x)\,dx \quad \text{and} \quad \lim_{a\to-\infty} \int_a^0 f(x)\,dx$$

exist. In this case we write

$$\int_{-\infty}^{\infty} f(x)\,dx := \lim_{a\to-\infty} \int_a^0 f(x)\,dx + \lim_{b\to\infty} \int_0^b f(x)\,dx.$$

Example 12.35.

1. $$\int_{-\infty}^{\infty} \frac{1}{1+x^2}\,dx = \lim_{a\to-\infty,\ b\to\infty} \bigl(\arctan x \,\big|_a^b\bigr) = \pi.$$

2. The limit $\int_{-\infty}^{\infty} e^{-x^2}\,dx$ exists because $e^{-x^2} \in O\bigl(\frac{1}{1+x^2}\bigr)$. In the third semester we will prove
$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}.$$

13 Taylor expansions

13.1 Taylor polynomial

Let $f\colon I \to \mathbb{R}$ be an n–times continuous differentiable function, $x_0 \in I$. We want f to approximate f near x_0 by a polynomial. The "best" approximation by a linear polynomial is the tangent

$$L(x) = f(x_0) + f'(x_0) \cdot (x - x_0).$$

If we allow higher degree polynomials, we get the Taylor polynomials.

Definition 13.1. Let $f\colon I \to \mathbb{R}$ be an n–times continuous differentiable function, $x_0 \in I$. Then

$$T^n_{x_0} f := \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k$$

is called the n–th **Taylor polynomial** of f in x_0.

Apparently, $T^n_{x_0} f$ is the unique polynomial of degree $\leq n$ which has the same derivatives up to order n in x_0 as f.

Theorem 13.2 (Taylor formula)**.** *Let $f\colon I \to \mathbb{R}$ be an $(n+1)$–times continuous differentiable function, and let $x_0 \in I$ be a point. Then*

$$f(x) = (T^n_{x_0} f)(x) + R_{n+1}(x) \ \big(= \sum_{k=0}^{n} \frac{f^{(k)}(x_0)(x - x_0)^k}{k!} + R_{n+1}(x)\big)$$

with error term

$$R_{n+1}(x) = \int_{x_0}^{x} f^{(n+1)}(t) \frac{(x - t)^n}{n!} \, dt.$$

In details:

$$\begin{aligned} f(x) \quad &= \quad f(x_0) + f'(x_0)(x - x_0) + \frac{f^{(2)}(x_0)}{2}(x - x_0)^2 + \cdots \\ &\qquad + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \int_{x_0}^{x} f^{(n+1)}(t) \frac{(x - t)^n}{n!} dt. \end{aligned}$$

Proof. We use induction by n. The case $n = 0$ holds by the fundamental theorem of calulus

$$f(x) = f(x_0) + \int_{x_0}^{x} f'(t) \, dt.$$

For the induction step $n-1 \to n$ we apply integration by parts to the error term:

$$\begin{aligned}
R_n(x) &= \int_{x_0}^{x} \underbrace{f^{(n)}(t)}_{f} \underbrace{\frac{(x-t)^{n-1}}{(n-1)!}}_{g'}\, dt \\
&\overset{g=-\frac{(x-t)^n}{n!}}{=} -f^{(n)}(t) \cdot \frac{(x-t)^n}{n!}\Big|_{x_0}^{x} + \int_{x_0}^{x} f^{(n+1)}(t)\frac{(x-t)^n}{n!}\, dt \\
&= f^{(n)}(x_0)\frac{(x-x_0)^n}{n!} + R_{n+1}(x).
\end{aligned}$$

Since

$$T_{x_0}^{n} f - T_{x_0}^{n-1} f = f^{(n)}(x_0)\frac{(x-x_0)^n}{n!}$$

holds, the assertion follows. □

Theorem 13.3 (Lagrange's form of the error term)**.** *Let $f\colon I \to \mathbb{R}$ be an $(n+1)$–times continuous differentiable function, and let $x_0 \in I$ be a point. Then $\exists\ \xi \in [x_0, x]$ if $x > x_0$ or $\xi \in [x, x_0]$ if $x < x_0$ such that*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{n+1}.$$

Proof. We apply the mean value theorem of integral calculus to

$$R_{n+1}(x) = \int_{x_0}^{x} f^{(n+1)}(t)\frac{(x-t)^n}{n!}\, dt$$

and obtain

$$R_{n+1}(x) = f^{(n+1)}(\xi)\int_{x_0}^{x} \frac{(x-t)^n}{n!} dt = f^{(n+1)}(\xi)\frac{(x-x_0)^{n+1}}{(n+1)!}.$$

□

Example 13.4. $f(x) = \sin x$, $x_0 = 0$.

$$(T_0^{2n+1}\sin)(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^n\frac{x^{2n+1}}{(2n+1)!},$$

since

$$\sin^{(k)}(0) = \begin{cases} 0, & k \text{ gerade}, \\ (-1)^{\frac{k-1}{2}}, & k \text{ ungerade}. \end{cases}$$

