Mathematics

for Biomedical Engineering Science

Benedikt Annweiler

Lehrstuhl II für Mathematik

The lecture and its lecture notes are mainly based upon
the 2014/15 lecture of this course by Jan Simon M.Sc., RWTH Aachen
and on
the courses Higher Mathematics I-III by Prof. Dr. Eberhard Triesch, RWTH Aachen

Contents

1 Basic concepts in mathematics

1.1 Set theory and number systems

Sets are some of the basic concepts in mathematics. They are the collections of elements. Unlike other structures it is inevitable that the elements are distinguishable.

Definition 1.1. (set)
A *set* is a collection of distinguishable elements.

Sets are uniquely determined by their elements. In particular, two sets are equal if and only if they contain the same elements. It is of no importance how often one element is contained in the set. It is only mandatory that that element ist in the set. Furthermore the order of the elements is unimportant.

Example 1.2.

$$\{1,2,3\} = \{1,1,2,3,3\} = \{3,1,2\}$$

From now on we only place each number only once in a set.

There are two ways to describe sets. One may describe a set by *explicitly* stating the elements like

$$A := \{1,3,4,5,6\}$$

or by *implicitly* describing the properties which the elements need to satisfy like

$$B := \{n \mid n \text{ is an odd number on a dice}\}.$$

In this case the set B may als be stated as $\{1,3,5\}$.

But not all sets need to contain elements.

Definition 1.3. (empty set)
The set without any elements is called *empty set* and will be denoted by $\{\}$ or $\emptyset$.

The empty set is an element itself. Therefore the set with the empty set exists. So $\{\emptyset\}$ is not empty but contains exactly one element.
We want to define relationships between sets.

Definition 1.4. (subset)
Let A and B be sets. If all elements of A are also elements of B, then A is a *subset* of B and will be denoted by $A \subseteq B$

We can now describe the equality of sets with the notation of subsets.

Lemma 1.5.
Let A and B be two sets. A equals B e.g. $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$ holds.

To define relationships between elements and sets we introduce the following important notations along with basic logic statements.

Definition 1.6.
Let a be an element and A be a set. Then the following notations describe:

i) "$a \in A$" means *a is an element of the set A*,

ii) "$a \notin A$" means *a is not an element of the set A*,

iii) "$\exists$" means *there is* or *there exists*,

iv) "$\nexists$" means *there exists no ...*,

v) "$\forall$" means *for all*,

vi) "$\Rightarrow$" means *implies*,

vii) "$\Leftrightarrow$" means *is equivalent to* or *if and only if*.

Example 1.7.
With the newly defined notations we can rewrite the subset symbol as follows:

$$(A \subseteq B) \;\Leftrightarrow\; (\forall a \in A \;:\; a \in B)$$

Sets may not only be included in others. We can also intersect and unite sets as well as take one set out of another.

Definition 1.8. (union, intersection and difference of sets)
Let A and B be sets.

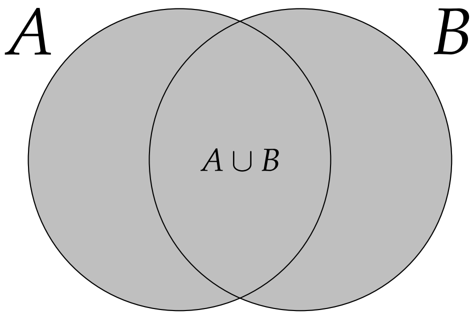
i) Then the *union of A and B* is denoted by $A \cup B := \{x \mid x \in A \text{ or } x \in B\}$.

ii) Then the *intersection of A and B* is denoted by $A \cap B := \{x \mid x \in A \text{ and } x \in B\}$. If $A \cap B = \emptyset$ holds, then the two sets A and B are called *disjoint* .

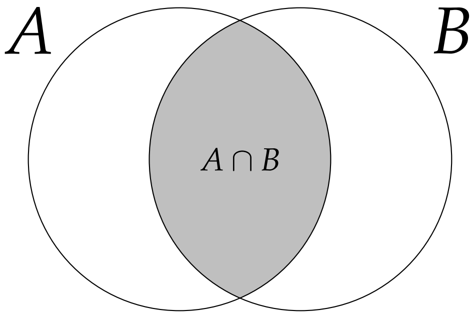
iii) Then the *difference of A and B* is denoted by $A \setminus B := \{x \mid x \in A \text{ and } x \notin B\}$.

Example 1.9.
The definition above can be illustrated by the following pictures:

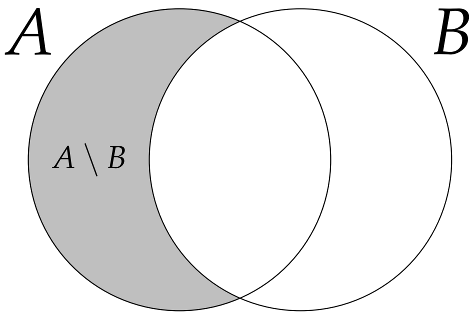
i) The union of sets A and B.



ii) The intersection of the sets A and B.



iii) The difference between the sets A and B.



Up to know we only looked at sets that consisted of distinct elements like single numbers. We can expand our view by taking in ordered pairs and the more general n-tuples.

Definition 1.10. (ordered pairs and n-tuples)
Two *ordered pairs* (a, b) and (c, d) are equal if and only if $a = c$ and $b = d$ holds. If an ordered pair consists of more than two numbers, we call that object $(a_1, a_2, ..., a_n)$ a *n-tuple*. Equality of n-tuples is defined analogously.

There are some operations defined on ordered pairs and n-tuples. We have a closer look on the operation called the cartesian product.

Definition 1.11. (cartesian product)
Let A and B be two sets. Then the *cartesian product* is defined as $A \times B := \{(a,b) \mid a \in A,\ b \in B\}$.
The cartesian product is generalized for sets $A_1, A_2, ..., A_n$ as $A_1 \times A_2 \times ... \times A_n := \{(a_1, a_2, ..., a_n) \mid a_1 \in A_1,\ a_2 \in A_2, ...,\ a_n \in A_n\}$.

We want to illustrate the concepts that we just learned with some numbers.

Example 1.12.
Let $A = \{1,3,5,7,9\}$ and $B = \{1,2,3,4\}$ be two sets.

$$\begin{aligned}
\Rightarrow A \cup B &= \{1,2,3,4,5,7,9\}, \\
A \cap B &= \{1,3\}, \\
A \setminus B &= \{5,7,9\}, \\
B \setminus A &= \{2,4\}, \\
A \times B &= \{(1,1),(1,2),(1,3),(1,4), \\
&\quad (3,1),(3,2),(3,3),(3,4), \\
&\quad ..., \\
&\quad (9,1),(9,2),(9,3),(9,4)\}.
\end{aligned}$$

The well known basic number systems are also an example of sets. We have the set of natural numbers $\mathbb{N} = \{1,2,3,...\}$, the set of integers $\mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$, the set of rationals $\mathbb{Q} = \{\frac{a}{b} \mid a \in \mathbb{Z},\ b \in \mathbb{N}\}$ and the set of all real numbers, that include both rational and irrational numbers. As we know by definiton $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q}$ holds. Furthermore $\mathbb{Q} \subseteq \mathbb{R}$ holds, since $\mathbb{R}$ contains $\mathbb{Q}$ and all irrational numbers. We want to proof that $\mathbb{R} \neq \mathbb{Q}$.

Theorem 1.13.
$\sqrt{2}$ is an irrational number, that is there is no rational number solving the equation $x^2 = 2$.

Proof:
The proof will be indirect. That means we make an assumption, i.e. the negation of the statement and show that this can't be true. Assume there is a fraction $\frac{p}{q}$ with $(\frac{p}{q})^2 = 2$ $(p, q \in \mathbb{N})$. We may assume that the fraction is reduced, that is the greatest common denominator is 1. By $p^2 = 2q^2$ follows that p is even. Denote $p = 2p'$ $(p' \in \mathbb{N})$. Hence

$$\begin{aligned}
p^2 = (2p')^2 = 4p'^2 &= 2q^2 \\
\Rightarrow 2p'^2 &= q^2.
\end{aligned}$$

Therefore q is also even. But then the greatest common denominator is at least 2, which yields an contradiction. $\square$

The real numbers have an important property, that the rational numbers are missing. The existing of an infimum and a supremum of any bounded set.

Definition 1.14. (supremum/infimum, maximum/minimum)
Let $A \subseteq \mathbb{R}$ be a subset. If A has an upper bound (a lower bound), i.e. there is some $M \in \mathbb{R}$ with $a \leq M$ ($a \geq M$) $\forall a \in A$, then A also has a unique smallest upper bound (greatest lower bound), called the *supremum* (*infimum*) of A which is denoted by $\sup(A)$ ($\inf(A)$). If $\sup(A) \in A$ ($\inf(A) \in A$) then $\max(A) := \sup(A)$ ($\min(A) := \inf(A)$) is called the *maximum* (*minimum*) of A.

We can also rewrite the definition for a supremum (infimum):

$$\sup(A) : a \leq \sup(A)\ \forall a \in A \text{ and } \sup(A) - \varepsilon \text{ is no upper bound } \forall \varepsilon > 0.$$
$$\inf(A) : a \geq \inf(A)\ \forall a \in A \text{ and } \inf(A) + \varepsilon \text{ is no lower bound } \forall \varepsilon > 0.$$

In the last definition it was stated, that all upper bounded (lower bounded) subsets of $\mathbb{R}$ have a supremum (infimum).

Example 1.15.
i) $\sup\{1,2,3\} = \max\{1,2,3\} = 3$,

ii) $\inf\{\frac{1}{n} \mid n \in \mathbb{N}\} = 0$,

iii) $\sup\{x \in \mathbb{R} \mid x^2 < 2\} = \sqrt{2}$ and $\sqrt{2} \notin \{x \in \mathbb{R} \mid x^2 < 2\}$.

Note that the boundaries of examples *i*) and *ii*) are also rational numbers, yet the supremum of *iii*) is irrational. Therefore the supremum won't exists in $\mathbb{Q}$, that is $\sup\{x \in \mathbb{Q} \mid x^2 < 2\}$ does not exist.

1.2 Functions

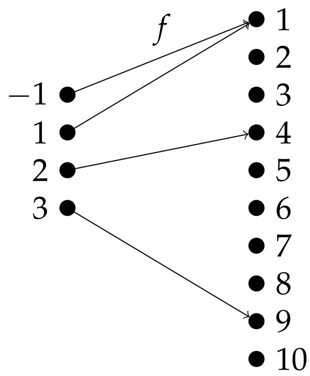
Functions express relations between two sets. They are used in all branches of mathematics.

Definition 1.16. (function)
Let A, B be non-empty sets. A *function* assigns to every element $a \in A$ precisely one element $f(a) \in B$. We write

$$f: A \to B: a \mapsto f(a).$$

A is called the *domain* of f, B is called the *co-domain* of f, $f(a)$ is called the *image of a under f* and $f(A) := \{f(a) \mid a \in A\}$ is called the *image of f*.

Example 1.17.
$A := \{-1, 1, 2, 3\}$, $B := \{1, 2, 3, ..., 10\}$, $f: A \to B: a \mapsto a^2$



Definition 1.18. (injective, surjective, bijective)
Let A, B be non-empty sets and $f : A \to B : a \mapsto f(a)$ a function. Then we define the following:

i) The function f is called *injective*, if for all $a_1, a_2 \in A$ we have if $f(a_1) = f(a_2)$ holds, then $a_1 = a_2$ follows.

ii) The function f is called *surjective*, if $f(A) = B$, that is the co-domain equals the image of f.

iii) The function f which is both injective and surjective is called *bijective*.

Injective means in simple words, that there are no two elements of the domain, that maps to the same element of the co-domain. Surjective may be described as every element of the co-domain is hit by an element of the domain.

By this definition the last example is neither injective nor surjective. One important funtion, the identity, is the most trivial example of a bijective function.

Definition 1.19. (identity function)
The function

$$id_A : A \to A : a \mapsto a$$

is called the *identity on A*.

Definition 1.20. (composition of functions)
Let A, B and C be non-empty sets and $f : A \to B$ and $g : B \to C$ be two functions. Then $g \circ f : A \to C : a \mapsto g(f(a))$ is called the *composition of f and g*.

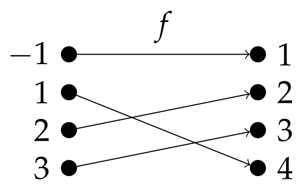
$$\begin{tikzcd} A \arrow[r, "f"] \arrow[rr, bend right, "g \circ f"'] & B \arrow[r, "g"] & C \end{tikzcd}$$

Remark 1.21.
Usually the composition of g and f and the composition of f and g yield different result, that is $g \circ f \neq f \circ g$.

Definition 1.22. (inverse function)
Let A, B be non-empty sets and $f : A \to B$ be a bijective function. Then there is a unique function $f^{-1} : B \to A$, called the *inverse of* f, with the property

$$f(a) = b \Leftrightarrow f^{-1}(b) = a, \ \forall a \in A, b \in B.$$

This means actually that the composition of f and f^{-1} and the composition of $f^{-}1$ and f both yield the identity function. That is $f \circ f^{-1} = id_A$ and $f^{-1} \circ f = id_B$.



Lemma 1.23.
Let A, B and C be non-empty sets and $f : A \to B$ and $g : B \to C$ be two bijective functions. Then the composition of f and g, that is $g \circ f : A \to C$, has the inverse $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Our functions are usually algebraic expressions like $x \mapsto \frac{x}{x^2+1}$. The non-empty set that we do operate on with our functions will be usually some subsets of $\mathbb{R}$. We call the following specific subsets of $\mathbb{R}$ intervals.

Definition 1.24. (interval)
An *interval* $I \subseteq \mathbb{R}$ is a subset of the following form:

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

A one sided unbounded interval is denoted by $(-\infty, b]$ and $[a, \infty)$.

Remark 1.25.
The interval of the real numbers may be denoted as $(-\infty, \infty) = \mathbb{R}$. Analogously the interval of the non-negative real numbers is $[0, \infty) = \mathbb{R}_+$.

Next we define operations for functions if the co-domain is $\mathbb{R}$.

Definition 1.26.
Let A be a non-empty set, $\lambda \in \mathbb{R}$ and $f, g : A \to \mathbb{R}$ functions. Then we define the following operations for all $x \in A$:

- $(f \pm g)(x) := f(x) \pm g(x)$,
- $(\lambda \cdot f)(x) := \lambda \cdot f(x)$,
- $(f \cdot g)(x) := f(x) \cdot g(x)$,
- $(\frac{f}{g})(x) := \frac{f(x)}{g(x)}$ if $0 \notin g(A)$.

Lastly we want to define the monotonic functions on the ordered set $\mathbb{R}$. $\mathbb{R}$ is ordered by the equivalent relation "$\leq$".

Definition 1.27. (monotonically increasing/decreasing)
Let $f : A \to \mathbb{R}$ be a function with $A \subseteq \mathbb{R}$. Furthermore let $I \subseteq A$ be an interval. f is called *monotonically increasing (decreasing) on* I, if

$$x \leq y \Rightarrow f(x) \leq f(y)$$
$$(x \leq y \Rightarrow f(x) \geq f(y))$$

holds for all $x, y \in I$. We call a function *stricktly monotonically increasing (decreasing)* if all "$\leq$","$\geq$" are replaced by "$<$" and "$>$" respectively.

