Formulas from the Manuscipt Mathematics for AI

written by Prof. Mario Ullrich

Version 25.01.2021

(formulas until section 9 might be outdated)

CG

Contents

1 Sets, numbers and functions

1.1 Sets

Definition 1.1. Let $P(\cdot)$ be a predicate for the class M.
The **universal quantifier** builds a proposition

$$\forall m \in M : P(m),$$

which is true if and only if *for all* $m \in M$ the proposition $P(m)$ is true.

The **existential quantifier** builds a proposition

$$\exists m \in M : P(m),$$

which is true if and only if there exists *at least one* $m \in M$ such that $P(m)$ is true.

The **uniqueness quantifier** for which

$$\exists!\, m \in M : P(m)$$

is true if and only if there exists *exactly one* $m \in M$ such that $P(m)$ is true.

Definition 1.4. Let M, N be sets in an universal set Ω. We define

- the **union** of M and N by
$$M \cup N := \{x : x \in M \text{ or } x \in N\},$$
- the **intersection** of M and N by
$$M \cap N := \{x : x \in M \text{ and } x \in N\},$$
- the **difference** of M and N
$$M \setminus N := \{x \in M : x \notin N\}$$
- and the **complement** of M (in Ω)
$$M^c := \{x \in \Omega : x \notin M\}.$$

Definition 1.6. Let M be a set and $n \in \mathbb{N}$. If the elements of M can be labeled by the numbers $\{1, \ldots, n\}$, then we say M **has cardinality** n, and we write

$$|M| = n \qquad \text{or} \qquad \#M = n.$$

Such sets are **finite**.

If M can be labeled by $\mathbb{N}$, we call M **countable**. (e.g. $\mathbb{Z}$)

If M is not countable, it is called **uncountable**.

1.2 Real numbers

Lemma 1.10. *Let $n \in \mathbb{N}$. Then we have n is even $\iff$ n^2 is even.*

Theorem 1.11. *There is no rational number x, such that $x^2 = 2$.*

Next we want to treat the real numbers axiomatically, which means that we appoint them certain structural characteristics. The following **field axioms** are valid for any real numbers:

$$
\begin{array}{rll}
\textbf{Commutativity}: & x + y = y + x, & x \cdot y = y \cdot x \\
\textbf{Associativity}: & x + (y + z) = (x + y) + z, & x \cdot (y \cdot z) = (x \cdot y) \cdot z \\
\textbf{Distributivity}: & x \cdot (y + z) = (x \cdot y) + (x \cdot z) & \\
\textbf{Identity elements}: & \exists 0, 1 \in \mathbb{R}\ \forall x \in \mathbb{R}\colon x + 0 = 0 + x = x, & x \cdot 1 = 1 \cdot x = x \\
\textbf{Inverse elements}: & \forall x \in \mathbb{R}\, \exists y \in \mathbb{R}\colon x + y = y + x = 0, & \\
 & \forall x \in \mathbb{R}\backslash\{0\}\, \exists y \in \mathbb{R}\colon x \cdot y = y \cdot x = 1 &
\end{array}
$$

1.3 Relations and functions

Definition 1.14. A **relation** R between two sets M and N is a subset of the cartesian product of M and N, i.e. $R \subseteq M \times N$.

Definition 1.16. Let $M, N \neq \varnothing$. We call $f : M \to N$ a **function** from M to N, if and only if each $x \in M$ is assigned exactly one $f(x) \in N$. The **mapping rule** is written in the following way $x \mapsto f(x)$.
M is called **domain** (of definition) and N **codomain** of f.
Let $S \subseteq M$. We define the **image** of S under f as

$$f(S) := \{f(x) : x \in S\} \subseteq N,$$

and the **range** of f as

$$f(M) := \{f(x) : x \in M\} \subseteq N.$$

Remark 1.17. Note, that all functions are relations, but not vice versa.

Definition 1.18. Let $R \subseteq M \times N$ be a relation. R is called

- **injective** if and only if

$$\forall (x_1, y_1), (x_2, y_2) \in R \colon x_1 \neq x_2 \Rightarrow y_1 \neq y_2,$$

  which is equivalent to

$$\forall (x_1, y_1), (x_2, y_2) \in R \colon y_1 = y_2 \Rightarrow x_1 = x_2.$$

- **surjective** if and only if

$$\forall y \in N \, \exists x \in M \colon (x, y) \in R.$$

- **bijective** if and only if it is injective and surjective.

If $f \colon M \to N$ is a function we can rewrite the definition above in the following way:

$$
\begin{aligned}
f \text{ is } \textbf{injective} &\iff \forall x_1, x_2 \in M \colon x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2) \\
f \text{ is } \textbf{surjective} &\iff \forall y \in N \, \exists x \in M \colon f(x) = y \\
f \text{ is } \textbf{bijective} &\iff \forall y \in N \, \exists! \, x \in M \colon f(x) = y.
\end{aligned}
$$

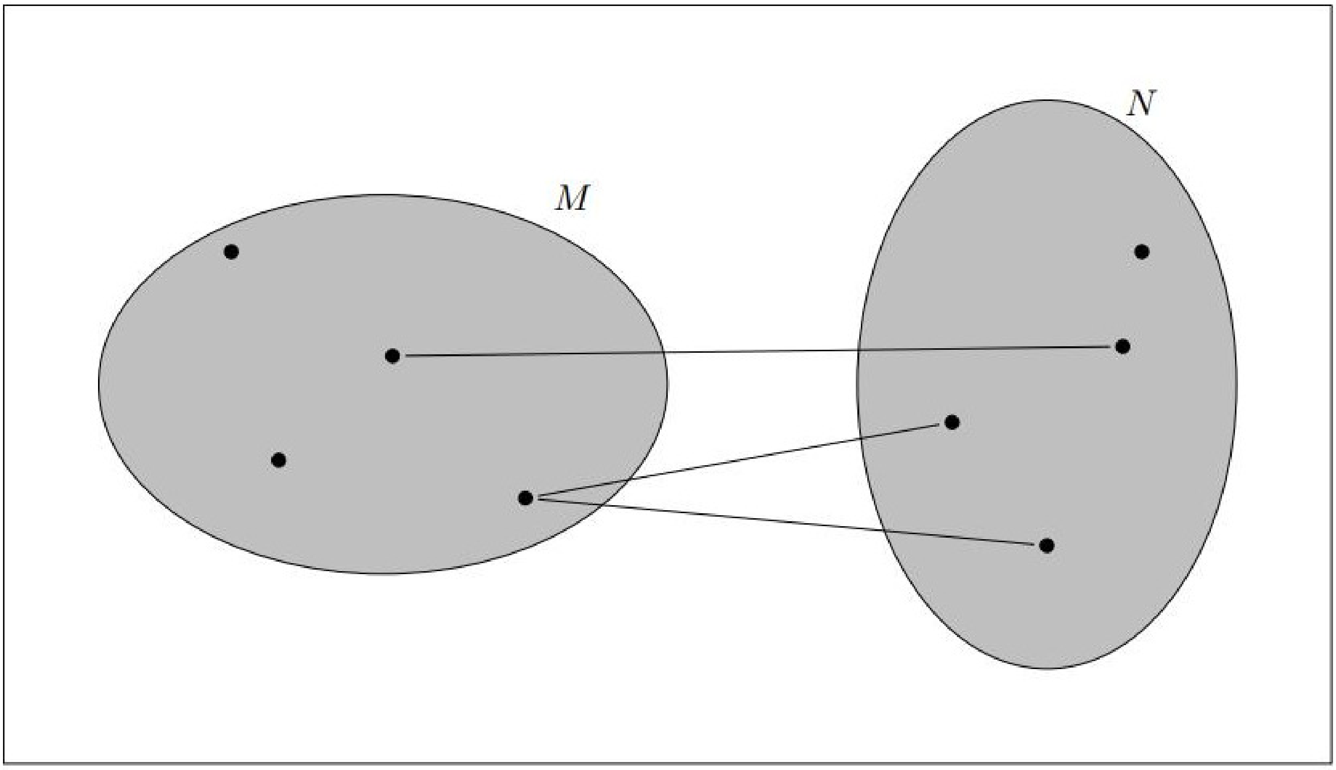

Figure 1: injective relation

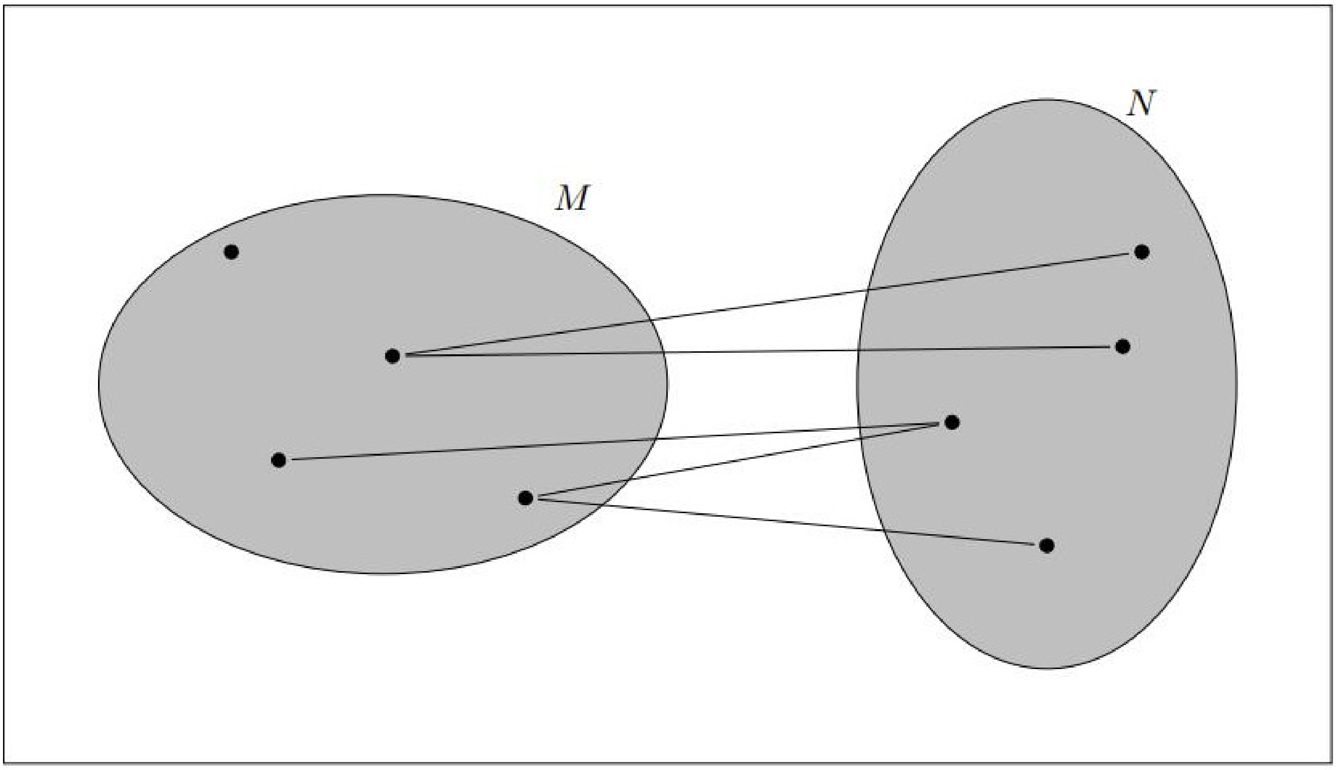

Figure 2: surjective relation

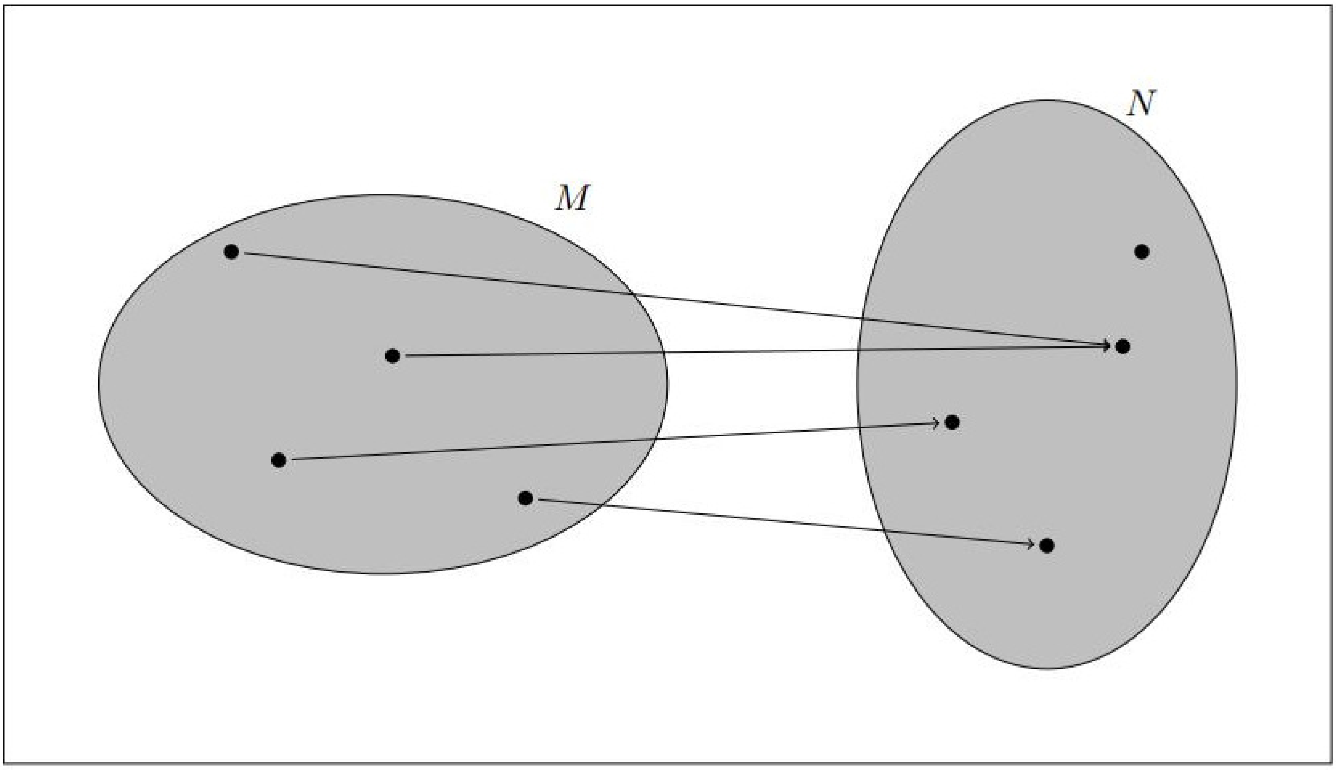

Figure 3: a function (not injective and not surjective)

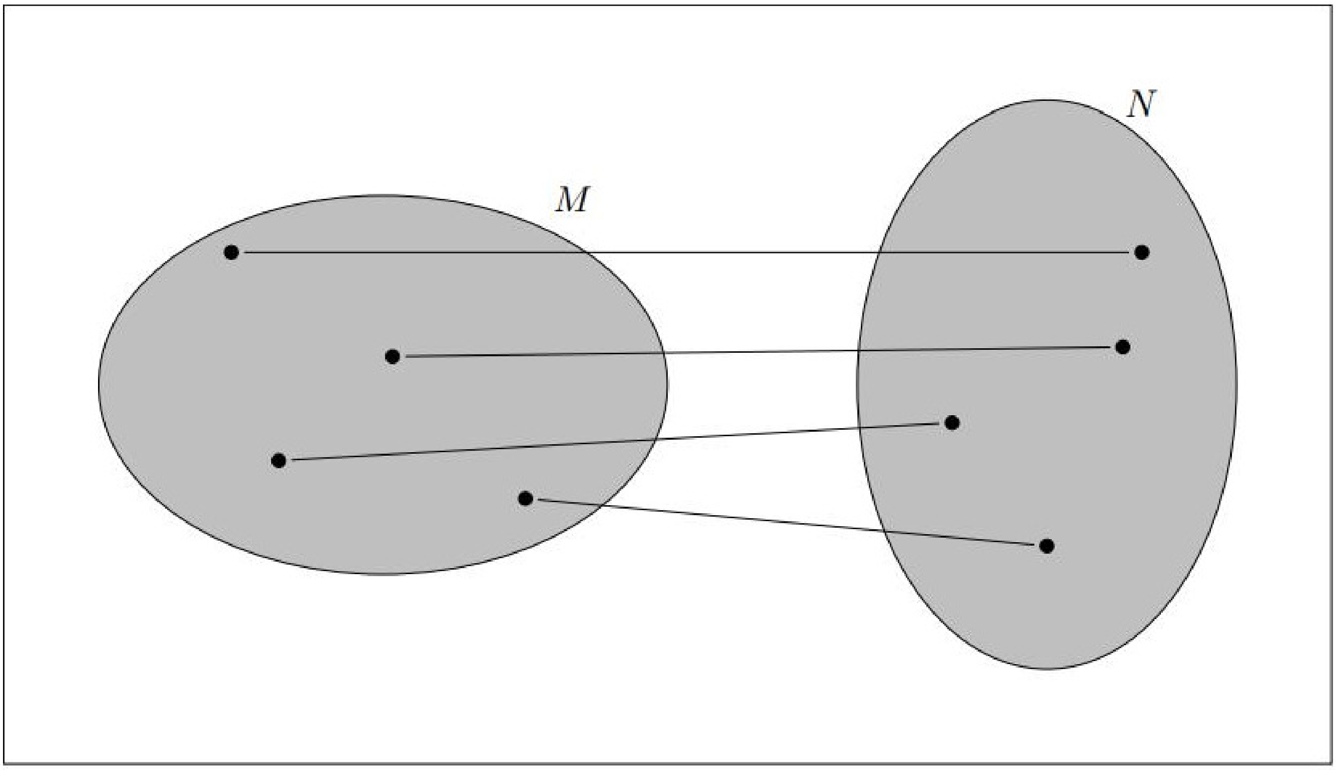

Figure 4: bijective function

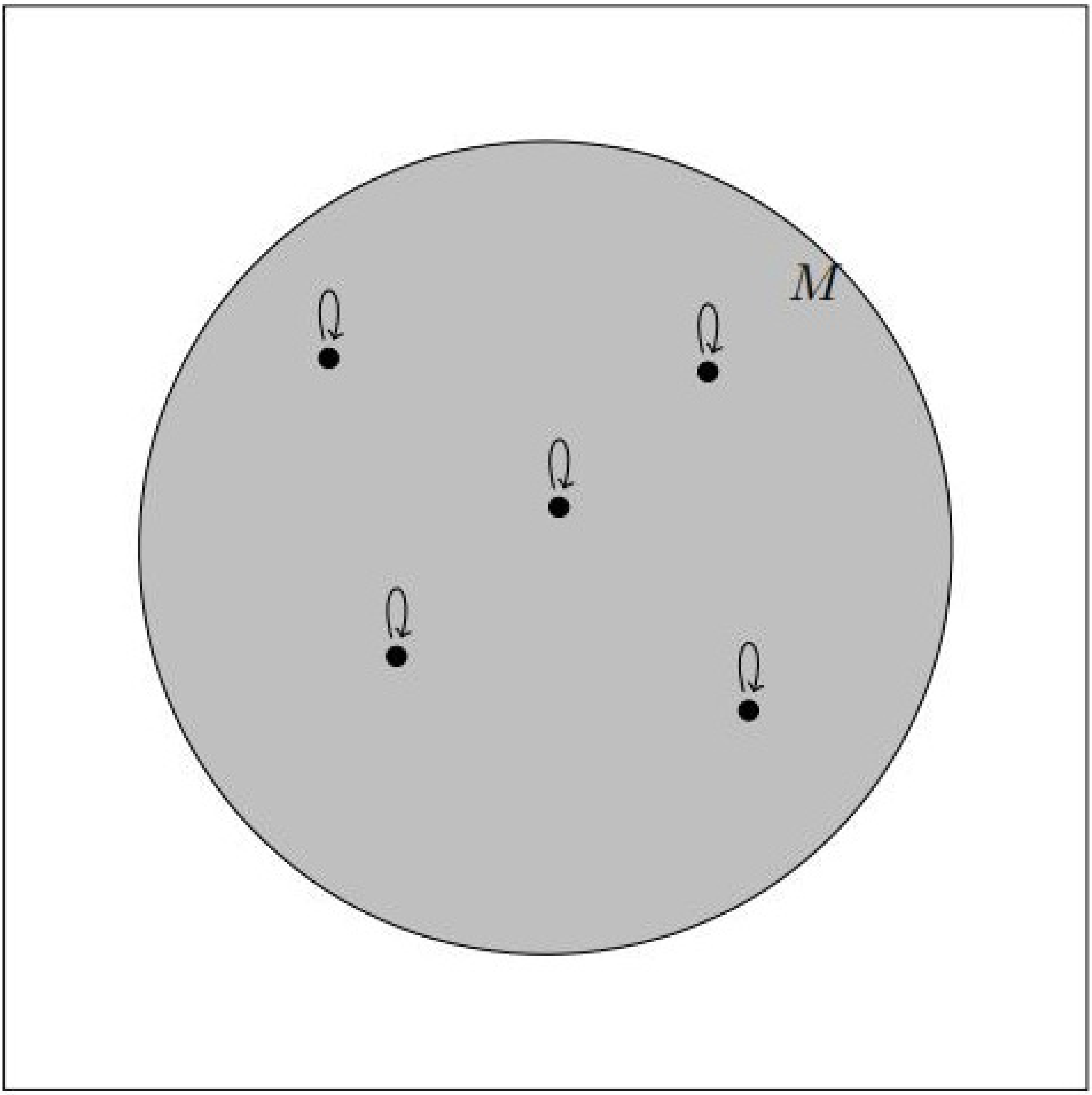

Figure 5: The identity $Id_M : M \to M$

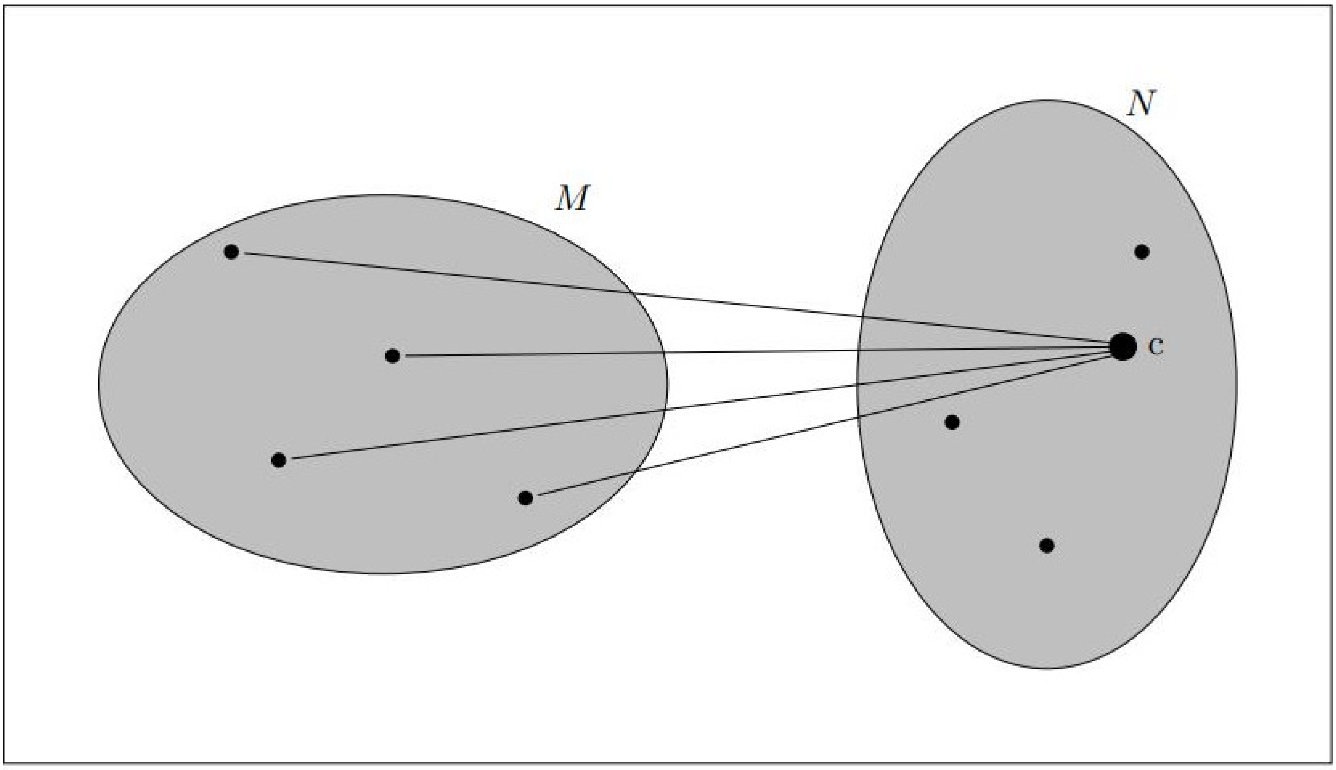

Figure 6: The constant function $f \equiv c$

Definition 1.19. Let $f : M \to N$ be a function. We define the **graph** of f as

$$G_f := \{(x, f(x)) \; : x \in M\} \subseteq M \times N.$$

Definition 1.20. Let $f \colon M \to N$ and $g \colon N \to M$ be functions with the properties

$$\forall x \in M \colon g(f(x)) = x$$

and

$$\forall y \in N \colon f(g(y)) = y,$$

then f and g are **inverses** of each other.
In this case we write $f^{-1} := g$ and $g^{-1} := f$ and call f (or g) **invertible**.

Theorem 1.23. *Let $f \colon M \to N$ be a function. Then,*

$$f \text{ is invertible} \quad \Longleftrightarrow \quad f \text{ is bijective.}$$

We now discuss the concatenation of functions. Let X, Y, Z be nonempty sets, $f\colon X \to Y$ and $g\colon Y \to Z$ be functions. We then define a function $(g \circ f)\colon X \to Z$ by **composing** f and g, i.e., $(g \circ f)(x) := g(f(x))$. As an exercise check that $g \circ f$ is indeed a function.