We estimate the error

$$|R_{n+1}(x)| \;=\; |f^{(n+1)}(\xi)| \cdot \frac{|x|^{n+1}}{(n+1)!} \leq \frac{|x|^{n+1}}{(n+1)!} \leq \frac{R^{n+1}}{(n+1)!}$$

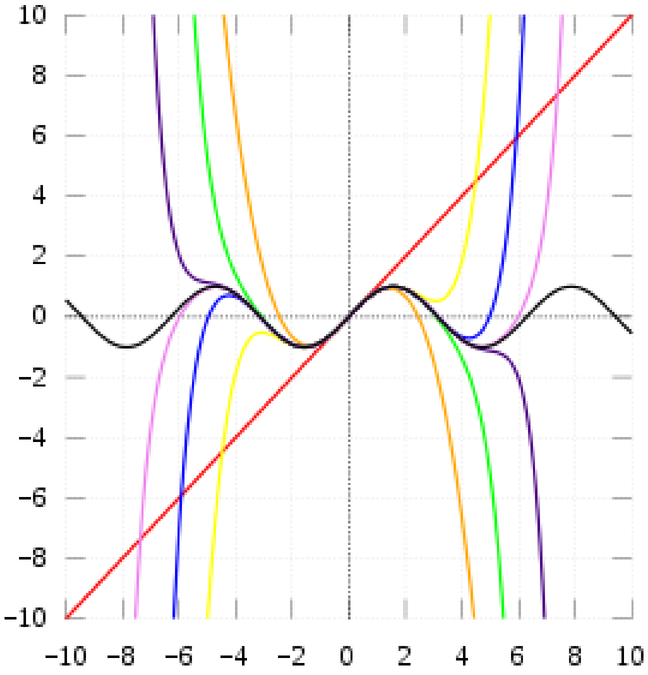
for $|x| \leq R$. Since $\frac{R^{n+1}}{(n+1)!} \leq \varepsilon < 1 \iff$

$$(n+1)\ln R \leq \ln\varepsilon + \sum_{k=1}^{n+1} \ln k \leq \ln\varepsilon + \int_1^{n+1} \ln x \, dx.$$

We have

$$\begin{aligned} (n+1)\ln R &\leq \ln\varepsilon + (x\ln x - x)\Big|_1^{n+1} \\ \iff \ln R &\leq \frac{\ln\varepsilon}{n+1} + \ln(n+1) + \frac{1}{n+1} \\ \iff R &\leq (e\varepsilon)^{\frac{1}{n+1}} \cdot (n+1) \end{aligned}$$

The n-th Taylor polynomial is a good approximation of $\sin(x)$ in a range, which gets larger for larger n.



Example 13.5. Consider the function $f(x) = (1+x)^\alpha$, for example $\alpha = \frac{1}{2}$. Since

$$f^{(k)}(x) = \alpha \cdot (\alpha - 1) \cdots (\alpha - k + 1) \cdot (1 + x)^{\alpha - k},$$

we deduce

$$T_0^n f(x) = \sum_{k=1}^{n} \binom{\alpha}{k} x^k,$$

where for $\alpha \in \mathbb{R}$ the **binomial coefficients** are defined by

$$\binom{\alpha}{k} := \frac{\alpha}{1} \cdot \frac{\alpha - 1}{2} \cdots \frac{\alpha - k + 1}{k}.$$

Definition 13.6. Let $f\colon I \to \mathbb{R}$ be an infinitely often differentiable function, and let $x_0 \in I$ be a point. Then

$$T_{x_0} f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k$$

is called the **Taylor series** of f with **expansion point** x_0. The Taylor series $T_{x_0} f$ is a power series in $(x - x_0)$.

Example 13.7.

1. $(T_0 \exp)(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ is the defining power series of $\exp$.

2. The function $f(x) = (1 + x)^\alpha$ has the Taylor series

$$(T_0 f)(x) = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k.$$

Question 13.8. Does the Taylor series of f converge towards f?

The answer is negative in general:

1. A Taylor series has not necessarily a positive radius R of convergence.

2. Even if $R > 0$, the Taylor series does not necessarily converge to f on $]-R + x_0, x_0 + R[$.

Example 13.9. We give an example of the second kind. Consider

$$f\colon \mathbb{R} \to \mathbb{R},\ x \mapsto f(x) = \begin{cases} e^{-\frac{1}{x^2}}, & x \neq 0, \\ 0, & \text{else.} \end{cases}$$

We show: f is ∞–often differentiable and $f^{(n)}(0) = 0\ \forall n$. In particular the Taylor series is 0,and does not converge to f.