Lemma 1.28.
Let $f : I \to \mathbb{R}$ be a function on the Interval $I \subseteq \mathbb{R}$. If f is either stricktly increasing or decreasing on I then f is injective.

1.3 Induction

After the indirect proof, used in $\sqrt{2}$ is not a rational number earlier, the concept of proof by induction is another way of proving mathematical theorems. A mathematical proof of a statement is in generala series of implications that leads to the inevitable conclusion, that this statement is true. In mathematics we can use proofs to establish facts for an infinite number of cases. The proof by induction is a method to show and infinite number of cases of a statement $S = S(n)$ where each case is dependant on a natural number $n \in \mathbb{N}$. Induction proves all those cases in one stroke for all $n \in \mathbb{N}$.

Theorem 1.29. (Induction Principle)
Let $S(n)$ be any statement depending on $n \in \mathbb{N}$. Furthermore the following two conditions hold:

i) $S(1)$ is true (base case),

ii) $\forall n \in \mathbb{N}$ holds: $S(n)$ is true $\Rightarrow S(n+1)$ is true. (inductive step)

Then $S(n)$ is true for all $n \in \mathbb{N}$.

Example 1.30.
We prove the **Gauß' Sum Formula**:

$$\sum_{i=1}^{n} i = 1 + 2 + ... + n = \frac{n(n+1)}{2}, \ \forall n \in \mathbb{N}.$$

Proof:
The proof will be done by induction. The statement is

$$S(n) \qquad \sum_{i=1}^{n} i = \frac{n(n+1)}{2}.$$

base case:

$$S(1) \qquad \underbrace{\sum_{i=1}^{1} i}_{=1} = 1 = \underbrace{\frac{1(1+1)}{2}}_{=1}.$$

inductive step: We assume that $S(n)$ is true for some $n \in \mathbb{N}$. This is called the induction hypothesis (I.H.). Then we will show:

$$S(n+1) \qquad \sum_{i=1}^{n+1} i = \frac{(n+1)(n+1+1)}{2}.$$

So

$$\begin{aligned}
\sum_{i=1}^{n+1} i &= \left(\sum_{i=1}^{n} i\right) + n + 1 \\
&\overset{I.H.}{=} \frac{n(n+1)}{2} + n + 1 \\
&= \frac{n(n+1) + 2(n+1)}{2} \\
&= \frac{(n+2)(n+1)}{2}.
\end{aligned}$$

Thus by proof by induction the Gau"s Sum Formula is proven. □

We now note some other well known examples, that should be proven with induction as an exercise.

Example 1.31.
- **Bernoulli's inequality**: For all $n \in \mathbb{N}$ and all $a \in \mathbb{R}$ with $a \geq -1$ we have

$$(a+1)^n \geq n \cdot a + 1.$$

- **Geometric Sum Formula**: For all $n \in \mathbb{N}$ and all $q \in \mathbb{R} \setminus \{1\}$ we have

$$\sum_{i=0}^{n} q^i = \frac{q^{n+1} - 1}{q - 1}.$$

As the next central important equation we want to show the Binomial Theorem. But we need some basic notations first:

Definition 1.32. (binomial coefficient)
Let $n \in \mathbb{N} \cup \{0\}$ and $k \in \{0, 1, ..., n\}$. Then the number of ways a k-element subset may be chosen in an n-element set is denoted by the *binomial coefficient* $\binom{n}{k}$.

Example 1.33.
Let $A := \{1, 2, 3\}$ be as set. Then $\binom{3}{2} = 3$ holds because $\{1, 2\}, \{1, 3\}, \{2, 3\}$ are all subsets of cardinality 2 in the 3-element set A.

The binomial coefficient may also be written as a specific formula.

Theorem 1.34.
Let $n \in \mathbb{N} \cup \{0\}$ and $k \in \{0, 1, ..., n\}$. Then

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

where the *factorials* are defined as $n! := 1 \cdot 2 \cdot 3 \cdot ... \cdot n$.

The binomial coefficient suffices the following basic properties.

Theorem 1.35.
Let $n \in \mathbb{N} \cup \{0\}$ and $k \in \{0, 1, ..., n\}$. Then the following holds:

i) $\binom{n}{k} = \binom{n}{n-k}$,

ii) $\binom{n}{0} = \binom{n}{n} = 1$,

iii) $\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}$, if $k \geq 1$.

The property *iii)* is actually the well known property to calculate the binomial coefficients recursivley in the Pascal's triangle.

Now we can note the Binomial theorem:

Theorem 1.36.
For all $n \in \mathbb{N} \cup \{0\}$ and all $a, b \in \mathbb{R}$ we have

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

Proof:
Proof by induction.

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

base case:

$$S(0) \qquad (a+b)^0 = 1 = \sum_{k=0}^{0} \binom{0}{k} a^k b^{0-k}.$$

inductive step: We assume that $S(n)$ is true for some $n \in \mathbb{N} \cup \{0\}$. Then we will show:

$$S(n+1) \qquad (a+b)^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} a^k b^{n+1-k}.$$

So

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

We now use the concept of the *index shift*, where we move the start and end value of the sum into one direction by an amount x and the index in the argument into the other direction by the exactly same amount.

$$\begin{aligned}
&= \sum_{k=1}^{n+1} \binom{n}{k-1} a^k b^{n-(k-1)} + \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k+1} \\
&= \binom{n}{n+1-1} a^{n+1} + \sum_{k=1}^{n} \binom{n}{k-1} a^k b^{n-k+1} + \sum_{k=1}^{n} \binom{n}{k} a^k b^{n-k+1} + \binom{n}{0} b^{n+1} \\
&= a^{n+1} + \sum_{k=1}^{n} \left[\binom{n}{k-1} + \binom{n}{k} \right] a^k b^{n-k+1} + b^{n+1} \\
&= a^{n+1} + \sum_{k=1}^{n} \binom{n+1}{k} a^k b^{n-k+1} + b^{n+1} \\
&= \sum_{k=0}^{n+1} \binom{n+1}{k} a^k b^{n-k+1}
\end{aligned}$$

Thus by proof by induction the Binomial theorem is proven. □

Remark 1.37.
The case $n = 2$ is well known as $(a+b)^2 = a^2 + 2ab + b^2$.

Remark 1.38.
Besides the notation for a sum, we also have a short notation for products:

$$\prod_{i=1}^{n} a_i = a_1 \cdot a_2 \cdot \ldots \cdot a_n.$$

A short example is $\prod_{i=1}^{5} i = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$.

1.4 Complex numbers

We have seen so far, that there is no real number that may solve the equation $x^2 = -1$, since we are not able to calculate the square root of a negative number. As before when we added the irrational numbers to the rational numbers to receive the real numbers, we will add the solution of the equation $x^2 = -1$ the real numbers to get an even bigger number system. We denote this solution with the imaginary unit.

Definition 1.39. (imaginary unit)
The solution to the equation $x^2 + 1 = 0$ is called *imaginary unit* and will be denoted by i.

Furthermore the imaginary unit gives us in combination with the real numbers the complex numbers.

Definition 1.40. (complex numbers)
The number system of the *complex numbers* will be denoted by $\mathbb{C} := \{x + y \cdot i \mid x, y \in \mathbb{R}\}$. For a complex number $z := x + yi,\ x, y \in \mathbb{R}$ we call x the *real part of* z denoted by $Re(z) = x$ and y the *imaginary part of* z denoted by $Im(z) = y$.
We define the addition and multiplication of two complex numbers $z = x + iy$ and $w = u + iv$ as

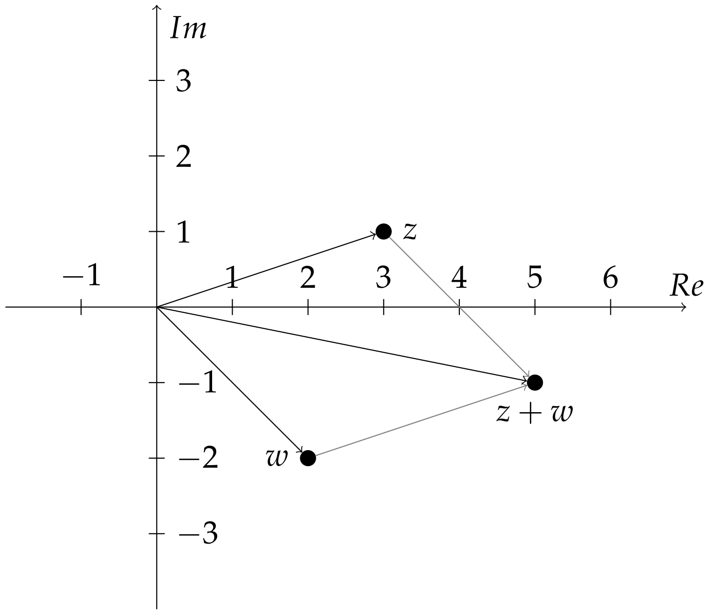
$$z + w = x + u + (y + v)i \text{ and } z \cdot w = xu - yv + (xv + yu)i.$$

Remark 1.41.
By the definition of the complex numbers we have $\mathbb{R} \subset \mathbb{C}$. That is $\mathbb{R} = \{x + y \cdot i \mid x \in \mathbb{R}, y = 0\} \subseteq \mathbb{C}$. The imaginary unit is in $\mathbb{C}$ but not in $\mathbb{R}$ since $x^2 + 1 = 0$ has no solution in $\mathbb{R}$.

We take a closer look at the addition of complex numbers. Let $z := x + iy$ and $w := u + iv$ be two complex numbers. Then we have:

$$z + w = x + iy + u + iv = x + u + i(v + y).$$

Let $z = 3 + i$ and $w = 2 - 2i$. Then we can visualize $z + w$ as:



The multiplication of complex numbers is more complicated. Let $z := x + iy$ and $w := u + iv$ be two complex numbers. Then we have:

$$\begin{aligned}
z \cdot w &= (x + iy)(u + iv) \\
&= xu + ixv + iyu + i^2 yv \\
&= xu + (-1)yv + i(xv + uy) \\
&= xu - vy + i(xv + uy).
\end{aligned}$$

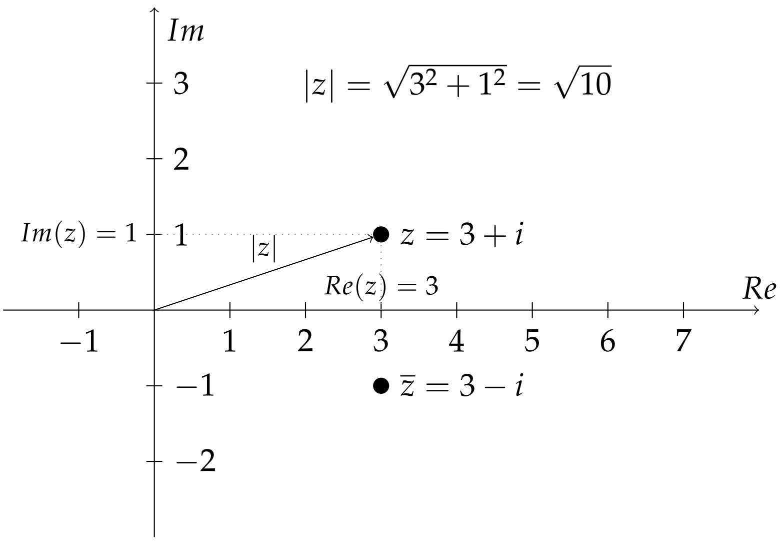
Later on we will understand the multiplication geometrically by using the polar coordinates of complex numbers.

We want to define two important objects for the complex numbers.

Definition 1.42. (complex conjugate, absolute value)
Let $z := x + yi$ be a complex number. Then we define $\overline{z} := x - yi$ as the *complex conjugate of* z. Furthermore we call $|z| := \sqrt{x^2 + y^2} \in \mathbb{R}$ the *absolute value of* z.

The absolute value may be seen as the distance of the complex number to the origin of the complex plane. The complex plain may be deriven from the 2-dimensional space if we interpretate the real part $Re(z)$ and the imaginary part $Im(z)$ of some complex number z as coordinates in a 2 dimensional plain. The complex conjugate $\overline{z}$ may be visualized as the reflection of z about the real axis.

Let $z = 3 + i$ be a complex number. Then we can visualize $|z|$ and $\overline{z}$ as:



Next we are going to establish some basic rules that hold for complex numbers.

Theorem 1.43.
Let z, w be two complex numbers. Then the follwing holds:

i) $\overline{z+w} = \overline{z} + \overline{w}$, $\overline{zw} = \overline{z} \cdot \overline{w}$, $\overline{\left(\frac{1}{z}\right)} = \frac{1}{\overline{z}}$ if $z \neq 0$,

ii) $\overline{\overline{z}} = z$,

iii) $z \in \mathbb{R} \;\Leftrightarrow\; z = \overline{z}$,

iv) $|z| \geq 0$, and furthermore $|z| = 0 \;\Leftrightarrow\; z = 0$,

v) $|z| = |\overline{z}|$, $z \cdot \overline{z} = |z|^2$,

vi) $\left|\frac{z}{w}\right| = \frac{|z|}{|w|}$, if $w \neq 0$,

vii) $||z| - |w|| \leq |z + w| \leq |z| + |w|$. (triangle inequality)

Now if we divide one complex number z by another complex number w and write this as a fraction $\frac{z}{w}$, we do not know what that fraction is as an actual complex number. We do not have a real part or an imaginary part of a complex number if we have some complex number in the denominator of the fraction. But we may solve the fraction to get a compley number with a distinct real and imaginary part. So let $z = x + iy$ and

$w = u + iv$ be two complex numbers.

$$\begin{aligned}\frac{z}{w} &= \frac{x+iy}{u+iv} \\ &= \frac{x+iy}{u+iv} \cdot \underbrace{\frac{u-iv}{u-iv}}_{=\frac{\overline{w}}{\overline{w}}} \\ &= \frac{xu - ivx + iuy - i^2vy}{u^2 - i^2v^2} \\ &= \frac{xu + vy + i(uy - vx)}{u^2 + v^2} \\ &= \frac{xu + vy}{u^2 + v^2} + i \cdot \frac{uy - vx}{u^2 + v^2}\end{aligned}$$

Example 1.44.
For two actual complex numbers $z = 3 + 5i$ and $w = 1 + 2i$ we have

$$\begin{aligned}\frac{z}{w} &= \frac{3+5i}{1+2i} \\ &= \frac{3+5i}{1+2i} \cdot \frac{1-2i}{1-2i} \\ &= \frac{3 - 6i + 5i - 10i^2}{1 - 4i^2} \\ &= \frac{3 + 10 + i(-6+5)}{1+4} \\ &= \frac{13}{5} - \frac{1}{5}i\end{aligned}$$

By the previous consideration we may deduce the following lemma.

Lemma 1.45. (Inversion Formula)
Let $z = x + iy$ be a complex number with $z \neq 0$. Then

$$\frac{1}{z} = \frac{1}{x+iy} = \frac{x}{x^2+y^2} + \frac{-y}{x^2+y^2}i$$

holds.