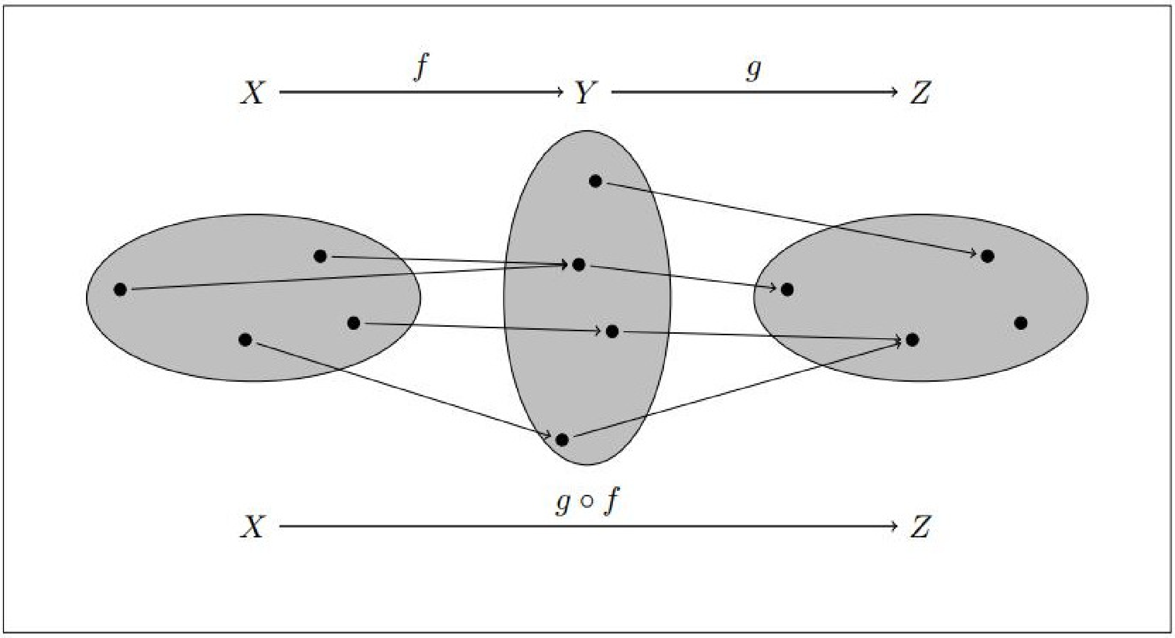

Figure 9: The composition $(g \circ f)(x) := g(f(x))$

Theorem 1.25. *Let $f : X \to Y$ and $g : Y \to Z$ be invertible functions. Then $g \circ f$ is invertible and*

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

1.4 Induction and combinatorics

Theorem 1.26 (Mathematical induction). *A predicate $P(n)$ is true for all $n \in \mathbb{N}$ if and only if the following two steps hold:*

a) ***Induction basis****: $P(1)$ is true.*

b) ***Induction step****: if $P(n)$ is true for arbitrary $n \in \mathbb{N}$, then also $P(n+1)$ is true.*

Definition 1.27. Let $a_k \in \mathbb{R}$ for $k \in \mathbb{N}$. Then we define sum and product as follows:

$$\sum_{k=1}^{1} a_k := a_1 \qquad \sum_{k=1}^{n+1} a_k := a_{n+1} + \sum_{k=1}^{n} a_k$$

$$\prod_{k=1}^{1} a_k := a_1 \qquad \prod_{k=1}^{n+1} a_k := a_{n+1} \cdot \prod_{k=1}^{n} a_k$$

k is called index.
A special case of products: $a^1 := a$ and $a^{n+1} := a \cdot a^n$.

Definition 1.28. The **factorial** $n!$ of a natural number $n \in \mathbb{N}$ is defined inductively as follows:

$$1! := 1 \qquad \text{and} \qquad \forall n \in \mathbb{N}\colon\ (n+1)! := (n+1) \cdot n!.$$

In addition, we set $0! := 1$.

The **binomial coefficient** $\binom{n}{k}$ (we say "n choose k") for $n, k \in \mathbb{N}_0$ with $n \geq k$ is defined by

$$\binom{n}{k} := \frac{n \cdot (n-1) \cdot (n-2) \cdots (n-k+2) \cdot (n-k+1)}{k!} = \frac{n!}{k!(n-k)!}.$$

Clearly, we have $\binom{n}{0} = 1$, $\binom{n}{1} = n$ and $\binom{n}{k} = \binom{n}{n-k}$.

Theorem 1.32. *Let $n, k \in \mathbb{N}$ with $k \leq n$. Then we have*

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

Theorem 1.33 (Binomial Theorem)**.** *Let $x, y \in \mathbb{R}$ and $n \in \mathbb{N}$. Then we have*

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

1.5 Special relations, infimum and supremum

Definition 1.36. Let $R \subseteq M^2$ be a relation for an arbitrary $M \neq \varnothing$. We call R

- **reflexive** if and only if
$$\forall x \in M\colon (x,x) \in R,$$
- **symmetric** if and only if
$$\forall (x,y) \in R\colon (y,x) \in R,$$
- **antisymmetric** if an only if
$$\forall x,y \in M\colon (x,y),(y,x) \in R \Longrightarrow x = y$$
- **transitive** if and only if
$$\forall x,y,z \in M\colon (x,y),(y,z) \in R \Longrightarrow (x,z) \in R$$
- **total** if an only if
$$\forall x,y \in M\colon (x,y) \in R \text{ or } (y,x) \in R.$$

Definition 1.37. A relation R is called

- **equivalence relation** if it is reflexive, symmetric and transitive,
- **partial order** if it is reflexive, antisymmetric and transitive,
- **total or linear order** if it is a partial order and total.

Definition 1.42. Let $A \subseteq \mathbb{R}$. We say A is

- **bounded from above** if and only if
$$\exists C \in \mathbb{R}\, \forall a \in A\colon a \leq C$$
and such a C an **upper bound** of A.
- **bounded from below** if and only if
$$\exists c \in \mathbb{R}\, \forall a \in A\colon c \leq a$$
and such a c a **lower bound** of A.
- **bounded** if and only if A is bounded from above and from below.

Definition 1.43. Let $a, b \in \mathbb{R}$. Then we define the **closed interval**

$$[a, b] := \{x \in \mathbb{R} \colon a \leq x \leq b\},$$

half open intervals

$$[a, b) := \{x \in \mathbb{R} \colon a \leq x < b\}$$

and

$$(a, b] := \{x \in \mathbb{R} \colon a < x \leq b\},$$

and **open interval**

$$(a, b) := \{x \in \mathbb{R} \colon a < x < b\}$$

between a and b.
Moreover, we write $[a, \infty) := \{x \in \mathbb{R} \colon a \leq x\}$ and $(\infty, b) := \{x \in \mathbb{R} \colon x < b\}$ etc.

Definition 1.45. Let $A \subset \mathbb{R}$ be nonempty and $T \in \mathbb{R}$. Then, T is called

- greatest lower bound or **infimum** of A, denoted by $\inf A := T$, if
  - $T \leq A$, i.e., T is a lower bound and
  - $x \leq A \implies x \leq T$, i.e., there is no greater lower bound.
- least upper bound or **supremum** of A, denoted by $\sup A := T$, if
  - $A \leq T$, i.e., T is an upper bound and
  - $A \leq x \implies T \leq x$, i.e., there is no smaller upper bound.

If A is not bounded from above (below) we set $\sup A := \infty$ ($\inf A := -\infty$).

Theorem 1.49. *The following assertions hold:*

a) ***Archimedean property:*** $\mathbb{N}$ *has no upper bound* $\iff \forall \varepsilon > 0 \; \exists n \in \mathbb{N} : \frac{1}{n} < \varepsilon$.

b) For $x, y \in \mathbb{R}$ *with* $x < y$ *there exists a rational number* $\frac{m}{n} \in \mathbb{Q}$ *with* $x \leq \frac{m}{n} \leq y$.

1.6 Special functions on R

Lemma 1.50. *Let $x, y, z \in \mathbb{R}$ with $z \geq 0$. Then, we have the following properties*

1. $|x| \geq 0$
2. $|x| = 0 \iff x = 0$
3. $|x| \geq x$
4. $|x \cdot y| = |x| \cdot |y|$

Theorem 1.52 (Triangle inequality). *Let $x, y \in \mathbb{R}$. Then we have that*

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

Corollary 1.53. *Let $x, y \in \mathbb{R}$. Then we have*

$$\Big||x| - |y|\Big| \leq |x - y|.$$

The **trigonometric functions** play a crucial role in analytic geometry. The most important are:

$\sin(x)$	(**sine**)
$\cos(x)$	(**cosine**)
$\tan(x)$	(**tangent**)
$\cot(x)$	(**cotangent**)

Figure 11 shows the 'geometric definition' of these functions in the unit circle.

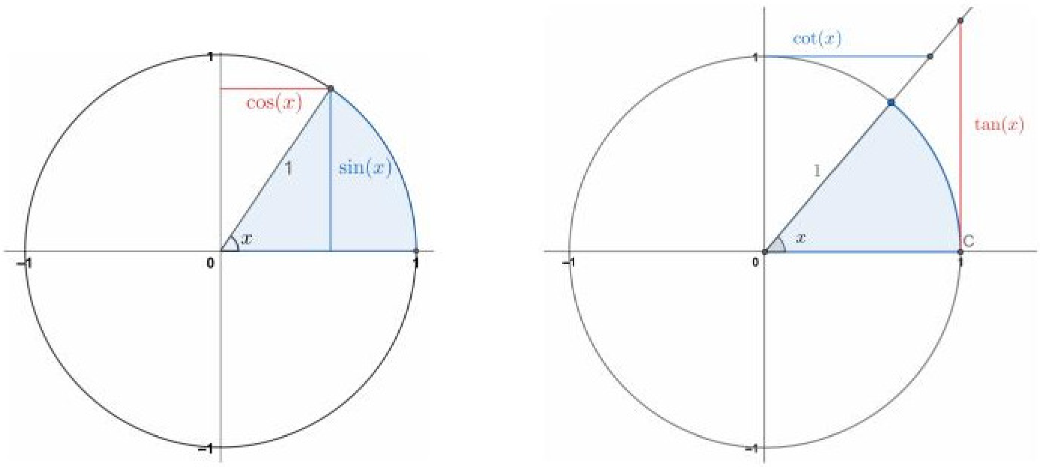

Figure 11: Illustration of trigonometric functions

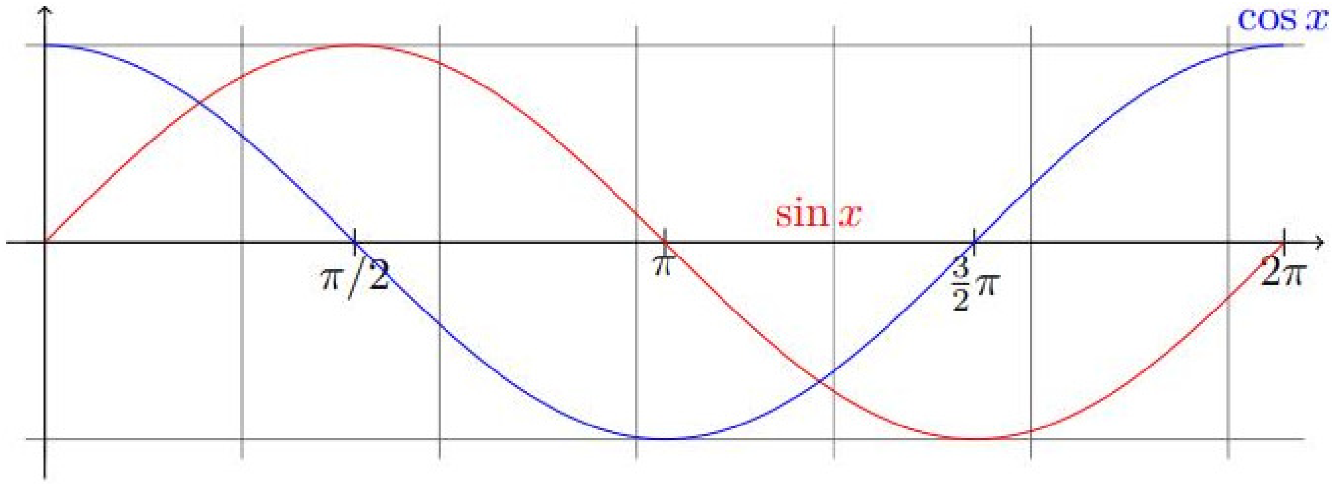

Figure 12: The graphs of sin and cos

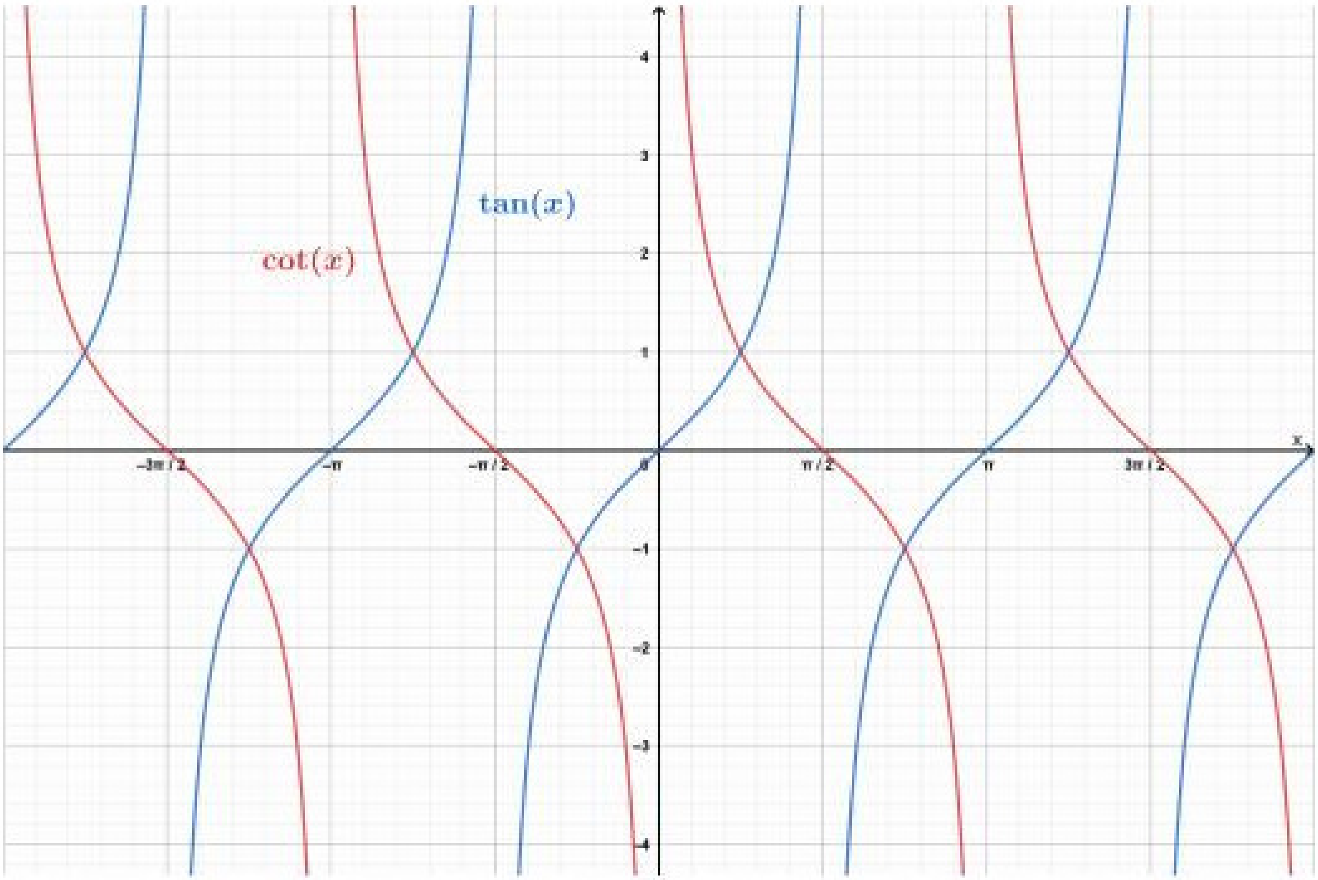

Figure 13: The graphs of tan and cot

Additionally, we have that most important **trigonometric identity**

$$\sin^2 x + \cos^2 x \; = \; 1,$$

and the **trigonometric addition formulas**:

$$\begin{aligned}
\sin(x+y) &= \sin x \cos y + \cos x \sin y \\
\cos(x+y) &= \cos x \cos y - \sin x \sin y \\
\sin x + \sin y &= 2 \sin\left(\frac{x+y}{2}\right) \cos\left(\frac{x-y}{2}\right) \\
\sin x - \sin y &= 2 \cos\left(\frac{x+y}{2}\right) \sin\left(\frac{x-y}{2}\right) \\
\cos x + \cos y &= 2 \cos\left(\frac{x+y}{2}\right) \cos\left(\frac{x-y}{2}\right) \\
\cos x - \cos y &= -2 \sin\left(\frac{x+y}{2}\right) \sin\left(\frac{x-y}{2}\right)
\end{aligned}$$

From them the following useful formulas may be deduced:

$$\begin{aligned}
\cos(0) &= \sin\left(\frac{\pi}{2}\right) = 1 & \qquad \cos(\pi) &= \sin\left(\frac{3\pi}{2}\right) = -1 \\
\cos(2x) &= \cos^2(x) - \sin^2(x) & \sin(2x) &= 2\sin(x)\cos(x) \\
1+\cos(2x) &= 2\cos^2(x) & 1+\sin(2x) &= (\sin(x)+\cos(x))^2 \\
1-\cos(2x) &= 2\sin^2(x) & 1-\sin(2x) &= (\sin(x)-\cos(x))^2
\end{aligned}$$

Finally a special combination of exponential functions yield the **hyperbolic functions**, defined in $\mathbb{R}$ by:

$$\sinh x := \frac{e^x - e^{-x}}{2} \qquad \textbf{(hyperbolic sine)}$$

$$\cosh x := \frac{e^x + e^{-x}}{2} \qquad \textbf{(hyperbolic cosine)}$$

$$\tanh x := \frac{\sinh x}{\cosh x} \qquad \textbf{(hyperbolic tangent)}$$

$$\coth x := \frac{\cosh x}{\sinh x} \qquad \textbf{(hyperbolic cotangent)}$$

It obviously holds that $\cosh^2 x - \sinh^2 x = 1$.

1.7 Complex numbers

Definition 1.54. The set of all **complex numbers** is defined by

$$\mathbb{C} := \{z = x + iy : x, y \in \mathbb{R}\}.$$

The representation of the term $z = x + iy$ is called the **canonical representation**.

For a complex number $z = x + iy$ we call x the **real part** of z and y the **imaginary part**. We write $\operatorname{Re} z = x$ and $\operatorname{Im} z = y$.

Let $z = x + iy \in \mathbb{C}$. We define the **complex conjugate** of z by

$$\overline{z} := x - iy.$$

In $\mathbb{C}$ we have the following calculation rules for $z = x + iy$ and $w = u + iv$:

- $z + w = (x + u) + i(y + v)$
- $zw = (xu - yv) + i(xv + uy)$

Definition 1.56. Let $z = x + iy \in \mathbb{C}$. We define the **absolute value** of z by

$$|z| := \sqrt{z\overline{z}} = \sqrt{x^2 + y^2}.$$

Lemma 1.57. *Let $z, w \in \mathbb{C}$. Then the following holds*

1. $|z| \geq 0$
2. $|z| = 0 \Leftrightarrow z = 0$
3. $|z| \geq |\operatorname{Re} z|$
4. $|z| \geq |\operatorname{Im} z|$
5. $|zw| = |z||w|$
6. $|z + w| \leq |z| + |w|$ *(triangle inequality)*
7. $\Big| |z| - |w| \Big| \leq |z - w|$

An easy and useful way of writing complex numbers can be obtained by using **Euler's formula**

$$e^{i\varphi} = \cos\varphi + i\sin\varphi.$$

To work with this formula, it is essential to mind the following important values:

φ	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{3\pi}{4}$	$\frac{5\pi}{6}$	π
$\sin(\varphi)$	$\frac{\sqrt{0}}{2}$	$\frac{\sqrt{1}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{4}}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{1}}{2}$	$\frac{\sqrt{0}}{2}$
$\cos(\varphi)$	$\frac{\sqrt{4}}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{1}}{2}$	$\frac{\sqrt{0}}{2}$	$\frac{-\sqrt{1}}{2}$	$\frac{-\sqrt{2}}{2}$	$\frac{-\sqrt{3}}{2}$	$\frac{-\sqrt{4}}{2}$

All remaining important values can be obtained from

$$\sin(x+\pi) = -\sin(x) \qquad \text{and} \qquad \cos(x+\pi) = -\cos(x)$$

Using these values of the trigonometric functions at the points $\frac{\pi}{2}, \pi, \frac{3\pi}{2}, 2\pi$ we obtain

$$e^{\frac{i\pi}{2}} = i \qquad e^{i\pi} = -1 \qquad e^{i\frac{3\pi}{2}} = -i \qquad e^{i2\pi} = e^{i0} = 1.$$

Theorem 1.59 (Fundamental theorem of algebra). *Let $a_k \in \mathbb{C}$ for $k = 0, 1, 2, \ldots, n$ such that $a_n \neq 0$. Then, we have the equality*

$$\sum_{k=0}^{n} a_k z^k = a_n \cdot \prod_{k=1}^{n} (z - z_k)$$

for some (not necessarily different) $z_1, \ldots, z_n \in \mathbb{C}$.
*Clearly, each z_k is a **root** of the polynomial $\sum_{k=0}^{n} a_k z^k$, i.e., $\sum_{k=0}^{n} a_k z_\ell^k = 0$ for all $\ell = 1, \ldots, n$.*

2 Sequences

Definition 2.1 (Sequence). Let $M \neq \emptyset$ be a set. An **(infinite) sequence** (in M) is a mapping $a\colon \mathbb{N} \to M$. With the notation $a_n := a(n)$, we can write the sequence as

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

For a sequence $(a_n)_{n\in\mathbb{N}}$ its **range** is given by

$$\{a_n\colon n \in \mathbb{N}\}.$$

The domain of a sequence (here mostly $\mathbb{N}$) is called the **index set** of the sequence.

We write $(a_n)_{n\in\mathbb{N}} \subset M$ to say that $\forall n \in \mathbb{N}\colon a_n \in M$, and write $M^{\mathbb{N}}$ for set of all sequences in M.

2.1 Convergence of sequences

Definition 2.4 (Neighborhood). Let $a \in \mathbb{R}$ and $\varepsilon > 0$. We define the ε-**neighborhood** of a as

$$U_\varepsilon(a) := \{x \in \mathbb{R} : |x - a| < \varepsilon\}.$$

Definition 2.5 (Convergence). Let $(a_n)_{n\in\mathbb{N}}$ be a sequence and $a \in \mathbb{C}$.
We say that the sequence $(a_n)_{n\in\mathbb{N}}$ **converges** to a if and only if

$$\forall \varepsilon > 0\, \exists n_0 \in \mathbb{N}\, \forall n \geq n_0\colon |a_n - a| < \varepsilon,$$

or equivalently

$$\forall \varepsilon > 0\, \exists n_0 \in \mathbb{N}\, \forall n \geq n_0\colon a_n \in U_\varepsilon(a).$$

In this case we call a the **limit** of the sequence and write

$$a = \lim_{n\to\infty} a_n \quad \text{or} \quad a_n \xrightarrow[n\to\infty]{} a \quad \text{or simply} \quad a_n \to a.$$

$(a_n)_{n\in\mathbb{N}}$ is called **convergent**, or we say that **the limit of** $(a_n)_{n\in\mathbb{N}}$ **exists**, if there exists some $a \in \mathbb{C}$ such that $a_n \to a$, otherwise $(a_n)_{n\in\mathbb{N}}$ is called **divergent**.

Definition 2.8 (Null sequence). Let $(a_n)_{n\in\mathbb{N}}$ be a real sequence such that

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

Then we call $(a_n)_{n\in\mathbb{N}}$ a **null sequence**.

Definition 2.12. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence.
We call the sequence **bounded** if

$$\exists R > 0 \ \forall n \in \mathbb{N} \colon |a_n| \le R.$$

Moreover, if $(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}$, we call the sequence **bounded from above** if and only if

$$\exists C \in \mathbb{R} \ \forall n \in \mathbb{N} \colon a_n \le C,$$

and **bounded from below** if and only if

$$\exists c \in \mathbb{R} \ \forall n \in \mathbb{N} \colon a_n \ge c.$$

Theorem 2.14. *Let $(a_n)_{n\in\mathbb{N}}$ be a convergent sequence. Then $(a_n)_{n\in\mathbb{N}}$ is bounded.*

2.2 Calculation rules for limits

Lemma 2.18. *Let $x \ge 0$. Then, for any $k \in \{1, \dots, n\}$, we have*

$$(1+x)^n \ge 1 + \binom{n}{k} x^k$$

Theorem 2.21. *Let $(a_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}}$ be convergent sequences, and let $\lambda \in \mathbb{C}$. Moreover, let $a = \lim_{n\to\infty} a_n$ and $b = \lim_{n\to\infty} b_n$. Then, we have*

(i) $\lim_{n\to\infty}(a_n + b_n) = a + b$

(ii) $\lim_{n\to\infty}(\lambda \cdot a_n) = \lambda \cdot a$

(iii) $\lim_{n\to\infty}(a_n \cdot b_n) = a \cdot b$

(iv) If $b \neq 0 (\Rightarrow \exists N_0 \in \mathbb{N} \forall n \ge N_0 : b_n \neq 0)$, then $\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{a}{b}$.

Theorem 2.25 (Sandwich rule). *Let $(a_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ be convergent real-valued sequences and let $(b_n)_{n\in\mathbb{N}}$ be a sequence such that*

$$a_n \le b_n \le c_n \qquad \text{for all } n \in \mathbb{N}.$$

If additionally,

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

then $(b_n)_{n\in\mathbb{N}}$ is convergent with

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

Definition 2.29. Let $(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}$. The sequence $(a_n)_{n\in\mathbb{N}}$ **tends to** ∞ $(= +\infty)$ if

$$\forall A > 0\ \exists n_0 \in \mathbb{N}\ \forall n \geq n_0\colon\ a_n > A.$$

We write $a_n \to \infty$ or $\lim\limits_{n\to\infty} a_n = \infty$, and call ∞ the **improper limit**.

The tendency to $-\infty$ is defined analogously with $a_n < -A$.

If the sequence (a_n) tends to $\pm\infty$, it is called **definitely divergent**.

The following (easy to prove) calculation rules hold for definite divergence:

$$\begin{aligned}
&a_n \to \infty,\ b_n \to \infty && \Longrightarrow && a_n + b_n \to \infty \text{ and } a_n \cdot b_n \to \infty\\
&a_n \to \infty,\ b_n \to b && \Longrightarrow && a_n + b_n \to \infty\\
&a_n \to \infty,\ \alpha \in \mathbb{R} && \Longrightarrow && \frac{\alpha}{a_n} \to 0\\
&a_n \to \infty,\ \alpha > 0 && \Longrightarrow && \alpha \cdot a_n \to \infty\\
&a_n \to \infty,\ \alpha < 0 && \Longrightarrow && \alpha \cdot a_n \to -\infty
\end{aligned}$$

If $a_n \to \infty$ and $b_n \to \infty$, no general rule can be given for $(a_n - b_n)$ and $\left(\frac{a_n}{b_n}\right)$. Therefore, also the limit of $(a_n b_n)$ for $a_n \to 0$ and $b_n \to \infty$ needs more care, and these limits do not have to exist nor be definitely divergent, consider e.g. $a_n = (-1)^n/n$ and $b_n = n$.

Definition 2.31 (Monotone sequences)**.** A real-valued sequence $(a_n)_{n\in\mathbb{N}}$ is called

- **increasing** if and only if
$$\forall n \in \mathbb{N} : a_{n+1} > a_n,$$
- **non-decreasing** if and only if
$$\forall n \in \mathbb{N} : a_{n+1} \geq a_n,$$
- **decreasing** if and only if
$$\forall n \in \mathbb{N} : a_{n+1} < a_n,$$
- **non-increasing** if and only if
$$\forall n \in \mathbb{N} : a_{n+1} \leq a_n.$$

Moreover, we say that a sequence is **monotone** if it is non-increasing or non-decreasing, or **strictly monotone** if it is either increasing or decreasing.