For this we prove by induction

$$f^{(n)}(x) = \begin{cases} p_n(\frac{1}{x}) \cdot e^{-\frac{1}{x^2}}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0, \end{cases}$$

where p_n is a polynomial. The base of the induction is clear. For the induction step we first consider the case $x \neq 0$:

$$\begin{aligned} \big(p_n\big(\frac{1}{x}\big) \cdot e^{-\frac{1}{x^2}}\big)' &= \big(p_n\big(\frac{1}{x}\big)\big)' \cdot e^{-\frac{1}{x^2}} + p_n\big(\frac{1}{x}\big) \cdot \frac{2}{x^3} \cdot e^{-\frac{1}{x^2}}. \\ &= \big(p_n'\big(\frac{1}{x}\big) \cdot \frac{-1}{x^2} + p_n\big(\frac{1}{x}\big) \cdot \frac{2}{x^3}\big) \cdot e^{-\frac{1}{x^2}}. \end{aligned}$$

Hence

$$p_{n+1}(t) = -p_n'(t) \cdot t^2 + 2p_n(t) \cdot t^3.$$

At $x = 0$ we obtain

$$f^{(n+1)}(0) = \lim_{x \to 0} \big(\frac{1}{x} p_n\big(\frac{1}{x}\big) \cdot e^{-\frac{1}{x^2}}\big) = 0,$$

since $\exp$ grows faster than any polynomial. □

By the answers of Question 13.8 the following result is less trivial than might look like on the first glance.

Theorem 13.10 (Binomial series)**.** *For $|x| < 1$ the following holds:*

$$(1 + x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k.$$

Proof. For $\alpha \notin \mathbb{N}$ the radius of convergence is $R = 1$. This follows from the quotient test because

$$\left|\frac{\binom{\alpha}{k+1}}{\binom{\alpha}{k}}\right| = |\frac{\alpha-k}{k+1}| \underset{k\to\infty}{\longrightarrow} 1.$$

To see that on for $|x|<1$ the Taylor series converges to $(1+x)^\alpha$ we have to prove that

$$R_{n+1}(x) = \frac{1}{n!}\int_0^x (x-t)^n f^{(n+1)}(t)\,dt = (n+1)\binom{\alpha}{n+1}\int_0^x (x-t)^n(1+t)^{\alpha-n-1}\,dt$$

converges for $n\to\infty$ to zero if $|x|<1$. We distinguish by the sign of x.
First case is $0\le x<1$. Set $C=\max(1,(1+x)^\alpha)$. Then for $0\le t\le x$ we have

$$0\le(1+t)^{\alpha-n+1}\le(1+t)^\alpha\le C.$$

Hence

$$\begin{aligned}|R_{n+1}(x)| &= (n+1)|\binom{\alpha}{n+1}|\int_0^{|x|}|(x+t)^n(1-t)^{\alpha?n?1}|dt\\ &\le (n+1)|\binom{\alpha}{n+1}|C\int_0^x(x-t)^n dt = C|\binom{\alpha}{n+1}|x^{n+1}.\end{aligned}$$

Since $\sum_{n=0}^\infty\binom{\alpha}{n}x^n$ converges, we have $|\binom{\alpha}{n+1}|x^{n+1}\longrightarrow 0$ for $n\to\infty$.

The second case, $-1<x<0$, is the difficult case because the function $(1+x)^\alpha$ or its derivatives can have a pole at $x=-1$.

$$\begin{aligned}|R_{n+1}(x)| &= (n+1)|\binom{\alpha}{n+1}|\int_0^x|(x-t)^n(1-t)^{\alpha-n-1}|dt\\ &\le (n+1)|\binom{\alpha}{n+1}|\int_0^{|x|}(\frac{|x|-t}{1-t})^n(1-t)^{\alpha-1}dt.\end{aligned}$$

The function $t\mapsto\frac{|x|-t}{1-t}$ is monotonously decreasing on $[0,|x|]$ since the derivative is negative

$$\frac{(1-t)(-1)-(-1)(|x|-t)}{(1-t)^2} = \frac{|x|-1}{(1-t)^2} < 0.$$

So

$$(\frac{|x|-t}{1-t})^n \le |x|^n \text{ for } 0\le t\le|x|.$$

With $C = |\alpha| \int_0^{|x|}(1-t)^{\alpha-1}dt$ we obtain

$$|R_{n+1}(x)| \leq |\alpha\binom{\alpha-1}{n}|\,|x|^n \int_0^{|x|}(1-t)^{\alpha-1}dt \leq C|\binom{\alpha-1}{n}||x|^n,$$

and this expressions converges to 0 for $n \to \infty$ because also the series $\sum_{n=0}^{\infty}\binom{\alpha-1}{n}x^n$ converges. □

13.2 Uniform convergence

Let

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

be a power series with radius of convergence $R > 0$. We want to prove that f is ∞–often differentiable on $]-R, R[$ and that the power series coincides with the Taylor series of f in $x_0 = 0$. Already continuity is not obvious.

Definition 13.11. Let $f_n \colon I \to \mathbb{R}$ be a sequence of functions on an interval. The sequence (f_n) **converges** (more precisely: **converges pointwise**) if for every x the sequence $(f_n(x))$ converges. In this case we call

$$f \colon I \to \mathbb{R},\ f(x) = \lim_{n\to\infty} f_n(x).$$

the **limit function** and write: $\lim_{n\to\infty} f_n = f$.

Question 13.12. Is $\lim f_n$ continuous if all f_n continuous?