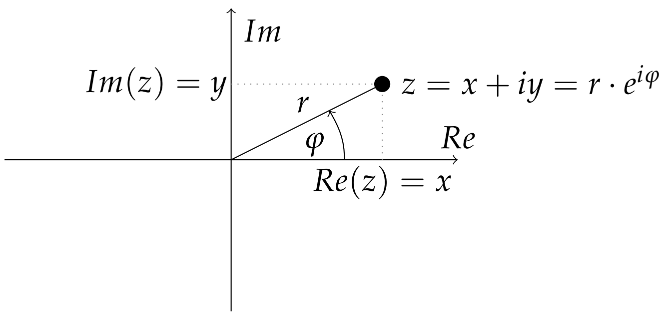
Now we may interpretate the real part and imaginary part of complex numbers as coordinates in the complex plane. But there is another way to look at complex numbers. We can visualize them not only by its real and imaginary part but we can define them with the help of the absolute value and an angle, called the argument.

Definition 1.46. (polar coordinates)
The *polar coordinates* of any complex number $z \in \mathbb{C} \setminus \{0\}$ are the pair (r, φ) with $r > 0,\ \varphi \in [0, 2\pi)$, such that

$$\begin{aligned} z &= r \cdot (\cos\varphi + i\sin\varphi) \\ &= r \cdot e^{i\varphi}. \end{aligned}$$

$r = |z|$ is the absolute value of z and $\varphi =: arg(z)$ is called the *argument* of z, that is the angle between the real axis and the line through 0 and z.

So for $z = x + iy = r \cdot e^{i\varphi} \in \mathbb{C}$ we have:

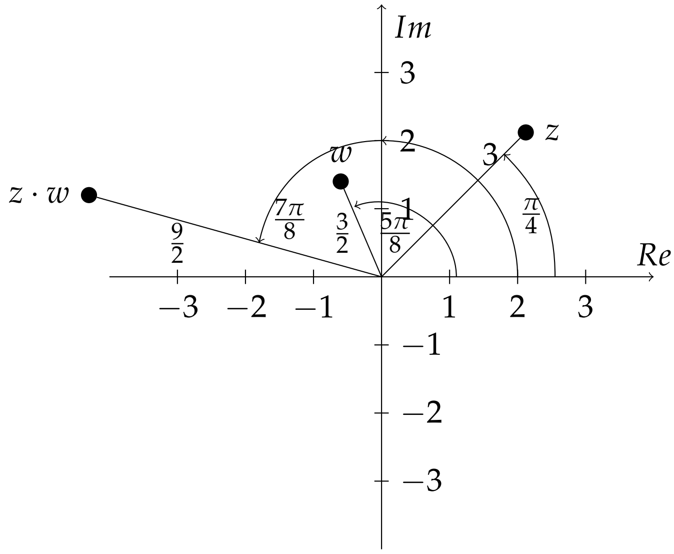


The usefulness of the polar coordinates can be shown by multiplication of complex numbers. Let $z := x + iy$ and $w := u + iv$ be two complex numbers.

$$\begin{aligned} z \cdot w &= r \cdot (\cos\varphi + i\sin\varphi) \cdot s \cdot (\cos\theta + i\sin\theta) \\ &= r \cdot s \cdot (\cos\varphi\cos\theta + i\cos\varphi\sin\theta + i\sin\varphi\cos\theta + \underbrace{i^2}_{=-1}\sin\varphi\sin\theta) \\ &= r \cdot s \cdot (\underbrace{(\cos\varphi\cos\theta - \sin\varphi\sin\theta)}_{=\cos(\varphi+\theta)} + i\underbrace{(\cos\varphi\sin\theta + \sin\varphi\cos\theta)}_{=\sin(\varphi+\theta)}) \\ &= r \cdot s \cdot (\cos(\varphi+\theta) + i\sin(\varphi+\theta)) \\ &= r \cdot s \cdot e^{i(\varphi+\theta)}. \end{aligned}$$

The equalities for $\sin(\varphi + \theta)$ and $\cos(\varphi + \theta)$ used are called the *trigonometric angle sum identities*. They hold for any two angles φ and θ. So the multiplication of complex numbers is the multiplication of its absolute values along with the addition of its arguments.

For $z = 3 \cdot e^{\frac{\pi}{4}i}$ and $w = \frac{3}{2} \cdot e^{\frac{5\pi}{8}i}$ the complex number $z \cdot w = \frac{9}{2} \cdot e^{i\frac{7\pi}{8}}$ may be visualized as followed:



So we might add some basic rules for working with polar coordinates.

Theorem 1.47.
Let $z, w \in \mathbb{C}$, $z \neq 0$, $w \neq 0$, $z = re^{i\varphi}$, $w = se^{i\theta}$. Then the following holds:

i) $z \cdot w = r \cdot s \cdot e^{i(\varphi+\theta)}$,

ii) $\frac{1}{z} = \frac{1}{r} \cdot e^{-i\varphi}$,

iii) $\overline{z} = r \cdot e^{-i\varphi}$,

iv) $z^n = r^n \cdot e^{in\varphi}$.

Part *iv*) of the last theorem yields the formula of De Moivre. With its help we may calculate any root of some complex number.

Theorem 1.48. (De Moivre Formula)
For all complex numbers $w = s \cdot e^{i\theta}$ there are exactly n numbers $z_0, z_1, ..., z_{n-1}$, that solve the equation $z_k^n = w$ for all $k \in 0, 1, ..., n-1$. Those numbers are

$$z_k = \sqrt[n]{s} \cdot e^{i\left(\frac{\varphi+2k\pi}{n}\right)},\ 0 \leq k \leq n-1 \text{ and } \varphi := \frac{\theta}{n}.$$

Those roots are equidistantly distributed on a circle with center $(0, 0)$ and radius $\sqrt[n]{s}$.

In the next chapter we will see some important facts about complex numbers in regard to polynomials.

1.5 Polynomials

In this chapter we will talk about polynomials. We will consider those functions, that have real coefficients.

Definition 1.49. (polynomial)
Let $n \in \mathbb{Z}_{\geq 0}$, $a_0, a_1, ..., a_n \in \mathbb{R}$, $a_n \neq 0$ and x be a variable. Then

$$p(x) := a_n x^n + a_{n-1} x^{n-1} + ... + a_2 x^2 + a_1 x + a_0$$

is called a *polynomial*. $n =: deg(p)$ is called the *degree* of $p(x)$. It is the largest exponent of x appearing in $p(x)$. For $deg(p) = 0$ we call p the *constant polynomial*. The numbers $a_0, a_1, ..., a_n$ are called the *coefficients* of $p(x)$.

Example 1.50.
Here is a polynomial written in a few different ways:

$$(x^2 - 4) \cdot (x + 1) = x^3 + x^2 - 4x - 4 = x^3 + x^2 - 4(x + 1).$$

The addition, substraction and multiplication of polynomials is again a polynomial. In particular for the multiplication of two polynomials we have the following lemma.

Lemma 1.51.
Let p and q be two polynomials. Then

$$deg(p \cdot q) = deg(p) + deg(q)$$

holds.

We do not have such a lemma for the addition or substraction of polynomials, because the coefficient of the largest exponent of x may be or may not be reduced from the polynomial with that operation.

If we divide one polynomial by another we usually do not get a polynomial as the result. A part of the divisor may appear in the denominator. So in general there will be a remainder. We call the act of dividing one polynomial by another the *polynomial division with remainder*.

Theorem 1.52. (polynomial division with remainder)
For any two polynomials $p(x), q(x), q(x) \neq 0$, there are polynomials $a(x)$ and $r(x)$ such that $deg(r) < deg(q)$ and

$$p = q \cdot a + r.$$

r is called the remainder.

Example 1.53.
Let $p(x) = x^3 - x^2 + 3x - 1$ and $q(x) = x^2 - 1$ be two polynomials. Then the polynomial division with remainder of $\frac{p(x)}{q(x)}$ is:

$$\begin{array}{l} \quad x^3 - x^2 + 3x - 1 = \left(x^2 - 1\right)\left(x - 1\right) + 4x - 2 \\ \underline{-x^3 \qquad\quad + x} \\ \qquad -x^2 + 4x - 1 \\ \qquad \underline{\;\; x^2 \qquad - 1} \\ \qquad\qquad 4x - 2 \end{array}$$

So $a(x) = x - 1$ and the remainder is $r = 4x - 2$. Note that $deg(r) = 1 < 2 = deg(q)$ holds. There is also a second common way to visualize the polynomial division with remainder:

$$\begin{array}{l} \left(\quad x^3 - x^2 + 3x - 1\right) : \left(x^2 - 1\right) = x - 1 + \dfrac{4x - 2}{x^2 - 1} \\ \underline{-x^3 \qquad\quad + x} \\ \qquad -x^2 + 4x - 1 \\ \qquad \underline{\;\; x^2 \qquad - 1} \\ \qquad\qquad 4x - 2 \end{array}$$

Definition 1.54. (factor, linear factor, irreducible polynom)
If a polynomial $p(x)$ can be divided by another polynomialn $q(x)$ without a remainder, that is $r = 0$, and $deg(b) \geq 1$ holds, then q is called a *factor* of p. A factor q with $deg(q) = 1$ is called a *linear factor* of p. A polynomial with no proper factor is called *irreducible*. All polynomials p with $deg(p) = 1$ are irreducible.

If our co-domain are the real numbers, polynomials may be factorized. But those irreducible factors may be of some huge degree. But if we have the complex numbers as our co-domain we the follwing strong theorem.

Theorem 1.55. (fundamental theorem of algebra)
Let $p(z) = a_n z^n + a_{n-1} z^{n-1} + ... + a_2 z^2 + a_1 z + a_0$ be a polynomial with complex coefficients and $a_n \neq 0$. Then there are exactly n numbers $z_1, z_2, ..., z_n$ in $\mathbb{C}$, such that the follwoing holds:

$$p(z) = a_n (z - z_1) \cdot (z - z_2) \cdot ... \cdot (z - z_n).$$

The numbers $z_1, z_2, ..., z_n$ do not need to be distinct.

So the fundamental theorem of algebra states, that any polynomialmay be factorized into linear factors, if the co-domain is $\mathbb{C}$.

Corollary 1.56.
Let p be a polynomial. Then the only irreducible factors of p over $\mathbb{C}$ are linear.

Now we want to talk about *zeroes* and their important connection to factored polynomials.

Definition 1.57. (zeroes)
Let $p(x)$ be a polynomial. A number $x \in \mathbb{R}$ with $p(x) = 0$ is called a *zero*.

There is an important connection between linear factors of a polynomialand the zeroes of that polynomial. We have:

Theorem 1.58.
Let $p(x)$ be a polynomial and $a \in \mathbb{R}$. Then the following equivalence holds:

$$p(a) = 0 \qquad \Leftrightarrow \qquad x - a \text{ is a linear factor of } p(x).$$

Proof:
"$\Rightarrow$": Assume that a is a zero of p, that is $p(a) = 0$. Now divide $p(x)$ by $x - a$ with remainder $r(x)$. This yields

$$p(x) = (x - a) \cdot q(x) + r(x)$$

for some $q(x)$ and $deg(r) < deg(x - a)$. Since $deg(x - a) = 1$ holds, we have that $r(x)$ is a constant, that is $r(x) = r_0 \in \mathbb{R}$. Therefore we have

$$p(x) = (x - a) \cdot q(x) + r_0.$$

With $p(a) = (a - a) \cdot q(a) + r_0$ we get $p(a) = r_0$. By the assumption that a is a zero of p this holds only for $r_0 = 0$. Therefore $x - a$ is a linear factor of $p(x)$.
"$\Leftarrow$": Assume that $x - a$ is a linear factor of $p(x)$, that is there is some $q(x)$ with

$$p(x) = (x - a) \cdot q(x).$$

Then $p(a) = (a - a) \cdot q(x) = 0$ yields, that a is a zero of $p(x)$. $\square$

This theorem implies the following.

Corollary 1.59.
A polynomial of degree n can have at most n zeroes.

Furthermore under certain conditions to the images of some numbers under that polynomial, we may deduce the equality of polynomials.

Corollary 1.60.
Let p, q be polynomials with $deg(p) \leq n$ and $deg(q) \leq n$. If there are $n+1$ different numbers $x_i,\ \ i \in \{0, 1, 2, ..., n\}$ with $p(x_i) = q(x_i)$ for all $i \in \{0, 1, 2, ..., n\}$, then $p = q$ holds.

The corollary states, that if we have enough numbers, whose images are the same under p and under q, then those two polynomials p and q need to be equal. By this we get to one very important application of polynomials, called the *Lagrange interpolation*. For this we assume, that we have $n+1$ different numbers and their images. Now we want to find a polynomialof degree at most n that satisfies those ordered pairs of numbers and their images.

Theorem 1.61. (Lagrange Interpolation Theorem)
Let $x_0 < x_1 <_x 2 < ... < x_n$ be real numbers and $y_0, y_1, ..., y_n \in \mathbb{R}$. Then there exists a unique polynomial $p(x)$ of degree n such that

$$p(x_i) = y_i,\ \forall i \in \{0, 1, 2, ..., n\}$$

holds. Furthermore that polynomial is given by

$$p(x) = \sum_{i=0}^{n} y_i p_i(x)$$

with

$$p_i(x) = \frac{\prod\limits_{j:j\neq i} (x - x_j)}{\prod\limits_{j:j\neq i} (x_i - x_j)}.$$

For all p_i holds $p_i(x_i) = 1$.

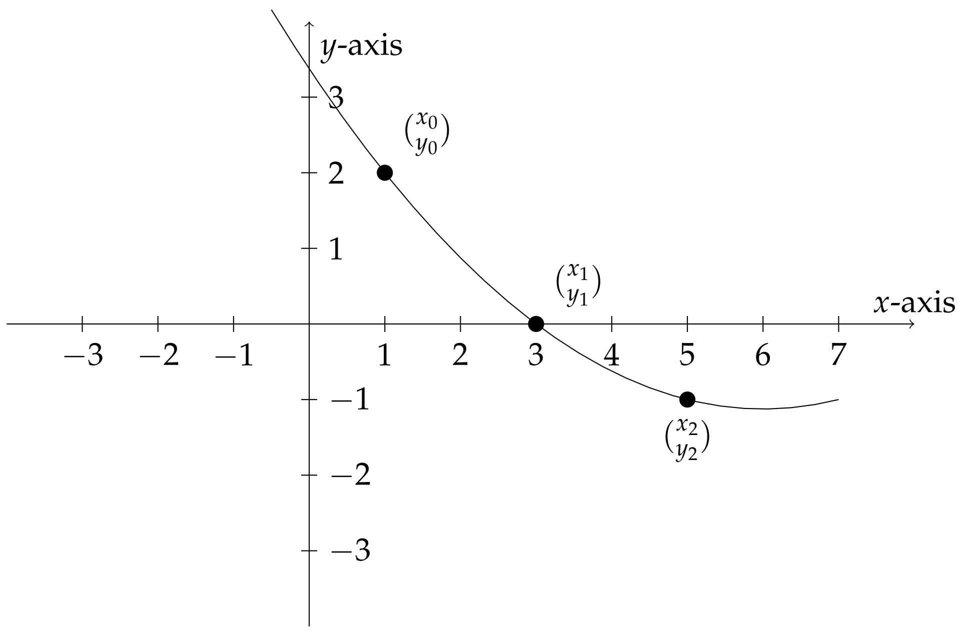
So with the Lagrange Interpolation Theorem we may find to a set of numbers $x_0 < x_1 < ... < x_n$ and their respected images $p(x_i)$ a polynomial p that fits all points $\binom{x_i}{p(x_i)}$ in the 2-dimensional plane. We conclude this chapter with an example about constructing a polynomial from its ordered pairs.