Theorem 2.34 (Monotonicity principle). *Let $(a_n)_{n\in\mathbb{N}}$ be a monotone sequence. Then,*

$$(a_n) \text{ is convergent} \iff (a_n) \text{ is bounded.}$$

In particular, if $(a_n)_{n\in\mathbb{N}}$ is non-decreasing (or increasing) then

$$\lim_{n\to\infty} a_n = \sup\{a_n : n \in \mathbb{N}\} =: \sup(a_n),$$

and if $(a_n)_{n\in\mathbb{N}}$ is non-increasing (or decreasing) then

$$\lim_{n\to\infty} a_n = \inf\{a_n : n \in \mathbb{N}\} =: \inf(a_n).$$

2.3 Subsequences and accumulation points

Definition 2.38. Let $(n_1, n_2, n_3, \dots)$ be an increasing sequence of natural numbers and $(a_n)_{n=1}^{\infty}$ be a sequence. Then we call

$$(a_{n_k})_{k=1}^{\infty} = (a_{n_1}, a_{n_2}, \dots)$$

a **subsequence** of $(a_n)_{n=1}^{\infty}$.

Definition 2.40. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence. We call $a \in \mathbb{C}$ an **accumulation point** of $(a_n)_{n\in\mathbb{N}}$ if there exists a subsequence $(a_{n_k})_{k\in\mathbb{N}}$ of $(a_n)_{n\in\mathbb{N}}$ with

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

Equivalently, we may use the definitions

$$\forall \varepsilon > 0\colon \#\{n\colon a_n \in U_\varepsilon(a)\} = \infty$$

or

$$\forall \varepsilon > 0 \; \forall n_0 \in \mathbb{N} \; \exists n \geq n_0\colon |a - a_n| < \varepsilon.$$

(Note the interchanged quantifiers compared to the definition of convergence.)

Theorem 2.42 (Bolzano-Weierstrass). *Let $(a_n)_{n=1}^{\infty}$ be a bounded sequence. Then $(a_n)_{n=1}^{\infty}$ has at least one convergent subsequence.*

Theorem 2.43 (Bolzano-Weierstrass). *Let $(a_n)_{n=1}^{\infty}$ be a bounded sequence. Then $(a_n)_{n=1}^{\infty}$ has at least one accumulation point.*

Definition 2.47. Let $(a_n)_{n\in\mathbb{N}}$ be a real-valued sequence.
Then we define the **limes inferior** of $(a_n)_{n\in\mathbb{N}}$ by

$$\liminf_{n\to\infty} a_n := \sup_{n\in\mathbb{N}} \inf_{k\geq n} a_k,$$

and the **limes superior** by

$$\limsup_{n\to\infty} a_n := \inf_{n\in\mathbb{N}} \sup_{k\geq n} a_k.$$

If $(a_n)_{n\in\mathbb{N}}$ is bounded and A is the set of all accumulation points, then we can also use

$$\liminf_{n\to\infty} a_n = \inf A,$$

and

$$\limsup_{n\to\infty} a_n = \sup A.$$

Corollary 2.52. *A sequence $(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ is convergent (or definitely divergent) if and only if*

$$\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n$$

In this case,

$$\lim_{n\to\infty} a_n = \liminf_{n\to\infty} a_n \quad \left(= \limsup_{n\to\infty} a_n\right).$$

2.4 Cauchy criterion

Definition 2.54. A sequence $(a_n)_{n\in\mathbb{N}}$ is called a **Cauchy sequence** if

$$\forall \varepsilon > 0\ \exists n_0 \in \mathbb{N}\ \forall n, m \geq n_0 \colon |a_n - a_m| \leq \varepsilon.$$

That is, the terms of a Cauchy sequence are *pairwise* close to each other for large n.

Theorem 2.56 (Cauchy criterion). *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence. Then,*

$$(a_n)_{n\in\mathbb{N}} \text{ is convergent} \quad \Longleftrightarrow \quad (a_n)_{n\in\mathbb{N}} \text{ is a Cauchy sequence.}$$

3 Series

Definition 3.1. Let $(a_n)_{n=1}^{\infty}$ be a sequence and

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

Then we call s_n the **n-th partial sum** of the **(infinite) series**

$$\sum_{k=1}^{\infty} a_k \qquad \text{or} \qquad a_1 + a_2 + \dots \qquad \text{or just} \qquad \sum a_k.$$

If the sequence of partial sums (s_n) converges to $s \in \mathbb{C}$, i.e., $s_n \to s$, then we call the series **convergent**, call s the **sum** of the series, and write

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

Otherwise we call the series **divergent**.

If $s_n \to \pm\infty$ we also write $\sum_{k=1}^{\infty} a_k = \pm\infty$, and say that the series is **definitely divergent**.

Theorem 3.9. *Let $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=1}^{\infty} b_k$ be convergent series, and let $c \in \mathbb{C}$. Then we have*

$$\sum_{k=1}^{\infty}(a_k + b_k) = \sum_{k=1}^{\infty} a_k + \sum_{k=1}^{\infty} b_k$$

and

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

Theorem 3.10. *Let $(a_n)_{n=1}^{\infty}$ be a non-negative sequence, i.e. $a_k \geq 0$ for all $k \in \mathbb{N}$. Then, the sequence of partial sums is bounded, i.e.,*

$$\exists C \in \mathbb{R}\, \forall n \in \mathbb{N}\colon |s_n| \leq C,$$

if and only if the series $\sum_{k=1}^{\infty} a_k$ converges.

Theorem 3.12 (Cauchy criterion). *Let $\sum_{k=1}^{\infty} a_k$ be a series. Then we have that $\sum_{k=1}^{\infty} a_k$ is convergent if and only if*

$$\forall \varepsilon > 0\, \exists n_0 \in \mathbb{N}\, \forall m > n \geq n_0\colon \left| \sum_{k=n+1}^{m} a_k \right| < \varepsilon.$$

In other words, the series $\sum_{k=1}^{\infty} a_k$ is convergent if and only if the sequence of partial sums is a Cauchy sequence.

Corollary 3.13. *We have*

$$\sum_{k=1}^{\infty} a_k \quad \text{is convergent} \qquad \Longrightarrow \qquad \lim_{k\to\infty} a_k = 0,$$

i.e., the terms $(a_k)_{k\in\mathbb{N}}$ of a convergent series are a null sequence.

In particular, if (a_n) is not a null sequence, then $\sum a_k$ is divergent.

3.1 Convergence tests

Definition 3.16. We say that the series $\sum_{k=1}^{\infty} a_k$ is **absolutely convergent** if and only if the series of absolute values of a_n is convergent, i.e., $\sum_{k=1}^{\infty} |a_k|$ is a convergent series. In this case we write

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

Theorem 3.19. *Absolutely convergent series are also convergent.*

3.1.1 Comparison test

Theorem 3.20. *Let $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=1}^{\infty} b_k$ be series.*

(i) If $\sum b_k$ is absolutely convergent, and $|a_k| \leq |b_k|$ for all $k \in \mathbb{N}$.
Then, $\sum a_k$ is also absolutely convergent.

(ii) If $\sum b_k = \infty$, and $0 \leq b_k \leq a_k$ for all $k \in \mathbb{N}$, then also $\sum a_k = \infty$.

3.1.2 Root test

Theorem 3.21 (root test). *Let $\sum_{k=1}^{\infty} a_k$ be a series.*

(i) If there exists some $k_0 \in \mathbb{N}$ and $q \in [0,1)$ such that for all $k \geq k_0$ we have

$$\sqrt[k]{|a_k|} \leq q,$$

then $\sum_{k=1}^{\infty} a_k$ is absolutely convergent.

(ii) Conversely, if there exists some $k_0 \in \mathbb{N}$ and $q > 1$ such that for all $k \geq k_0$ we have

$$\sqrt[k]{|a_k|} \geq q,$$

then $\sum a_k$ is not absolutely convergent (and therefore divergent if $a_k \geq 0$).

3.1.3 Ratio test

Theorem 3.26 (ratio test). *Let $\sum_{k=1}^{\infty} a_k$ be a series.*

(i) If there exists some $k_0 \in \mathbb{N}$ and $q \in [0,1)$ such that for all $k \geq k_0$ we have

$$a_k \neq 0 \qquad \text{and} \qquad \left|\frac{a_{k+1}}{a_k}\right| \leq q,$$

then $\sum_{k=1}^{\infty} a_k$ is absolutely convergent.

(ii) Conversely, if there exists some $k_0 \in \mathbb{N}$ and $q > 1$ such that for all $k \geq k_0$ we have

$$a_k \neq 0 \qquad \text{and} \qquad \left|\frac{a_{k+1}}{a_k}\right| \geq q,$$

then $\sum a_k$ is not absolutely convergent (and therefore divergent if $a_k \geq 0$).

3.1.4 Cauchy's condensation test

Theorem 3.32 (Cauchy's condensation test). *Let $\sum_{k=1}^{\infty} a_k$ be a series with $0 \leq a_{k+1} \leq a_k$ for all k. Then,*

$$\sum_{k=1}^{\infty} a_k \quad \text{is convergent} \qquad \Longleftrightarrow \qquad \sum_{k=1}^{\infty} 2^k \, a_{2^k} \quad \text{is convergent} \; .$$

3.1.5 Leibniz criterion

Theorem 3.34 (Leibniz criterion). *Let $(a_k)_{k\in\mathbb{N}}$ be monotone with $a_k \to 0$. Then we have that*

$$\sum_{k=1}^{\infty} (-1)^k a_k \qquad \text{is convergent.}$$

4 Continuous functions and limits

Definition 4.1. Let $D \subset \mathbb{R}$, $x_0 \in D$ and $f\colon D \to \mathbb{R}$. We call f **continuous at** x_0 if for all sequences $(x_n)_{n\in\mathbb{N}}$ with $x_n \to x_0$ we have that $\lim_{n\to\infty} f(x_n)$ exists and

$$\lim_{n\to\infty} f(x_n) = f(x_0) = f\left(\lim_{n\to\infty} x_n\right).$$

If $U \subset D$ and f is continuous at all $x \in U$ we call f continuous on U, and

if f is continuous at all $x \in D$, then we just call f **continuous**.

Theorem 4.5 (ε-δ-criterion). *Let $f\colon D \to \mathbb{R}$ and $x_0 \in D$. Then, f is continuous at x_0 if and only if*

$$\forall \varepsilon > 0\, \exists \delta > 0\, \forall x \in D\colon |x - x_0| < \delta \implies |f(x) - f(x_0)| < \varepsilon.$$

In words: Given $x_0 \in D$. For all (fixed) $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x \in D$ with $|x - x_0| < \delta$ we have that $|f(x) - f(x_0)| < \varepsilon$.

Definition 4.8. Let $f\colon D \to \mathbb{R}$ be a real function. We call f **uniformly continuous** if for all sequences (x_n),$(y_n) \subset D$ with $\lim_{n\to\infty} |x_n - y_n| = 0$ we have that

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

Theorem 4.10 (ε-δ-criterion for uniform continuity). *Let $f\colon D \to \mathbb{R}$. Then, f is uniformly continuous if and only if*

$$\forall \varepsilon > 0\, \exists \delta > 0\, \forall x, y \in D\colon |x - y| < \delta \implies |f(x) - f(y)| < \varepsilon.$$

In words: For all (fixed) $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x, y \in D$ with $|x - y| < \delta$ we have that $|f(x) - f(y)| < \varepsilon$.

Theorem 4.12. *Let $a, b \in \mathbb{R}$ with $a < b$. A continuous function $f\colon [a, b] \to \mathbb{R}$ on a closed interval $[a, b]$ is uniformly continuous.*

Definition 4.14. Let $f\colon D \to \mathbb{R}$. We call f **Lipschitz continuous** if there exists some $L > 0$ such that

$$\forall x, y \in D\colon\ |f(x) - f(y)| \leq L|x - y|.$$

The constant L is called **Lipschitz constant**.

Theorem 4.17. *Let $f : D \to \mathbb{R}$. Then,*

$$f \text{ is Lipschitz continuous} \implies f \text{ is uniformly continuous} \implies f \text{ is continuous}$$

4.1 Calculation rules of continuous functions

Theorem 4.20 (Calculation rules for continuous functions). *Let $f, g : D \to \mathbb{R}$ be continuous in $x_0 \in D$ and $c \in \mathbb{R}$. Then $f + g$, $f \cdot g$ and $c \cdot f$ are continuous in x_0.*
If additionally $g(x_0) \neq 0$, then $\frac{f}{g}$ is also continuous in x_0.

Theorem 4.24. *Let $D, E \subset \mathbb{R}$. Moreover, let $f \colon D \to E$ be continuous at $x_0 \in D$, and $g \colon E \to \mathbb{R}$ be continuous at $y_0 = f(x_0) \in E$. Then, $g \circ f$ is continuous at x_0.*

Theorem 4.28. *Let $f \colon [a, b] \to D \subset \mathbb{R}$ be a bijective function. If f is continuous on $[a, b]$, then the inverse function f^{-1} is continuous on D.*

4.2 Intermediate and extreme value theorem

Theorem 4.30 (Intermediate value theorem). *Let $I = [a, b]$ be a closed interval and $f \colon I \to \mathbb{R}$ be a continuous function. Then, for every $y \in \mathbb{R}$ with*

$$\min\{f(a), f(b)\} \leq y \leq \max\{f(a), f(b)\},$$

there exists some $x \in I$ such that

$$f(x) = y.$$

Theorem 4.35 (Extreme value theorem). *Let $I = [a, b] \subset \mathbb{R}$ be a closed interval and $f : I \to \mathbb{R}$ be a continuous function. Then there exist $x_{\min}, x_{\max} \in I$ such that*

$$\begin{aligned} f(x_{\min}) &= \inf_{x \in I} f(x) \\ f(x_{\max}) &= \sup_{x \in I} f(x). \end{aligned}$$

In other words, continuous functions attain their extreme values on closed intervals.

In particular, continuous functions on closed intervals are bounded, i.e., $\sup_{x \in I} |f(x)| < \infty$.

4.3 Limits of functions

Definition 4.40 (Accumulation point of a set). Let $M \subset \mathbb{R}$ be a non-empty set. We call $x_0 \in \mathbb{R} \cup \{-\infty, +\infty\}$ an **accumulation point of M** if there exists a sequence $(x_n)_{n\in\mathbb{N}}$ in M such that

$$x_0 = \lim_{n\to\infty} x_n \quad \text{and} \quad x_n \neq x_0 \text{ for all } n \in \mathbb{N}.$$

An equivalent definition for $x_0 \in \mathbb{R}$ is

$$\forall \varepsilon > 0\colon \left(B_\varepsilon(x_0) \setminus \{x_0\}\right) \cap M \neq \varnothing,$$

i.e., every ε-neighborhood $B_\varepsilon(x_0)$ around x_0 contains a point of M different from x_0.

Definition 4.44 (Limit of functions). Let $D \subset \mathbb{R}$ and $f\colon D \to \mathbb{R}$. Moreover, let $y \in \mathbb{R}$ and $x_0 \in \mathbb{R} \cup \{\infty, -\infty\}$ be an accumulation point of D. We call y the **limit of f as $x \to x_0$**, if for arbitrary sequences (x_n) in D such that $x_n \to x_0$ and $x_n \neq x_0$ for all $n \in \mathbb{N}$, we have

$$f(x_n) \to y.$$

In this case we use the notation

$$\lim_{x\to x_0} f(x) = y.$$

The case $y = \pm\infty$ is called **improper limit**.

Definition 4.54 (One-sided limits). Let $D \subset \mathbb{R}$, $y \in \mathbb{R}$ and $f\colon D \to \mathbb{R}$.

- Let x_0 be an accumulation point of $D_+ := D \cap (x_0, \infty)$.
  We say that f has a **right-handed limit y as $x \to x_0$**, if for arbitrary sequences $(x_n) \subset D_+$ with $x_n \to x_0$ we have that

  $$f(x_n) \to y.$$

  We use the notation $\lim_{x\searrow x_0} f(x) = y$ or $\lim_{x\to x_0^+} f(x) = y$.

- Let x_0 be an accumulation point of $D_- := D \cap (-\infty, x_0)$.
  We say that f has a **left-handed limit y as $x \to x_0$**, if for arbitrary sequences $(x_n) \subset D_-$ with $x_n \to x_0$ we have that

  $$f(x_n) \to y.$$

  We use the notation $\lim_{x\nearrow x_0} f(x) = y$ or $\lim_{x\to x_0^-} f(x) = y$.

Again, the case $y \in \pm\infty$ is also allowed and called **improper limit**.

Theorem 4.58. *Let $D \subset \mathbb{R}$, $y \in \mathbb{R}$ and $f\colon D \to \mathbb{R}$.*
Moreover, let x_0 be an accumulation point of $(-\infty, x_0) \cap D$ and of $(x_0, \infty) \cap D$.
Then,

$$\lim_{x \to x_0} f(x) \text{ exists} \quad \Longleftrightarrow \quad \lim_{x \searrow x_0} f(x) \text{ and } \lim_{x \nearrow x_0} f(x) \text{ exist and are equal.}$$

In this case, we have $\lim_{x \to x_0} f(x) = \lim_{x \nearrow x_0} f(x) = \lim_{x \searrow x_0} f(x)$.

Corollary 4.59. *Let $D \subset \mathbb{R}$, $f\colon D \to \mathbb{R}$ and $x_0 \in D$ be an accumulation point of D. It holds*

$$f \text{ is continuous at } x_0 \quad \Longleftrightarrow \quad \lim_{x \to x_0} f(x) = f(x_0).$$

If both one-sided limits of f at x_0 are well defined, then

$$f \text{ is continuous at } x_0 \quad \Longleftrightarrow \quad \lim_{x \nearrow x_0} f(x) = \lim_{x \searrow x_0} f(x) = f(x_0)$$

4.4 Precise definitions and properties of special functions

Theorem 4.61. *Let $x \in \mathbb{R}$ and (x_n) be a sequence in $\mathbb{Q}$ with $\lim_{n\to\infty} x_n = x$.*
Then (a^{x_n}) converges. Moreover, the limit only depends on x but not on the sequence (x_n).

Definition 4.62. Let $x \in \mathbb{R} \setminus \mathbb{Q}$. Then we define

$$a^x := \lim_{n \to \infty} a^{x_n},$$

where (x_n) is a sequence of rational numbers such that $x_n \to x$.

Lemma 4.63. *For all $x < 1$ we have*

$$1 + x \leq e^x \leq \frac{1}{1-x}.$$

The lower bound holds for all $x \in \mathbb{R}$.

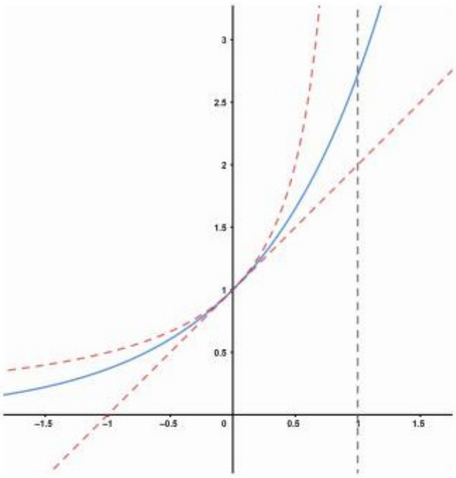

Figure 28: Inequality from Lemma 4.63

5 Differential calculus

Definition 5.1. Let $I = (a, b)$ and $f : I \to \mathbb{R}$. We call f **differentiable at** $x_0 \in I$ if

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \qquad \text{exists.}$$

In this case we call this limit **derivative of** f **at** x_0, and write $f'(x_0)$ or $\frac{d}{dx} f(x_0)$or $\frac{df}{dx}(x_0)$.

If f is differentiable at every point of I, we call f differentiable (in I) and denote by f' or $\frac{d}{dx} f$ the derivative of f.

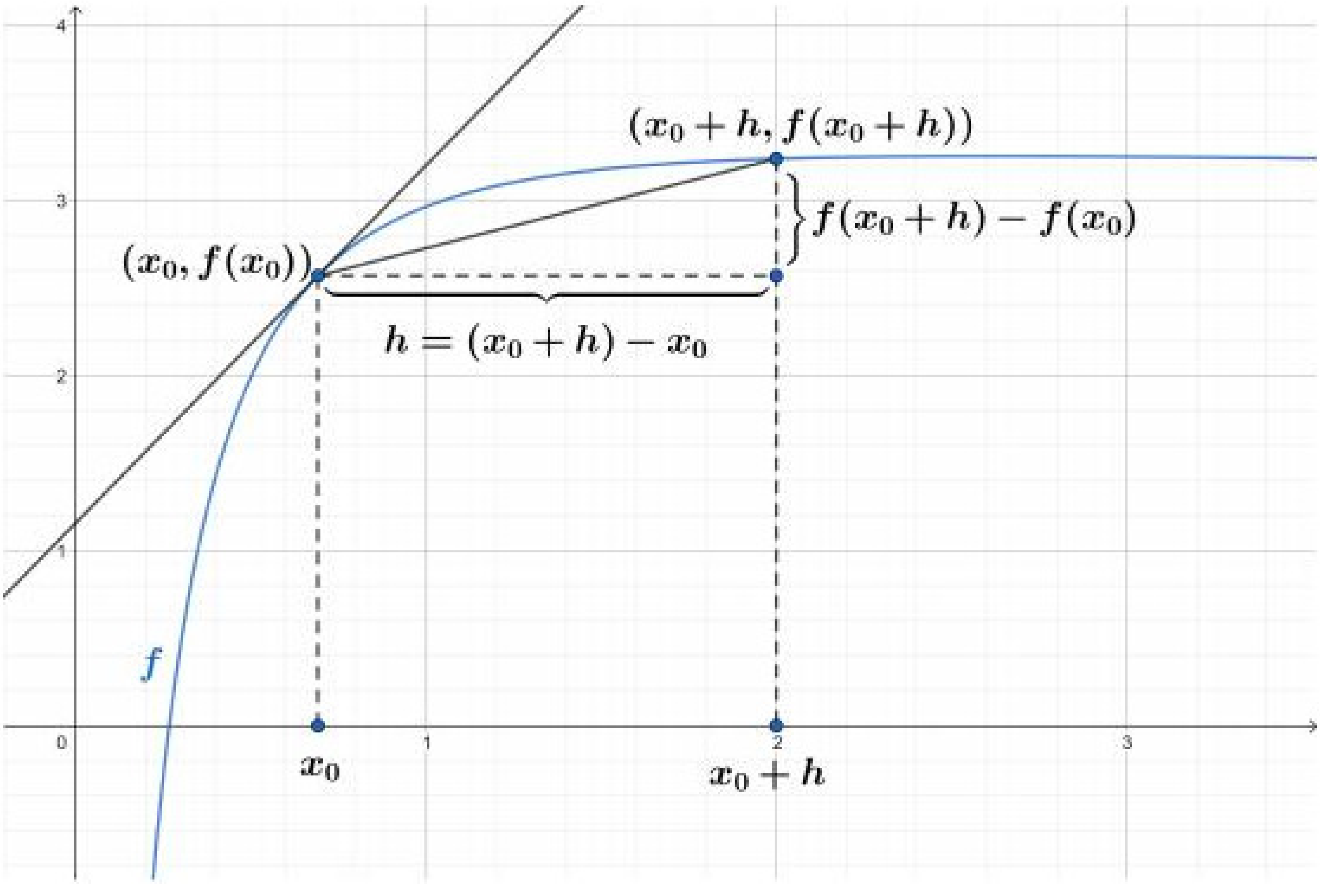

Figure 31: Difference quotient

Theorem 5.9. *Let $f : (a, b) \to \mathbb{R}$ be differentiable at x_0. Then f is continuous at x_0.*

5.1 Calculation rules for differentiable functions

Theorem 5.10 (Linearity of derivatives). *Let f, g be differentiable at x_0. Then,*

- *$(f+g)'(x_0)$ exists and $(f+g)'(x_0) = f'(x_0) + g'(x_0)$, and*
- *for any $c \in \mathbb{R}$ we have $(c \cdot f)'(x_0) = c \cdot f'(x_0)$.*

Theorem 5.12 (Product rule). *Let f, g be differentiable at x_0, then $(fg)'(x_0)$ exists and*

$$(fg)'(x_0) = f'(x_0)g(x_0) + g'(x_0)f(x_0).$$

In short, $(fg)' = f'g + g'f$.

Theorem 5.13 (Chain rule). *Let $f\colon I \to J$ and $g\colon J \to \mathbb{R}$ such that f is differentiable at x_0 and g is differentiable at $f(x_0)$. Then, $(g \circ f)'(x_0)$ exists and*

$$(g \circ f)'(x_0) = g'(f(x_0))\, f'(x_0).$$

In short, $(g \circ f)' = (g' \circ f) \cdot f'$.

Theorem 5.14 (Quotient rule). *Let f, g be differentiable at x_0, and assume that $g(x_0) \neq 0$. Then, $\frac{f}{g}$ is differentiable at x_0 and*

$$\left(\frac{f}{g}\right)'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g(x_0)^2}.$$

In short, $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$.

Theorem 5.17. *Let $f\colon I \to \mathbb{R}$ be strictly monotone and continuous. If f is differentiable at $x_0 \in I$, and $f'(x_0) \neq 0$, then f^{-1} is differentiable at $y_0 = f(x_0)$ and*

$$\left(f^{-1}\right)'(y_0) = \frac{1}{f'(x_0)} = \frac{1}{f'(f^{-1}(y_0))}.$$

In short, $(f^{-1})' = \frac{1}{f' \circ f^{-1}}$.

5.2 Global and local extrema

Definition 5.22. Let $I = (a, b)$ and $f : I \to \mathbb{R}$.
Then, f has a **(global) minimum at** $x_0 \in I$ if

$$f(x) \geq f(x_0) \qquad \text{for all } x \in I,$$

and f has a **(global) maximum at** $x_0 \in I$ if

$$f(x) \leq f(x_0) \text{ for all } x \in I.$$

The point x_0 is called **(global) maximum/minimum point**, or **global extreme point**.

The value $f(x_0)$ is called **maximum/minimum value**, or, collectively, **extremum**.

Definition 5.24. Let $I = (a, b)$ and $f : I \to \mathbb{R}$.
Then, f has a **local minimum at** $x_0 \in I$ if there exists $\varepsilon > 0$ such that

$$f(x) \geq f(x_0) \qquad \text{for all } x \in I \cap (x_0 - \varepsilon, x_0 + \varepsilon),$$

and f has a **local maximum at** $x_0 \in I$ if there exists $\varepsilon > 0$ such that

$$f(x) \leq f(x_0) \qquad \text{for all } x \in I \cap (x_0 - \varepsilon, x_0 + \varepsilon).$$

The point x_0 is called **local maximum/minimum point**, or **(local) extreme point**.

Theorem 5.25 (Necessary condition for an extreme point). *Let $I = (a, b)$ and $f : I \to \mathbb{R}$. If $x_0 \in I$ is a local extreme point of f and f is differentiable at x_0, then*

$$f'(x_0) = 0.$$

We call $x_0 \in I$ with $f'(x_0) = 0$ a ***stationary point*** *of f.*

Theorem 5.29 (Second derivative test). *Let $I = (a, b)$, $x_0 \in I$ and $f : I \to \mathbb{R}$ be twice continuously differentiable at x_0, i.e., f'' exists and is continuous at x_0.
Moreover, assume that x_0 is a stationary point of f, i.e., $f'(x_0) = 0$. Then,*

$$f''(x_0) > 0 \qquad \Longrightarrow \qquad f \text{ has a local minimum at } x_0,$$

and

$$f''(x_0) < 0 \qquad \Longrightarrow \qquad f \text{ has a local maximum at } x_0.$$

If $f''(x_0) = 0$, we do not gain any information about the possible extremum.