The answer is no in general. For example, for $f_n \colon [0,1] \to \mathbb{R},\ f_n(x) = x^n$ the limit $f = \lim f_n$ exists, but

$$f(x) = \begin{cases} 0, & \text{if } x \in [0,1[, \\ 1, & \text{if } x = 1. \end{cases}$$

f is not continuous. One needs a stronger assumption on the convergence.

Definition 13.13. Let $(f_n \colon I \to \mathbb{R})_{n\in\mathbb{N}}$ be a sequence of functions on an interval I. (f_n) **converges uniformly** to a limit function $f \colon I \to \mathbb{R}$, if

$$\forall \varepsilon > 0\ \exists\, n_0 :\ |f_n(x) - f(x)| < \varepsilon\ \forall n \geq n_0\ \forall\, x \in I.$$

Theorem 13.14 (Uniform limit of continuous functions)**.** *If $(f_n\colon I \to \mathbb{R})$ is a sequence of continuous functions, which converges uniformly to f, then also f is continuous.*

Proof. Let $x_0 \in I$ and $\varepsilon > 0$ be given. For $\frac{\varepsilon}{3}$ there exists an n_0 with

$$|f_n(x) - f(x)| < \frac{\varepsilon}{3}\ \forall n \geq n_0\ \forall\, x \in I.$$

Since f_{n_0} is continuous, $\exists\, \delta > 0$, such that

$$|f_{n_0}(x) - f_{n_0}(x_0)| < \frac{\varepsilon}{3}\ \forall x \in I \text{ with } |x - x_0| < \delta.$$

Hence

$$\begin{aligned} |f(x) - f(x_0)| &\leq |f(x) - f_{n_0}(x)| + |f_{n_0}(x) - f_{n_0}(x_0)| + |f_{n_0}(x_0) - f(x_0)| \\ &< \varepsilon\ \forall x \text{ with } |x - x_0| < \delta. \end{aligned}$$

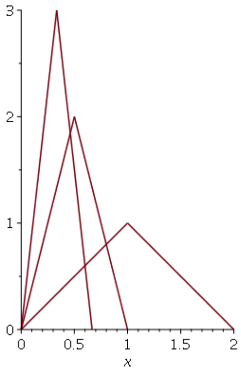
□

Question 13.15. Can one interchange integration with taking the limit?

No, not necessarily.

Example 13.16. If $f_n : [0,1] \to \mathbb{R}$ is defined by

$$f(x) = \begin{cases} n^2 x & \text{if } 0 \leq x \leq \frac{1}{n} \\ n^2(\frac{2}{n} - x) & \text{if } \frac{1}{n} < x \leq \frac{2}{n} \\ 0 & \text{else .} \end{cases}$$



Then $\int_0^1 f_n(x)\,dx = 1 \;\forall\, n \geq 2$ and $\lim_{n\to\infty} f_n = 0$ pointwise, but

$$\lim_{n\to\infty}\int_0^1 f_n(x)\,dx = 1 \neq 0 = \int_0^1 \lim_{n\to\infty} f_n(x)\,dx.$$

Theorem 13.17. *Let $f_n\colon [a,b] \to \mathbb{R}$ be a sequence of continuous functions on a closed bounded interval which converges uniformly to $f\colon [a,b] \to \mathbb{R}$. Then*

$$\int_a^b f(x)\,dx = \lim_{n\to\infty}\int_a^b f_n(x)\,dx.$$

Proof. Since f is also continuous, it is integrable. Moreover

$$\Big|\int_a^b f(x)\,dx - \int_a^b f_n(x)\,dx\Big| \leq \int_a^b \big|f(x) - f_n(x)\big|\,dx \leq \varepsilon(b-a),$$

if we choose n large enough, such that $|f(x) - f_n(x)| < \varepsilon \;\forall x \in [a,b]$. The assertion follows.

□

Remark 13.18. For improper integrals one needs additional assumptions as the example

$$f(x) = \begin{cases} \frac{1}{n^2}(n - |x|) & \text{if } |x| < n \\ 0 & \text{else .} \end{cases}$$

shows. Apparently, $\lim f_n = 0$ is uniform, but

$$\int_{-\infty}^{\infty} f_n(x)\,dx = 1 \neq 0 = \int_{\infty}^{\infty} 0\,dx.$$

Corollary 13.19. *Let $f_n \colon [a,b] \to \mathbb{R}$ be a sequence of continuous differentiable functions, which converges pointwise towards $f \colon [a,b] \to \mathbb{R}$. If the sequence of derivatives (f_n') converges uniformly, then f is differentiable and we have*

$$f' = \lim_{n\to\infty} f_n'.$$

Proof. Let $f^* = \lim f_n'$. Then f^* is continuous on $[a,b]$ by Theorem 13.14. Since

$$f_n(x) = f_n(a) + \int_a^x f_n'(t)\, dt$$

for $x \in [a,b]$, we deduce

$$f(x) = f(a) + \int_a^x f^*(t)\, dt.$$

from Theorem 13.17. Hence the fundamental theorem of calculus implies that f is differentiable with $f' = f^*$. □

13.3 Application to power series

Theorem 13.20. *Let $f(x) = \sum_{n=0}^\infty a_n x^n$ be a power series with radius of convergence $R > 0$, hence $f \colon]-R,R[\to \mathbb{R}$. The power series*

1. $\sum_{n=1}^\infty n a_n x^{n-1}$,
2. $\sum_{n=0}^\infty a_n \frac{x^{n+1}}{n+1}$,

which we obtain by termwise differentiation and integration, have the same radius of convergence and converge on $]-R,R[$ to

1. $f'(x)$,
2. $\int_0^x f(x)\, dx$.