Example 1.62.
The points $\binom{1}{2}$, $\binom{3}{0}$ and $\binom{5}{-1}$ are given in the plane. So we have $x_0 = 1,\ x_1 = 3$ and $x_2 = 5$ with their images $y_0 = 2,\ y_1 = 0$ and $y_2 = -1$. First we calculate the p_i.

$$\begin{aligned}
p_0(x) &:= \frac{x-3}{1-3} \cdot \frac{x-5}{1-5} = \frac{1}{8}(x-3)(x-5) \\
p_1(x) &:= \frac{x-1}{3-1} \cdot \frac{x-5}{3-5} = \frac{-1}{4}(x-1)(x-5) \\
p_2(x) &:= \frac{x-1}{5-1} \cdot \frac{x-3}{5-3} = \frac{1}{8}(x-1)(x-3)
\end{aligned}$$

Notice that $p_0(x_0) = p_0(1) = 1$ but $p_0(x_1) = p_0(x_2) = 0$. For p_1 and p_2 this works repectively. So we deduce our polynom as

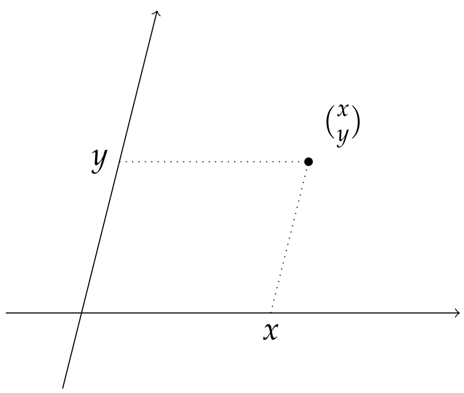
$$\begin{aligned} p(x) &= 2 \cdot p_0(x) + 0 \cdot p_1(x) + (-1) \cdot p_2(x) \\ &= \frac{1}{8}x^2 - \frac{3}{2}x + \frac{27}{8}. \end{aligned}$$



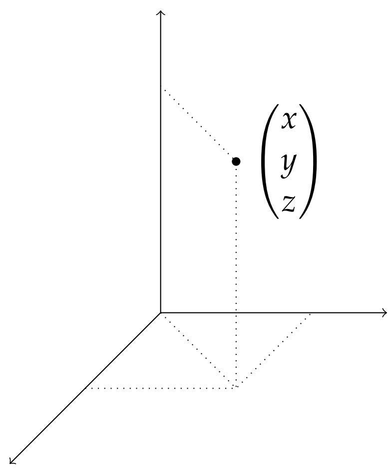
2 Linear Algebra

2.1 Vector spaces

As we have already seen, we can display the numbers of some ordered number system on number line. Furthermore we displayed points into the complex plane. For us, the complex plane was a two dimensional, that might be interpretated with cartesian coordiantes as the cartesian product of the two sets $\mathbb{R}$ and $\mathbb{R}$. So it yielded a right-angled coordinate system. Yet not all coordinate systems for two dimensional planes need to right-angled.



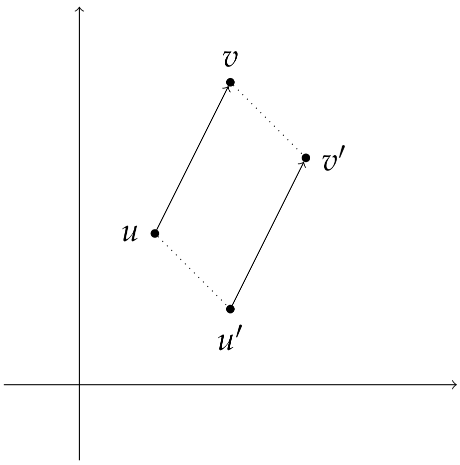
Now we know that we can also display points into the 3-dimensional space. Analoguesly we can have 3 axis for our coordinates and create the coordinate system for a 3-dimensional space.



This may be generalized up to n dimensions. So a point in an n-dimensional space is a n-tuple of real numbers and the n-dimensional space may be described by

$$\mathbb{R}^n = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \middle| \; x_i \in \mathbb{R}, \; 1 \leq i \leq n \right\}$$

Thus far we talked about points in some n-dimensional space. But now we want to describe and define vectors and vector spaces. In $\mathbb{R}^n$ with $n \leq 3$ we may interpretate vectors as directed lines between points in the space. for two points u and v the vector $\overrightarrow{uv}$ might be the line from u to v.



We do not differentiate between one directed line and another, that was derived by the first line through translation. Therefore in our picture the two lines $\overrightarrow{uv}$ and $\overrightarrow{u'v'}$ are the same vectors, that is $\overrightarrow{uv} = \overrightarrow{u'v'}$.
So vectors itself might be understood as translation. For example if you have a moving point at a given time in space, the vector of that point might be the velocity vector. It tells us, where the moving point will be for a specific point in time later on.

We will define the vector space in an axiomatic way, that is we define a structure by demanding conditions.

Definition 2.1. (vector space)
A *vector space* is a set V with specific defined operations: an addition of elements in V and multiplication with elements of a field F (i.e. $F = \mathbb{C}$ or $F = \mathbb{R}$), called the *scalar multiplication*. For all $u, v, w \in V$ and all $\alpha, \beta \in F$ the rules for those operations are defined as:

i) $u + v = v + u$ (*commutative property*)

ii) $(u + v) + w = u + (v + w)$ (*associative property*)

iii) $\exists\ 0 = 0_V \in V$ (called *zero vector*) with $u + 0_V = u$

iv) $\forall\ u \in V\ \exists\ (-u) \in V$ (called *inverse element*) with $u + (-u) = 0_V$

v) $\alpha \cdot (\beta \cdot u) = (\alpha\beta) \cdot u$

vi) $\alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v$

vii) $(\alpha + \beta) \cdot u = \alpha \cdot u + \beta \cdot u$

viii) $1 \cdot u = u$

The elements of the vector space V are called *vectors*.

Example 2.2.
i) Let

$$V = F^n = \left\{ \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} \middle| \ z_i \in \mathbb{R} \text{ or } \mathbb{C},\ 1 \leq i \leq n \right\}$$

be a vector space over a field F. The addition and scalar multiplication in a vector space is defined componentwise:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}, \qquad \alpha \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \alpha \cdot x_1 \\ \alpha \cdot x_2 \\ \vdots \\ \alpha \cdot x_n \end{pmatrix}$$

The zero vector is defined as

$$0_V = \left.\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}\right\} n \text{ rows.}$$

With those definitions of an addition and a scalar multiplication it can be verified that V is indeed a vector space.

ii) For all n

$$V = \text{Poly}(n) = \{p(x) = a_n x^n + a_{n-1}x^{n-1} + ... + a_1 x + a_0 |\ a_i \in F,\ 0 \leq i \leq n\}$$

is the vector space of all polynomials up to degree n over the field F. For F we used in the chapter polynomials $\mathbb{R}$. The zero vektor is the polynomial $p(x) = 0$, that is all coefficients a_i are zero. The addition and multiplication for two polynomials $p(x) = \sum_{i=0}^{n} \alpha_i x^i$ and $q(x) = \sum_{i=0}^{n} \beta_i x^i$ are defined as:

$$p(x) + q(x) = \sum_{i=0}^{n} (\alpha_i + \beta_i) x^i, \qquad \alpha \cdot p(x) = \sum_{i=0}^{n} (\alpha \cdot \alpha_i) x^i.$$

With a given vector space and its vectors, we may want to combine the vectors. Furthermore we might want to create subsets of a vector space V that are vector spaces itself, too. We will call those structures subspaces of V.

Definition 2.3. (linear combination, subspace, span)
Let V be a vector space with $v_1, v_2, ..., v_n \in V$ and $\alpha_1, \alpha_2, ..., \alpha_n \in F$. Then the sum

$$\sum_{i=1}^{n} \alpha_i v_i$$

is called the *linear combination* of $v_1, v_2, ..., v_n$.
Now let $U \subset V$ be a subset. U is called a *subspace of* V, if U including the operations inherited by V is a vector space itself, that is if

$$\begin{aligned} 0_V &\in U, \\ u, v \in U &\Rightarrow u + v \in U, \\ u \in U, \alpha \in F &\Rightarrow \alpha \cdot u \in U \end{aligned}$$

holds.
Furthermore $\text{Span}(v_1, ..., v_n) := \{\sum_{i=1}^{n} \alpha_i v_i |\ \alpha_i \in F,\ 1 \leq i \leq n\}$ is called the *span* of $v_1, v_2, ..., v_n$.

Remark 2.4.
The span of some vectors is a vector space itself.

With linear combination, we can create vectors from combining other vectors. But when is one vector not generatable out of a set of other vectors? If it is not generatable, then it is called linearly independent.

Definition 2.5. (linearly independent)
Let V be a vector space with $v_1, v_2, ..., v_n \in V$. The vectors $v_1, ..., v_n$ are called *linearly independent* if $\sum_{i=1}^{n} \alpha_i v_i = 0$ always implies $\alpha_1 = \alpha_2 = ... = \alpha_n = 0$. Otherwise they are called *linearly dependent*.

In other words it is not possible to generate the zero vector with a set of linearly independent vectors, except if all coefficients are already zero. Or a set of vectors is linearly independent if there is no vector or a combination of vectors of that set that is reproducable by the remaining vectors.

Definition 2.6. (finitely generated, generating set)
A vector space V is called *finitely generated*, if $V = Span(v_1, ..., v_n)$ for a finite set $v_1, v_2, ..., v_n \in V$. Then $v_1, ..., v_n$ is called a *generating set*.

Now a set of vectors can generate a vector space. But the mathematical interesting question is, how many vectors are actually needed to generate that vector space. The next theorem will give us the equivalence to a minimal set of vectors, that generates the vector space.

Theorem 2.7.
Let V be a finitely generated vector space. Then the following are equivalent:

i) $v_1, v_2, ..., v_n$ is a *minimal generating set*,

ii) $v_1, v_2, ..., v_n$ is a *maximal linearly independent set*,

iii) Each vector $v \in V$ is generated by a unique linear combination of $v_1, v_2, ..., v_n$.

Instead of showing the equivalence of each statement to each other statementment, one would use the method *cycle of implications* to show the equivalence of this three implications. That means, it needs to be shown that $i) \Rightarrow ii)$, $ii) \Rightarrow iii)$ and $iii) \Rightarrow i)$.

The concept of the last theorem also has a specific name and is a very important structure in mathematics.

Definition 2.8. (basis)
A set of vectors, that meets any of the three conditions in the previous theorem is called a *basis* of V.

Example 2.9.
The canonical basis for $\mathbb{R}^n$ would be the vectors

$$\begin{pmatrix}1\\0\\0\\\vdots\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\\\vdots\\0\end{pmatrix}, \dots, \begin{pmatrix}0\\0\\\vdots\\0\\1\end{pmatrix}.$$

Those are also called the *unit vectors* of $\mathbb{R}^n$.

Definition 2.10. (dimension)
Let V be a finitely generated vector space. Then the number of elements in a basis of V is called the *dimension* of V.

Remark 2.11.
All bases of a certain vector space have the same dimension.

2.2 Systems of linear equations and matrices

Now we want solve systems of linear equations. As we have seen earlier, we can solve two linear equations with two unknown variables by adding and subtracting them from each other. We were also able to solve one equation for one variable and then use that solution in the second equation to calculate the variables. The question is now, what are we going to do if we have huge systems of linear equation, that is we have lots of equations and lots of unknown variables? In this chapter we learn how to solve this with the *Gaussian elimination*. But first we look at an example for systems of linear equations.

Example 2.12.
There are 3 persons of unknown age. The persons are Lina, Arne and Oliver. 4 years ago Lina was twice as old as Arnes' and Olivers' age combined. Next year Oliver will be one fifth as old as Linas' and Arnes' age combined. In ten years Arne will be half as old as Linas' and Oliver' ages combined. The question is: How old are Lina, Arne and Oliver today?
This actually yields us the follwing system of linear equations with A, O and L as their ages today:

$$\begin{aligned} 2A + 2O - L &= 12 \\ -A + 5O - L &= -3 \\ 2A - O - L &= 0. \end{aligned}$$

At the end of the chapter we will be able to show, that Lina is 14, Arne is 9 and Oliver is 4 years old.

Definition 2.13. (matrix)
A rectangular table

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

with numbers $a_{ij} \in F$, $1 \leq i \leq m$, $1 \leq j \leq n$ is called a $m \times n$-*matrix* over F. i represents the *row index* and j represents *column index* of some entry a_{ij}. The set of all $m \times n$-matrices over F is denoted by $F^{m \times n}$. The elements of $F^{1 \times n}$ are called *row vectors* and the elements of $F^{m \times n}$ are called *column vectors*.

We can now define the addition and multiplication of matrices. Let $A, B \in \mathbb{R}^{m \times n}$, $C \in \mathbb{R}^{n \times o}$ and $b \in \mathbb{R}^{n \times 1}$ be some matrices. Then we have

$$A + B = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

$$A \cdot b = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ \vdots \\ b_{m1} \end{pmatrix} = \begin{pmatrix} a_{11} \cdot b_1 + ... + a_{1n} \cdot b_n \\ \vdots \\ a_{m1} \cdot b_1 + ... + a_{mn} \cdot b_n \end{pmatrix} \in \mathbb{R}^{n \times 1}$$

$$A \cdot C = A \cdot (C_1 | C_2 | ... | C_n) = (A \cdot C_1 | A \cdot C_2 | ... | A \cdot C_n) \in \mathbb{R}^{n \times o}$$

with $C_1, C_2, ..., C_n$ are the columns of C.

Remark 2.14.
If we multiply two matrices with each other, the number of columns of the first matrix needs to match the number of rows of the second matrix. That is, we can only multiply two matrices A, B with $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{s \times t}$ if $n = s$ holds. Then we have $A \cdot B \in \mathbb{R}^{m \times t}$.

The matrix multiplication follows the laws of associativity.

Lemma 2.15.
Let $A, B, C \in \mathbb{R}^{n \times n}$ be some matrices. Then we have

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

The previous lemma works for any matrices that can be multiplied with each other. They do not need to be of quadratic shape. While the associativity works for matrix multiplication, the commutativity does not.

Remark 2.16.
In general the matrix multiplication is not commutative. That is for two matrices $A, B \in \mathbb{R}^{n\times n}$, we have in general

$$A \cdot B \neq B \cdot A.$$

Example 2.17.
Let

$$A = \begin{pmatrix} 1 & 3 & 3 \\ 2 & 0 & 0 \\ 1 & 3 & 1 \end{pmatrix} \in \mathbb{R}^{3\times 3} \text{ and } B = \begin{pmatrix} 3 & 2 \\ \frac{1}{2} & \frac{1}{2} \\ 4 & 2 \end{pmatrix} \in \mathbb{R}^{3\times 2}$$

be some real matrices. Then we have

$$A^2 = A \cdot A = \begin{pmatrix} 10 & 12 & 6 \\ 2 & 6 & 6 \\ 8 & 6 & 4 \end{pmatrix} \in \mathbb{R}^{3\times 3} \text{ and } A \cdot B = \begin{pmatrix} \frac{33}{2} & \frac{19}{2} \\ 6 & 4 \\ \frac{17}{2} & \frac{11}{2} \end{pmatrix} \in \mathbb{R}^{3\times 2}.$$

The matrix products of $B \cdot B = B^2$ or $B \cdot A$ do not exist.

We add one quick notation, that will simplify the notations of matrices and vectors in certain scenarios.