5.3 Mean value theorem and l'Hospital's rule

Theorem 5.32 (Rolle). *Let $f\colon [a,b] \to \mathbb{R}$ be continuous and differentiable on (a,b). Furthermore, assume that $f(a) = f(b)$. Then there exists some $\xi \in (a,b)$ such that $f'(\xi) = 0$.*

Theorem 5.34 (Mean value theorem). *Let $f\colon [a,b] \to \mathbb{R}$ be continuous in $[a,b]$ and differentiable in (a,b). Then, there exists some $\xi \in (a,b)$ such that*

$$f'(\xi) = \frac{f(b) - f(a)}{b - a}.$$

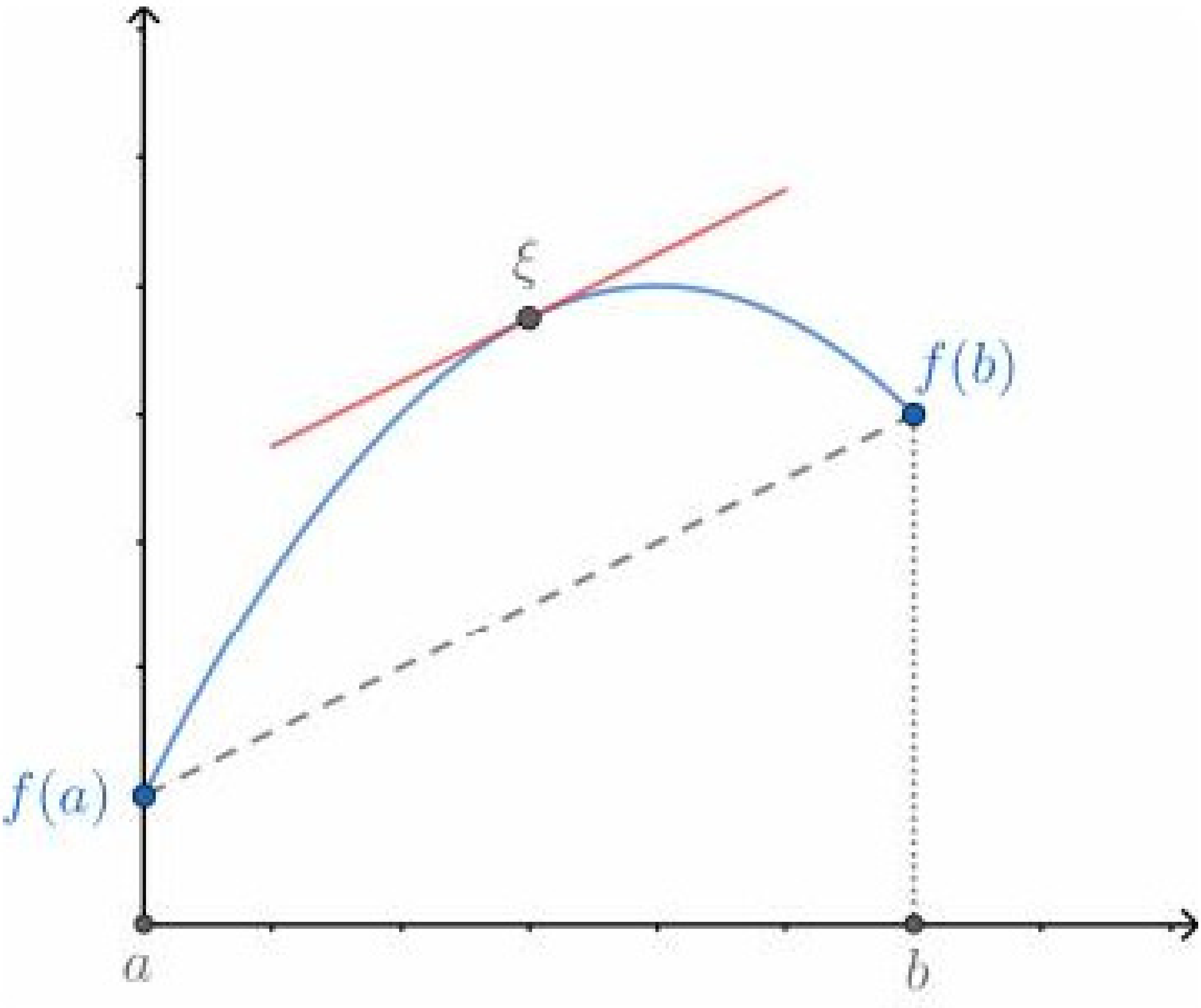

Figure 36: Mean value theorem

Theorem 5.35 (General mean value theorem). *Let $f, g\colon [a,b] \to \mathbb{R}$ be continuous in $[a,b]$ and differentiable in (a,b). Then, there exists some $\xi \in (a,b)$ such that*

$$f'(\xi)(g(b) - g(a)) = g'(\xi)(f(b) - f(a)).$$

In particular, if $g' \neq 0$ on (a,b), then there exists some $\xi \in (a,b)$ such that

$$\frac{f'(\xi)}{g'(\xi)} = \frac{f(b) - f(a)}{g(b) - g(a)}.$$

Theorem 5.36 (l'Hospital). *Let $I = (a,b)$ and $x_0 \in [a,b]$. Let $f, g\colon I\backslash\{x_0\} \to \mathbb{R}$ be differentiable on $I\backslash\{x_0\}$ and $g' \neq 0$. Furthermore assume that either $\lim_{x\to x_0} f(x) = \lim_{x\to x_0} g(x) = 0$ or $\lim_{x\to x_0} f(x) = \pm\infty, \lim_{x\to x_0} g(x) = \pm\infty$ holds. Then we have*

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} = \lim_{x\to x_0} \frac{f'(x)}{g'(x)},$$

if the right hand side exists.

5.4 Monotonicity and convexity

Definition 5.43 (Monotonicity). Let $f\colon (a,b) \to \mathbb{R}$.
We call f **(strictly) increasing** if

$$\forall\, x_1, x_2 \in (a,b)\colon x_1 < x_2 \implies f(x_1) < f(x_2),$$

or **(strictly) decreasing** if

$$\forall\, x_1, x_2 \in (a,b)\colon x_1 < x_2 \implies f(x_1) > f(x_2).$$

If we replace '$<$' by '$\leq$', or '$>$' by '$\geq$', then we say f is **non-decreasing** or **non-increasing**, respectively.

Theorem 5.44. *Let f be differentiable on $I = (a,b)$.*
Then,

$$f \text{ non-decreasing} \iff f' \geq 0$$

and

$$f \text{ non-increasing} \iff f' \leq 0.$$

Definition 5.48. Let I be an interval and $f\colon I \to \mathbb{R}$. We say that f is **convex** in I if $\forall \lambda \in (0,1)$ and $x_0, x_1 \in I$ there holds

$$f((1-\lambda)x_0 + \lambda x_1) \leq (1-\lambda)f(x_0) + \lambda f(x_1).$$

We call f **concave** if $\forall \lambda \in (0,1)$ and $x_0, x_1 \in I$ we have

$$f((1-\lambda)x_0 + \lambda x_1) \geq (1-\lambda)f(x_0) + \lambda f(x_1).$$

Theorem 5.52. *Let f be twice differentiable on an open interval I. Then f is convex if and only if $f''(x) \geq 0$, or concave if and only $f''(x) \leq 0$.*

5.5 Taylor's theorem

Theorem 5.53 (Taylor's theorem). *Let $f\colon (a,b) \to \mathbb{R}$ be $(n+1)$-times differentiable and let $x_0 \in (a,b)$. Then, for all $x \in (a,b)$ there is a ξ between x_0 and x such that*

$$f(x) \;=\; \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k \;+\; \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{n+1}.$$

We call

$$T_n(x) \;:=\; \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k,$$

*the **Taylor polynomial of f of order** n **(at** x_0**)**.*

The term

$$R_n(x) \;:=\; f(x) - T_n(x) \;=\; \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{n+1}$$

*is called the **remainder** of the Taylor polynomial (in Lagrange form).*

Corollary 5.57. *In the setting of Theorem 5.53, assume additionally that $f\colon (a,b) \to \mathbb{R}$ satisfies $|f^{(n+1)}(x)| \leq M$ for some $M < \infty$ and all $x \in (a,b)$, then*

$$|f(x) - T_n(x)| \;\leq\; \frac{M(b-a)^{n+1}}{(n+1)!} \qquad \text{for all} \quad x \in (a,b).$$

Theorem 5.60. *Let $f\colon I \to \mathbb{R}$ (with $I = (a,b)$ or $I = \mathbb{R}$) be an infinitely often differentiable function, and let $x_0 \in I$. If $r > 0$ is such that*

$$\lim_{n\to\infty} \frac{r^n}{n!} \cdot \sup_{\xi \in U_r(x_0)} |f^{(n)}(\xi)| \;=\; 0,$$

where $U_r(x_0) := (x_0 - r, x_0 + r) \subset I$, then f can be written by its Taylor series

$$f(x) \;=\; \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k \qquad \text{for all} \quad x \in U_r(x_0).$$

In particular, if, for every fixed $\xi \in I$, we have

$$\lim_{n\to\infty} \frac{\sqrt[n]{|f^{(n)}(\xi)|}}{n} \;=\; 0,$$

then the above holds for all $x, x_0 \in I$.

6 Basic integration theory

6.1 Antiderivatives

Definition 6.2. Let $\Omega \subset \mathbb{R}$. Then, we call Ω an **open set**, if

$$\forall x \in \Omega \, \exists \varepsilon > 0 \colon \; U_\varepsilon(x) \subset \Omega,$$

where $U_\varepsilon(x) = (x - \varepsilon, x + \varepsilon)$ is the ε-neighborhood of x.
That is, around every point there is a small open interval, that is completely contained in Ω.

Moreover, let $\Omega^c := \mathbb{R} \setminus \Omega$ denote the **complement** of Ω.

Then, we call Ω a **closed set**, if Ω^c is an open set.

Definition 6.4. Let $\Omega \subset \mathbb{R}$ be an open set. If $F \colon \Omega \to \mathbb{R}$ is a differentiable function such that

$$F'(x) \;=\; f(x) \qquad \text{for all } \; x \in \Omega,$$

then we call F an **antiderivative** or **indefinite integral** of f.

We also use the notation

$$F \;=\; \int f(x)\,dx \;=\; \int f\,dx$$

to say that F is a antiderivative of f, and call f the **integrand**.

Next we provide a list of antiderivatives which we will use from now on. All of them follow by differentiating the right hand side. (Do this again as an exercise!)

$$
\begin{aligned}
\int a^x\,dx &= \frac{a^x}{\ln a}, \quad a>0,\ a\neq 1\\
\int x^a\,dx &= \frac{x^{a+1}}{a+1}, \quad a\neq -1\\
\int \frac{dx}{x} &= \ln|x|\\
\int \cos x\,dx &= \sin x\\
\int \sin x\,dx &= -\cos x\\
\int \frac{dx}{\cos^2 x} &= \tan x\\
\int \frac{dx}{\sin^2 x} &= -\cot x\\
\int \frac{dx}{1+x^2} &= \arctan x\\
\int \frac{dx}{\sqrt{1-x^2}} &= \arcsin x\\
\int \frac{dx}{\sqrt{x^2-1}} &= \operatorname{arcosh} x\\
\int \frac{dx}{\sqrt{x^2+1}} &= \operatorname{arsinh} x
\end{aligned}
$$

All the antiderivatives $\int f\,dx$ above exist on the whole domain where f is defined.

6.2 Calculation rules for antiderivatives

Lemma 6.9 (Linearity). *Let $F=\int f(x)\,dx$ and $G=\int g(x)\,dx$. Then, for all $\alpha,\beta\in\mathbb{R}$,*

$$\alpha F+\beta G \;=\; \alpha\int f(x)\,dx+\beta\int g(x)\,dx \;=\; \int \alpha f(x)+\beta g(x)\,dx.$$

If F and G have different domains, then $F+G$ is understood as a function on the intersection.

Lemma 6.13 (Integration by parts). *Let f and g be differentiable functions. Then,*

$$\int f'g\,dx \;=\; fg-\int fg'\,dx.$$

Lemma 6.19 (Substitution rule). *Let $F=\int f\ x$ and g be a differentiable function. Then,*

$$F(g(x)) \;=\; \int f(g(x))\,g'(x)\,dx.$$

Corollary 6.23. *Let $f : \Omega \to \mathbb{R} \setminus \{0\}$ be a differentiable function. Then,*

$$\int \frac{f'(x)}{f(x)}\, dx \;=\; \ln |f(x)|.$$

6.3 A first definition of the integral

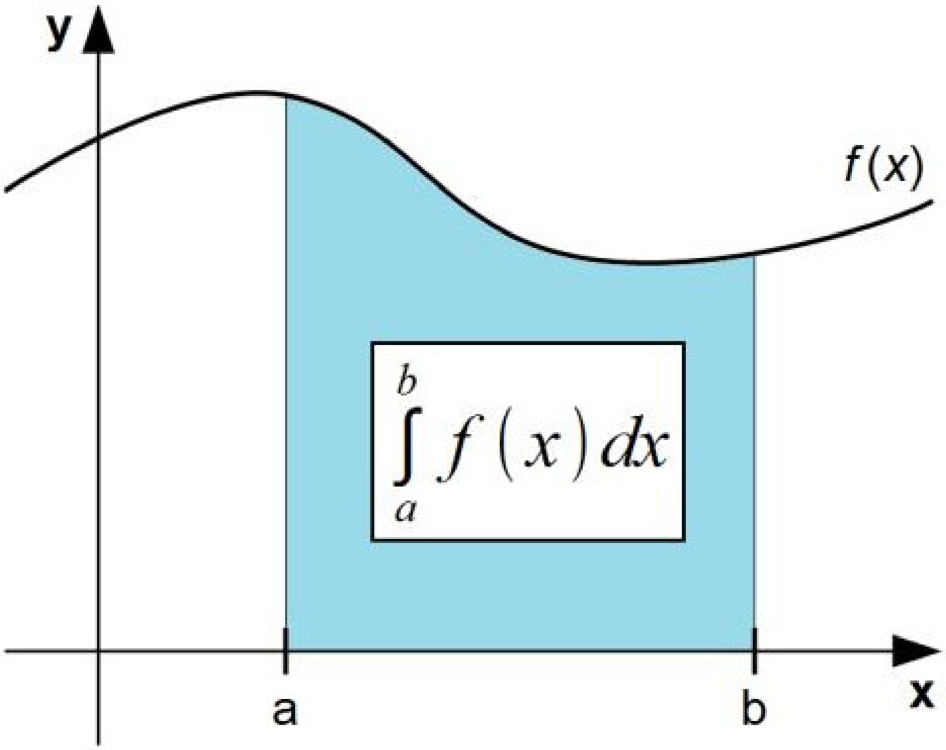

Figure 38: The area of the shaded region is denoted by $\int_a^b f(x)\, dx$

Although the resulting approximation of the integral might be quite different, this difference gets smaller and smaller if we increase the number of subintervals, see Figure 40.

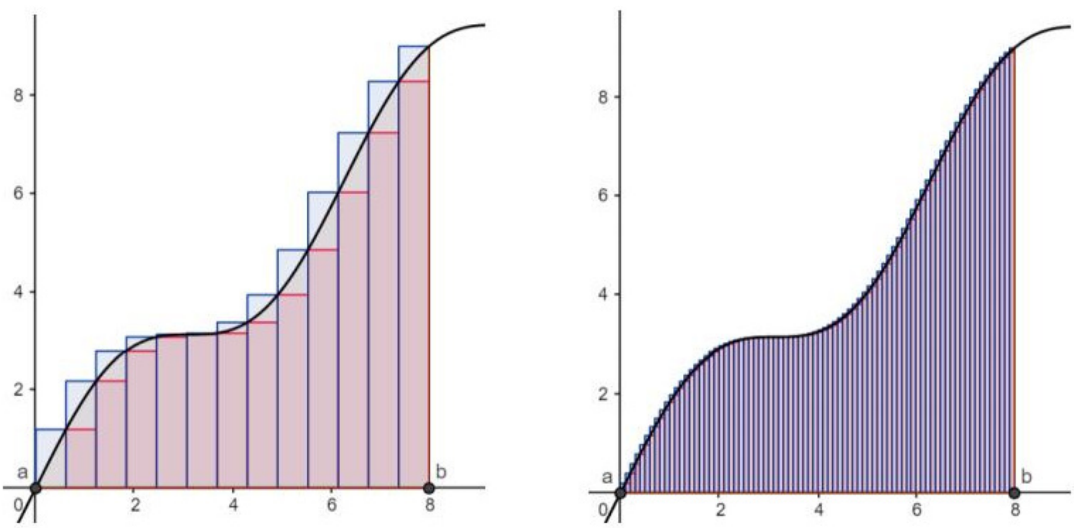

Figure 40: Smallest and the largest rectangles

Lemma 6.28. *Let $f\colon [0,1] \to \mathbb{R}$ be continuous. Then,*

$$\lim_{n\to\infty} L_n(f) \;=\; \lim_{n\to\infty} U_n(f).$$

In particular, both limits exist. Therefore, also the sequence $\left(Q_n(f)\right)_{n\in\mathbb{N}}$ converges.

Definition 6.29. Let $f\colon \Omega \to \mathbb{R}$ be continuous, and consider an interval $[a,b] \subset \Omega$. Then, we define by

$$\int_a^b f(x)\,dx \;:=\; \lim_{n\to\infty} \frac{b-a}{n} \sum_{k=0}^{n-1} f\left(a + \frac{k(b-a)}{n}\right)$$

the **(definite) integral of f over** $[a,b]$. We call a and b the **limits of the integral**.

Lemma 6.31. *Let $f, g\colon \Omega \to \mathbb{R}$ be continuous functions, and let $[a,b] \subset \Omega$. Then,*

1) $f = 0$ *on* $[a,b] \implies \int_a^b f\,dx = 0$

2) $\int_a^a f\,dx = 0$ *(That's the area over an interval with length 0.)*

3) $\int_a^b f + g\,dx = \int_a^b f\,dx + \int_a^b g\,dx$ *(Linearity)*

4) $\int_a^b \lambda \cdot f(x)\,dx = \lambda \cdot \int_a^b f\,dx$ *for* $\lambda \in \mathbb{R}$. *(Homogeneity)*

5) $c \in [a,b] \implies \int_a^b f\,dx = \int_a^c f\,dx + \int_c^b f\,dx$ *(Splitting the area in two parts.)*

6) $f \le g \implies \int_a^b f\,dx \le \int_a^b g\,dx$ *(Monotonicity w.r.t. the function)*

7) $f \ge 0$ *and* $[c,d] \subset [a,b] \implies \int_c^d f\,dx \le \int_a^b f\,dx$ *(Monotonicity w.r.t. the limits)*

There is one case of the above lemma that is particularly important. In analogy to the very similar inequality for (finite) sums, this is also called **triangle inequality**.

Corollary 6.32. *Let $f\colon \Omega \to \mathbb{R}$ be a continuous function, and let $[a,b] \subset \Omega$. Then,*

$$\int_a^b f\,dx \;\le\; \int_a^b |f|\,dx.$$

6.4 The fundamental theorem of calculus

Theorem 6.37 (Mean value theorem for definite integrals). *Let $f, g\colon [a,b] \to \mathbb{R}$ be continuous functions, and assume that $g \geq 0$. Then, $\int_a^b f(x)g(x)\,dx$ exists and there exists some $\xi \in [a,b]$ such that*

$$\int_a^b f(x)g(x)\,dx \;=\; f(\xi)\int_a^b g(x)\,dx.$$

In particular, we have (for $g = 1$) that

$$\frac{1}{b-a}\int_a^b f(x)\,dx \;=\; f(\xi)$$

for some $\xi \in [a,b]$.

Theorem 6.38 (Fundamental theorem of calculus). *Let f be continuous on some open interval $I \subset \mathbb{R}$, and $a \in I$. Then, the function $F\colon I \to \mathbb{R}$ defined by*

$$F(x) \;=\; \int_a^x f(y)\,dy$$

is an antiderivative of f, i.e., $F' = f$.

Moreover, for any $a, b \in I$ and any antiderivative F of f, we have

$$\int_a^b f(x)\,dx \;=\; F(b) - F(a),$$

and we write $[F]_a^b := F(b) - F(a)$.

Corollary 6.43 (Integration by parts). *Let f, g be continuously differentiable on $[a,b]$. Then,*

$$\int_a^b f'(x)g(x)\,dx = [fg]_a^b - \int_a^b f(x)g'(x)\,dx.$$

Corollary 6.47 (Substitution rule). *Let $I = [a,b]$, f be continuous and g be continuously differentiable on I. Then,*

$$\int_a^b f(g(x))\,g'(x)\,dx \;=\; \int_{g(a)}^{g(b)} f(y)\,dy$$

6.5 Improper integrals

Let us start with unbounded intervals. In this case, we define the integrals

$$\begin{aligned}
\int_a^\infty f\,dx &:= \lim_{b\to\infty} \int_a^b f\,dx, \\
\int_{-\infty}^b f\,dx &:= \lim_{a\to-\infty} \int_a^b f\,dx, \\
\int_{-\infty}^\infty f\,dx &:= \int_{-\infty}^0 f\,dx + \int_0^\infty f\,dx
\end{aligned}$$

whenever the integrals and limits on the right hand side exist. We then say that the integrals on the left *converge*.

From the fundamental theorem of calculus, we know how to compute the finite integrals on the right hand side. That is, if F is an antiderivative of f (on the corresponding interval), then $\int_a^b f\,dx = F(b) - F(a)$. To compute the above limits, it is therefore enough to compute the limits for the antiderivative. That is, if we denote

$$F(-\infty) := \lim_{a\to-\infty} F(a) \quad \text{and} \quad F(\infty) := \lim_{b\to\infty} F(b),$$

then we obtain

$$\begin{aligned}
\int_a^\infty f\,dx &= [F]_a^\infty = F(\infty) - F(a), \\
\int_{-\infty}^b f\,dx &= [F]_{-\infty}^b = F(b) - F(-\infty), \\
\int_{-\infty}^\infty f\,dx &= [F]_{-\infty}^\infty = F(\infty) - F(-\infty),
\end{aligned}$$

whenever the corresponding limits exist.

Lemma 6.53. *Let $a \in \mathbb{R}$ and $f\colon (a,\infty) \to \mathbb{R}$ be a continuous function. Then,*

- *$\int_a^\infty f\,dx$ is convergent if $|f(x)| \le \frac{c}{x^\beta}$ for some $c < \infty$ and $\beta > 1$.*
- *$\int_a^\infty f\,dx$ is divergent if $f(x) \ge \frac{c}{x}$ for some $c > 0$.*

6.6 Piecewise continuous functions

Definition 6.61. Let $I = [a,b]$. We say that a function $f\colon I \to \mathbb{R}$ is **piecewise continuous** if and only if there are exist a finite number of points $x_1, \dots, x_m \in I$ such that

1. f is continuous on every subinterval $[a, x_1)$, $(x_m, b]$ and (x_k, x_{k+1}) for $k = 1, \dots, m-1$,
2. the limits $\lim_{x\nearrow x_k} f(x)$ and $\lim_{x\searrow x_k} f(x)$ exist and are finite.

We call $x_1, \dots, x_m$ the **(finite) discontinuities** of f.

7 Fourier series

7.1 Periodic functions and trigonometric polynomials

Definition 7.2. A function $f\colon \mathbb{R} \longmapsto \mathbb{C}$ is called **periodic**, or **1-periodic**, if

$$f(x+1) = f(x) \quad \text{for all } x \in \mathbb{R}.$$

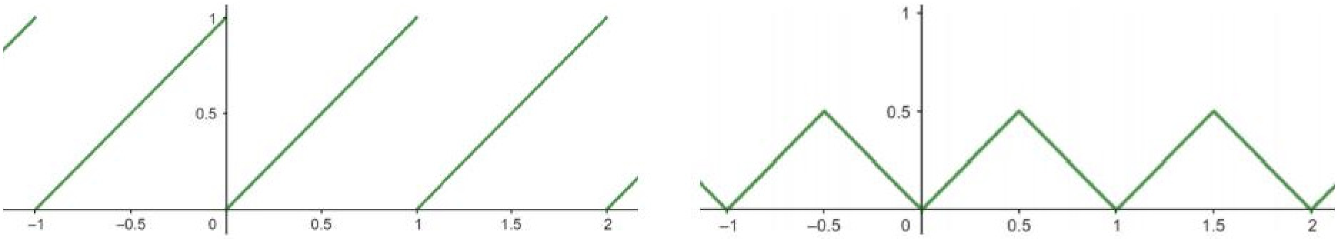

Figure 41: The sawtooth wave and the triangle wave.

Definition 7.3. We say that a periodic function $f\colon [0,1) \to \mathbb{C}$ has a property if and only if its periodic extension to $\mathbb{R}$, i.e., the function $f\colon \mathbb{R} \to \mathbb{C}$ with $f(x) = f(\{x\})$, has this property.

Definition 7.8. A **trigonometric polynomial** p is a periodic function of the form

$$p(x) \;=\; \sum_{k=-n}^{n} c_k\, e^{2\pi i k x}\,,$$

where $n \in \mathbb{N}$ and $c_{-n}, \ldots, c_n \in \mathbb{C}$ are called the **coefficients** of the trigonometric polynomial.

Note that this very short notation for a trigonometric polynomial can clearly be written out by using cosine and sine terms. In particular, by using Euler's formula, we obtain that

$$\begin{aligned} p(x) &= \sum_{k=-n}^{n} c_k\, e^{2\pi i k x} = \sum_{k=-n}^{n} c_k \Big(\cos(2\pi k x) + i\sin(2\pi k x)\Big) \\ &= \sum_{k=-n}^{n} c_k \cos(2\pi k x) + i \sum_{k=-n}^{n} c_k \sin(2\pi k x) \end{aligned}$$

However, this representation can be simplified further by using that cosine is **even**, i.e., $\cos(-x) = \cos(x)$, and that sine is an **odd function**, i.e., $\sin(-x) = -\sin(x)$. From this we get

$$\begin{aligned} p(x) &= \sum_{k=-n}^{n} c_k \cos(2\pi k x) + i \sum_{k=-n}^{n} c_k \sin(2\pi k x) \\ &= c_0 + \sum_{k=1}^{n} (c_k + c_{-k}) \cos(2\pi k x) + i \sum_{k=1}^{n} (c_k - c_{-k}) \sin(2\pi k x). \end{aligned}$$

(Check this in detail!) We may therefore write trigonometric polynomials, as in Definition 7.8, in the form

$$p(x) = \sum_{k=0}^{n} a_k \cos(2\pi k x) + \sum_{k=1}^{n} b_k \sin(2\pi k x).$$

if we set $a_0 = c_0$ and, for $k \geq 1$,

$$\begin{aligned} a_k &:= c_k + c_{-k}, \\ b_k &:= i(c_k - c_{-k}). \end{aligned}$$

With this, we recover the explicit formulas for a_k and b_k from page 152.