In particular f is ∞–often differentiable and

$$f^{(n)}(0) = a_n \cdot n!.$$

Proof. By the formula of Cauchy–Hadamard 7.7 we know $\sum a_n x^n$ converges for $|x| < R$ if and only if

$$\limsup_{n\to\infty} \sqrt[n]{|a_n x^n|} = \big(\limsup_{n\to\infty} \sqrt[n]{|a_n|}\big) \cdot R \leq 1,$$

i.e.,

$$R = \frac{1}{\limsup_{n\to\infty} \sqrt[n]{|a_n|}}.$$

Since

$$\lim_{n\to\infty} \sqrt[n]{n} = \lim_{n\to\infty} n^{\frac{1}{n}} = \lim_{n\to\infty} e^{\frac{\ln n}{n}} = e^0 = 1,$$

we deduce that $\sum_{n=1}^{\infty} n a_n x^{n-1}$ and $\sum_{n=0}^{\infty} a_n \frac{x^{n+1}}{n+1}$ have the same radius of convergence. The sequence of partial sums of f and f' converges on each proper subintervall $[-r, r]$ for $r < R$ uniformly. Hence the assertion follows on $[-r, r]$ by Theorem 13.17 and Corollary 13.19. □

Example 13.21.

1. The logarithm $\ln(1 + x)$ is the antiderivative of

   $$f(x) = \frac{1}{1+x} = \sum_{k=0}^{\infty} (-1)^n x^n.$$

   Termwise integration and $\ln(1) = 0$ gives

   $$\ln(1 + x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{n+1}}{n+1}$$

   for $|x| < 1$.

2. $\arctan$ is the antiderivative of

   $$\frac{1}{1+x^2} = f(x) = \sum_{n=0}^{\infty} (-1)^n x^{2n}.$$

   Integration and $\arctan(0) = 0$ give

   $$\arctan(x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1} \text{ for } |x| < 1.$$

3. Since

$$\sum_{n=0}^{\infty} nx^n = x \cdot \sum_{n=1}^{\infty} nx^{n-1},$$

we conclude

$$\sum_{n=0}^{\infty} nx^n = x \cdot \big(\frac{1}{1-x}\big)' = \frac{x}{(1-x)^2} \text{ for } |x| < 1.$$

For example:

$$\sum_{n=1}^{\infty} n\big(\frac{1}{2}\big)^n = \frac{\frac{1}{2}}{(1-\frac{1}{2})^2} = 2.$$

In example 1. and 2. of 13.21 the series converge also for $x = 1$. This suggests

$$\sum_{n=0}^{\infty}(-1)^n \frac{1}{n+1} = \ln(1+1) = \ln 2,$$
$$\sum_{n=0}^{\infty}(-1)^n \frac{1}{2n+1} = \arctan(1) = \frac{\pi}{4}$$

(since $\tan\frac{\pi}{4} = 1$). That this is really true, follows from the following theorem.

Theorem 13.22 (Abel's theorem)**.** *Let $\sum_{n=0}^{\infty} a_n$ be a convergent series. Then the power series $f(x) = \sum_{n=0}^{\infty} a_n x^n$ has radius of convergence ≥ 1, and the limit function $f\colon]-1,1[\to \mathbb{R}$ satisfies*

$$\lim_{x\to 1} f(x) = \sum_{n=0}^{\infty} a_n.$$

Proof. Set

$$s_n = \sum_{k=n+1}^{\infty} c_k.$$

Then $s_{-1} = f(1)$ and $s_n - s_{n-1} = -c_n$ for all $n \in \mathbb{N}$, and $\lim_{n\to\infty} s_n = 0$. Hence there is a constant $K > 0$ such that $|s_n| \leq K$ for all n. The series

$$\sum_{n=0}^{\infty} s_n x^n$$

converges for $|x| < 1$, since $\sum_{n=0}^{\infty} Kx^n$ is majorising.

$$\begin{aligned}(1-x)\sum_{n=0}^{\infty} s_n x^n &= \sum_{n=0}^{\infty} s_n x^n - \sum_{n=0}^{\infty} s_n x^{n+1} \\ &= \sum_{n=0}^{\infty} s_n x^n - \sum_{n=0}^{\infty} s_{n-1} x^n + s_{-1} = -\sum_{n=0}^{\infty} c_n x^n + f(1).\end{aligned}$$