Definition 2.18. (transpose)
Let $A \in \mathbb{R}^{m\times n}$ be a matrix with

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Then we call A^T the *transpose* of matrix A, that is defined as

$$A^T := \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix}.$$

The transpose A^T of a matrix A is actually the reflection of the entries of A at the axis that is represented by the diagonal elements of A.

Example 2.19.
So for some matrices we actually have

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \text{ with } A^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

or

$$b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \text{ with } b^T = \begin{pmatrix} b_1 & b_2 & ... & b_n \end{pmatrix}.$$

Now we have the structure of a matrix and we know how to multiply matrices and vectors. With that we can actually define our systems of linear equations and simplify a possible solving algorithm to the Gaussian Algorithm over the extended matrix form.

Definition 2.20. (system of linear equations, extended matrix)
A system of m equations with n variables $x_1, x_2, ..., x_n$ of the form

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

with $a_{11}, ...a_{mn}, b_1, ..., b_n \in \mathbb{R}$ is called a *system of liner equations*. For

$$A := \begin{pmatrix} a_{11} & ... & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & ... & a_{mn} \end{pmatrix} \text{ and } b := \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

we have that $Ax = b$ is equivalent to the system of linear equations, where $x = \begin{pmatrix} x_1 & ... & x_n \end{pmatrix}^T$ is the vector of the variables of the system. We call

$$\left(A\middle|b\right) = \left(\begin{array}{ccc|c} a_{11} & ... & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{m1} & ... & a_{mn} & b_n \end{array}\right)$$

the *extended matrix*. Furthermore b is called the *right-hand side* of the system.

In general we distinguish between two types of systems of linear equations.

Definition 2.21. (homogeneous, inhomogeneous)
Let $(A|b)$ be an extended matrix. The system is called *homogeneous*, if $b = (0, ..., 0)^T$. Otherwise we call the system *inhomogeneous*.

Solving systems of linear equations is to find the set

$$S = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^{n\times 1} \,\middle|\, Ax = b \right\}.$$

For that we use the Gaussian algorithm. The algorithm itself is based on three ***elementary row operations***. Those are

i) Swapping two equations,

ii) Multiplying an equation with some number $\lambda \in \mathbb{R} \setminus \{0\}$,

iii) Adding a multiple of some equation to another.

These operations translate for an extended matrix $(A|b)$ into

i) Swapping two rows,

ii) Multiplying a row with some number $\lambda \in \mathbb{R} \setminus \{0\}$,

iii) Adding a multiple of some row to another.

Theorem 2.22.
Elementary row operations do not change the solution of a system of linear equations.

We do not change the solution of a system of linear equations because we actually manipulate the extended matrix with a bijective map.

Definition 2.23. (simple form)
A matrix $A \in \mathbb{R}^{m\times n}$ is called in *simple form*, if there is a number $r \in \{0, 1, ..., m\}$ with

i) The first r rows contain at least one non-zero entry each, the rows $r + 1$ to m contain zeroes only.

ii) If $s_i := \min\{j|\ a_{ij} \neq 0\}$ for all $i \leq r$ then $s_1 < s_2 < ... < s_r$.

iii) $a_{is_i} = 1$ for all $i \in \{1, 2, ..., m\}$.

iv) $a_{ij} = 0$ for all $j \neq s_i$.

A matrix in simple form is basicly a triangular or stair-like shaped matrix, that ensures that each step has a 1 as an entry and below und above that entry are only zeroes.

Example 2.24.
The following matrix is in simple form:

$$\begin{pmatrix} 1 & * & 0 & 0 & * & * & 0 & * \\ 0 & 0 & 1 & 0 & * & * & 0 & * \\ 0 & 0 & 0 & 1 & * & * & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The $*$ represents any random real number.

We are able to get the solution of any system of linear equations right from the simple form. So we need the following theorem:

Theorem 2.25.
Each matrix can be reduced to a simple form with elementary row transformations.

The Gaussian algorithm gives us a step by step algorithm about how we are able to get out of any extended matrix a simple form. Basicly we use elementary row transformations to achieve the simple form.

Algorithm 2.26. (Gaussian Algorithm)

Input:	A linear system $Ax = b$
Output:	The set S of all solutions
1.	Use elementary row transformations on $(A\|b)$ to bring A into simple form A'.
2.	If A' is in diagonal form, read off the unique solution S.
3.	Erase all zero rows in $(A'\|b')$, if there are any.
4.	Add for each variable x_i whose column does not have a step in A' an equation $1 \cdot x_i = \alpha_i$ with $\alpha_i \in \mathbb{R}$. By that the extended matrix $(A'\|b')$ without zero rows will be transformed into an uniquely defined system $(A''\|b'')$ with variables $\alpha_1, ..., \alpha_k \in \mathbb{R}$. In particular A'' has a quadratic shape.
5.	Use elementary row transformations on $(A''\|b'')$ to bring A'' into simple form.
6.	Read off the set S of solutions with variables $\alpha_1, ..., \alpha_k \in \mathbb{R}$.

Remark 2.27.
If $Ax = b$ has one unique solution then it can be read off directly from the simple form without any additional conditions, since it is a diagonal matrix.

Remark 2.28.
If at any point in the Gaussian algorithm one zero row on the then transformed matrix A equals a non negative number on the right-hand side, then this system has no solution.

Example 2.29.
We use the problem at the beginning of this chapter. So we have the system of linear equations

$$\begin{aligned} 2A + 2O - L &= 12 \\ -A + 5O - L &= -3 \\ 2A - O - L &= 0. \end{aligned}$$

We can transform this into the extended matrix with $A = x_1$, $O = x_2$ and $L = x_3$.

$$\left(\begin{array}{ccc|c} 2 & 2 & -1 & 12 \\ -1 & 5 & -1 & -3 \\ 2 & -1 & -1 & 0 \end{array}\right)$$

Now we try to create as many zeroes as possible and shape the matrix in a stair-like form, that is the simple form.

$$\left(\begin{array}{ccc|c} 2 & 2 & -1 & 12 \\ -1 & 5 & -1 & -3 \\ 2 & -1 & -1 & 0 \end{array}\right) \overset{I\cdot\frac{1}{2}}{\rightsquigarrow} \left(\begin{array}{ccc|c} 1 & 1 & -\frac{1}{2} & 6 \\ -1 & 5 & -1 & -1 \\ 2 & -1 & -1 & 0 \end{array}\right) \underset{III-2\cdot I}{\overset{II+I}{\rightsquigarrow}} \left(\begin{array}{ccc|c} 1 & 1 & -\frac{1}{2} & 6 \\ 0 & 6 & -\frac{3}{2} & 3 \\ 0 & -3 & 0 & -12 \end{array}\right)$$

$$\overset{II\updownarrow III}{\rightsquigarrow} \left(\begin{array}{ccc|c} 1 & 1 & -\frac{1}{2} & 6 \\ 0 & -3 & 0 & -12 \\ 0 & 6 & -\frac{3}{2} & 3 \end{array}\right) \overset{II\cdot(-\frac{1}{3})}{\rightsquigarrow} \left(\begin{array}{ccc|c} 1 & 1 & -\frac{1}{2} & 6 \\ 0 & 1 & 0 & 4 \\ 0 & 6 & -\frac{3}{2} & 3 \end{array}\right) \underset{III-6\cdot II}{\overset{I-II}{\rightsquigarrow}} \left(\begin{array}{ccc|c} 1 & 0 & -\frac{1}{2} & 2 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & -\frac{3}{2} & 21 \end{array}\right)$$

$$\overset{III\cdot(-\frac{2}{3})}{\rightsquigarrow} \left(\begin{array}{ccc|c} 1 & 0 & -\frac{1}{2} & 2 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 14 \end{array}\right) \overset{I+\frac{1}{2}III}{\rightsquigarrow} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 9 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 14 \end{array}\right)$$

Hence we can read off the solution as $A = 9$, $O = 4$ and $L = 14$.

Next we have system of linear equations in an extended matrix that is already in simple form. We differentiate two cases:

Example 2.30.

i) The extended matrix is of an homogeneous system: We have

$$\left(\begin{array}{cccccc|c} 1 & 0 & 4 & 5 & 0 & 7 & 0 \\ 0 & 1 & 2 & 8 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{array}\right).$$

Now we add one equation for all missing variables:

$$\left(\begin{array}{cccccc|c} 1 & 0 & 4 & 5 & 0 & 7 & 0 \\ 0 & 1 & 2 & 8 & 0 & 3 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \alpha_1 \\ 0 & 0 & 0 & 1 & 0 & 0 & \alpha_2 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \alpha_3 \end{array}\right) \text{ with } \alpha_1, \alpha_2, \alpha_3 \in \mathbb{R} \setminus \{0\}.$$

We derive the simple form again, which is now a diagonal matrix without the right-hand side.

$$\left(\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & 0 & 0 & -4\alpha_1 - 5\alpha_2 - 7\alpha_3 \\ 0 & 1 & 0 & 0 & 0 & 0 & -2\alpha_1 - 8\alpha_2 - 3\alpha_3 \\ 0 & 0 & 1 & 0 & 0 & 0 & \alpha_1 \\ 0 & 0 & 0 & 1 & 0 & 0 & \alpha_2 \\ 0 & 0 & 0 & 0 & 1 & 0 & -\alpha_3 \\ 0 & 0 & 0 & 0 & 0 & 1 & \alpha_3 \end{array}\right) \text{ with } \alpha_1, \alpha_2, \alpha_3 \in \mathbb{R} \setminus \{0\}.$$

The solution is 3-dimensional. So derive 3 linearly independent vectors from that system. The easiest way is with $\alpha_1 = 1, \alpha_2 = \alpha_3 = 0$ for the first vector, $\alpha_2 = 1, \alpha_1 = \alpha_3 = 0$ for the second vector and $\alpha_3 = 1, \alpha_1 = \alpha_2 = 0$ for the last vector. So we get for our solution:

$$S = Span(v_1, v_2.v_3) = Span\left(\begin{pmatrix} -4 \\ -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -5 \\ -8 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -7 \\ -3 \\ 0 \\ 0 \\ -1 \\ 1 \end{pmatrix}\right)$$

ii) The extended matrix is of an inhomogeneous system:

$$\left(\begin{array}{cccccc|c} 1 & 0 & 4 & 5 & 0 & 7 & 3 \\ 0 & 1 & 2 & 8 & 0 & 3 & 2 \\ 0 & 0 & 0 & 0 & 1 & 1 & 4 \end{array}\right).$$

First we derive a single special solution. Since we have an equation for x_1, x_2 and x_5 we set $x_3 = x_4 = x_6 = 0$. Then we get

$$v_s = \begin{pmatrix} 3 \\ 2 \\ 0 \\ 0 \\ 4 \\ 0 \end{pmatrix}$$

as one solution for our system. We still need to get all solutions. So as a second part we need to get the solution of the homogeneous system. Since we have the same matrix as in $i)$ we get the three vectors:

$$v_1 = \begin{pmatrix} -4 \\ -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, v_2 = \begin{pmatrix} -5 \\ -8 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, v_3 = \begin{pmatrix} -7 \\ -3 \\ 0 \\ 0 \\ -1 \\ 1 \end{pmatrix}$$

So the solution to our inhomogeneous system is:

$$S = v_s + Span(v_1, v_2, v_3) = \{v_s + v \mid v \in Span(v_1, v_2, v_3)\}$$
$$= \left\{ \begin{pmatrix} 3 \\ 2 \\ 0 \\ 0 \\ 4 \\ 0 \end{pmatrix} + v \;\middle|\; v \in Span\left(\begin{pmatrix} -4 \\ -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -5 \\ -8 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -7 \\ -3 \\ 0 \\ 0 \\ -1 \\ 1 \end{pmatrix} \right) \right\}$$

There is a quick way to get the set of all solutions, if the solution does not consist of just one element.

Theorem 2.31.
Let $Ax = b$ be a system of linear equations and $v_s \in \mathbb{R}^n$ is one solution, that is $Av_s = b$. If A is in simple form, then

$$v_j := e_j - \sum_{k=1}^{n} a_{kj} e_{jk},$$

for all j, that do not have a step in their column, form a basis for the set of solutions S_h of the homogeneous system $Ax = 0$. The e_j are the unit vectors of $\mathbb{R}^n$. Furthermore the set of solutions for $Ax = b$ is

$$S = \{v_s + v \mid v \in S_h\}.$$

$n - r$, with r is the number of steps in the simple form of A, is called the *defect* of the linear system $Ax = b$.

The set of all solutions might be a single vector or in other cases it might be some subset of vectors of the vector space V. In general this subset is not a subspace of V. It is rather a so called *affine subspace of V*.

Definition 2.32. (affine subspace)
Let $A \subseteq V$ be a subset of the vector space V. Then A is called an *affine subspace of* V if A is a translated subspace of V, that is if there is a subspace $U_A \subseteq V$ and a vector $v \in V$ such that

$$\begin{aligned} A =& v + U_A \\ :=& \{v + u \mid u \in U_A\}. \end{aligned}$$

U_A is called the *subspace associated to* A.

In general an affine subspace is just a subspace that is translated into some direction by some vector $v \in V$.

Example 2.33.

i) In $\mathbb{R}^2$ all lines and all points are affine subspaces of $\mathbb{R}^2$.

ii) In $\mathbb{R}^3$ all points, lines and planes are affine subspaces of $\mathbb{R}^3$.

The affine subspace is always parallel to its associated subspace, since the affine subspace is only a translated subspace. It is not rotated, reflected at some point or axis, etc..

Lemma 2.34.
All subspaces $A \subseteq V$ are also affine subspaces of V.

This is because any subspace of V may be written as $0_V + A = \{0_V + u \mid u \in A\}$.

Definition 2.35. (dimension of an affine subspace)
The *dimension of an affine subspace* $A \subseteq V$ is just the dimension of its associated subspace U_A.

2.3 The scalar product

We already introduced the vector space R^n. That is the so called *Euclidean Space*. In this chapter we will define some maps and structures for the Euclidean Space.

Definition 2.36. (n-dimensional Euclidean space)
$\mathbb{R}^n$ is called the n-dimensional Euclidean space.

The 2- and 3-dimensional space can be visualized with our cartesian coordinates. The 2-dimensional space is just a plain, whereas the 3-dimensional space may be visualized as our 3-dimensional world as we see and feel it. Then the vectors of the 2-dimensional and 3-dimensional Euclidean space are points on the plane or in the space. We want to define some basic functions in an Euclidean space.

Definition 2.37. (scalar product)
The function

$$\cdot \ : \ \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} : \ (u,v) \mapsto u \cdot v := u_1 \cdot v_1 + u_2 \cdot v_2 + ... + u_n \cdot v_n$$

is called the *(canonical) scalar product on* $\mathbb{R}^n$.

The scalar product is often referred to as $< \, , \, >$. That is for two vectors $u, v \in \mathbb{R}^n$ we have $< u, v > := u \cdot v$. The scalar product itself allows us to calculate the length of vectors and even angles between two vectors.