Lemma 7.11. *Let p be a trigonometric polynomial in the form given in Definition 7.8. Then, p is ...*

- *real-valued, i.e., $p\colon \mathbb{R} \to \mathbb{R}$, if and only if*

$$\operatorname{Re}(c_k) = \operatorname{Re}(c_{-k}) \quad \text{and} \quad \operatorname{Im}(c_k) = -\operatorname{Im}(c_{-k}),$$

- *even, i.e., $p(x) = p(-x)$, if and only if $c_k = c_{-k}$,*
- *odd, i.e., $p(x) = -p(-x)$, if and only if $c_k = -c_{-k}$,*

for all $k \in \mathbb{Z}$.

By the above lemma and the representation of a trigonometric polynomial using only cosine and sine, we see that even and odd trig. polynomials can be written solely as a cosine or a sine series. That is, an **even trigonometric polynomial** can be written as

$$p(x) \;=\; \sum_{k=-n}^{n} c_k\, e^{2\pi i k x} \;=\; \sum_{k=0}^{n} a_k \cos(2\pi k x),$$

while an **odd trigonometric polynomial** can be written as

$$p(x) \;=\; \sum_{k=-n}^{n} c_k\, e^{2\pi i k x} \;=\; \sum_{k=1}^{n} b_k \sin(2\pi k x).$$

(Note that odd functions do not contain the "constant term" for $k = 0$.)
These formulas will sometimes be helpful in actual computations.

7.2 Fourier coefficients and Fourier series

Definition 7.14. Let $f\colon [0,1) \to \mathbb{C}$. Then, for a given integer $k \in \mathbb{Z}$, we call

$$\hat{f}(k) \;:=\; \int_0^1 f(x) e^{-2\pi i k x}\, dx$$

the **k-th Fourier coefficient of f**.

For $n \geq 0$, we call the trigonometric polynomial

$$S_n f(x) \;:=\; \sum_{k=-n}^{n} \hat{f}(k)\, e^{2\pi i k x}$$

the **n-th partial sum of the Fourier series** of f.

Moreover, we call the series

$$Sf(x) \;:=\; \lim_{n\to\infty} S_n f(x) \qquad \left(=\; \sum_{k=-\infty}^{\infty} \hat{f}(k)\, e^{2\pi i k x}\right)$$

the **Fourier series** of f. (We use this notation also if the limit does not exist.)

We say that the **Fourier series equals** f (pointwise) if $f(x) = Sf(x)$ for all $x \in [0,1)$.

Lemma 7.15. *Let $k \in \mathbb{Z} \setminus \{0\}$. Then,*

$$\frac{d}{dx}\, e^{2\pi i k x} \;=\; (2\pi i k)\, e^{2\pi i k x} \qquad \text{and} \qquad \int e^{2\pi i k x}\, dx \;=\; \frac{e^{2\pi i k x}}{2\pi i k}$$

for all $x \in \mathbb{R}$. In particular, we obtain

$$\int_0^1 e^{2\pi i k x}\, dx \;=\; 0$$

for $k \neq 0$, and $\int_0^1 e^{2\pi i 0 x}\, dx = 1$.

7.3 First convergence theorems

Definition 7.22. Let $(f_n)_{n=1}^{\infty}$ be a sequence of functions on a set D, and let $f\colon D \to \mathbb{C}$.

(i) If

$$\lim_{n\to\infty} f_n(x) \;=\; f(x) \qquad \text{for all } \; x \in D,$$

then we say that (f_n) **converges pointwise** to f. We use the notation $f_n \xrightarrow{\text{pw}} f$ or "$f_n \to f$ pointwise".

(ii) If

$$\lim_{n\to\infty} \sup_{x\in D} |f_n(x) - f(x)| \;=\; 0,$$

then we say that (f_n) **converges uniformly** to f. We write "$f_n \to f$ uniformly".

Lemma 7.25. *Let $(f_n)_{n=1}^{\infty}$ be a sequence of continuous functions on a set D, and let $f\colon D \to \mathbb{C}$ be such that $f_n \to f$ uniformly. Then,*

(i) *f is continuous, and*

(ii) *for all $a, b \in D$ with $[a,b] \subset D$, we have*

$$\int_a^b f\,dx \;=\; \lim_{n\to\infty} \int_a^b f_n\,dx.$$

Theorem 7.26. *Let $f\colon [0,1) \to \mathbb{C}$ be continuous. If $\sum_{k=-\infty}^{\infty} |\hat{f}(k)| < \infty$ holds, then*

$$Sf(x) \;=\; \lim_{n\to\infty} S_n f(x) \;=\; f(x)$$

for every $x \in [0,1)$. Moreover, $S_n f \to f$ uniformly.

Lemma 7.27. *Let $f, g\colon [0,1) \to \mathbb{C}$ be continuous functions such that $\hat{f}(k) = \hat{g}(k)$ for all $k \in \mathbb{Z}$. Then,*

$$f(x) \;=\; g(x) \qquad \text{for all } \; x \in [0,1).$$

In particular, $\hat{f}(k) = 0$ for all $k \in \mathbb{Z}$ implies $f = 0$.

Corollary 7.29. *We have that*

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}.$$

Lemma 7.30. *Let $(a_k)_{k\in\mathbb{Z}}$ be a (complex-valued) sequence such that*

$$\sum_{k=-\infty}^{\infty} |a_k| < \infty.$$

Then, the sequence of trigonometric polynomials $(g_n)_{n\geq 0}$ with

$$g_n(x) = \sum_{k=-n}^{n} a_k e^{2\pi i k x}$$

converges uniformly to a continuous periodic function $g\colon [0,1) \to \mathbb{C}$ with

$$\hat{g}(k) = a_k \qquad \textit{for all } k \in \mathbb{Z}.$$

Theorem 7.32. *Let $f\colon [0,1) \to \mathbb{C}$ be periodic and twice continuously differentiable. Then,*

$$Sf(x) = \lim_{n\to\infty} S_n f(x) = f(x)$$

for every $x \in [0,1)$. Moreover, $S_n f \to f$ uniformly.

Lemma 7.33. *Let $s \in \mathbb{N}$ and $f\colon [0,1) \to \mathbb{C}$ be periodic and s-times continuously differentiable. Then, we have*

$$|\hat{f}(k)| \leq \frac{M}{|2\pi k|^s} \qquad \textit{for all } k \neq 0,$$

where

$$M := \max_{x\in[0,1)} |f^{(s)}(x)|.$$

Lemma 7.35. *Let $h\colon [0,\infty) \to [0,\infty)$ be a continuous and non-increasing function. Then, for all $m, n \in \mathbb{N}_0$ with $m < n$, we have*

$$\sum_{k=m+1}^{n} h(k) \leq \int_m^n h(x)\,dx \leq \sum_{k=m}^{n-1} h(k).$$

In particular, if H is an antiderivative of h, then,

$$\sum_{k=m+1}^{\infty} h(k) \leq H(\infty) - H(m),$$

where $H(\infty) := \lim_{x\to\infty} H(x)$.

Corollary 7.37. *Let $f\colon [0,1) \to \mathbb{C}$ be continuous such that for some $s > 1$ and $B < \infty$ we have*

$$|\hat{f}(k)| \;\le\; \frac{B}{|k|^s}$$

for all $k \neq 0$. Then,

$$|f(x) - S_n f(x)| \;\le\; \frac{2B}{s-1}\,\frac{1}{n^{s-1}}$$

for all $x \in [0,1)$ and $n \in \mathbb{N}$.

7.4 The theorem of Dirichlet

Theorem 7.40. *Let $f\colon [0,1) \to \mathbb{C}$ be periodic and piecewise continuous. Then, if f is differentiable at $x_0 \in [0,1)$, we have*

$$Sf(x_0) \;=\; \lim_{n\to\infty} S_n f(x_0) \;=\; f(x_0).$$

Moreover, if f is differentiable, then $S_n f \to f$ uniformly.

8 Linear Algebra I

8.1 Matrices

Definition 8.1. Let $n, m \in \mathbb{N}$ and $a_{ij} \in \mathbb{R}$ for $1 \le i \le n$ and $1 \le j \le m$.
A (real) $m \times n$ **matrix** (read "m by n matrix") is an array given by

$$A = (a_{ij})_{i,j=1}^{m,n} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

In this case we use the notation $A \in \mathbb{R}^{m\times n}$, and call m and n the **dimensions** of A.
A $m \times 1$ matrix is called a **column vector**, a $1 \times n$ matrix is called a **row vector**, and if $m = n$, then the matrix is called **quadratic**, or a **square matrix**.

Lemma 8.9. *Let $A \in \mathbb{R}^{m\times p}$ and $B \in \mathbb{R}^{p\times n}$. Then,*

$$(AB)^T = B^T A^T.$$

(Note that the order changed.)

8.2 Systems of linear equations

Definition 8.16. Let $n, m \in \mathbb{N}$, $a_{ij} \in \mathbb{R}$ and $b_j \in \mathbb{R}$ for $1 \le i \le n$ and $1 \le j \le m$.
A **system of linear equations** or **linear system** with real coefficients is given by

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & b_m \end{array}$$

The x_i with $1 \le i \le n$ are called **variables**, or *unknowns*.
The a_{ij} are called the **coefficients** of the system.
The matrix $A = (a_{ij})_{i=1,j=1}^{m,n}$ is the **matrix of coefficients**.
The tuple $b := (b_1, \dots, b_m)$ is called the **right hand side** (RHS) of the system.

If there exist such numbers $x_1, \dots, x_n \in \mathbb{R}$ that fulfill all the equations, then we call the tuple $x = (x_1, \dots, x_n)$ a **solution** to the linear system.

If there is no solution, then we call the linear system **inconsistent**.

Definition 8.17. Given a linear system $Ax = b$ with coefficient matrix A and RHS b, then we denote the **set of solutions** by

$$L(A, b) := \Big\{x \in \mathbb{R}^n \colon\ Ax = b\Big\}.$$

Definition 8.24. Let $A \in \mathbb{R}^{m \times n}$ and $x \in \mathbb{R}^n$. A linear system of the form

$$Ax = 0,$$

i.e., the RHS is the zero vector $0 := 0_{m1}$, is called a **homogeneous** system of equations.

To a given linear system $Ax = b$, we call $Ax = 0$ the corresponding homogeneous system.

Lemma 8.25. *Let $A \in \mathbb{R}^{m \times n}$, $x, y \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$, such that $Ax = b$ and $Ay = 0$. Then, $x + y$ also solves the linear system, i.e., $A(x + y) = b$.*

In particular, if there is a solution $y \neq 0$ to $Ay = 0$, and at least one solution x to $Ax = b$, then there are infinitely many solutions.

Moreover, if $Ay = 0$ has only the trivial solution $y = 0$, and there is at least one solution x to $Ax = b$, then this solution x is unique.

8.3 Gaussian elimination

Definition 8.27. A matrix $C = (c_{ij}) \in \mathbb{R}^{m\times n}$ of the form

$$C = \begin{pmatrix} 0 \cdots\cdots 0 & c_{1j_1} & * \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots * \\ 0 \cdots\cdots\cdots\cdots\cdots 0 & c_{2j_2} & * \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots * \\ 0 \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots 0 & c_{3j_3} & * \cdots\cdots\cdots\cdots\cdots * \\ \vdots & & \vdots \\ 0 \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots 0 & c_{kj_k} & * \cdots\cdots * \\ 0 \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots 0 \\ \vdots & & \vdots \\ 0 \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots 0 \end{pmatrix},$$

where $*$ stands for an arbitrary entry, is in **row echelon form** (ger. 'Treppenform'). That is, there exist numbers $k \leq m$ and $1 \leq j_1 < \cdots < j_k \leq n$ such that for all $1 \leq i \leq k$:

- $c_{ij_i} \neq 0$,
- $c_{ij} = 0$ for all $j < j_i$, i.e. c_{ij_i} is the first non-zero element in the i-th row, and
- $c_{\ell j_i} = 0$ for all $\ell > i$, i.e. c_{ij_i} is the last non-zero element in the j_i-th column.

The number k is called **rank** of the matrix, and we write $k = \operatorname{rank}(C)$.

If, in addition,

- $c_{ij_i} = 1$ for all $i \leq k$, and
- $c_{\ell j_i} = 0$ for all $\ell < i$, i.e. c_{ij_i} is the only non-zero element in the j_i-th column,

then the matrix is in **reduced row echelon form** (ger. 'Treppennormalform').

Lemma 8.38. *Let $A \in \mathbb{R}^{n\times n}$ be a square matrix.*

- *If $\operatorname{rank}(A) = n$, then the linear system $Ax = b$ has a unique solution for any $b \in \mathbb{R}^n$.*
- *If $\operatorname{rank}(A) < n$, then the homogeneous system $Ax = 0$ has infinitely many solutions.*
- *If $\operatorname{rank}(A) < n$, then the linear system $Ax = b$ has either no or infinitely many solutions, depending on $b \in \mathbb{R}^n$.*

8.4 The determinant

Definition 8.41. A **permutation** is a bijective mapping $\sigma\colon \{1, 2, \dots, n\} \to \{1, 2, \dots, n\}$. We use the notation

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{pmatrix}$$

to describe a permutation.

Moreover, we denote

$$S_n = \{\sigma\colon \sigma \text{ is a permutation of } \{1, 2, \dots, n\}\}.$$

Definition 8.46. Let $\sigma, \sigma' \in S_n$. We define the product of σ and σ', write $\sigma \circ \sigma'$, as

$$\sigma \circ \sigma'(i) = \sigma(\sigma'(i)),$$

where $1 \leq i \leq n$. If $\sigma \circ \sigma' = \sigma' \circ \sigma = \mathrm{id}$, then we say that σ, σ' are inverse to each other. In this case we write $\sigma^{-1} = \sigma'$.

Lemma 8.49. *Let σ be a permutation. Then there holds*

$$\operatorname{sgn}(\sigma) = \prod_{i<j} \frac{\sigma(j) - \sigma(i)}{j - i},$$

where the product is over all pairs (i, j) with $i, j \in \{1, \dots, n\}$ and $i < j$.

Definition 8.51. Let $A \in \mathbb{R}^{n \times n}$ be a square matrix. We define the **determinant** of A by

$$\det A = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

Lemma 8.56. *For any square matrix $A \in \mathbb{R}^{n \times n}$ we have*

$$\det(A^T) = \det(A).$$

Lemma 8.57. *For any square matrices $A, B \in \mathbb{R}^{n \times n}$ we have*

$$\det(A \cdot B) = \det A \cdot \det B.$$

Theorem 8.61. *Let $A \in \mathbb{R}^{n\times n}$, and denote the rows of A by $a_1, \dots, a_n \in \mathbb{R}^{1\times n}$, and the columns by $c_1, \dots, c_n \in \mathbb{R}^n$. Moreover, let $\lambda, \mu \in \mathbb{R}$, $v \in \mathbb{R}^n$ and $w \in \mathbb{R}^{1\times n}$.*

1. *For any $1 \le j \le n$ the determinant is linear in the j-th column of A, i.e.,*

$$\begin{aligned}&\det\Big(c_1, \dots, (\lambda c_j + \mu v), \dots, c_n\Big)\\ &\qquad = \lambda \det(A) + \mu \det\Big(c_1, \dots, v, \dots, c_n\Big)\end{aligned}$$

2. *For any $1 \le i \le n$, the determinant is linear in the i-th row of A, i.e.,*

$$\det\begin{pmatrix} a_1 \\ \vdots \\ \lambda a_i + \mu w \\ \vdots \\ a_n \end{pmatrix} = \lambda \det(A) + \mu \det\begin{pmatrix} a_1 \\ \vdots \\ w \\ \vdots \\ a_n \end{pmatrix}$$

3. *If there exist $i \neq j$ such that $a_i = a_j$ or $c_i = c_j$, then $\det(A) = 0$.*

4. *The identity matrix $I_n \in \mathbb{R}^{n\times n}$ satisfies*

$$\det(I_n) = 1.$$

Corollary 8.63. *Let $A \in \mathbb{R}^{n\times n}$. Then,*

1. *If the matrix B is obtained by multiplying one column of A by a scalar $\lambda \in \mathbb{R}$, then $\det B = \lambda \det A$. In particular, $\det(\lambda A) = \lambda^n \det(A)$.*

2. *If the matrix B is obtained by interchanging two columns of A, then $\det B = -\det A$.*

3. *Adding a multiple of one column to another column does not change the determinant.*

4. *The points above also hold if we replace "column" by "row".*

Lemma 8.66. *Let $A \in \mathbb{R}^{n\times n}$ be an **upper triangular matrix**, which is a matrix of the form*

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ & \ddots & \vdots \\ 0 & & a_{nn} \end{pmatrix},$$

i.e., all entries below the diagonal are zero. Then,

$$\det A = \prod_{i=1}^{n} a_{ii}.$$

This formula holds also for lower triangular matrices, which are zero above the diagonal.

Theorem 8.68. *Let $A \in \mathbb{R}^{n\times n}$. Then,*

$$\det A \neq 0 \iff \operatorname{rank} A = n \iff A \text{ is invertible.}$$

Definition 8.69. Let $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ and $1 \leq i, j, \leq n$.
Then, we define the (i,j)**-minor of** A by

$$M_{ij} = \det \begin{pmatrix} a_{1,1} & \cdots & a_{1,j-1} & a_{1,j+1} & \cdots & a_{1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{i-1,1} & \cdots & a_{i-1,j-1} & a_{i-1,j+1} & \cdots & a_{i-1,n} \\ a_{i+1,1} & \cdots & a_{i+1,j-1} & a_{i+1,j+1} & \cdots & a_{i+1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,j-1} & a_{n,j+1} & \cdots & a_{n,n} \end{pmatrix},$$

i.e., M_{ij} is the determinant of the $(n-1)\times(n-1)$-submatrix of A obtained by deleting the i-th row and the j-th column. We call $M = (M_{ij})_{i,j=1}^n$ the **matrix of minors** of A.

Moreover, we define the (i,j)**-cofactor** by $C_{ij} := (-1)^{i+j} M_{ij}$,
and call $C = (C_{ij})_{i,j=1}^n$ the **cofactor matrix** of A.

Theorem 8.70 (Laplace expansion). *Let $A = (a_{ij}) \in \mathbb{R}^{n\times n}$. Then, we can compute the determinant of A by expansion along ...*

- *the i-th row:*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} M_{ij} = \sum_{j=1}^{n} a_{ij} C_{ij} \qquad (\text{fixed } i)$$

- *the j-th column:*

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} M_{ij} = \sum_{i=1}^{n} a_{ij} C_{ij} \qquad (\text{fixed } j)$$

8.5 Cramer's rule

Theorem 8.74 (Cramer's rule). *Let $A \in \mathbb{R}^{n\times n}$ with $\det A \neq 0$, and $b \in \mathbb{R}^n$. Then, the linear system $Ax = b$ has the unique solution $x = (x_1, \ldots, x_n)^T$ given by*

$$x_k = \frac{\det(A_k)}{\det(A)},$$

where A_k is given by

$$A_k := \begin{pmatrix} a_{1,1} & \ldots & a_{1,k-1} & b_1 & a_{1,k+1} & \ldots & a_{1,n} \\ a_{2,1} & \ldots & a_{2,k-1} & b_2 & a_{1,k+1} & \ldots & a_{1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n,1} & \ldots & a_{n,k-1} & b_n & a_{n,k+1} & \ldots & a_{n,n} \end{pmatrix}.$$

8.6 Inverse matrices

Definition 8.77. Let $A \in \mathbb{R}^{n\times n}$ and assume that there exists some $A' \in \mathbb{R}^{n\times n}$ with the property that

$$A \cdot A' = A' \cdot A = I_n.$$

Then, we say that A is **invertible** or **regular**, and we write $A^{-1} := A'$ to denote the inverse.

Lemma 8.78. *For any regular matrix $A \in \mathbb{R}^{n\times n}$ we have*

$$\det(A^{-1}) = \frac{1}{\det A}.$$

Theorem 8.79. *Let $A \in \mathbb{R}^{n\times n}$ with $\det(A) \neq 0$, and let $C = (C_{ij}) \in \mathbb{R}^{n\times n}$ be the cofactor matrix of A. Then,*

$$A^{-1} = \frac{1}{\det A} C^T,$$

i.e.,

$$(A^{-1})_{ij} = \frac{C_{ji}}{\det A},$$

where $(A^{-1})_{ij}$ denotes the ij-th entry of A^{-1}.

Theorem 8.82. *Let $A \in \mathbb{R}^{n\times n}$ be an invertible matrix, i.e.,* $\det(A) \neq 0$. *Then, the reduced row echelon form of*

$$\left(\begin{array}{ccc|ccc} a_{11} & \dots & a_{1n} & 1 & & 0 \\ \vdots & & \vdots & & \ddots & \\ a_{n1} & \dots & a_{nn} & 0 & & 1 \end{array}\right)$$

has the form

$$\left(\begin{array}{ccc|ccc} 1 & & 0 & a'_{11} & \dots & a'_{1n} \\ & \ddots & & \vdots & & \vdots \\ 0 & & 1 & a'_{n1} & \dots & a'_{nn} \end{array}\right)$$

and it holds that $A^{-1} = (a'_{ij})_{i,j=1}^{n}$.

8.7 Eigenvalues and eigenvectors

Definition 8.85. Let $A \in \mathbb{R}^{n\times n}$. A vector $v \in \mathbb{C}^n$ with $v \neq 0_n$ is called **eigenvector** of A to the **eigenvalue** $\lambda \in \mathbb{C}$ if

$$Av = \lambda v.$$

We call the pair (λ, v) an **eigenpair** of A.

Moreover, the set of all eigenvalues of A, i.e.,

$$\sigma(A) := \Big\{\lambda \in \mathbb{C}\colon\ \exists x \in \mathbb{C}^n \setminus \{0\} \text{ with } Ax = \lambda x\Big\},$$

is called the **spectrum** of A.

Definition 8.88 (Eigenspace). Let $A \in \mathbb{R}^{n\times n}$ and $\lambda \in \mathbb{R}$. Then, we call the set

$$E(A,\lambda) := \Big\{x \in \mathbb{R}^n\colon Ax = \lambda x\Big\}$$

the **eigenspace** of A associated with λ, or just eigenspace of λ if the matrix is clear.

If $\lambda \in \mathbb{C} \setminus \mathbb{R}$, we have to consider $x \in \mathbb{C}^n$ in the above definition.

Definition 8.89 (Kernel). Let $A \in \mathbb{R}^{n\times n}$. Then, we call the set

$$N(A) := \Big\{x \in \mathbb{R}^n\colon Ax = 0\Big\}$$

the **kernel** (or **nullspace**) of A.

Lemma 8.95. *Let $A \in \mathbb{R}^{n\times n}$. Then,*

$$\lambda \in \sigma(A) \iff \det(A - \lambda I_n) = 0.$$

Recall that $\lambda \in \sigma(A)$ just means that λ is an eigenvalue of A.

Definition 8.96. For a given matrix $A \in \mathbb{R}^{n\times n}$ we define

$$p(\lambda) = p_A(\lambda) := \det(A - \lambda I_n),$$

which is called the **characteristic polynomial** of A.
(We omit the subscript A in p_A if the matrix is clear.)

Theorem 8.103. *Let $A \in \mathbb{R}^{n\times n}$ with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then,*

$$\det A = \prod_{k=1}^{n} \lambda_k.$$

Corollary 8.104. *Let $A \in \mathbb{R}^{n\times n}$. Then, A is invertible (aka. regular) if and only if $0 \notin \sigma(A)$, i.e., 0 is not an eigenvalue of A.*

Lemma 8.105. *Let $A \in \mathbb{R}^{n\times n}$ be an invertible matrix and λ be an eigenvalue of A, i.e., $\lambda \in \sigma(A)$. Then, λ^{-1} is an eigenvalue of A^{-1}, i.e.,*

$$\frac{1}{\lambda} \in \sigma\left(A^{-1}\right).$$

Moreover, if (λ, v) is an eigenpair of A, then $(\frac{1}{\lambda}, v)$ is an eigenpair of A^{-1}.

Definition 8.106. Let $A \in \mathbb{R}^{n\times n}$. Then, we can write the characteristic polynomial of A in the form

$$p(\lambda) = (\lambda_1 - \lambda)^{\mu_1}(\lambda_2 - \lambda)^{\mu_2} \cdots (\lambda_k - \lambda)^{\mu_k}$$

where all λ_i's are different.
For $i = 1, \ldots, k$, the integer μ_i is called **(algebraic) multiplicity** of the eigenvalue λ_i.
If $\mu_i = 1$ for some $1 \le i \le k$, then we say that λ_i is a **simple eigenvalue**.

Lemma 8.108. *Let $A \in \mathbb{R}^{n\times n}$ and $\lambda \in \mathbb{R}$ be a simple eigenvalue of A. Then, there is some $v \in \mathbb{R}^n$ with $v \neq 0$, such that*

$$E(A, \lambda) = \{\alpha v \colon \alpha \in \mathbb{R}\},$$

i.e., all eigenvectors to a simple eigenvalue are multiples of one vector.

8.8 Diagonalization

Lemma 8.109. *Let $A \in \mathbb{R}^{n\times n}$ be symmetric, i.e., $A^T = A$, and let (λ_1, v_1) and (λ_2, v_2) be two eigenpairs of A with $\lambda_1 \neq \lambda_2$. Then,*

$$\langle v_1, v_2\rangle = v_1^T v_2 = 0.$$

That is, eigenvectors to different eigenvalues are orthogonal.

Definition 8.111. Let $O \in \mathbb{R}^{n\times n}$ be such that

$$O^T O = I_n,$$

then we call O an **orthogonal matrix**.
(Equivalently, O is an orthogonal matrix if and only if $O^{-1} = O^T$.)

Lemma 8.112. *A matrix $O = (c_1, \dots, c_n) \in \mathbb{R}^{n\times n}$ is orthogonal if and only if all columns of O are orthogonal, i.e.,*

$$\langle c_i, c_j\rangle = 0 \qquad \textit{for all} \quad i \neq j,$$

and all columns are ***normalized****, i.e.,*

$$\langle c_i, c_i\rangle = 1 \qquad \textit{for all} \quad i = 1, \dots, n.$$

That is, O is orthogonal if and only if $\langle c_i, c_j\rangle = \delta_{ij}$.

From the above, we know that eigenvectors to different eigenvalues are orthogonal. To build up an orthogonal matrix from these vectors, we also need them to be normalized. That is, we need $\langle v, v\rangle = 1$ ($\Leftrightarrow \|v\| = 1$) for every eigenvector v. Such a vector, i.e., $v \in \mathbb{R}^n$ with $Av = \lambda v$ and $\|v\| = 1$, is called **normalized eigenvector** to the eigenvalue λ.

Luckily, we know that every multiple of an eigenvector is also an eigenvector to the same eigenvalue. We can therefore just divide the vector by its norm. That is, if (λ, v) is an eigenpair of A, then

$$\overline{v} := \frac{1}{\|v\|} \cdot v$$

is a normalized eigenvector to λ.

With this, we can now give the following important result.

Theorem 8.115. *Let $A \in \mathbb{R}^{n\times n}$ be symmetric with n different eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ and corresponding normalized eigenvectors $v_1, \ldots, v_n \in \mathbb{R}^n$. Then,*

$$A = VDV^T,$$

where

$$V := (v_1, \ldots, v_n)$$

and

$$D = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}.$$

Corollary 8.118. *Let $\ell \in \mathbb{Z}$ and $A = VDV^T \in \mathbb{R}^{n\times n}$ be as in Theorem 8.115, then*

$$A^\ell = VD^\ell V^T,$$

where $D^\ell = \operatorname{diag}(\lambda_1^\ell, \ldots, \lambda_n^\ell)$. (For $\ell < 0$ we need here that A is invertible, i.e., $\lambda_k \neq 0$.)