Hence

$$f(1) - f(x) = (1-x)\sum_{n=0}^{\infty} s_n x^n.$$

Let $\varepsilon > 0$ be given. We choose N so large, such that $|s_n| \leq \frac{\varepsilon}{2}$ for all $n \geq N$ and set $\delta = \frac{\varepsilon}{2KN}$. Then for $x \in [0,1[$ with $1 - x < \delta$ we obtain

$$\begin{aligned}|f(1) - f(x)| &\leq (1-x)\sum_{n=0}^{N-1} |s_n| x^n + (1-x)\sum_{n=N}^{\infty} |s_n| x^n \\ &\leq (1-x)KN + (1-x)\frac{\varepsilon}{2}\sum_{n=0}^{\infty} x^n \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,\end{aligned}$$

as desired. □

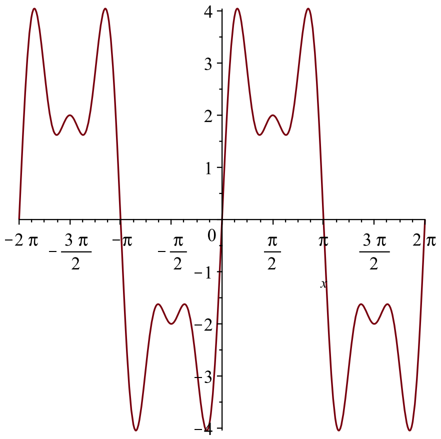
14 Fourier expansion

Let $f : \mathbb{R} \to \mathbb{R}$ be a periodic function with period L, i.e. $f(x+L) = f(x)$. By replacing x by $\frac{2\pi}{L}x$ we may assume that f is 2π-periodic. Examples of 2π-periodic functions are $\sin(x), \cos(x)$ and more general $\sin(kx), \cos(\ell x)$ for $k, \ell \in \mathbb{N}$.

A sum

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{n} a_k \cos(kx) + b_k \sin(kx)$$

with coefficients $a_0, \dots, a_n, b_1, \dots, b_n \in \mathbb{R}$ is called a **Fourier** or **trigonometric polynomial**.



$$f(x) = 3\sin(x) + 2\sin(3x) + \sin(5x)$$

The coefficients can be recovered from f, since

$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos(kx)dx \text{ for } k = 0, 1, \dots$$

and

$$b_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin(kx)dx \text{ for } k = 1, 2, \dots.$$

holds. Indeed,

$$\int_0^{2\pi} \cos(kx)\sin(\ell x)dx = 0 \quad \forall k \in \mathbb{N}_0, \ell \in \mathbb{N},$$

$$\int_0^{2\pi} \sin(kx)\sin(\ell x)dx = 0 = \int_0^{2\pi} \cos(kx)\cos(\ell x)dx \quad \forall k \in \mathbb{N}_0, \ell \in \mathbb{N}, k \neq \ell$$

as one can see by applying integration by parts twice.

$$\int_0^{2\pi} \cos^2(kx)dx = \pi = \int_0^{2\pi} \sin^2(kx)dx \quad \forall k \geq 1,$$

since

$$\int \sin^2(kx)dx = \frac{1}{2}(x - \frac{1}{k}\sin(kx)\cos(kx)) \text{ and } \int \cos^2(kx)dx = \frac{1}{2}(x + \frac{1}{k}\sin(kx)\cos(kx)).$$

Finally,

$$\int_0^{2\pi} \cos(0x)dx = \int_0^{2\pi} 1\,dx = 2\pi,$$

which explains our choice of $\frac{a_0}{2}$ for the constant term in f.

Frequently, it is useful to allow complex valued functions

$$f : \mathbb{R} \to \mathbb{C},$$

since

$$\cos x = \frac{1}{2}(e^{ix} + e^{-ix}),\ \sin x = \frac{1}{2}(e^{ix} - e^{-ix}).$$

We can write the trigonometric polynomial f as

$$f(x) = \sum_{k=-n}^{n} c_k e^{ikx}$$

with

$$c_0 = \frac{1}{2}a_0,$$

$$c_k = \frac{1}{2}(a_k - ib_k), c_{-k} = \frac{1}{2}(a_k + ib_k).$$

To recover the Fourier coefficients c_k directly via integration, we define for $g : [a,b] \to \mathbb{C},\ g(x) = u(x) + iv(x)$ with real and imaginary part $u, v : [a,b] \to \mathbb{R}$ the integral

$$\int_a^b g(x)dx := \int_a^b u(x)dx + i\int_a^b v(x)dx$$

in case both integrals exists. For example, the functions $e_m(x) = e^{imx}, m \neq 0$ have the integrals

$$\int_a^b e_m(x)dx = \frac{1}{im} e^{imx}\Big|_a^b.$$

In particular, we obtain

$$\int_0^{2\pi} e_m(x)dx = 0 \quad \forall m \in \mathbb{Z} \setminus \{0\}.$$

The coefficients of the trigonometric polynomial

$$f(x) = \sum_{k=-n}^{n} c_k e^{ikx}$$

are given by

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-ikx} dx \text{ for } k = 0, \pm 1, \pm 2, \dots, \pm n$$

since

$$f(x)e^{-ikx} = \sum_{m=-n}^{n} c_m e^{i(m-k)x} dx$$

Definition 14.1. Let $f : \mathbb{R} \to \mathbb{C}$ be a 2π-periodic function, which is integrable over $[0, 2\pi]$. Then the numbers