Definition 2.38. (lenght, distance, angles betweens vectors, orthogonal)
The *length* of a vector $v \in \mathbb{R}^n$ is

$$|v| := \sqrt{v \cdot v} = \sqrt{v_1^2 + v_2^2 + ... + v_n^2}.$$

The *distance* between two points $u, v \in \mathbb{R}^n$ is

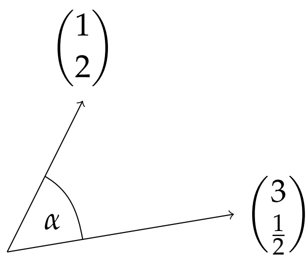
$$|u - v| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + ... + (u_n - v_n)^2}.$$

Let $u, v \in \mathbb{R}^n$ be two vectors. If the angle $\alpha := \measuredangle(u, v)$ between u and v is between 0 and π, then we have

$$\cos \alpha = \frac{u \cdot v}{|u| \cdot |v|}.$$

Two vectors $u, v \in V$ are called *orthogonal* if $u \cdot v = 0$ holds.

Example 2.39.
For two vectors $\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ \frac{1}{2} \end{pmatrix}$, the angle between them is given as $\alpha = \arccos \frac{1 \cdot 3 + 2 \cdot \frac{1}{2}}{\sqrt{1^2+2^2} + \sqrt{3^2 + \frac{1}{2}^2}} = \arccos \frac{4}{\sqrt{5} + \sqrt{\frac{37}{4}}} \approx 40,7^\circ$.



2.4 Linear functions

In linear algebra we work with linear functions and linear maps. Every map we have seen so far in this chapter, except the section with the scalar product, was a linear map. We define such a map as follows:

Definition 2.40. (linear map)
Let V, W be vector spaces. A map $f : V \to W$ is called *linear* if and only if the two following conditions hold:

i) For all $\lambda \in K, v \in V$ we have
$$f(\lambda \cdot v) = \lambda \cdot f(v).$$

ii) For all $v, w \in V$ we have
$$f(v + w) = f(v) + f(w).$$

A linear map is a map that satisfies some special properties. In particular the zero vector needs to map to zero. Therefore it is always an easy counterexample to show that the zero vector does not map to zero.

Example 2.41.
i)
$$f : \mathbb{R}^3 \to \mathbb{R}^3 : (x, y, z)^T \mapsto \begin{pmatrix} 2x + y \\ 2y + z \\ 2z \end{pmatrix}$$

f is a linear function since the conditions hold:
$$f(\lambda x, \lambda y, \lambda z) = \begin{pmatrix} 2\lambda x + \lambda y \\ 2\lambda y + \lambda z \\ 2\lambda z \end{pmatrix} = \lambda \cdot \begin{pmatrix} 2x + y \\ 2y + z \\ 2z \end{pmatrix} = \lambda f(x, y, z)$$

and
$$\begin{aligned} f(x + r, y + s, z + t) &= \begin{pmatrix} 2(x + r) + y + s \\ 2(y + s) + z + t \\ 2(z + t) \end{pmatrix} = \begin{pmatrix} 2x + y \\ 2y + z \\ 2z \end{pmatrix} + \begin{pmatrix} 2r + s \\ 2s + t \\ 2t \end{pmatrix} \\ &= f(x, y, z) + f(r, s, t). \end{aligned}$$

ii)

$$g : \mathbb{R}^2 \to \mathbb{R}^2 : (x,y)^T \mapsto \begin{pmatrix} y - x \\ x + 2 \end{pmatrix}$$

g is not a linear function because the first condition is not satisfied. In particular the zero vector does not map to the zero.

$$f(0 \cdot x, 0 \cdot y) = f(0,0) = \begin{pmatrix} 0 - 0 \\ 0 + 2 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0 \cdot f(x,y).$$

Theorem 2.42.
Let V, W be vector spaces. A map $f : V \to W$ is linear if and only if there exists a matrix A with $f(v) = A \cdot v$ for all $v \in V$.

Basicly we can rewrite each linear function into a matrix vector product and vice versa.

Example 2.43.
The matrix A for f in the previous example is

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

Definition 2.44. (identity matrix)
The matrix

$$I_n := \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} = (\delta_{ij})_{1 \leq i,j \leq n} \in \mathbb{F}^{n \times n}$$

with $\delta_{ij} = \begin{cases} 1 & , i = j, \\ 0 & , i \neq j \end{cases}$ is called the identity matrix.

The identity matrix is the neutral element of the matrix multiplication. Hence we have $A \cdot I_n = I_n \cdot A = A$ for all matrices $A \in \mathbb{F}^{n \times n}$.

Definition 2.45. (inverse matrix)
If for $A \in \mathbb{F}^{n \times n}$ exists a $B \in \mathbb{F}^{n \times n}$ with $A \cdot B = I_n$, then A is called *invertible*. Furthermore $B \cdot A = I_n$ holds and B is invertible, too. B is called the *inverse* of A denoted by A^{-1}.

An invertible matrix is often referred to as a *regular* matrix. A matrix, that is not regular is referred to as a *singular*.

Example 2.46.
Let $A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 1 & 1 & 1 \end{pmatrix}$ be a matrix. To get the inverse we need to get a matrix A^{-1} that holds

$$A \cdot A^{-1} = I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

This actually means, that we have 3 vectors in A^{-1} that yield their respected unit vectors of $\mathbb{R}^3$ if multiplied with A. So we use the Gaussian algorithm to solve a system of linear equations for the unit vectors.

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 2 & 1 & 3 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right) \rightsquigarrow \ldots \rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -2 & -1 & 5 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 & 1 & -3 \end{array}\right)$$

If the Gaussian algorithm yields a unique solution, then we have an inverse for A. In this case the inverse is

$$A^{-1} = \begin{pmatrix} -2 & -1 & 5 \\ 1 & 0 & -1 \\ 1 & 1 & -3 \end{pmatrix}$$

and we have

$$\begin{pmatrix} -2 & -1 & 5 \\ 1 & 0 & -1 \\ 1 & 1 & -3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} -2 & -1 & 5 \\ 1 & 0 & -1 \\ 1 & 1 & -3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The following lemma is already known for bijective functions.

Lemma 2.47.
Let $A, B \in \mathbb{F}^{n \times n}$ be two invertible matrices. Then $A \cdot B$ is invertible and we have

$$(A \cdot B)^{-1} = B^{-1} \cdot A^{-1}.$$

Proof:
The result follows directly from the associativity of matrix multiplication. We have

$$\begin{aligned} (A \cdot B) \cdot (B^{-1} \cdot A^{-1}) &= A \cdot B \cdot B^{-1} \cdot A^{-1} = A \cdot I_n \cdot A^{-1} \\ &= A \cdot A^{-1} = I_n. \end{aligned}$$

Hence $(A \cdot B)^{-1} = B^{-1} \cdot A^{-1}$ follows. □

We want to continue with some linear functions, that reflect or rotate points in the plane. Therefore we will observe the following functions in the 2-dimensional Euclidean space.

Definition 2.48. (rotation, reflection, scaling)
Let $v \in \mathbb{R}^2$ be a vector.
Then for $\alpha \in [0, 2\pi)$ we have the *rotation of v by an angle α* about the origin defined as

$$R_\alpha \cdot v := \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \cdot v.$$

The *reflection of v at the vertical axis* is defined as

$$S \cdot v := \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \cdot v.$$

The *scaling of v by a scalar $\lambda \in \mathbb{R}$* is defined as

$$D_\lambda \cdot v := \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \cdot v.$$

The three basic functions may also be combined to for example first rotate and then scale etc.. In higher dimension we can also have rotations and scalings. Furthermore we have rotations about some axis that may be somewhere in the space, but that do not need to be some axis of the coordinate system. We will not cover this in this lecture.

2.5 Determinants and characteristic space

We start this chapter with a function, that defined by some conditions, is fundamental for various concepts in linear algebra.

Definition 2.49. (determinant)
The function

$$det : \mathbb{F}^{n \times n} \to \mathbb{F}$$

is called a *determinant* if it satisfies the following conditions:

i) $det(I_n) = 1$.

ii) The determinant is *alternating* in the columns of the matrix, that is for $A = (a^1|a^2|...|a^n)$ with columns $a^1, a^2, ..., a^n \in \mathbb{F}^n$ we have: $det(A) = 0$ holds, if there are two indicies i, j with $1 \le i < j \le n$ such that $a^i = a^j$ holds.

iii) The determinant is *multilinear*, that is it is linear in each column of A: With $A = (a^1|...|a^i|...|a^n)$ and $a^i = \alpha\overline{a}^i + \beta\widetilde{a}^j$ we have

$$det(A) = \alpha \cdot det(a^1|...|\overline{a}^i|...|a^n) + \beta \cdot det(a^1|...|\widetilde{a}^i|...|a^n).$$

The determinant function is uniquely defined by those three conditions.

The concept of the determinant has its usefulness. With the determinant we are able to determine if a system of linear equations does have a unique solution or rather if a quadratic matrix is invertible.

Lemma 2.50.
Let $A \in \mathbb{R}^{n\times n}$ be a matrix. A is invertible if $det(A) \neq 0$.

Furthermore the determinant can be used to solve a system of linear equations. This algorithm is called *Cramer's rule*. Sadly we are not able to show Cramer's rule since our time in this lecture is limited. By that reason we will only show how to get the determinant of a matrix $A \in \mathbb{R}^{n\times n}$ with $n \leq 3$. So the case for $n = 1$ is the entry itself. For the case $n = 2$ we have

Lemma 2.51.
Let $A \in \mathbb{R}^{2\times 2}$ be a matrix. Then the determinant of A is given by

$$det(A) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11} \cdot a_{22} - a_{12} \cdot a_{21}.$$

Finally for $n = 3$ we have

Lemma 2.52. (Sarrus)
Let $A \in \mathbb{R}^{3\times 3}$ be a matrix. Then the determinant of A is given by

$$\begin{aligned} det(A) = &\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \\ =&a_{11} \cdot a_{22} \cdot a_{33} + a_{12} \cdot a_{23} \cdot a_{31} + a_{13} \cdot a_{21} \cdot a_{32} \\ &- a_{13} \cdot a_{22} \cdot a_{31} - a_{11} \cdot a_{23} \cdot a_{32} - a_{12} \cdot a_{21} \cdot a_{33}. \end{aligned}$$

For $n = 3$ the determinant is the result of the sum of the multiplied entries of each main diagonal and subtract the multiplied entries of all other diagonals. An example of calculating a determinant is on the next page.

For bigger matrices we could use the *Gaussian algorithm* or *Laplace's formula* to determine the determinant of a matrix A. But due to limited time we are not able to show it here.

We have seen, how to determine the determinant for small systems. For the case $n = 2$ we get a simple formula, to calculate the inverse of a function via the determinant.

Lemma 2.53.
Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2\times 2}$ be an invertible matrix. Then the inverse of A is given as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{det(A)} \cdot \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Now we are interested in a set of vectors, that are not generally transformed by any matrix but that are only scaled. We will call those vectors *eigenvectors*. They have an important role in mathematics. For example, if there is a given linear function, does this function have a set of vectors in their domain such that for each vector we get the same vector again? That is

$$A \cdot v = v.$$

More generally we can rephrase the question into: Is there a set of vectors that only map to a scalar multiple of the original vector? For some scalar $\lambda \in \mathbb{R}$ this might be noted as

$$A \cdot v = \lambda \cdot v.$$

So we define the characters of the previous equation.

Definition 2.54. (eigenvector, eigenvalue, characteristic space)
Let $A \in \mathbb{R}^{n\times n}$ be a matrix, $v \in \mathbb{R}^n \setminus \{0\}$ and $\lambda \in \mathbb{R}$. Then v is called an *eigenvector* of A with *eigenvalue* λ, if $A \cdot v = \lambda \cdot v$ holds. The *characteristic space* or *eigenspace* of some eigenvalue λ is the set of all eigenvectors of the eigenvalue λ.

An important function in the context of the characteristic space is the characteristic polynomial.

Definition 2.55. (characteristic polynomial)
The *characteristic polynomial* of a matrix $A \in \mathbb{R}^{n\times n}$ is a function $\mathbb{R} \to \mathbb{R} : \lambda \mapsto det(A - \lambda \cdot I_n)$ denoted by

$$\chi_A(\lambda) := det(A - \lambda \cdot I_n).$$

The characteristic polynomial yields the eigenvalues of the corresponding matrix.

Lemma 2.56.
The zeroes of the characteristic polynomial of A are the eigenvalues of A.

Example 2.57.
Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$ be a matrix. Then the characteristic polynomial is

$$\begin{aligned}
\chi_A(\lambda) =& det(A - \lambda \cdot I_3) = det(\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} - \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}) \\
=& det(\begin{pmatrix} 1-\lambda & 2 & 3 \\ 4 & 5-\lambda & 6 \\ 7 & 8 & 9-\lambda \end{pmatrix}) \\
=& (1-\lambda) \cdot (5-\lambda) \cdot (9-\lambda) + 2 \cdot 6 \cdot 7 + 3 \cdot 4 \cdot 8 \\
& - 3 \cdot (5-\lambda) \cdot 7 - (1-\lambda) \cdot 6 \cdot 8 - 2 \cdot 4 \cdot (9-\lambda) \\
& = -\lambda^3 + 15\lambda^2 + 18\lambda \\
& = -\lambda \cdot (\lambda^2 - 15\lambda - 18) \\
& = -\lambda \cdot (\lambda - \frac{15 + 3\sqrt{33}}{2}) \cdot (\lambda - \frac{15 - 3\sqrt{33}}{2}).
\end{aligned}$$

Hence the eigenvalues of A are $\lambda_1 = 0,\ \lambda_2 = \frac{15+3\sqrt{33}}{2},\ \lambda_3 = \frac{15-3\sqrt{33}}{2}$.

With the help of the characteristic polynomial we are able to determine the eigenvalues and eigenvectors of A. So to get the eigenvalues and eigenvectors of some matrix A, we proceed with the following steps:

1. Determine the characteristic polynomial $\chi_A(\lambda)$ of A

2. Find the zeroes of $\chi_A(\lambda)$, which are the eigenvalues $\lambda_1, ..., \lambda_s$ of A

3. For each eigenvalue λ_i, $i \in \{1, ..., s\}$ the characteristic space of the eigenvalue λ_i is the set of solutions to the homogeneous system

$$(A - \lambda_i \cdot I_n) \cdot v_i = 0.$$

Example 2.58.
Let $A = \begin{pmatrix} 1 & 3 & 0 \\ 0 & 0 & -2 \\ 0 & -2 & 0 \end{pmatrix}$ be a matrix. Determine the eigenvalues and their respected

characteristic spaces. So first we need to calculate the characteristic polynomial:

$$\begin{aligned}\chi_A(\lambda) &= det(\begin{pmatrix}1-\lambda & 3 & 0\\ 0 & -\lambda & -2\\ 0 & -2 & -\lambda\end{pmatrix})\\ &= (1-\lambda)\cdot(-\lambda)^2-(-2)^2\cdot(1-\lambda)\\ &= (1-\lambda)\cdot(\lambda^2-4)\\ &= (1-\lambda)\cdot(\lambda-2)\cdot(\lambda+2)\end{aligned}$$

So the eigenvalues of A are $\lambda_1 = 1$, $\lambda_2 = 2$ and $\lambda_3 = -2$. Now we can determine the characteristic spaces. For $\lambda_1 = 1$ we have

$$\left(\begin{array}{ccc|c}0 & 3 & 0 & 0\\ 0 & -1 & -2 & 0\\ 0 & -2 & -1 & 0\end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c}0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0\end{array}\right) \Rightarrow v_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}$$

For $\lambda_2 = 2$ we get

$$\left(\begin{array}{ccc|c}-1 & 3 & 0 & 0\\ 0 & -2 & -2 & 0\\ 0 & -2 & -2 & 0\end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c}1 & -3 & 0 & 0\\ 0 & 1 & 1 & 0\\ 0 & 0 & 0 & 0\end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c}1 & 0 & 3 & 0\\ 0 & 1 & 1 & 0\\ 0 & 0 & 0 & 0\end{array}\right) \Rightarrow v_2 = \begin{pmatrix}-3\\-1\\1\end{pmatrix}$$

For $\lambda_3 = -2$ we have

$$\left(\begin{array}{ccc|c}3 & 3 & 0 & 0\\ 0 & 2 & -2 & 0\\ 0 & -2 & 2 & 0\end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c}1 & 1 & 0 & 0\\ 0 & 1 & -1 & 0\\ 0 & 0 & 0 & 0\end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c}1 & 0 & 1 & 0\\ 0 & 1 & -1 & 0\\ 0 & 0 & 0 & 0\end{array}\right) \Rightarrow v_3 = \begin{pmatrix}-1\\1\\1\end{pmatrix}$$

So $\lambda_1 = 1$ has an eigenvector $v_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}$, $\lambda_2 = 2$ has an eigenvector $v_2 = \begin{pmatrix}-3\\-1\\1\end{pmatrix}$ and $\lambda_3 = -2$ has an eigenvector $v_3 = \begin{pmatrix}-1\\1\\1\end{pmatrix}$. The characteristic spaces are generated by the span of their respected eigenvectors.