In particular, if (λ, v) is an eigenpair of A, then (λ^ℓ, v) is an eigenpair of A^ℓ.

8.9 Singular value decomposition

Theorem 8.121. *Let $A \in \mathbb{R}^{n\times n}$ be symmetric with n different eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ and corresponding normalized eigenvectors $v_1, \ldots, v_n \in \mathbb{R}^n$. Then,*

$$A = \sum_{k=1}^{n} \lambda_k \cdot v_k v_k^T.$$

If the eigenvectors are not normalized, we can write

$$A = \sum_{k=1}^{n} \frac{\lambda_k}{\langle v_k, v_k\rangle} \cdot v_k v_k^T.$$

Recall that $v_k v_k^T \in \mathbb{R}^{n\times n}$, and that $\langle v_k, v_k\rangle = \|v_k\|^2 = v_k^T v_k$.

9 Multivariate Calculus

9.1 Sequences in R^d

Theorem 9.4 (Triangle inequality). *For all $x, y \in \mathbb{R}^d$ we have*

$$\|x + y\|_2 \leq \|x\|_2 + \|y\|_2 .$$

Lemma 9.5 (Cauchy-Schwarz inequality). *For $x, y \in \mathbb{R}^d$ it holds that*

$$|\langle x, y\rangle| \leq \|x\|_2 \, \|y\|_2 .$$

Moreover, we have the equality $|\langle x, y\rangle| = \|x\|_2 \, \|y\|_2$ if and only if $y = c \cdot x$ for some $c \in \mathbb{R}$.

Definition 9.7 (Convergence). Let $(x_k)_{k\in\mathbb{N}} \subset \mathbb{R}^d$ be a sequence of vectors. If there exists some $y = (y_1, \ldots, y_d) \in \mathbb{R}^d$ such that

$$x_{k,i} \to y_i \quad \text{for all} \quad i = 1, \ldots, d.$$

then we say (x_k) converges to y, and write $x_k \to y$ (as $k \to \infty$), or $\lim_{k\to\infty} x_k = y$. If there is no such vector y the sequence is not convergent, or divergent.

Lemma 9.10. *Let $(x_k)_{k\in\mathbb{N}} \subset \mathbb{R}^d$ be a sequence of vectors and $y \in \mathbb{R}^d$. Then,*

$$x_k \to y \quad \iff \quad x_k - y \to 0 \quad \iff \quad \|x_k - y\|_2 \to 0.$$

Definition 9.11 (Norm). Let $\|\cdot\| : \mathbb{R}^d \to \mathbb{R}^+$ is called norm if there holds

1) $\|x\| = 0 \iff x = 0$

2) For any $\lambda \in \mathbb{R}$ and any $x \in \mathbb{R}^d$ we have $\|\lambda x\| = |\lambda| \cdot \|x\|$.

3) For all $x, y \in \mathbb{R}^d$ we have $\|x + y\| \leq \|x\| + \|y\|$.

The properties 1) - 3) are called **definiteness, homogeneity** and **triangle inequality**.

Lemma 9.12. *For any $x \in \mathbb{R}^d$ we have*

1) $\|x\|_\infty \leq \|x\|_2 \leq \sqrt{d}\, \|x\|_\infty$,

2) $\frac{1}{\sqrt{d}} \|x\|_1 \leq \|x\|_2 \leq \sqrt{d}\, \|x\|_1$,

3) $\|x\|_\infty \leq \|x\|_1 \leq d\, \|x\|_\infty$.

9.2 Continuous functions

Definition 9.13. Let $\Omega \subset \mathbb{R}^d$, $f\colon \Omega \to \mathbb{R}^m$ and $x_0 \in \Omega$.
Then we call f **continuous at** x_0 if for any sequence $(x_k)_{k\in\mathbb{N}} \subset \mathbb{R}^d$, such that $x_k \in \Omega$ for all k and $x_k \to x_0$, we have

$$\lim_{k\to\infty} f(x_k) = f\left(\lim_{k\to\infty} x_k\right) = f(x_0).$$

If $U \subset \Omega$ and f is continuous at any $x_0 \in U$, then we say that f is continuous on U.

Definition 9.14 (Limit of functions). Let $\Omega \subset \mathbb{R}^d$ and $f\colon \Omega \to \mathbb{R}^m$. Moreover, let $y \in \mathbb{R}^m$ and $x_0 \in \mathbb{R}^d$ be an **accumulation point** of Ω, i.e., there is a sequence $(x_k)_{k\in\mathbb{N}} \subset \Omega \setminus \{x_0\}$ with $x_k \to x_0$. Then, we call y the **limit of f as $x \to x_0$**, if for all sequences $(x_k) \subset \Omega \setminus \{x_0\}$ with $x_k \to x_0$, we have

$$f(x_k) \to y.$$

In this case we use the notation

$$\lim_{x\to x_0} f(x) = y.$$

Lemma 9.15. *Let $f, g\colon \mathbb{R}^d \to \mathbb{R}$ be continuous. Then $f+g, f\cdot g$ are also continuous. If additionally $g \neq 0$, then $\frac{f}{g}$ is also continuous.*

Moreover, for a cont. function $h\colon D \to \mathbb{R}$ with $D \supset f(\mathbb{R}^d)$, we get that $h \circ f$ is continuous.

Theorem 9.22. *Let $f\colon \Omega \to \mathbb{R}$ and $p \in \{1, 2, \infty\}$. Then, f is continuous at $x_0 \in \Omega$ if and only if for any $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $x \in \Omega$, we have*

$$\|x - x_0\|_p < \delta \implies |f(x) - f(x_0)| < \varepsilon.$$

In a formula,

$$\forall \varepsilon > 0\, \exists \delta > 0\, \forall x \in \Omega\colon\ \|x - x_0\|_p < \delta \implies |f(x) - f(x_0)| < \varepsilon.$$

9.3 Differential calculus

Definition 9.23. Let $x \in \mathbb{R}^d$, $\varepsilon > 0$ and $\|\cdot\|$ be a norm on $\mathbb{R}^d$. Then we define

$$U_\varepsilon(x) = \{y \in \mathbb{R}^d\colon\ \|x - y\| < \varepsilon\}$$

as **open neighborhood** or **open ball** (w.r.t. $\|\cdot\|$) of radius ε around x.

A set $G \subset \mathbb{R}^d$ is called **open** if for every $x \in G$ there exists some $\varepsilon > 0$ such that $U_\varepsilon(x) \subset G$. Moreover, a set $C \subset \mathbb{R}^d$ is called **closed** (in $\mathbb{R}^d$) if $\mathbb{R}^d \setminus C$ is open.

9.3.1 Partial derivatives

Definition 9.24 (Partial derivative). Let $G \subset \mathbb{R}^d$ be an open set, $f\colon G \to \mathbb{R}$ and $x \in G$. If the limit

$$\frac{\partial f}{\partial x_i}(x) := \lim_{h \to 0} \frac{f(x + he_i) - f(x)}{h}$$

exists, then we call the function f **partially differentiable** at x w.r.t. the i-th coordinate.

If $\frac{\partial f}{\partial x_i}(x)$ exists for any $1 \leq i \leq n$ and all $x \in M \subset G$, then we call f **partially differentiable** in M and simply partially differentiable if $M = G$.

If f is partially differentiable (in a neighborhood of x) and all partial derivatives are continuous (at x), then we say f is **continuously partially differentiable** (at x).

Definition 9.30 (Gradient). Let $f\colon G \to \mathbb{R}$ be partially differentiable in G. Then, we call

$$(\operatorname{grad} f)(x) = \nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \ldots, \frac{\partial f}{x_d}(x)\right)$$

the **gradient of** f at the point $x \in G$.

Theorem 9.33 (Product rule). *Let $f, g\colon G \to \mathbb{R}$ be partially differentiable functions. Then we have*

$$\nabla(fg) = (\nabla f) \cdot g + (\nabla g) \cdot f.$$

9.3.2 (Total) differentiability

Definition 9.35. Let $G \subset \mathbb{R}^d$ be an open set, $f\colon G \to \mathbb{R}$ and $x \in G$.
We call f **(totally) differentiable** at x if there exists a linear mapping $df_x\colon \mathbb{R}^d \to \mathbb{R}$ such that

$$\lim_{y\to 0} \frac{f(x+y) - f(x) - df_x(y)}{\|y\|} = 0.$$

The mapping df_x is called **(total) derivative** (or differential) of f at x.

Equivalently, f is **(totally) differentiable** at x with derivative df_x if there exists a function $r\colon \mathbb{R}^d \to \mathbb{R}$ such that

$$f(x+y) = f(x) + df_x(y) + r(y)$$

(whenever $x + y \in G$) and

$$\lim_{y\to 0} \frac{r(y)}{\|y\|} = 0.$$

If f is differentiable at every point of G, then we simply say f is differentiable.

Theorem 9.39. *Let $G \subset \mathbb{R}^d$ be open and $f\colon G \to \mathbb{R}$ be (totally) differentiable at $x \in G$. Then,*

1) *f is continuous at x,*
2) *all partial derivatives of f exist at x, and*
3) *the (total) derivative of f is given by the gradient by*

$$df_x(y) = \nabla f(x) \cdot y = \langle \nabla f(x), y\rangle = \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(x) \cdot y_i.$$

Theorem 9.41. *Let $f\colon G \to \mathbb{R}$ be a continuously partially differentiable function at $x \in G$. Then, f is (totally) differentiable at x and the derivative is given by the gradient.*

9.3.3 Directional derivatives

Definition 9.45 (Directional derivatives). Let $G \subset \mathbb{R}^d$ be open, $f\colon G \to \mathbb{R}$ and $x \in G$. A vector $v \in \mathbb{R}^d$ with $\|v\|_2 = 1$ is called a **direction**, and we define the set of all directions, i.e., the $(d-1)$-dimensional **unit sphere**, by

$$\mathbb{S}^{d-1} := \{v \in \mathbb{R}^d \colon \|v\|_2 = 1\}.$$

The **directional derivative** of f at $x \in G$ w.r.t. v is given by

$$D_v f(x) \;=\; \lim_{h\to 0} \frac{f(x+hv) - f(x)}{h}$$

if the limit exists.

Theorem 9.49. *Let $G \subset \mathbb{R}^d$ be open, $f\colon G \to \mathbb{R}$ be differentiable at $x \in G$ and $v \in \mathbb{S}^{d-1}$. Then, the directional derivative at x w.r.t. v can be computed as*

$$D_v f(x) \;=\; df_x(v) \;=\; \langle \nabla f(x), v\rangle \;=\; \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(x) \cdot v_i.$$

Theorem 9.52. *Let $G \subset \mathbb{R}^d$ and $f\colon G \to \mathbb{R}$ be differentiable at $x \in G$. Then,*

$$|D_v f(x)| \;\le\; \|\nabla f(x)\| \qquad \text{for all} \quad v \in \mathbb{S}^{d-1} = \{v \in \mathbb{R}^d \colon \|v\| = 1\}.$$

Moreover, $D_v f(x) \;=\; \pm\|\nabla f(x)\|$ if and only if $v = \pm\frac{\nabla f(x)}{\|\nabla f(x)\|}$.

9.3.4 Higher order partial derivatives

Definition 9.53. Let $G \subset \mathbb{R}^d$ be open and $f\colon G \to \mathbb{R}$ be partially differentiable at $x \in G$. If for any $i, j \in \{1, 2, \dots, d\}$ we have that the **second-order partial derivatives**

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(x) := \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}(x)$$

exist, then we call f **twice partially differentiable** at x.

If all first-order partial derivatives are totally differentiable (at x), then we call f **twice differentiable** (at x).

If all second-order partial derivatives are continuous (at x), then we call f **twice continuously differentiable** (at x).

Definition 9.55 (Hessian). Let $G \subset \mathbb{R}^d$ and $f\colon G \to \mathbb{R}$ be twice partially differentiable at $x \in G$. The matrix

$$H_f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right)_{i,j=1}^{d} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_d}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_d}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1}(x) & \frac{\partial^2 f}{\partial x_d \partial x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_d^2}(x) \end{pmatrix}$$

is called the **Hessian (matrix)** (german: Hesse-Matrix) of f at the point x.

Theorem 9.57. *Let $G \subset \mathbb{R}^d$, $f\colon G \to \mathbb{R}$ be twice differentiable at $x \in G$ and $u, v \in \mathbb{S}^{d-1}$. Then, the second-order directional derivative at x w.r.t. u and v can be computed as*

$$D_u(D_v f)(x) = u^T H_f v = \sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \cdot u_i v_j,$$

where $H_f := H_f(x)$ is the Hessian of f at x.

In particular, $D_v^2 f(x) = v^T H_f v$.

Theorem 9.58 (Schwarz's theorem). *Let $G \subset \mathbb{R}^d$ and $f\colon G \to \mathbb{R}$ be twice continuously differentiable at $x \in G$. Then, for any $i, j \in \{1, 2, \dots, d\}$, it holds that*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x).$$

In particular, this shows that the Hessian of f is a symmetric matrix, i.e. $H_f(x) = (H_f(x))^T$.

Definition 9.61. Let $G \subset \mathbb{R}^d$ and $f\colon G \to \mathbb{R}$ be k-times partially differentiable (at $x \in G$). If all partial derivatives of each k-th order partial derivative exist (at $x \in G$), then we say that f is **$(k+1)$-times partially differentiable** (at $x \in G$). We use the notation

$$D_{i_{k+1}} \dots D_{i_2} D_{i_1} f(x) := \frac{\partial}{\partial x_{i_{k+1}}} \cdots \frac{\partial}{\partial x_{i_2}} \frac{\partial f}{\partial x_{i_1}}(x)$$

for $i_j \in 1, \dots, d$ with $j = 1, \dots, k+1$.

If all k-th order partial derivatives are totally differentiable (at x), then we call f **$(k+1)$-times differentiable** (at x).

If all $(k+1)$-st order partial derivatives are continuous (at x), then we call f **$(k+1)$-times continuously differentiable** (at x).

Theorem 9.62. *Let $G \subset \mathbb{R}^d$ and $f\colon G \to \mathbb{R}$ be k-times continuously differentiable. Then for any $i_1, i_2, \dots, i_k \in \{1, 2, \dots, d\}$ and any permutation σ of $\{1, \dots, k\}$ we have*

$$D_{i_{\sigma(k)}} \dots D_{i_{\sigma(2)}} D_{i_{\sigma(1)}} f(x) = D_{i_k} \dots D_{i_2} D_{i_1} f(x).$$

9.4 Extrema

Definition 9.63. Let $\Omega \subset \mathbb{R}^d$ and $f\colon \Omega \to \mathbb{R}$.
Then, a point $x_0 \in \Omega$ is called **local minimum**, if there exists some $\varepsilon > 0$ such that

$$f(x) \geq f(x_0) \qquad \forall x \in U_\varepsilon(x_0) \cap \Omega.$$

and if $f(x) > f(x_0)$ for $x \in U_\varepsilon(x_0) \cap \Omega \setminus \{x_0\}$, we call x_0 a **strict local minimum**.

Analogously, we call x_0 a **local maximum**, if

$$f(x) \leq f(x_0) \qquad \forall x \in U_\varepsilon(x_0) \cap \Omega,$$

and a **strict local maximum**, if $f(x) < f(x_0)$ for all $x \in U_\varepsilon(x_0) \cap \Omega \setminus \{x_0\}$.

If $f(x) \geq f(x_0)$ or $f(x) \leq f(x_0)$ for all $x \in \Omega$, then we call x_0 a **(global) minimum** or **(global) maximum** of f, respectively.

We say that x_0 is a **(local) extremum**, if x_0 is a (local) maximum or minimum.

The function value $f(x_0)$ at an extremum x_0 is called **minimum/maximum value** of f.

Theorem 9.67 (Necessary condition for an extreme point). *Let $G \subset \mathbb{R}^d$ be open and $f\colon G \to \mathbb{R}$ be partially differentiable. If $x_0 \in G$ is a local extremum, then*

$$\nabla f(x_0) = 0.$$

*A point $x_0 \in G$ such that $\nabla f(x_0) = 0$ is called **stationary point** of f.*

Definition 9.70. Let $A \in \mathbb{R}^{d\times d}$ be a symmetric matrix. We call A ...

- **positive-definite** if
$$x^T Ax > 0 \qquad \text{for all } x \in \mathbb{R}^d \setminus \{0\}.$$
- **positive semi-definite** if
$$x^T Ax \geq 0 \qquad \text{for all } x \in \mathbb{R}^d.$$
- **negative-definite** if
$$x^T Ax < 0 \qquad \text{for all } x \in \mathbb{R}^d \setminus \{0\}.$$
- **negative semi-definite** if
$$x^T Ax \leq 0 \qquad \text{for all } x \in \mathbb{R}^d.$$

If A is neither positive semi-definite nor negative semi-definite, then A is called **indefinite**.

Theorem 9.73 (Second (partial) derivative test)**.** *Let $f\colon G \to \mathbb{R}$ be twice continuously differentiable and let $x_0 \in G$ such that $\nabla f(x_0) = 0$. Then, we have*

1) $H_f(x_0)$ is positive-definite $\implies$ x_0 is a strict local minimum

2) $H_f(x_0)$ is negative-definite $\implies$ x_0 is a strict local maximum

3) $H_f(x_0)$ is indefinite $\implies$ x_0 is not an extremum of f

In all other cases (i.e., semi-definite but not definite), we do not gain information from the second derivative test.

Lemma 9.75. *Let $A \in \mathbb{R}^{d\times d}$ be a symmetric matrix. Then,*
$$\lambda_{\min} \leq \frac{x^T Ax}{x^T x} \leq \lambda_{\max} \qquad \text{for all } x \in \mathbb{R}^d \setminus \{0\},$$
where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalue of A, respectively.

In particular, A is ...

- *positive-definite if and only if all eigenvalues of A are positive, i.e. $\sigma(A) \subset (0, \infty)$.*
- *negative-definite if and only if all eigenvalues of A are negative, i.e. $\sigma(A) \subset (-\infty, 0)$.*
- *indefinite if and only if A has positive and negative eigenvalues.*

See Definition 8.85 for the definition of the spectrum $\sigma(A)$.

Lemma 9.80 (Sylvester's criterion). *Let $A = (a_{ij})_{i,j=1}^d \in \mathbb{R}^{d\times d}$ be a symmetric matrix, and let $A_k = (a_{ij})_{i,j=1}^k \in \mathbb{R}^{k\times k}$ be the (upper left) submatrices of A. Then, A is ...*

- *positive-definite if and only if* $\det(A_k) > 0$ *for all* $k = 1, \dots, d$.
- *negative-definite if and only if* $\det(A_k) > 0$ *for even* k *and* $\det(A_k) < 0$ *for odd* k.

Corollary 9.83 (Second derivative test for $d = 2$). *Let $G \subset \mathbb{R}^2$, $f\colon G \to \mathbb{R}$ be twice continuously differentiable and let $x_0 \in G$ such that $\nabla f(x_0) = 0$.*
Moreover, let $H := H_f(x_0)$ be the Hessian of f at x_0 with upper left entry H_{11}. Then, we have

1) $\det(H) > 0$ *and* $H_{11} > 0 \quad \Longrightarrow \quad$ *x_0 is a strict local minimum*

2) $\det(H) > 0$ *and* $H_{11} < 0 \quad \Longrightarrow \quad$ *x_0 is a strict local maximum*

3) $\det(H) < 0 \quad \Longrightarrow \quad$ *x_0 is not an extremum of f*

If $\det(H) = 0$, *we do not gain information from the second derivative test.*

9.4.1 Extrema subject to constraints

Theorem 9.86 (Necessary condition). *Let $f, g\colon \mathbb{R}^d \to \mathbb{R}$ be continuously differentiable functions, and x_0 be a local extremum of f subject to the constraint $g(x) = c$ for some $c \in \mathbb{R}$. Then, if $\nabla g(x_0) \neq 0$, there exists a constant λ such that*

$$\nabla f(x_0) = -\lambda \nabla g(x_0).$$

9.5 Taylor series

Lemma 9.91. *Let $G \subset \mathbb{R}^d$ be open, $f\colon G \to \mathbb{R}$ be k-times continuously differentiable and let $x, y \in G$ such that for any $t \in [0,1]$ we have that $x + ty \in G$.*
Then, $g\colon [0,1] \to \mathbb{R}$ with $g(t) := f(x+ty)$ is k-times continuously differentiable and we have

$$g^{(k)}(t) = \frac{d^k g}{dt^k}(t) = \sum_{|\alpha|=k} \frac{k!}{\alpha!} D^\alpha f(x+ty) \cdot y^\alpha.$$

Theorem 9.93 (Taylor's theorem). *Let $G \subset \mathbb{R}^d$ be open, $f\colon G \to \mathbb{R}$ be $(n+1)$-times continuously differentiable and $x, y \in G$ such that for any $t \in [0,1]$ the point $x + ty$ is also contained in G. Then there exists some $\theta \in (0,1)$ such that*

$$f(x+y) = \sum_{|\alpha|\le n} \frac{D^\alpha f(x)}{\alpha!} \cdot y^\alpha + \sum_{|\alpha|=n+1} \frac{D^\alpha f(x+\theta y)}{\alpha!} \cdot y^\alpha.$$

Corollary 9.94. *Let $G \subset \mathbb{R}^d$, $f\colon G \to \mathbb{R}$ be $(n+1)$-times continuously differentiable, $x_0 \in G$ and let $U = U(x_0)$ be a neighborhood of x_0 which is completely contained in G. Then, for any $x \in U$ we have the representation*

$$f(x) = \sum_{|\alpha|\le n} \frac{D^\alpha f(x_0)}{\alpha!}(x-x_0)^\alpha + \sum_{|\alpha|=n+1} \frac{D^\alpha f(\xi)}{\alpha!}(x-x_0)^\alpha$$

for some ξ between x_0 and x, i.e., $\xi = x_0 + \theta(x-x_0)$ for some $\theta \in (0,1)$. We call

$$T_n(x) := \sum_{|\alpha|\le n} \frac{D^\alpha f(x_0)}{\alpha!}(x-x_0)^\alpha,$$

the ***Taylor polynomial of f of order n (at x_0).***

The term

$$R_n(x) := f(x) - T_n(x) = \sum_{|\alpha|=n+1} \frac{D^\alpha f(\xi)}{\alpha!}(x-x_0)^\alpha$$

is called the ***remainder*** *of the Taylor polynomial.*

Corollary 9.95. *Let $G \subset \mathbb{R}^d$, $f\colon G \to \mathbb{R}$ be n-times continuously differentiable, $x_0 \in G$ and let $U = U(x_0)$ be a neighborhood of x_0 which is completely contained in G.*
Moreover, assume additionally that $|D^\alpha f(x)| \le M$ for some $M < \infty$, all $\alpha \in \mathbb{N}_0^d$ with $|\alpha| = n$, and all $x \in U(x_0)$. Then for any $x \in U(x_0)$ we have

$$|f(x) - T_n(x)| \le \frac{2M \cdot d^n}{n!} \cdot \|x - x_0\|^n.$$

Definition 9.98. Let $f\colon G \to \mathbb{R}$ be infinitely-often differentiable and let $x_0 \in G$.
The formal series given by

$$Tf(x) = \sum_{\alpha \in \mathbb{N}_0^d} \frac{D^\alpha f(x_0)}{\alpha!}(x-x_0)^\alpha$$

is called **Taylor series** of f at x_0.

Theorem 9.99. *Let $G \subset \mathbb{R}^d$ be open, $f\colon G \to \mathbb{R}$ be infinitely-often differentiable and $x_0 \in G$. If $r > 0$ is such that*

$$\lim_{n\to\infty} \frac{r^n}{n!} \cdot \max_{\alpha \in \mathbb{N}_0^d\colon |\alpha|=n} \sup_{\xi \in U_r(x_0)} |D^\alpha f(\xi)| = 0,$$

and $U_r(x_0) = \{x \in \mathbb{R}^d\colon \|x - x_0\| < r\}$ is completely contained in G.

Then, f can be written by its Taylor series

$$f(x) = \sum_{\alpha \in \mathbb{N}_0^d} \frac{D^\alpha f(x_0)}{\alpha!}(x-x_0)^\alpha \qquad \text{for all} \quad x \in U_r(x_0).$$

9.6 Differential calculus for vector-valued functions

Definition 9.101. Let $G \subset \mathbb{R}^d$ and $f\colon G \to \mathbb{R}^m$.
We call f **continuous, differentiable, etc.** if and only if for every $i = 1, \dots, m$ we have that f_i is continuous, differentiable, etc., respectively.

Definition 9.104. Let $G \subset \mathbb{R}^d$, $f\colon G \to \mathbb{R}^m$ and $x \in G$.
If for any $1 \le j \le d$ and any $1 \le i \le m$ the partial derivative

$$\frac{\partial f_i}{\partial x_j}(x) = \lim_{h \to 0} \frac{f_i(x + he_j) - f_i(x)}{h}$$

exists we call f **partially differentiable** at x.
If this is the case for any $x \in G$ we simply say that f is partial differentiable.

Moreover, if f is partial differentiable at $x \in G$, we call

$$J_f(x) := \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \dots & \frac{\partial f_1}{\partial x_d}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \dots & \frac{\partial f_2}{\partial x_d}(x) \\ \vdots & & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \frac{\partial f_m}{\partial x_2}(x) & \dots & \frac{\partial f_m}{\partial x_d}(x) \end{pmatrix}$$

Jacobian matrix or functional matrix of f at x.

Definition 9.107. Let $G \subset \mathbb{R}^d$, $f\colon G \to \mathbb{R}^m$ be a continuous function.
If there exists a linear mapping $D\colon \mathbb{R}^d \to \mathbb{R}^m$ and a mapping $r\colon \mathbb{R}^d \to \mathbb{R}^m$ such that

$$f(x + y) = f(x) + D(y) + r(y),$$

where r satisfies

$$\lim_{y \to 0} \frac{r(y)}{\|y\|} = 0,$$

then we call f **differentiable** at x. We call $D = df_x$ the (total) derivative of f at x.

Theorem 9.109. *Let $G \subset \mathbb{R}^d$ and $f\colon G \to \mathbb{R}^m$ be differentiable at $x \in G$ with derivative df_x. Then,*

1) f is continuous at x,

2) all partial derivatives of all components f_i exist at x, and

3) the (total) derivative of f is given by the Jacobian by

$$df_x(y) \;=\; J_f(x) \cdot y \;=\; \begin{pmatrix} \langle \nabla f_1(x), y\rangle \\ \langle \nabla f_2(x), y\rangle \\ \vdots \\ \langle \nabla f_m(x), y\rangle \end{pmatrix}.$$

Moreover, if $f\colon G \to \mathbb{R}^m$ is a mapping such that all partial derivatives of all components are continuous at $x \in G$, then f is also differentiable at x.

Theorem 9.111. *Let $G \subset \mathbb{R}^d$, $f\colon G \to \mathbb{R}^p$ and $g\colon \mathbb{R}^p \to \mathbb{R}^m$ for some $d, p, m \in \mathbb{N}$. If f is continuous at x and g is continuous at $y = f(x)$, then $g \circ f$ is continuous at x.*

Theorem 9.112 (Chainrule). *Let $f\colon \mathbb{R}^d \to \mathbb{R}^p$ be differentiable at $x \in \mathbb{R}^d$ and $g\colon \mathbb{R}^p \to \mathbb{R}^m$ be differentiable at $y = f(x) \in \mathbb{R}^p$.*
Then, their composition is also differentiable at x and

$$J_{g\circ f}(x) = J_g(f(x)) \cdot J_f(x).$$

In short, this can be written as $J_{g\circ f} = (J_g \circ f) \cdot J_f$.