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-ikx} dx, \quad k \in \mathbb{Z}$$

are called the **Fourier coefficients** of f, and

$$\sum_{k=-\infty}^{\infty} c_k e^{ikx},$$

i.e. the sequence of partial sums

$$s_n(x) = \sum_{k=-n}^{n} c_k e^{ikx},$$

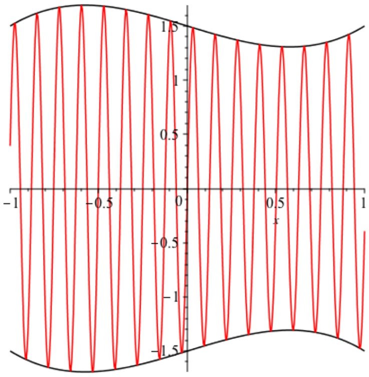
is called the **Fourier series** of f.

Theorem 14.2. *Let $f : [a,b] \to \mathbb{R}$ be a continuously differentiable function. Consider*

$$F(k) = \int_a^b f(x)\sin(kx)dx.$$

for $k \in \mathbb{R}$. Then

$$\lim_{k\to\pm\infty} F(k) = 0.$$



Proof. For $k \neq 0$ partial integration gives

$$F(k) = -f(x)\frac{\cos(kx)}{k}\Big|_a^b + \frac{1}{k}\int_a^b f'(x)\cos(kx)dx.$$

Hence $|F(k)| \leq \frac{2M}{|k|} + \frac{M(b-a)}{|k|}$ if $|f(x)| \leq M$ and $|f'(x)| \leq M$, and

$$|F(k)| \underset{k\to\pm\infty}{\longrightarrow} 0.$$

□

Example 14.3 (Sawtooth function)**.** Consider the sawtooth function $\sigma : \mathbb{R} \to \mathbb{R}$ defined by

$$\begin{aligned}
\sigma(0) &= 0, \\
\sigma(x) &= \frac{\pi - x}{2} \text{ for } x \in]0, 2\pi[, \text{ and} \\
\sigma(x) &= \sigma(x + 2\pi n)\ \forall x \in \mathbb{R} \text{ and } n \in \mathbb{Z}.
\end{aligned}$$

Its Fourier coefficients are

$$a_k = \frac{1}{\pi}\int_0^{2\pi} \sigma(x)\cos(kx)dx = 0$$

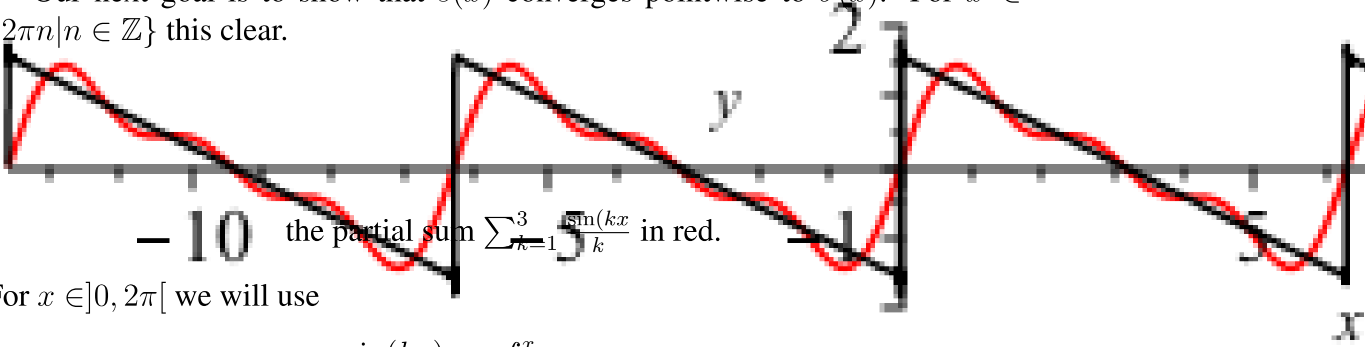
by symmetry and

$$\begin{aligned} b_k &= \frac{1}{\pi}\int_0^{2\pi} \frac{\pi - x}{2}\sin(kx)dx \\ &= \frac{1}{2\pi}\int_0 2\pi - x\sin(kx)dx \\ &= \frac{1}{2\pi}x\frac{1}{k}\cos(kx)\Big|_0^{2\pi} + \frac{1}{2\pi}\int_0^{2\pi}\frac{1}{k}\cos(kx)dx \\ &= \frac{1}{k} + 0 = \frac{1}{k}. \end{aligned}$$

Hence the Fourier series of σ is

$$s(x) = \sum_{k=1}^{\infty} \frac{\sin(kx}{k}.$$

Our next goal is to show that $s(x)$ converges pointwise to $\sigma(x)$. For $x \in \{2\pi n | n \in \mathbb{Z}\}$ this clear.



the partial sum $\sum_{k=1}^{3} \frac{\sin(kx}{k}$ in red.