Definition 2.59. (algebraic and geometric multiplicity)
The *(algebraic) multiplicity* of some eigenvalue λ_i is the multiplicity of the eigenvalue, that is the number of linear factors $(\lambda - \lambda_i)$ in the characteristic polynomial. The *geometric multiplicity* of an eigenvalue λ is the dimension of the characteristic space of λ.

A characteristic space may have a geometric multiplicity greater than 1 but it is always at most the algebraic multiplicity.

Example 2.60.
Let $A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ be the zero matrix. Then A has only the eigenvalue $\lambda_1 = \lambda_2 = \lambda_3 = 0$ with a multiplicity of 3. The eigenspace of $\lambda_1 = 0$ is $\mathbb{R}^3$. Therefore the geometric multiplicity of $\lambda_1 = 0$ is 3.

3 Limits

3.1 Sequences

Definition 3.1. (sequence)
A function $f : \mathbb{N} \to \mathbb{R}$ is called a *sequence*. We denote the n-th element with a_n rather than with $f(n)$, thus the sequence is

$$(a_1, a_2, a_3, ...) \text{ or } (a_n)_{n\in\mathbb{N}}.$$

Also the Domain of a sequence might be $\mathbb{N} \cup \{0\}$.

Remark 3.2.
A sequence can be defined explicitly for each element or *recursively*, that is each element a_n is defined by some previous elements.

Example 3.3.
Some examples of a sequence, that is given explicitly are:

i) $(2)_{n\in\mathbb{N}}$ a *constant sequence*

ii) $(a_0 + n \cdot d)_{n\in\mathbb{N}}$ an *arithmetic sequence* with *commomn difference d*

iii) $(a_0 \cdot q^n)_{n\in\mathbb{N}}$ an *geometric sequence* with *common ratio q*

iv) $((1 + \frac{1}{n})^n)_{n\in\mathbb{N}}$

v) $((-1)^n)_{n\in\mathbb{N}}$

vi) $(\frac{1}{n})_{n\in\mathbb{N}}$

Examples of sequences, that are stated recursively are

i) $a_0 = 1$, $a_{n+1} = \sqrt{1 + a_n}$

ii) $a_0 = 0$, $a_1 = 1$, $a_{n+2} = a_{n+1} + a_n$ the *Fibonacci-sequence*

As seen in the last example, a recursively defined sequence may be dependet on more than one previous element.

Definition 3.4. (monotonicially increasing/decreasing, bounded)
A sequence $(a_n)_n$ is called *monotonicially increasing* (*monotonicially decreasing*) if $a_n \leq a_{n+1}$ ($a_n \geq a_{n+1}$) holds for all n. A sequence $(a_n)_n$ is called *bounded*, if there exists some $M \in \mathbb{R}$ such that $|a_n| \leq M$ for all n holds.

Definition 3.5. (convergent with limit, divergent)
A sequence $(a_n)_n$ is called *convergent with limit a*, if for each $\varepsilon > 0$ there is some $n_0 = n_0(\varepsilon) \in \mathbb{N}$ such that

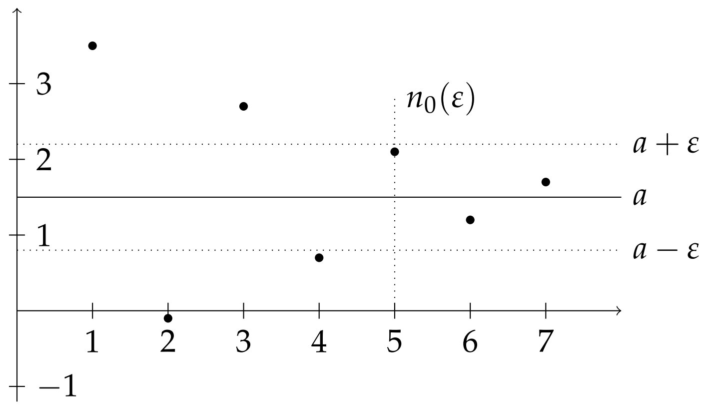
$$|a_n - a| < \varepsilon, \ \forall n \geq n_0$$

holds. Then the limit of the sequence is denoted by

$$\lim_{n\to\infty} a_n = a \text{ or } a_n \overset{n\to\infty}{\longrightarrow} a.$$

If no such limit a exists, then $(a_n)_n$ is called *divergent*.

Remark 3.6.
If a sequence $(a_n)_n$ is divergent, we denote its limit by $\lim\limits_{n\to\infty} = +\infty$ ($\lim\limits_{n\to\infty} = +\infty$) if for any $M \in \mathbb{R}$ there is some $n_0 = n_0(M) \in \mathbb{N}$ such that $a_n \geq M$ ($a_n \leq M$) holds for all $n \geq n_0$.



Remark 3.7.
If a sequence is convergent with limit a, then this limit is unique.

Remark 3.8.
A sequence that converges is bounded.

Example 3.9.
We want to show, that $(a_n)_{n\in\mathbb{N}} = (\frac{1}{n})_{n\in\mathbb{N}}$ converges with limit 0. By the definiton we need to show that for all $\varepsilon > 0$ exists some $n_0 = n_0(\varepsilon) \in \mathbb{N}$ such that

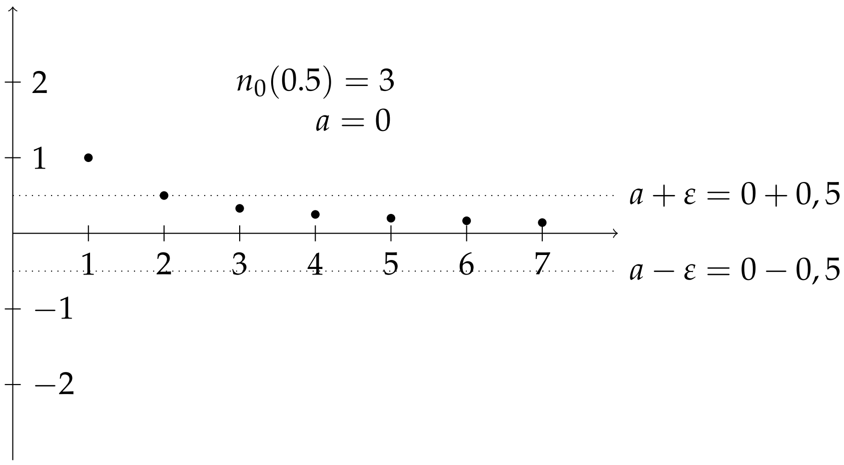
$$\begin{aligned} \left|\frac{1}{n} - 0\right| &= \frac{1}{n} < \varepsilon \\ \Leftrightarrow \qquad & \frac{1}{\varepsilon} < n \end{aligned}$$

for all $n \geq n_0$. So for small ε we get big $\frac{1}{\varepsilon}$ and therefore big a n_0. Since $\mathbb{N}$ is not bounded by any number we may choose any $n_0 > \frac{1}{\varepsilon}$. So we get for all $\varepsilon > 0$, that there exists an $n_0 := n_0(\varepsilon) > \frac{1}{\varepsilon}$ such that

$$\left|\frac{1}{n} - 0\right| < \varepsilon$$

holds for all $n \geq n_0$. Thus

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



Theorem 3.10.
Any sequence that is both bounded and monotonic is convergent.

Example 3.11.
With the last theorem we want to show that $(a_n)_{n\in\mathbb{N}} := (\frac{1}{n})_{n\in\mathbb{N}}$ converges. So we need to show that it is both bounded and monotonic. Furthermore we guess, that $\lim\limits_{n\to\infty} \frac{1}{n} = 0$. First we show that $(\frac{1}{n})_{n\in\mathbb{N}}$ is monotonicially decreasing. We have for any $n \in \mathbb{N}$

$$a_n = \frac{1}{n} > \frac{1}{n+1} = a_{n+1}.$$

So $(a_n)_{n\in\mathbb{N}}$ is monotonicially decreasing. Second, the sequence $(a_n)_{n\in\mathbb{N}}$ is bounded by $M = 1$, since $\frac{1}{n} \geq 0$ holds for all $n \in \mathbb{N}$. Furthermore $a_1 = 1$ and a_n is monotonicially decreasing, so we have $a_n \leq 1$. Then we deduce

$$-M = -1 \leq 0 \leq \frac{1}{n} \leq 1 = M.$$

We have shown, that $(\frac{1}{n})_{n\in\mathbb{N}}$ is both bounded and monotonic. So it converges.

In the previous example it was sufficent to show, that $(\frac{1}{n})_{n\in\mathbb{N}}$ has a lower bound and is monotonicially decreasing.

Theorem 3.12.
Let $(a_n)_n$ and $(b_n)_n$ be convergent sequences with limits a and b respectively. Let α and β be real or complex numbers. Then we have

i) $\lim\limits_{n\to\infty}(\alpha \cdot a_n + \beta \cdot b_n) = \alpha \cdot a + \beta \cdot b$,

ii) $\lim\limits_{n\to\infty} a_n \cdot b_n = a \cdot b$,

iii) $\lim\limits_{n\to\infty} \frac{a_n}{b_n} = \frac{a}{b}$, if $b_n \neq 0$ for all n and $b \neq 0$.

Theorem 3.13. (Sandwich's lemma)
Let $(a_n)_n, (b_n)_n, (c_n)_n$ be sequences with

$$a_n \le b_n \le c_n, \ \forall n \in \mathbb{N}$$

and

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

Then we have

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

We add a list of some convergent sequences including their respective limits:

i) $\lim\limits_{n\to\infty} \frac{1}{n} = 0$

ii) $\lim\limits_{n\to\infty} \frac{1}{\sqrt{n}} = 0$

iii) $\lim\limits_{n\to\infty} \frac{1}{n^k} = 0$ for all $k \in \mathbb{N}$

iv) $\lim\limits_{n\to\infty} q^{\frac{1}{n}} = 1$ for all $q > 0$

v) $\lim\limits_{n\to\infty} \sqrt[n]{n} = 1$

vi) $\lim\limits_{n\to\infty} q^n = \begin{cases} 0 & , -1 < q < 1 \\ 1 & , q = 1 \\ \infty & , q > 1 \end{cases}$

Theorem 3.14. (Euler's number)
The sequence

$$((1+\frac{1}{n})^n)_{n\in\mathbb{N}}$$

is monotonicially increasing and satisfies

$$2 \leq (1+\frac{1}{n})^n \leq 3$$

for all $n \in \mathbb{N}$. Thus it converges with

$$\lim_{n\to\infty}(1+\frac{1}{n})^n =: e.$$

e is called *Euler's* number. ($e \approx 2.71828$)

Proof:
Let $a_n := (1+\frac{1}{n})^n$. Then

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

We have

$$\begin{aligned}
(n+1)^2 &\geq 1 \\
\Leftrightarrow \qquad 1 &\geq \frac{1}{(n+1)^2} \\
\Leftrightarrow a := -\frac{1}{(n+1)^2} &\geq -1.
\end{aligned}$$

Hence we may employ Bernoulli's inequality with $(a+1)^{n+1} \geq (n+1)\cdot a + 1$:

$$\begin{aligned}
\frac{a_{n+1}}{a_n} &\geq \frac{n+1}{n}\left((n+1)\frac{-1}{(n+1)^2}+1\right) = \frac{n+1}{n}(1-\frac{1}{n+1}) \\
&= \frac{n+1}{n}\cdot\frac{n}{n+1} = 1 \\
\Leftrightarrow a_{n+1} &\geq a_n.
\end{aligned}$$

Thus $(a_n)_n$ is monotonicially increasing. Since $a_1 = 2$ holds, we have $2 \leq a_n$ for all n. Now we will prove $a_n \leq 3$ for all n:

The Binomial Theorem yields:

$$
\begin{aligned}
a_n &= \left(1+\frac{1}{n}\right)^n = \sum_{k=0}^{n}\binom{n}{k}\frac{1}{n^k} = 1+1+\sum_{k=2}^{n}\binom{n}{k}\frac{1}{n^k} \\
&\leq 2+\sum_{k=2}^{n}\frac{1}{2^{k-1}}
\end{aligned}
$$

with

$$
\binom{n}{k}\frac{1}{n^k} = \frac{n\cdot(n+1)\cdot\ldots\cdot(n-k+1)}{k!\cdot n^k} \leq \frac{1}{k!} \leq \frac{1}{2^{k-1}}.
$$

The Geometric Sum Formula $\sum_{i=0}^{n} q^i = \frac{q^{n+1}-1}{q-1}$ yields then:

$$
\begin{aligned}
a_n &\leq 2+\sum_{k=2}^{n}\frac{1}{2^{k-1}} = 2+\sum_{k=1}^{n-1}\frac{1}{2^k} = 1+\sum_{k=0}^{n-1}\frac{1}{2^k} \\
&= 1+\frac{\frac{1}{2^n}-1}{\frac{1}{2}-1} = 1+\frac{1-(\frac{1}{2})^2}{1-\frac{1}{2}} \\
&= 1+2\left(1-\left(\frac{1}{2}\right)^n\right) \leq 1+2 = 3.
\end{aligned}
$$

□

Example 3.15.
We will show an example of a rekursive defined sequence. Let $a_1 := 1$ and $a_{n+1} := \frac{1}{4}a_n^2 + 1$. We will use the theorem about bounded and monotonic squences. Besides those two conditions we need to calculate the limit.

i) If there exists a limit, that is $a = \lim\limits_{n\to\infty}$, then we have:

$$
\begin{aligned}
& & a_{n+1} &= \frac{1}{4}a_n^2+1 \\
&\overset{n\to\infty}{\longrightarrow} & a &= \frac{1}{4}a^2+1 \\
&\Leftrightarrow & 0 &= \frac{1}{4}a^2-a+1 = \left(\frac{1}{2}a-1\right)^2 \\
&\Leftrightarrow & a &= 2.
\end{aligned}
$$

ii) Now we show that $(a_n)_n$ is monotonic.

$$
a_{n+1}-a_n = \frac{1}{4}a_n^2+1-a_n = \left(\frac{1}{2}a_n-1\right)^2 \geq 0,\ \forall n\in\mathbb{N}.
$$

Therefore $(a_n)_n$ is monotonic.

iii) It remains to be shown, that $(a_n)_n$ $(S(n))$ is bounded. We will show by induction, that there is an upper bound with $a_n \leq 2$ for all $n \in \mathbb{N}$. *base case:*

$$S(1) \qquad a_1 = 1 \leq 2.$$

inductive step: We assume that $S(n)$ is true for some $n \in \mathbb{N}$.

$$a_{n+1} = \frac{1}{4}a_n^2 + 1 \overset{I.H.}{\leq} \frac{1}{4}2^2 + 1 = 2.$$

Thus by proof by induction, $(a_n)_n$ has an upper bound.