9.7 Line integrals

Definition 9.115. Let $[a, b] \subset \mathbb{R}$ be a non-empty closed interval.

Then, every continuous function $\gamma\colon [a, b] \to \mathbb{R}^d$ is called a **path**.

The points $\gamma(a)$ and $\gamma(b)$ are called **initial point** and **endpoint** of the path γ.

A **curve** C in $\mathbb{R}^d$ is the image of $[a, b]$ w.r.t. some path γ, i.e., $C = \gamma([a, b])$.
In this case we call γ a **parametrization** of C.

If $\gamma(a) = \gamma(b)$ the path γ and the curve w.r.t. γ are is called **closed**.

Definition 9.120. Let $\gamma\colon [a, b] \to \mathbb{R}^d$ and $\tilde\gamma\colon [c, d] \to \mathbb{R}^d$ be parametrizations of the curve $C = \gamma([a, b]) = \tilde\gamma([c, d])$. We call γ and $\tilde\gamma$ **equivalent**, write $\gamma \sim \tilde\gamma$, if there exists a continuous strictly increasing bijection $h\colon [a, b] \to [c, d]$ such that $\gamma(t) = \tilde\gamma(h(t))$.

Definition 9.125. Let $\gamma\colon [a,b] \to \mathbb{R}^d$ be a path. We say that γ is **rectifiable** if

$$L(\gamma) := \sup \sum_{i=1}^{n} \|\gamma(t_i) - \gamma(t_{i-1})\| < \infty,$$

where the supremum is over all partitions $a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b$ and all $n \in \mathbb{N}$. In this case we say that $L(\gamma)$ is the **length** of γ.

Lemma 9.127. *Let C be a curve, and let $\gamma_1\colon [a,b] \to \mathbb{R}^d$ and $\gamma_2\colon [c,d] \to \mathbb{R}^d$ be two equivalent parametrizations. Then γ_1 is rectifiable if and only if γ_2 is rectifiable, and in this case it holds*

$$L(\gamma_1) = L(\gamma_2).$$

Theorem 9.128. *Let $\gamma\colon [a,b] \to \mathbb{R}^d$ be a continuously differentiable path. Then γ is rectifiable and we have*

$$L(\gamma) = \int_a^b \|\gamma'(t)\| \, dt = \int_a^b \sqrt{\gamma_1'(t)^2 + \gamma_2'(t)^2 + \cdots + \gamma_d'(t)^2} dt.$$

9.7.1 Line integrals of scalar fields

Definition 9.131. Let $\gamma\colon [a,b] \to \mathbb{R}^d$ be a continuously differentiable and bijective parametrization of a curve C and let $f\colon \gamma([a,b]) \to \mathbb{R}$ be continuous. Then we define the **line integral of f over C** by

$$\int_C f \, ds := \int_a^b f(\gamma(t)) \cdot \|\gamma'(t)\| \, dt.$$

Lemma 9.134. *As in Definition 9.131, let C be a curve with a bijective parametrization $\gamma\colon [a,b] \to \mathbb{R}^d$. Then, every equivalent parametrization $\tilde{\gamma}\colon [c,d] \to \mathbb{R}^d$ is also bijective and we have*

$$\int_a^b f(\gamma(t)) \cdot \|\gamma'(t)\| \, dt = \int_c^d f(\tilde{\gamma}(t)) \cdot \|\tilde{\gamma}'(t)\| \, dt.$$

For such a curve C we define $L(C) := L(\gamma)$ for any bijective parametrization γ.

9.7.2 Line integrals of vector fields

Definition 9.137. Let $\gamma\colon [a,b] \to \mathbb{R}^d$ be a bijective and continuously differentiable parametrization of the curve C. Moreover, let $F\colon \mathbb{R}^d \to \mathbb{R}^d$ be a continuous vector field. Then we call

$$\begin{aligned}\int_C F\,d\gamma &:= \int_a^b \langle F(\gamma(t)), \gamma'(t)\rangle\; dt \\ &= \int_a^b F_1(\gamma(t))\cdot\gamma_1'(t) + F_2(\gamma(t))\cdot\gamma_2'(t) + \cdots + F_d(\gamma(t))\cdot\gamma_d'(t)\,dt\end{aligned}$$

the **line integral of the vector field F along C.**

9.8 Multiple integrals

Definition 9.140 (Riemann-integrable functions)**.** Let $R = \prod_{i=1}^d [a_i,b_i]$ be a box and $f\colon R\to\mathbb{R}$ be a bounded function. Then, if

$$\lim_{n\to\infty} L_n(f) = \lim_{n\to\infty} U_n(f),$$

we call f a **(Riemann-)integrable function** and define the **integral of f over R** by this common limit, i.e.,

$$\int_R f(x)dx := \int_R f(x_1,\ldots,x_d)\,d(x_1,\ldots,x_d) := \lim_{n\to\infty} L_n(f).$$

Lemma 9.141. *Let $R = \prod_{i=1}^d [a_i,b_i]$ and $f\colon R\to\mathbb{R}$ be continuous. Then, f is integrable and*

$$\int_R f(x)dx = \lim_{n\to\infty} Q_n(f).$$

Theorem 9.142 (Fubini)**.** *Let $R = \prod_{i=1}^d [a_i,b_i]$ be a box and $f\colon R\to\mathbb{R}$ be continuous. Then,*

$$\int_R f(x)dx = \int_{a_1}^{b_1}\left(\int_{a_2}^{b_2}\left(\cdots\left(\int_{a_d}^{b_d} f(x_1,\ldots,x_d)\,dx_d\right)\cdots\right)dx_2\right)dx_1.$$

The order of the iterated integrals can be chosen arbitrary.

Definition 9.146 (Jordan-measurable set). Let $A \subset \mathbb{R}^d$ be a bounded domain, i.e., there is a box $R \subset \mathbb{R}^d$ with $A \subset R$. We call A **Jordan-measurable**, if χ_A is Riemann-integrable.

In this case, we define the **volume of** A by

$$\mathrm{vol}_d(A) := \int_R \chi_A(x)\,dx.$$

Moreover, we call a bounded function $f\colon A \to \mathbb{R}$ **integrable over** A, if $f \cdot \chi_A$ is integrable, and we set

$$\int_A f(x)\,dx := \int_R f(x) \cdot \chi_A(x)\,dx.$$

(Here we set $(f \cdot \chi_A)(x) = 0$ for $x \notin A$.)

Definition 9.147 (Normal domains). A bounded set $A \subset \mathbb{R}^2$ of the form

$$A = \Big\{(x_1, x_2) \in \mathbb{R}^2 \colon x_1 \in [a,b], \quad \varphi_1(x_1) \le x_2 \le \psi_1(x_1)\Big\}$$

for some $a, b \in \mathbb{R}$ and continuous functions $\varphi_1, \psi_1\colon [a,b] \to \mathbb{R}$, is called a **normal domain**. (Strict inequalities are also allowed.)

In higher dimensions, we define inductively that $A \subset \mathbb{R}^d$ is called a normal domain if

$$A = \Big\{(x', x_d) \in \mathbb{R}^d \colon x' \in A', \quad \varphi_d(x') \le x_d \le \psi_d(x')\Big\}$$

for some normal domain $A' \in \mathbb{R}^{d-1}$ and continuous functions $\varphi_d, \psi_d\colon A' \to \mathbb{R}$.
(We used for a simpler notation $x' := (x_1, \dots, x_{d-1})$.)

Lemma 9.148. *Let $A \subset \mathbb{R}^2$ be a normal domain of the form*

$$A = \Big\{x \in \mathbb{R}^2 \colon x_1 \in [a,b], \quad \varphi(x_1) \le x_2 \le \psi(x_1)\Big\}$$

where $\varphi \le \psi$ and both are continuous functions. Then, the integral of an integrable function $f\colon A \to \mathbb{R}$ equals

$$\int_A f(x)\,dx = \int_a^b \int_{\varphi(x_1)}^{\psi(x_1)} f(x_1, x_2)\,dx_2\,dx_1.$$

In particular, the area of A equals

$$\mathrm{vol}_2(A) = \int_A 1\,dx = \int_a^b (\psi(t) - \varphi(t))\,dt.$$

Theorem 9.152 (Substitution rule). *Let $G \subset \mathbb{R}^d$ be open and $A \subset G$ be a bounded and Jordan-measurable set. Moreover, let $\Phi\colon G \to \mathbb{R}^d$ be a continuously differentiable and injective function such that either* $\det J_\Phi(u) > 0$ *or* $\det J_\Phi(u) < 0$ *for any $u \in G$, where $J_\Phi(u)$ is the Jacobi matrix of Φ at $u \in G$.*
Then, $\Phi(A)$ is also Jordan-measurable and, for any bounded and continuous function $f\colon \Phi(A) \to \mathbb{R}$, we have

$$\int_{\Phi(A)} f(x)\,dx = \int_A f(\Phi(u)) \cdot |\det J_\Phi(u)|\,du.$$

We say that we use the substitution $x = \Phi(u)$.

9.9 Basic measure theory and the Lebesgue integral

Definition 10.1. Let Ω be some set and $\mathcal{A} \subset \mathcal{P}(\Omega)$ be a family of subsets of Ω.
We call $\mathcal{A}$ a **σ-algebra** (over Ω) if the following properties hold

1) $\Omega \in \mathcal{A}$,
2) $A \in \mathcal{A} \implies A^c = \Omega \setminus A \in \mathcal{A}$,
3) $A_1, A_2, \ldots \in \mathcal{A} \implies \bigcup_{i\in\mathbb{N}} A_i \in \mathcal{A}$.

So, a σ-algebra is a set of sets which is closed under complement and countable unions.

For a set Ω and a σ-algebra $\mathcal{A} \subset \mathcal{P}(\Omega)$, we call the tuple $(\Omega, \mathcal{A})$ a **measurable space**, and the sets $A \in \mathcal{A}$ **measurable sets** (sometimes $\mathcal{A}$-measurable sets).

Definition 10.6. Let $(\Omega, \mathcal{A})$ be a measurable space.
A function $\mu\colon \mathcal{A} \to [0, \infty]$ is called **measure** if

1) $\mu(\varnothing) = 0$
2) For pairwise disjoint $A_1, A_2, \ldots \in \mathcal{A}$ we have $\mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \mu(A_i)$.

The second property is called σ-additivity.

Such a triple $(\Omega, \mathcal{A}, \mu)$ is called a **measure space**.

Definition 10.7 (Null set). Let $(\Omega, \mathcal{A}, \mu)$ be a measure space.

A set $N \in \mathcal{A}$ such that $\mu(N) = 0$ is called **null set**.

If a property holds for all $x \in \Omega \setminus N$, i.e., for all x except a set of measure 0, then we say the property holds **μ-almost everywhere** (μ-a.e.) or for **μ-almost all** $x \in \Omega$.

Lemma 10.8. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space.*
Then, for all $A, B, A_1, A_2, \ldots \in \mathcal{A}$, we have

1) $\mu(A) \leq \mu(B)$ *if $A \subset B$, and* *(Monotonicity)*
2) $\mu(B \setminus A) = \mu(B) - \mu(A)$ *if $A \subset B$ and $\mu(A) < \infty$,* *(Differences)*
3) $\mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) \leq \sum_{i\in\mathbb{N}} \mu(A_i)$ *for all $A_1, A_2, \ldots \in \mathcal{A}$.* *(Subadditivity)*

Theorem 10.14 (Caratheodory's extension theorem)**.**
Let $\mathcal{E} \subset \mathcal{P}(\Omega)$ and $\mu' \colon \mathcal{E} \to [0, \infty]$ be a function with $\mu'(\varnothing) = 0$ that satisfies

1. $A \cup B \in \mathcal{E}$ *for all $A, B \in \mathcal{E}$,*
2. $A \setminus B \in \mathcal{E}$ *for all $A, B \in \mathcal{E}$,*
3. *there exist $E_1, E_2, \ldots \in \mathcal{E}$ with $\Omega = \bigcup_{i=1}^{\infty} E_i$ and $\mu'(E_i) < \infty$ for all i, and*
4. $\mu'\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \mu'(A_i)$ *for all pairwise disjoint $A_1, A_2, \ldots \in \mathcal{E}$.*

Then there exists a unique measure $\mu \colon \sigma(\mathcal{E}) \to [0, \infty]$ with $\mu(E) = \mu'(E)$ for all $E \in \mathcal{E}$.

A function μ' as above is also called a ***pre-measure****, and μ the* ***extension*** *of μ'.*

Definition 10.15. Let $\Omega \subset \mathbb{R}^d$. The σ-algebra $\mathcal{B}(\Omega) := \sigma(\mathcal{E})$, i.e., the σ-algebra generated by all half-open boxes in Ω, is called the **Borel-σ-algebra**. Sets of the Borel-σ-algebra are called **Borel-measurable** or **Borel-sets**.

Corollary 10.16 (Borel-Lebesgue measure)**.** *Let $\mathcal{B}(\Omega)$ be the Borel-σ-algebra over $\Omega \subset \mathbb{R}^d$. Then there exists a unique measure $\mu \colon \mathcal{B}(\Omega) \to [0, \infty]$ which assigns each rectangle its volume. This measure is called the (d-dimensional)* ***Borel-Lebesgue measure****.*

Corollary 10.17 (Lebesgue measure)**.** *Let $\mathcal{L}(\Omega)$ be the Lebesgue-σ-algebra over $\Omega \subset \mathbb{R}^d$. Then there is a unique measure $\lambda_d \colon \mathcal{L}(\Omega) \to [0, \infty]$ which assigns each rectangle its volume, and zero to every set in $\mathcal{N}$. This measure is called the (d-dimensional)* ***Lebesgue measure****.*

Lemma 10.20. *A set $N \subset \mathbb{R}^d$ is a null set w.r.t. the Lebesgue measure λ_d, i.e., $N \in \mathcal{N}$, if and only if*

$$\forall \varepsilon > 0 \; \exists A_1, A_2, \ldots \in \mathcal{B}(\Omega)\colon \quad N \subset \bigcup_{i=1}^{\infty} A_i \quad \text{and} \quad \sum_{i=1}^{\infty} \lambda_d(A_i) \leq \varepsilon.$$

The sets A_i can chosen to be boxes.

9.9.1 Measurable functions

Definition 10.22. Let $(\Omega, \mathcal{A})$ be a measurable space and let $f: \Omega \to \mathbb{R}$ be a real-valued function. If

$$\forall B \in \mathcal{B}(\mathbb{R}): f^{-1}(B) \in \mathcal{A},$$

then we call f a **measurable function**.

We sometimes write $f: (\Omega, \mathcal{A}) \to \mathbb{R}$ or call f $\mathcal{A}$-measurable to indicate the σ-algebra.

A measurable function $f: (\Omega, \mathcal{L}(\Omega)) \to \mathbb{R}$ for $\Omega \in \mathbb{R}^d$ is called **(Lebesgue-)measurable**.

Lemma 10.24. *Let $(\Omega, \mathcal{A})$ be a measurable space and let $f: \Omega \to \mathbb{R}$ be a real-valued function. Then, f is measurable if and only if*

$$f^{-1}((-\infty, a)) = \{x \in \Omega: f(x) < a\} \in \mathcal{A},$$

for all $a \in \mathbb{R}$.

Even more, it is enough to check the condition for measurability in Definition 10.22 only for all $B \in \mathcal{E}$, where $\mathcal{E}$ is an arbitrary generator of $\mathcal{B}(\mathbb{R})$.

Lemma 10.26. *Let $(\Omega, \mathcal{A})$ be a measurable space and let $f: \Omega \to \mathbb{R}$ be a measurable function. Then, the **level sets***

$$f^{-1}(\{c\}) = \{x \in \Omega: f(x) = c\}$$

are measurable for any $c \in \mathbb{R}$, i.e., $f^{-1}(\{c\}) \in \mathcal{A}$.

Lemma 10.27. *Let $\Omega \subset \mathbb{R}^d$ and let $f: \Omega \to \mathbb{R}$ be continuous. Then f is Lebesgue-measurable.*

Lemma 10.29. *Let $(\Omega, \mathcal{A})$ be a measurable space and let f, g be measurable functions. Then $f + g$, $f \cdot g$, $\max\{f, g\}$, $\min\{f, g\}$, $f^+ = \max\{0, f\}$, $f^- = \min\{0, f\}$, $|f|$ and αf for $\alpha \in \mathbb{R}$ are also measurable functions.*

Moreover, if $(f_n)_n$ is a sequence of measurable functions, then $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$ and $\liminf_n f_n$ (considered point-wise) are also measurable functions.

Definition 10.30 (Simple functions). Let $(\Omega, \mathcal{A})$ be a measurable space, $A_1, \dots, A_n \in \mathcal{A}$ and $\alpha_1, \dots, \alpha_n \in \mathbb{R}$. We call

$$f(x) = \sum_{k=1}^{n} \alpha_k \, \chi_{A_k}(x)$$

a **simple function**. If the A_k are pairwise disjoint and $\alpha_i \neq \alpha_j$ for $i \neq j$, then we say that the simple function is in **canonical form**.

Lemma 10.31. *Let $(\Omega, \mathcal{A})$ be a measurable space and $f\colon \Omega \to \mathbb{R}$ be a non-negative measurable function. Then, there is a sequence of simple functions $(f_n)_{n\in\mathbb{N}}$ such that $f_1 \le f_2 \le f_3 \le \dots$ and*

$$f(x) = \lim_{n\to\infty} f_n(x).$$

That is, we can write f point-wise as a monotone limit of simple functions.

Corollary 10.32. *Let $(\Omega, \mathcal{A})$ be a measurable space and let f be a measurable function. Then, f is the point-wise limit of simple functions.*

9.9.2 The Lebesgue integral

Definition 10.33 (Integral for simple functions)**.** Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let f be a non-negative simple function, i.e.,

$$0 \le f(x) := \sum_{k=1}^{n} \alpha_k \, \chi_{A_k}(x), \qquad \alpha_k > 0, \; A_k \in \mathcal{A}.$$

Then, we define its **integral** over a measurable set $E \in \mathcal{A}$ by

$$\int_E f \, d\mu := \sum_{k=1}^{n} \alpha_k \, \mu(A_k \cap E).$$

The special choice of $\Omega \subset \mathbb{R}^d$, $\mathcal{A} = \mathcal{L}(\Omega)$ and $\mu = \lambda_d$ leads to the **Lebesgue integral** of f. In this case we use the notation

$$\int_E f(x) \, dx = \int_E f \, d\lambda_d.$$

Lemma 10.38. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, $E \in \mathcal{A}$ and f, g be non-negative simple functions. Then, we have*

1) $\nu\colon \mathcal{A} \to [0, \infty]$ with $\nu(E) := \int_E f \, d\mu$ defines a measure on $\mathcal{A}$.

2) $\int_E f + g \, d\mu = \int_E f \, d\mu + \int_E g \, d\mu$.

3) For $\alpha \in \mathbb{R}$ we have $\int_E \alpha \cdot f \, d\mu = \alpha \cdot \int_E f \, d\mu$.

4) If $f \le g$ then also $\int_E f \, d\mu \le \int_E g \, d\mu$.

Definition 10.39 (Integral for non-negative functions)**.** Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let f be a non-negative measurable function. For a measurable set $E \in \mathcal{A}$ we define

$$\int_E f \, d\mu := \sup \left\{ \int_E g \, d\mu \colon g \text{ is a simple function with } 0 \le g \le f \right\}.$$

Lemma 10.40. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, f, g be non-negative measurable functions and $E, F \in \mathcal{A}$ be measurable sets. Then,*

1) *If $f \leq g$, then $\int_E f\, d\mu \leq \int_E g\, d\mu$*
2) *If $E \subset F$, then $\int_E f\, d\mu \leq \int_F f\, d\mu$.*
3) *If $f = 0$ on E, then $\int_E f\, d\mu = 0$*
4) *If $\mu(E) = 0$, then $\int_E f\, d\mu = 0$.*

Theorem 10.42 (Monotone convergence theorem)**.** *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let $(f_n)_{n\in\mathbb{N}}$ be a sequence of non-negative measurable functions such that for all $x \in \Omega$,*

$$f_1(x) \leq f_2(x) \leq f_3(x) \leq \dots \qquad \text{and} \qquad \lim_{n\to\infty} f_n(x) = f(x).$$

Then, for any measurable E, we have

$$\int_E f\, d\mu = \int_E \lim_{n\to\infty} f_n\, d\mu = \lim_{n\to\infty} \int_E f_n\, d\mu.$$

Lemma 10.44. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, f, g be non-negative measurable functions and $\alpha \in \mathbb{R}$. Then,*

$$\int_E (f+g)\, d\mu = \int_E f\, d\mu + \int_E g\, d\mu$$

and

$$\int_E \alpha f\, d\mu = \alpha \int_E f\, d\mu,$$

for any measurable E.

Lemma 10.45. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, $E \subset \mathcal{A}$ and f be a non-negative measurable function. Then,*

$$\int_E f\, d\mu = \int_\Omega \chi_E \cdot f\, d\mu.$$

Lemma 10.46. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let $g_1, g_2, \dots$ be non-negative measurable functions. Then $g = \sum_{k=1}^{\infty} g_k$ is measurable and for measurable E we have*

$$\int_E g\, d\mu = \sum_{k=1}^{\infty} \int_E g_k\, d\mu.$$

Definition 10.48. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, $E \in \mathcal{A}$ and $f = f^+ - f^-$ be a measurable function. We call f **integrable over** E if

$$\int_E f^+ \, d\mu < \infty \quad \text{and} \quad \int_E f^- \, d\mu < \infty,$$

and we define

$$\int_E f \, d\mu := \int_E f^+ \, d\mu - \int_E f^- \, d\mu.$$

For $E = \Omega$ we call f just integrable.

Lemma 10.50. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, $E \in \mathcal{A}$, f, g be integrable functions and $\alpha \in \mathbb{R}$. Then, we have*

1) $f + g$ *is integrable with* $\int_E f + g \, d\mu = \int_E f \, d\mu + \int_E g \, d\mu$

2) $\int_E \alpha f \, d\mu = \alpha \int_E f \, d\mu$

3) $f \geq 0 \implies \int_E f \, d\mu \geq 0$

4) $|\int_E f \, d\mu| \leq \int_E |f| \, d\mu$

9.9.3 Lebesgue's theorem

Lemma 10.52 (Fatou). *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, $E \in \mathcal{A}$ and $(f_n)_{n \in \mathbb{N}}$ be a sequence of non-negative measurable functions. Then*

$$\int_E \liminf_n f_n \, d\mu \leq \liminf_n \int_E f_n \, d\mu.$$

If there exists a non-negative integrable function g such that $f_n \leq g$ for all $n \in \mathbb{N}$, then

$$\limsup_{n \to \infty} \int_E f_n d\mu \leq \int_E \limsup_{n \to \infty} f_n d\mu.$$

Definition 10.53. A measure space $(\Omega, \mathcal{A}, \mu)$ is called **complete** if

$$A \subset N \text{ for some } N \in \mathcal{A} \text{ with } \mu(N) = 0 \quad \Longrightarrow \quad A \in \mathcal{A}.$$

That is, arbitrary subsets of null sets are measurable in a complete measure space.

Lemma 10.55. *Let $(\Omega, \mathcal{A}, \mu)$ be a complete measure space and g be a measurable function. If $f\colon \Omega \to \mathbb{R}$ is such that $f = g$ almost everywhere, i.e., $f(x) = g(x)$ for all $x \in \Omega \setminus N$ for some $N \in \mathcal{A}$ with $\mu(N) = 0$, then, f is measurable.*

Moreover, if g is integrable, then f is integrable and

$$\int_E f \, d\mu = \int_E g \, d\mu$$

for any measurable $E \in \mathcal{A}$.

In particular, $\int_E f \, d\mu$ is well-defined if f is only defined almost everywhere.

Theorem 10.56 (Dominated convergence theorem). *Let $(\Omega, \mathcal{A}, \mu)$ be a complete measure space and let $(f_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions such that $f(x) = \lim_n f_n(x)$ exists for almost all $x \in \Omega$. If there exists a integrable function $g\colon \Omega \to [0, \infty)$ such that $\forall n \in \mathbb{N}\colon |f_n| \le g$, then all f_n and f are integrable and for any measurable $E \subset \Omega$ it holds*

$$\int_E f \, d\mu = \int_E \lim_{n\to\infty} f_n \, d\mu = \lim_{n\to\infty} \int_E f_n \, d\mu.$$

*We call g the **integrable majorant** of (f_n).*

Lemma 10.59. *Let $A \subset \mathbb{R}^d$ be a Jordan-measurable set and $f\colon A \to \mathbb{R}$ be Riemann-integrable, see Definition 9.146. Then, f is Lebesgue-integrable and the integrals coincide.*

Lemma 10.63 (Differentiation under the integral sign). *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and $I \subset \mathbb{R}$ be an open interval. If a function $f\colon \Omega \times I \to \mathbb{R}$ satisfies*

- *$x \mapsto f(x, t)$ is integrable for each fixed $t \in I$,*
- *$t \mapsto f(x, t)$ is differentiable for each fixed $x \in \Omega$, and*
- *there is some integrable function g with $|\frac{\partial f}{\partial t}(x, t)| \le g(x)$ for all $x \in \Omega$ and $t \in I$.*

Then,

$$\frac{d}{dt} \int_\Omega f(x, t) \, d\mu(x) = \int_\Omega \frac{\partial}{\partial t} f(x, t) \, d\mu(x) \qquad \text{for all } t \in I.$$

9.9.4 Product measures and Fubini's theorem

Definition 10.66. Let $(\Omega_1, \mathcal{A}_1, \mu_1)$, $(\Omega_2, \mathcal{A}_2, \mu_2)$ be measure spaces.
The **product σ-algebra**, denoted by $\mathcal{A}_1 \otimes \mathcal{A}_2$, is the smallest σ-algebra (over $\Omega_1 \times \Omega_2$) that contains all sets of the form $A_1 \times A_2$ with $A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$, i.e., $\mathcal{A}_1 \otimes \mathcal{A}_2 := \sigma(\mathcal{A}_1 \times \mathcal{A}_2)$.

A measure μ on $\mathcal{A}_1 \otimes \mathcal{A}_2$ which satisfies

$$\mu(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2)$$

is called **product measure** of μ_1 and μ_2.

Definition 10.67. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space.

If $\mu(\Omega) < \infty$, then μ is called a **finite measure** and $(\Omega, \mathcal{A}, \mu)$ is a **finite measure space**.

Moreover, if there exist $E_1 \subset E_2 \subset \ldots \in \mathcal{A}$ with $\Omega = \bigcup_{i=1}^{\infty} E_i$ and $\mu(E_i) < \infty$ for all i, then μ is called a **σ-finite measure** and $(\Omega, \mathcal{A}, \mu)$ is called **σ-finite measure space**.