For $x \in]0, 2\pi[$ we will use

$$\frac{\sin(kx)}{k} = \int_{\pi}^{x} \cos(kt)dt$$

and the following Proposition

Proposition 14.4.

$$\sum_{k=1}^{n} \cos(kt) = \frac{\sin((n+\frac{1}{2})t)}{2\sin(\frac{1}{2}t)} - \frac{1}{2}.$$

Proof.

$$\cos x = \frac{e^{ix} + e^{-ix}}{2}.$$

Hence

$$\begin{aligned}
\sum_{k=1}^{n} \cos(kt) &= \sum_{k=1}^{n} \frac{e^{ikx} + e^{-ikx}}{2} \\
&= \frac{1}{2}(\sum_{k=-n}^{n} e^{ikt}) - \frac{1}{2} \\
&= \frac{1}{2} e^{-int} (\sum_{\ell=0}^{2n} e^{i\ell t}) - \frac{1}{2} \\
&= \frac{1}{2} e^{-int} (\frac{1 - e^{i(2n+1)t}}{1 - e^{it}}) - \frac{1}{2} \\
&= \frac{1}{2} (\frac{e^{i(n+\frac{1}{2})t} - e^{-i(n+\frac{1}{2})t}}{e^{i\frac{t}{2}} - e^{-i\frac{t}{2}}}) - \frac{1}{2} \\
&= \frac{\sin((n+\frac{1}{2})t)}{2\sin(\frac{t}{2})} - \frac{1}{2}.
\end{aligned}$$

□

We conclude

$$\sum_{k=1}^{n} \frac{\sin(kx)}{k} = \int_{\pi}^{x} \sin((n+\frac{1}{2})t) \cdot \frac{1}{2\sin\frac{t}{2}} dt + \frac{\pi - x}{2}$$

Theorem 14.5.
$$\sum_{k=1}^{\infty} \frac{\sin(kx)}{k} = \frac{\pi - x}{2} \textit{ for } x \in]0, 2\pi[$$
and for $\delta > 0$ *the convergence is uniform on* $[\delta, 2\pi - \delta]$.

Proof. $\frac{1}{2\sin\frac{t}{2}}$ and its derivative is bounded on $[\delta, 2\pi-\delta]$. Hence

$$\int_{\pi}^{x} \sin((n+\frac{1}{2})t) \cdot \frac{1}{2\sin\frac{t}{2}} dt \underset{n\to\infty}{\longrightarrow} 0$$

by Theorem 14.2, and this convergence is uniform for $x \in [\delta, 2\pi - \delta]$. □

As an application we will prove the formula

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

Consider the series
$$F(x) = \sum_{k=1}^{\infty} \frac{\cos(kx)}{k^2}.$$

The series converges uniform on $\mathbb{R}$ because $\sum_{k=1}^{\infty} \frac{1}{k^2}$ majorizes $F(x)$. The sequence of derivatives
$$-\sum_{k=1}^{\infty} \frac{\sin(kx)}{k}$$
converges on $[\delta, 2\pi - \delta]$ uniformly. Hence $F(x)$ is continues, and its derivative on $]0, 2\pi[$ is
$$F'(x) = \frac{x - \pi}{2}$$
by Theorem 13.19. Hence

$$F(x) = (\frac{x-\pi}{2})^2 + c$$

for some constant c. To determine c we integrate F. On one hand, we have

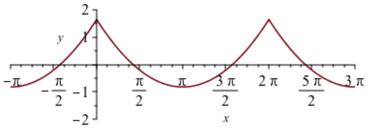
$$\int_0^{2\pi} F(x)dx = \int_0^{2\pi} ((\frac{x-\pi}{2})^2 + c)dx$$

$$= \frac{2}{3}(\frac{x-\pi}{2})^3\Big|_0^{2\pi} + 2\pi c = \frac{\pi^3}{6} + 2\pi c.$$

On the other hand,

$$\int_0^{2\pi} F(x)dx = \sum_{k=1}^{\infty} \int_0^{2\pi} \frac{\cos(kx)}{k^2} dx = 0$$

since the convergence is uniform. Hence $c = -\frac{\pi^2}{12}$.



$$F(x) = \sum_{k=1}^{\infty} \frac{\cos(kx)}{k^2} = (\frac{x-\pi}{2})^2 - \frac{\pi^2}{12}.$$

Evaluating F at $x = 0$, we obtain

$$\sum_{n=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6},$$

since $\frac{1}{4} - \frac{1}{12} = \frac{1}{6}$.

Theorem 14.6. *Let $f : \mathbb{R} \to \mathbb{R}$ be continuous 2π-periodic function, which is piecewise continuously differentiable. Then its Fourier series converges uniformly to f.*

This will be a topic next term, when we have introduced vector spaces and inner products on them.