Lemma 10.70. *Let $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ be measurable spaces. Then, for any $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$, $x \in \Omega_1$ and $y \in \Omega_2$, we have $A^x \in \mathcal{A}_2$ and $A_y \in \mathcal{A}_1$.*

Theorem 10.71. *Let $(\Omega_1, \mathcal{A}_1, \mu_1)$, $(\Omega_2, \mathcal{A}_2, \mu_2)$ be σ-finite measure spaces. Then, there is a unique product measure of μ_1 and μ_2 on $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$, which is denoted by $\mu_1 \otimes \mu_2$.*

For every $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$, this measure satisfies

$$(\mu_1 \otimes \mu_2)(A) = \int_{\Omega_1} \mu_2(A^x)\, d\mu_1(x) = \int_{\Omega_2} \mu_1(A_y)\, d\mu_2(y).$$

Lemma 10.73. *Let $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ be measurable spaces. Then, for any $x \in \Omega_1$, $y \in \Omega_2$ and any $(\mathcal{A}_1 \otimes \mathcal{A}_2)$-measurable function $f\colon \Omega_1 \times \Omega_2 \to \mathbb{R}$, we have that f^x and f_y are measurable.*

Theorem 10.74 (Tonelli). *Let $(\Omega_1, \mathcal{A}_1, \mu_1)$, $(\Omega_2, \mathcal{A}_2, \mu_2)$ be σ-finite measure spaces. Let $f\colon \Omega_1 \times \Omega_2 \to [0, \infty)$ be measurable (precisely, $(\mathcal{A}_1 \otimes \mathcal{A}_2)$-measurable). Then,*

- $\int_{\Omega_2} f^x \, d\mu_2$ *is $\mathcal{A}_1$-measurable (as function of $x \in \Omega_1$),*
- $\int_{\Omega_1} f_y \, d\mu_2$ *is $\mathcal{A}_2$-measurable (as function of $y \in \Omega_2$),*

and we have

$$\int_{\Omega_1\times\Omega_2} f \, d(\mu_1 \otimes \mu_2) = \int_{\Omega_1} \left(\int_{\Omega_2} f \, d\mu_2 \right) d\mu_1 = \int_{\Omega_2} \left(\int_{\Omega_1} f \, d\mu_1 \right) d\mu_2,$$

or, written out,

$$\int_{\Omega_1\times\Omega_2} f \, d(\mu_1 \otimes \mu_2) = \int_{\Omega_1} \left(\int_{\Omega_2} f(x,y) \, d\mu_2(y) \right) d\mu_1(x) = \int_{\Omega_2} \left(\int_{\Omega_1} f(x,y) \, d\mu_1(x) \right) d\mu_2(y).$$

Again, all integrals might be ∞ here.

Theorem 10.76 (Fubini). *Let $(\Omega_1, \mathcal{A}_1, \mu_1)$, $(\Omega_2, \mathcal{A}_2, \mu_2)$ be σ-finite measure spaces. Let $f\colon \Omega_1 \times \Omega_2 \to \mathbb{R}$ be integrable on $\Omega_1 \times \Omega_2$. Then,*

- f^x *is μ_2-integrable for almost every $x \in \Omega_1$,*
- f_y *is μ_1-integrable for almost every $x \in \Omega_2$,*

and we have

$$\int_{\Omega_1\times\Omega_2} f \, d(\mu_1 \otimes \mu_2) = \int_{\Omega_1} \left(\int_{\Omega_2} f \, d\mu_2 \right) d\mu_1 = \int_{\Omega_2} \left(\int_{\Omega_1} f \, d\mu_1 \right) d\mu_2.$$

Lemma 10.77. *Let $A \in \mathcal{L}(\mathbb{R}^2)$ be a (Lebesgue-)measurable set. Then, the sections $A^x, A_y \subset \mathbb{R}$ are measurable, i.e., $A^x, A_y \in \mathcal{L}(\mathbb{R})$, for almost every $x, y \in \mathbb{R}$.*

Moreover, we have that $N \subset \mathbb{R}^2$ is a null set if and only if N^x is a null set for a.e. $x \in \mathbb{R}$. (Similar for y.)

Theorem 10.79 (Fubini for the Lebesgue measure). *Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be Lebesgue-measurable such that either*

- $\int_{\mathbb{R}} \int_{\mathbb{R}} |f(x,y)|\, dy\, dx < \infty$ *or*
- $\int_{\mathbb{R}} \int_{\mathbb{R}} |f(x,y)|\, dx\, dy < \infty$,

then, f is integrable and we have

$$\int_{\mathbb{R}^2} f\, d\lambda_2 = \int_{\mathbb{R}^2} f(x,y)\, d(x,y) = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x,y)\, dy\, dx = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x,y)\, dx\, dy.$$

Moreover, we have that f^x and f_y are integrable for almost every x and y.

In addition, for every measurable set $A \subset \mathbb{R}^2$, we have

$$\int_A f\, d\lambda_2 = \int_{\mathbb{R}} \int_{A^x} f(x,y)\, dy\, dx = \int_{\mathbb{R}} \int_{A_y} f(x,y)\, dx\, dy$$

if one of the iterated integrals exist.

Theorem 10.83 (Fubini). *Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be Lebesgue-integrable. Then, we have*

$$\int_{\mathbb{R}^d} f\, d\lambda_d = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \left(\cdots \left(\int_{\mathbb{R}} f(x_1, \ldots, x_d)\, dx_d \right) \cdots \right) dx_2 \right) dx_1$$

with an arbitrary order of integration on the right.
Moreover, f is integrable if one of the integrals over $|f|$ is finite.

9.9.5 Connection to probability theory

Definition 10.88. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space such that

$$\mu(\Omega) = 1.$$

Then, μ is called a **probability measure**, and $(\Omega, \mathcal{A}, \mu)$ is called a **probability space**.

A set $A \in \mathcal{A}$ is usually called an **event**.

Definition 10.94. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X a random variable.
We define the **expected value** (or **expectation** or **mean**) of X as

$$\mathbb{E}X := \mathbb{E}(X) := \int_\Omega X\, d\mathbb{P} = \int_\Omega X(\omega)\, d\mathbb{P}(\omega).$$

The **variance of** X is defined as

$$\mathrm{Var}(X) := \mathbb{E}\left(|X - \mathbb{E}(X)|^2 \right).$$

By writing $\mathbb{E}X$ and $\mathrm{Var}(X)$ we generally assume that the expectations (aka. integrals) exist.

Lemma 10.97 (Chebyshev's inequality). *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X a random variable such that expectation $\mathbb{E}X$ and variance $\operatorname{Var}(X)$ exist. Then,*

$$\mathbb{P}\Big(|X - \mathbb{E}X| > a\Big) \leq \frac{\operatorname{Var}(X)}{a^2} \qquad \text{for all } a > 0.$$

Lemma 10.99. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $X_1, \ldots, X_n$ be random variables. Then,*

$$\mathbb{E}(\overline{X}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(X_i)$$

and if we have additionally that $\mathbb{E}(X_i \cdot X_j) = \mathbb{E}(X_i)\,\mathbb{E}(X_j)$ for all $i \neq j$, then

$$\operatorname{Var}(\overline{X}) = \frac{1}{n^2} \sum_{i=1}^{n} \operatorname{Var}(X_i).$$

Corollary 10.100. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $X_1, \ldots, X_n$ be uncorrelated random variables with $\mathbb{E}X_i =: Z \in \mathbb{R}$ and $\operatorname{Var}(X_i) =: \sigma^2$ for all $i = 1, \ldots, n$. Then,*

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^{n} X_i - Z\right| > \frac{\sigma}{\sqrt{\delta n}}\right) \leq \delta \qquad \text{for all } \delta > 0.$$

9.10 Basic functions analysis

9.10.1 Vector spaces

Definition 11.1. Let $\mathbb{F}$ be a field and V be a non-empty set.
We call V a **vector space over** $\mathbb{F}$ (or **linear space over** $\mathbb{F}$ or $\mathbb{F}$**-vector space**), if the following properties hold:
There exists a **vector addition** $+\colon V \times V \to V$ such that

1) Associativity: For all $x, y, z \in V$ there holds $x + (y + z) = (x + y) + z$
2) Commutativity: For any vectors $x, y \in V$ we have $x + y = y + x$
3) Neutral element: there exists an element $0 \in V$ such that $x + 0 = x$ for all $x \in V$
4) Inverse element: For any $x \in V$ there exists an element $-x \in V$ such that $x - x = 0$.

Moreover, there is a **scalar multiplication** $\cdot\colon \mathbb{F} \times V \to V$ such that for all $x, y \in V$ and $\mu, \lambda \in \mathbb{F}$ the following properties are valid:

5) Distributivity w.r.t. vector addition: $\lambda \cdot (x + y) = \lambda \cdot x + \lambda \cdot y$
6) Distributivity w.r.t. field addition: $(\lambda + \mu) \cdot x = \lambda \cdot x + \mu \cdot x$
7) Associativity: $(\lambda \cdot \mu) \cdot x = \lambda \cdot (\mu \cdot x)$
8) Neutral element: $1 \cdot x = x$

The elements $x \in V$ are called **vectors** and the elements of $\mathbb{F}$ are called **scalars**.

In the case $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, we call V a **real or complex vector space**, respectively.

Definition 11.14 (Subspaces). Let V be a $\mathbb{F}$-vector space and suppose that $U \subset V$.

If U is a $\mathbb{F}$-vector space, then we call U a **(linear) subspace** of V.

If additionally $U \subsetneq V$, then we say it is a **proper subspace**.

Definition 11.19. A set of vectors $\{v_1, v_2, \ldots, v_n\} \subset V$, where V is a $\mathbb{F}$-vector space is called **linearly dependent** if there exist $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$, not all zero, such that

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = 0.$$

Otherwise, i.e., if

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = 0 \qquad \Longleftrightarrow \qquad \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0,$$

then we call the set $\{v_1, v_2, \ldots, v_n\}$ **linearly independent**.

A vector of the form

$$y = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n,$$

for some $\alpha_k \in \mathbb{F}$ is called **linear combination** (in V) of the vectors $v_1, \ldots v_n$.

Lemma 11.20. *Let V be a vector space and $\{v_1, \ldots, v_n\} \subset V$ be linearly dependent if and only if at least one of the v_k's is a linear combination of the other vectors.*

Definition 11.26 (Basis). Let V be a vector space over some field $\mathbb{F}$ and $n \in \mathbb{N}$.

For a set $G := \{b_1, \ldots, b_n\} \subset V$ we define the **(linear) span of** G (in V) by

$$\operatorname{span}(G) := \left\{ v \in V : v = \sum_{k=1}^{n} c_k b_k \quad \text{for some } \ c_1, \ldots, c_n \in \mathbb{F} \right\},$$

i.e., the set of all linear combinations of elements of G.

We call a finite set of vectors $G := \{b_1, \ldots, b_n\} \subset V$ a **generating set of** V if each element in V is a linear combination of elements in G, i.e., $V = \operatorname{span}(G)$.

A set $B \subset V$ is called a **basis of** V if B is a generating system of V and all elements in B are linearly independent.

Theorem 11.30. *Let V be a vector space and $B = \{b_1, b_2, \ldots, b_n\}$ be any (finite) basis of V.*
Then,

1) *for every $v \in V$ there are unique $c_1, \ldots, c_n \in F$ such that $v = \sum_{i=1}^{n} c_i b_i$, these c_i are called the **coordinates of** v **w.r.t. the basis** B.*

2) *If $B' \subsetneq B$ then B' is no longer a generating system of V.*

3) *For any $x \in V \setminus B$ we have that $B \cup \{x\}$ is linearly dependent.*

Lemma 11.31. *Let V be a vector space which is generated by a finite set $\{b_1, b_2, \ldots, b_m\}$. Then, $\{b_1, b_2, \ldots, b_m\}$ contains a basis. In particular, every such V has a basis.*

Lemma 11.32 (Steinitz). *Let V be a vector space which is generated by $\{b_1, b_2, \ldots, b_n\} \subset V$, i.e., $V = \operatorname{span}\{b_1, \ldots, b_n\}$. ($n = \infty$ is allowed.)*

If $\{a_1, a_2, \ldots, a_k\} \subset V$ are linearly independent, then $k \leq n$.

Moreover, there is a renumbering of the b_i's, such that the set $\{a_1, \ldots, a_k, b_{k+1}, \ldots, b_n\}$ is a generating system of V.

Corollary 11.33. *Let V be a vector space which is generated by a finite set.*

If $\{a_1, a_2, \ldots, a_n\}$ and $\{b_1, b_2, \ldots, b_m\}$ are bases of V, then $m = n$.

This implies that a set $\{x_1, x_2, \ldots, x_n\} \subset V$, where n is the length of one basis, forms a basis, if one of the following is fulfilled:

1) *$\{x_1, x_2, \ldots, x_n\}$ is linear independent, or*

2) *$\{x_1, x_2, \ldots, x_n\}$ is a generating set.*

Definition 11.34. Let V be a vector space with basis $\{b_1, b_2, \ldots, b_d\}$.

Then, we define the **dimension of** V by $\dim(V) := d = \#\{b_1, b_2, \ldots, b_d\}$.

If $\dim(V) < \infty$, then we call V a **finite-dimensional** vector/function space.

In the case that there are infinitely many linearly independent vectors, we write $\dim(V) = \infty$, and call V an **infinite-dimensional** vector/function space.

9.10.2 Normed spaces

Definition 11.39. Let X be a $\mathbb{F}$-vector space and $\|\cdot\| : X \to \mathbb{R}$ be a real-valued mapping. We say that $\|\cdot\|$ is a **norm (on X)** if the following holds:

1) For any $x \in X$ with $x \neq 0$ we have $\|x\| > 0$. (**positive definiteness**)

2) For $x \in X, \lambda \in \mathbb{F}$ it holds $\|\lambda x\| = |\lambda| \|x\|$. (**homogeneity**)

3) If $x, y \in X$ then we have $\|x + y\| \leq \|x\| + \|y\|$. (**triangle inequality**)

The tuple $(X, \|\cdot\|)$ is then called a **normed space**.

9.10.3 The Lp-spaces

Definition 11.52. Let $(\Omega, \mathcal{A}, \mu)$ be measure space and let $0 < p < \infty$. Then, we define

$$\mathscr{L}_p(\Omega, \mu) := \left\{ f : \Omega \to \mathbb{C} : \ f \text{ is measurable and } \int_\Omega |f|^p \, d\mu < \infty \right\}.$$

Lemma 11.57. *Let $(\Omega, \mathcal{A}, \mu)$ be measure space and let $0 < p < \infty$. Then, $\mathscr{L}_p(\Omega, \mu)$ together with the point-wise addition and the usual scalar multiplication is a vector space.*

Lemma 11.58. *Let $a, b \geq 0$ and $r \in (0, 1)$. Then*

$$a^r b^{1-r} \leq r\, a + (1 - r) b.$$

Theorem 11.59 (Hölder's inequality). *Let $1 < p, q < \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$. If $f \in \mathscr{L}_p(\Omega, \mu)$ and $g \in \mathscr{L}_q(\Omega, \mu)$, then $fg \in \mathscr{L}_1(\Omega, \mu)$ and*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q .$$

Corollary 11.60. *Let $1 < p, q < \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then, for all sequences $(x_n)_{n\in\mathbb{N}}, (x_n)_{n\in\mathbb{N}} \in \mathbb{C}^{\mathbb{N}}$, we have*

$$\sum_{k=1}^{\infty} |x_k y_k| \leq \left(\sum_{k=1}^{\infty} |x_k|^p\right)^{\frac{1}{p}} \left(\sum_{k=1}^{\infty} |y_k|^q\right)^{\frac{1}{q}}$$

Theorem 11.61 (Minkowski's inequality). *Let $f, g \in \mathscr{L}_p(\Omega, \mu)$ with $p \in [1, \infty)$. Then,*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p .$$

Proposition 11.62. *Let I be a countable set and $p \in [1, \infty)$. Then,*

$$\|a\|_p := \left(\sum_{k\in I} |a_k|^p\right)^{1/p}$$

defines a norm on $\ell_p(I) := \{a = (a_i)_{i\in I} \colon \|a\|_p < \infty\}$.

Definition 11.63. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and $0 < p < \infty$. Moreover, for each $f \in \mathscr{L}_p(\Omega, \mu)$, we define the set

$$[f] := \Big\{g \in \mathscr{L}_p(\Omega, \mu) \colon f = g \text{ almost everywhere}\Big\},$$

which is called the **equivalence class of** f.

Then, we define the L_p**-spaces**

$$L_p(\Omega, \mu) := \Big\{[f] \colon f \in \mathscr{L}_p(\Omega, \mu)\Big\}.$$

We write $L_p(\Omega)$ for $L_p(\Omega, \mu)$ if $\Omega \subset \mathbb{R}^d$ and μ is the Lebesgue measure.

Theorem 11.67. *For any $1 \leq p < \infty$ we have that $L_p(\Omega, \mu)$ together with*

$$\|f\|_p := \left(\int_{\Omega} |f|^p d\mu\right)^{\frac{1}{p}}$$

is a normed space.

Definition 11.69. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. Then, we define

$$\mathscr{L}^\infty(\Omega, \mu) := \{f: \Omega \to \mathbb{R}: f \text{ is measurable and } \|f\|_\infty < \infty\}.$$

If we consider again the equivalence classes $[f] = \{g \in \mathscr{L}_\infty(\Omega, \mu): f = g \text{ a.e.}\}$, then we define

$$L_\infty(\Omega, \mu) = \{[f]: f \in \mathscr{L}^\infty(\Omega, \mu)\}.$$

We write $L_\infty(\Omega)$ for $L_p(\Omega, \mu)$ if $\Omega \subset \mathbb{R}^d$ and μ is the Lebesgue measure, and $\ell_\infty(I)$ if I is a countable set and μ is the counting measure.

Theorem 11.72. *$L_\infty(\Omega, \mu)$ together with $\|\cdot\|_\infty$ is a normed space.*

Theorem 11.73 (Hoelder's inequality continued)**.** *Let $1 \le p, q \le \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$. If $f \in L_p(\Omega, \mu)$ and $g \in L_q(\Omega, \mu)$, then $fg \in L_1(\Omega, \mu)$ and*

$$\|fg\|_1 \le \|f\|_p \|g\|_q.$$

For $p = \infty$ we have $q = 1$ and therefore $\|fg\|_1 \le \|f\|_\infty \|g\|_1$.

9.10.4 Sequences in normed spaces and Banach spaces

Definition 11.74 (Convergence in normed spaces.)**.** Let $(F, \|\cdot\|)$ be a normed space. A sequence $(f_n)_{n\in\mathbb{N}}$ is called **convergent in** $(F, \|\cdot\|)$ if there exists some $f \in F$ such that

$$\lim_{n\to\infty} \|f - f_n\| = 0.$$

If this is the case we write $f = \lim_{n\to\infty} f_n$ or $f_n \to f$ in $(F, \|\cdot\|)$.

If the norm is fixed, we just say $(f_n)_{n\in\mathbb{N}}$ is **convergent in** F and write $f_n \to f$ in F.

Definition 11.80 (Cauchy sequences)**.** Let $(F, \|\cdot\|)$ be a normed space and let $(f_n)_{n\in\mathbb{N}} \subset F$. If for any $\varepsilon > 0$ there exists some $n_0 \in \mathbb{N}$ such that for any $n, m \ge n_0$ we have that

$$\|f_n - f_m\| < \varepsilon,$$

then we say $(f_n)_{n\in\mathbb{N}}$ is a **Cauchy sequence (w.r.t.** $\|\cdot\|$**).**

If all Cauchy sequences of a normed space $(F, \|\cdot\|)$ are convergent sequences, then we call $(F, \|\cdot\|)$ a **complete normed space** or a **Banach space**.

Lemma 11.82. *Let $(F, \|\cdot\|)$ be a normed space. If $(f_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, which has a convergent subsequence, then $(f_n)_{n\in\mathbb{N}}$ is a convergent sequence.*

Theorem 11.85. *Let $I = [a, b]$ be a closed interval. Then, $(C(I), \|\cdot\|_\infty)$ is a Banach space.*

Theorem 11.86. *For any $1 \leq p \leq \infty$, the space $L_p(\Omega, \mu)$ is a Banach space.*

That is, any Cauchy sequence in $(L_p(\Omega, \mu), \|\cdot\|_p)$ is convergent in this space.

Corollary 11.87. *The spaces $(\ell_p, \|\cdot\|_p)$ are Banach spaces for $1 \leq p \leq \infty$.*

9.10.5 Metric spaces

Definition 11.90. Let X be a set and $d\colon X \times X \to \mathbb{R}$ with the properties

1) For any $x, y \in X$ it holds $d(x, y) \geq 0$ and $d(x, y) = 0 \iff x = y$.
2) $d(x, y) = d(y, x)$ holds for all $x, y \in X$.
3) If $x, y, z \in X$ then we have $d(x, z) \leq d(x, y) + d(y, z)$.

The properties $1), 2)$ and $3)$ are called **positive definiteness, symmetry** and **triangle inequality**, and the tuple (X, d) is called **metric space**.

9.10.6 Inner products and Hilbert spaces

Definition 11.92. Let X be a $\mathbb{R}$-vector space. A mapping $\langle\cdot,\cdot\rangle\colon X \times X \to \mathbb{R}$ which is

1) **linear in both arguments**, i.e., $\langle \lambda x + \mu y, z\rangle = \lambda\langle x, z\rangle + \mu\langle y, z\rangle$ and $\langle x, \lambda y + \mu z\rangle = \lambda\langle x, y\rangle + \mu\langle x, z\rangle$,
2) **symmetric**, i.e., $\langle x, y\rangle = \langle y, x\rangle$, and
3) **positive definite**, i.e., $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0 \iff x = 0$,

where $x, y, z \in X$ and $\lambda, \mu \in \mathbb{R}$ are arbitrary, is called **(real) inner product**.

The tuple $(X, \langle\cdot,\cdot\rangle)$ is called **inner product space**.

Theorem 11.98 (Cauchy-Schwarz). *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space. Then, for all $x, y \in X$ it holds*

$$|\langle x, y\rangle|^2 \;\leq\; \langle x, x\rangle \cdot \langle y, y\rangle.$$

Theorem 11.99. *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space.*

Then, the mapping $\|\cdot\|\colon X \to \mathbb{R}$ with $\|x\| := \sqrt{\langle x, x\rangle}$ is a norm, called the ***induced norm****.*

In other words, every inner product space is a normed space.

Corollary 11.100. *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space. Then for all $x, y \in X$ we have*

$$\langle x, y\rangle \leq \|x\| \cdot \|y\|,$$

where $\|\cdot\| = \sqrt{\langle x, x\rangle}$ is the induced norm.

Definition 11.101 (Hilbert spaces). Let $(H, \langle \cdot, \cdot \rangle)$ be an inner product space.

If $(H, \|\cdot\|)$ is a Banach space, where $\|f\| := \sqrt{\langle f, f \rangle}$ is the induced norm, then we call $(H, \|\cdot\|)$ (or just H) a **Hilbert space**.

Definition 11.106. Let $(H, \langle \cdot, \cdot \rangle)$ be an inner product space.

We say that $x, y \in H$ are **orthogonal** if

$$\langle x, y \rangle = 0.$$

If x and y additionally satisfy $\langle x, x \rangle = \langle y, y \rangle = 1$, then we call them **orthonormal**.

Moreover, we say **a set** $A \subset H$ **is orthogonal/orthonormal**, if arbitrary $x, y \in H$ with $x \neq y$ are orthogonal/orthonormal.

For a set $A \subset H$ we define the **orthogonal complement** of A by

$$A^{\perp} := \Big\{ x \in H \colon \langle x, y \rangle = 0 \ \text{ for all } \ y \in A \Big\}.$$

Definition 11.109. Let H be a Hilbert space.

A countable orthonormal set $\{e_1, e_2, e_3, \dots\}$ is called **orthonormal basis (ONB) of H** if each element in $x \in H$ can be written as

$$x = \sum_{k=1}^{\infty} \alpha_k e_k$$

for some $\alpha_k \in \mathbb{R}$ (or $\mathbb{C}$). This means that, for some sequence $(\alpha_k)_{k \in \mathbb{N}}$, we have

$$\lim_{n \to \infty} \left\| \sum_{k=1}^{n} \alpha_k e_k - x \right\| = 0.$$

Theorem 11.110. *Let H be a Hilbert space and $\{e_1, e_2, e_3, \dots\}$ be an orthonormal set.*

Then, the following statements are equivalent:

1) *If for $x \in H$ we have that $\langle x, e_k \rangle = 0$ for all $k \in \mathbb{N}$, then $x = 0$.*

2) *The set $\{e_1, e_2, e_3, \dots\}$ is an orthonormal basis of H.*

3) *For any $x \in H$ we have* $x = \sum_{k=1}^{\infty} \langle x, e_k \rangle e_k$.

4) *For any $x, y \in H$ we have* $\langle x, y \rangle = \sum_{k=1}^{\infty} \langle x, e_k \rangle \langle e_k, y \rangle$.

5) *For any $x \in H$ we have* $\|x\|^2 = \sum_{k=1}^{\infty} |\langle x, e_k \rangle|^2$

9.10.7 Reproducing kernel Hilbert spaces

Definition 11.115. Let H be a Hilbert space of functions on some set Ω, i.e., any $f \in H$ has the form $f\colon \Omega \to \mathbb{C}$.

If there exists some function $k\colon \Omega \times \Omega \to \mathbb{C}$ with the properties

1) for any $x \in \Omega$ we have that $k(x, \cdot) \in H$, and

2) for any $x \in \Omega$ and any $f \in H$ we have that $f(x) = \langle f, k(x, \cdot)\rangle$,

then we call k a **reproducing kernel** and the space H is called a **reproducing kernel Hilbert space (RKHS)** on Ω.

Lemma 11.116. *Let H be a RKHS on Ω with kernel k.*

If a sequence $(f_n)_{n\in\mathbb{N}}$ converges in H to some f, that is

$$\lim_{n\to\infty} \|f_n - f\| = 0,$$

then $f_n(x) \to f(x)$ as $n \to \infty$ for every $x \in X$.

Theorem 11.117. *Let H be a RKHS on Ω with kernel k.*

If $\{e_1, e_2, \dots\}$ is an orthonormal basis in H, then we have that

$$k(x, y) = \sum_{i=1}^{\infty} e_i(x) e_i(y),$$

where the series converges point-wise for every $x, y \in \Omega$.

Lemma 11.119. *Let H be a RKHS on Ω with kernel k. Then for any $x, y \in \Omega$*

1) $k(x, y) = \langle k(x, \cdot), k(y, \cdot)\rangle$

2) k is conjugate symmetric, i.e., $k(x, y) = \overline{k(y, x)}$

3) k is unique, i.e., there is not other reproducing kernel in H.

Definition 11.120. Let Ω be a non-empty set.

We say that a symmetric function $k\colon \Omega \times \Omega \to \mathbb{R}$ is **positive definite** if the matrix $K := (k(x_i, x_j))_{i,j=1}^n \in \mathbb{R}^{n\times n}$ is positive semi-definite for any $n \in \mathbb{N}$ and any choice $x_1, \dots, x_n \in \Omega$. That is, if we have for any $c_1, \dots c_n \in \mathbb{R}$ that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j k(x_i, x_j) \geq 0.$$

Lemma 11.122. *Let H be a RKHS on Ω with kernel k. Then, k is positive definite.*

Theorem 11.123 (Moore-Aronszajn). *Let Ω be a non-empty set and let $k\colon \Omega \times \Omega \to \mathbb{R}$ be a positive definite function on Ω.*

Then, there exists a unique Hilbert space H such that k is its reproducing kernel.