By the theorem for monotonic and bounded sequences, we have that $(a_n)_n$ is convergent with

$$\lim_{n\to\infty} a_n = 2.$$

Remark 3.16.
The considerations done about sequences of single numbers can be extended to points $x \in \mathbb{R}^n$. The limit, convergence and divergence will be defined identically. The theorem about adding and scalar multiplication of limits still holds then.

3.2 Series

In this chapter we will consider infinite sums, like $0+1+2+3+...$ and others. Those will be called series.

Definition 3.17. (series)
Let $(a_k)_{k\in\mathbb{N}}$ be a sequence. Then for all $n \in \mathbb{N}$

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

is called the *n-th partial sum*. The sequence $(s_n)_{n\in\mathbb{N}}$ is called a *series*. It is denoted by

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

A series *converges* if the sequence of its partial sums, that is $(s_n)_{n\in\mathbb{N}}$ converges. In this case we write

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

for the limit. If a series does not converge, then it is called *divergent*. A series *converges absolutely*, if $\sum_{k=1}^{\infty} |a_k|$ converges.

Lemma 3.18.
If $\sum_{k=1}^{\infty} a_k$ converges absolutely, then it also converges.

Theorem 3.19. (necessary condition)
If $\sum_{k=1}^{\infty} a_k$ converges, then $\lim\limits_{k\to\infty} a_k = 0$ holds.

Example 3.20.

- **Geometric Series**
$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}, \text{ for } |q| < 1.$$
We have with the Geometric Sum Formula
$$\sum_{k=0}^{\infty} q^k = \lim_{n\to\infty} \sum_{k=0}^{n} q^k \overset{q\neq 1}{=} \lim_{n\to\infty} \frac{q^{n+1}-1}{q-1} = \lim_{n\to\infty} \frac{1-q^{n+1}}{1-q} \overset{|q|<1}{=} \frac{1}{1-q}.$$

- $$\sum_{k=1}^{\infty} \frac{1}{k(k+1)} = 1.$$
We have
$$\begin{aligned}\sum_{k=1}^{\infty} \frac{1}{k(k+1)} &= \lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k(k+1)} = \lim_{n\to\infty} \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k+1}\right) \\ &= \lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=1}^{n} \frac{1}{k+1} = \lim_{n\to\infty} \left(\frac{1}{1} - \frac{1}{n+1}\right) = 1.\end{aligned}$$
The sum is a so called *telescoping series*. The proof is a method of differences. In the series only the very first and the very last term will remain. All other summands are erased after cancellation.

- **Harmonic Series**
$$\sum_{k=1}^{\infty} \frac{1}{k} = +\infty$$
The Harmonic Series diverges because
$$\begin{aligned}\sum_{k=1}^{\infty} \frac{1}{k} &= \lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k} \\ &= \lim_{n\to\infty} \frac{1}{1} + \frac{1}{2} + \underbrace{\frac{1}{3} + \frac{1}{4}}_{\geq 2\cdot\frac{1}{4}=\frac{1}{2}} + \underbrace{\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}}_{\geq 4\cdot\frac{1}{8}=\frac{1}{2}} + \ldots + \frac{1}{n} \\ &= \infty\end{aligned}$$

The Harmonic Series is an example, that the necessary condition is not sufficient for a series to converge. The necessary condition is satisfied with $\lim_{k\to\infty} a_k = \lim_{k\to\infty} \frac{1}{k} = 0$, yet the series diverges.

Next we state the linearity of series.

Theorem 3.21.
Let $\sum_{k=1}^{\infty} a_k = a$ and $\sum_{k=1}^{\infty} b_k = b$ be two convergent series and $\alpha, \beta \in \mathbb{R}$. Then

$$\sum_{k=1}^{\infty} (\alpha a_k + \beta b_k) = \alpha a + \beta b$$

holds.

Example 3.22.
Consider the series $\sum_{k=1}^{\infty} \frac{7}{10^k}$. Then we have

$$\sum_{k=1}^{\infty} \frac{7}{10^k} = 7 \cdot \sum_{k=1}^{\infty} \frac{1}{10^k} = 7 \cdot \left(-1 + \sum_{k=0}^{\infty} \frac{1}{10^k} \right) \overset{geom.series}{=} 7 \left(-1 + \frac{1}{1 - \frac{1}{10}} \right) = ... = \frac{1}{7}.$$

Theorem 3.23. (Cauchy Product of Series)
Let $\sum_{k=0}^{\infty} a_k = a$ and $\sum_{k=0}^{\infty} b_k = b$ be two absolutely convergent series. Then the *Cauchy product* converges with

$$\left(\sum_{k=0}^{\infty} a_k \right) \cdot \left(\sum_{k=0}^{\infty} b_k \right) := \sum_{k=0}^{\infty} \left(\sum_{j=0}^{k} a_j b_{k-j} \right) = a \cdot b.$$

Now we state a couple of sufficient conditions for series to be convergent:

Theorem 3.24. (comparison test)
Let $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=1}^{\infty} b_k$ be two series with $b_k \geq 0$ for all k.

i) If there is some $k_0 \in \mathbb{N}$ such that $|a_k| \leq b_k$ for all $k \geq k_0$ and $\sum_{k=1}^{\infty} b_k$ converges, then $\sum_{k=1}^{\infty} a_k$ converges absolutely.

ii) If there is some $k_0 \in \mathbb{N}$ such that $|a_k| \geq b_k$ for all $k \geq k_0$ and $\sum_{k=1}^{\infty} b_k$ diverges, then $\sum_{k=1}^{\infty} a_k$ diverges as well.

Example 3.25.
Show that $\sum_{k=1}^{\infty} \frac{1}{k^2}$ converges. We have

$$\frac{1}{k^2} \leq \frac{2}{k(k+1)}$$

and

$$\sum_{k=1}^{\infty} \frac{2}{k(k+1)} = 2\underbrace{\sum_{k=1}^{\infty} \frac{1}{k(k+1)}}_{=1} = 2$$

converges. Hence by the comparison test $\sum_{k=0}^{\infty} \frac{1}{k^2}$ converges as well.

Since we have shown that the harmonic series diverges but $\sum_{k=1}^{\infty} \frac{1}{k^2}$ converges, we want to state the following lemma

Lemma 3.26.
The series

$$\sum_{k=1}^{\infty} \frac{1}{k^{\alpha}}$$

converges for $\alpha > 1$, but diverges for $0 < \alpha \leq 1$.

Theorem 3.27. (ratio test)
Let $\sum_{k=1}^{\infty} a_k$ be a series with $a_k \neq 0$ for all k. Furthermore assume that the limit

$$a := \lim_{k\to\infty} \left| \frac{a_{k+1}}{a_k} \right|$$

exists. Then $\sum_{k=1}^{\infty} a_k$ converges absolutely for $a < 1$, but diverges for $a > 1$. There is no statement for $a = 1$.

Example 3.28.
Show that $\sum_{k=1}^{\infty} \frac{k^2}{2^k}$ converges. We have

$$\left| \frac{a_{k+1}}{a_k} \right| = \left| \frac{\frac{(k+1)^2}{2^{k+1}}}{\frac{k^2}{2^k}} \right| = \left| \frac{(k+1)^2}{k^2} \cdot \frac{2^k}{2^{k+1}} \right| = \left(\frac{k+1}{k} \right)^2 \cdot \frac{1}{2} = \frac{1}{2} \left(1 + \frac{1}{k} \right)^2 \xrightarrow{k\to\infty} \frac{1}{2} < 1.$$

Therefore by the ratio test, the series $\sum_{k=1}^{\infty} \frac{k^2}{2^k}$ converges.

Theorem 3.29. (root test)
Let $\sum_{k=1}^{\infty} a_k$ be a series. If the limit

$$a := \lim_{k\to\infty} \sqrt[k]{|a_k|}$$

exists, then $\sum_{k=1}^{\infty} a_k$ converges for $a < 1$, but diverges for $a > 1$. There is no statement for $a = 1$.

Example 3.30.
Show that $\sum_{k=1}^{\infty} \frac{k^2+1}{3^k}$ converges. We have

$$\lim_{k\to\infty} \sqrt[k]{|a_k|} = \lim_{k\to\infty} \sqrt[k]{\frac{k^2+1}{3^k}} = \lim_{k\to\infty} \frac{\sqrt[k]{k^2+1}}{3} = \frac{1}{3} < 1.$$

Therefore the series $\sum_{k=1}^{\infty} \frac{k^2+1}{3^k}$ converges by the root test.

The last example could been shown with the ratio test, too.

Theorem 3.31. (alternating series test, Leibniz's test)
Let $\sum_{k=1}^{\infty}(-1)^k a_k$ be a series with $a_k \geq 0$ for all k. Furthermore $(a_k)_k$ is monotonicially decreasing with $\lim\limits_{k\to\infty} a_k = 0$. Then $\sum_{k=1}^{\infty}(-1)^k a_k$ converges.

We have seen that the harmonic series diverges. But alternating the summands of the harmonic series creates a convergent series.

Example 3.32.
Show that $\sum_{k=1}^{\infty}(-1)^k \frac{1}{k}$ converges. We have that $\frac{1}{k} \geq 0$ holds for all k. Furthermore $\left(\frac{1}{k}\right)_k$ is monotonicially decreasing with limit $\lim\limits_{k\to\infty} a_k = 0$, as we have shown before. Therefore by the alternating series test the series $\sum_{k=1}^{\infty}(-1)^k \frac{1}{k}$ converges.

3.3 Continuous Functions

We can derive the limit of functions from our previous chapters about limits of sequences.

Definition 3.33. (limit)
Let $D \subseteq \mathbb{R}^n$ be a subset, $f : D \to \mathbb{R}$ a function and $x_0 \in \mathbb{R}^n$ such that there is a sequence $(y_m)_m$ with $y_m \in D \setminus \{x_0\}$ for all m and $\lim\limits_{m\to\infty} y_m = x_0$. If for all such sequences $(y_m)_m$ the limit

$$\lim_{m\to\infty} f(y_m)$$

exists and is the same, then we write

$$\lim_{x\to x_0} f(x) = a \text{ or } f(x) \overset{x\to x_0}{\longrightarrow} a.$$

We call a the *limit of f in x_0*.
If D contains an interval $(-\infty, s)$ $((s,\infty))$ for some $s \in \mathbb{R}$, then the limit

$$\lim_{x\to -\infty} f(x) \qquad \left(\lim_{x\to\infty} f(x)\right)$$

is defined analoguesly.

Yet there is another defintion of the limit, that works without the sturcture of sequences. The two definitions are the same.

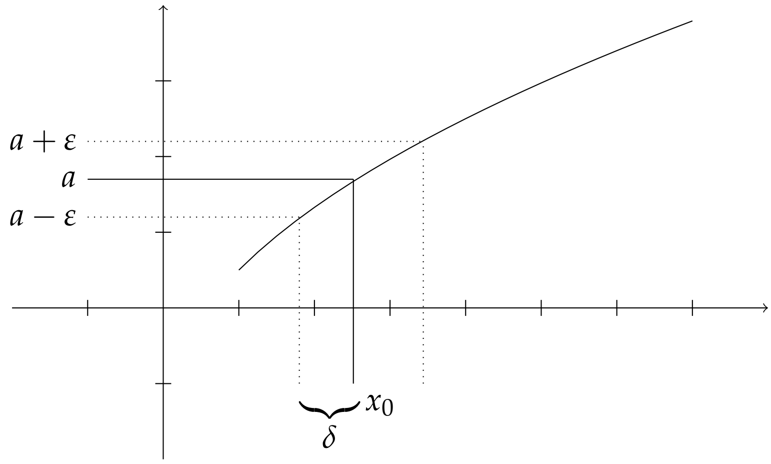
Definition 3.34. (limit)
Let $f : I \to \mathbb{R}$ be a function on the interval $I \subseteq \mathbb{R}$ and $x_0 \in I$ or x_0 is a boundary point of I. Then there exists the limit

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

if the following holds:
For all $\varepsilon > 0$ exists a $\delta = \delta(\varepsilon) > 0$, such that

$$|f(x) - a| < \varepsilon$$

for all $x \in I$ with $|x - x_0| < \delta,\ x \neq x_0$.



From the limits of sequences we can derive the following rules for the limits of functions.

Theorem 3.35.
Let $D \subseteq \mathbb{R}^n$ be a subset, $f, g : D \to \mathbb{R}$ functions and $x_0 \in \mathbb{R}^n$ a point such that limits $\lim\limits_{x \to x_0} f(x) = a$ and $\lim\limits_{x \to x_0} g(x) = b$ exist. Then we have

i) $\lim\limits_{x \to x_0} (f(x) + g(x)) = a + b$

ii) $\lim\limits_{x \to x_0} (f(x) \cdot g(x)) = a \cdot b$

iii) $\lim\limits_{x \to x_0} \frac{f(x)}{g(x)} = \frac{a}{b}$, if $b \neq 0$.

Now we want to use the limit about functions to define continuity:

Definition 3.36. (continuous)
Let $D \subseteq \mathbb{R}^n$ be a subset, $f : D \to \mathbb{R}$ a function and $x_0 \in D$ such there is a sequence $(y_m)_m$ with $y_m \in D \setminus \{x_0\}$ for all m and $\lim\limits_{m\to\infty} y_m = x_0$. If

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

that is, the limit exists and equals the value of x_0, then f is called *continuous at* x_0. If f is continuous at all $x_0 \in D$, then f is called *continuous*.

We may define analoguesly the continuous functions without the sequences but just with the limits of functions via the ε-δ-Definition.

Definition 3.37. (continuous)
Let $f : I \to \mathbb{R}$ a function with an interval I and $x_0 \in I$.

i) f is called *continuous in* x_0, if

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

holds, that is for all $\varepsilon > 0$ exists a $\delta > 0$ with: $|f(x) - f(x_0)| < \varepsilon$, if $|x - x_0| < \delta$.

ii) f is called *continuous on* I, if f is continuous in all points $x_0 \in I$.

Example 3.38.
Here we give some functions, that are continuous or not.

i) Polynomials are continuous.

ii) The function

$$f : \mathbb{R} \to \mathbb{R} : x \mapsto \begin{cases} 0 & , x < 0 \\ 1 & , x \geq 0 \end{cases}$$

is not continuous in 0 because the limit $\lim\limits_{x\to 0}$ does not exist. The one sided limits exists, with $\lim\limits_{x\searrow 0} f(x) = 1$ and $\lim\limits_{x\nearrow 0} f(x) = 0$, but they do not match.

iii) The vector length

$$|\cdot| : \mathbb{R}^n \to \mathbb{R} : v \mapsto |v|$$

is a continuous function.

iv) The *Dirichlet function*

$$f : \mathbb{R} \to \mathbb{R} : x \mapsto \begin{cases} 0 & , x \text{ irrational} \\ 1 & , x \text{ rational} \end{cases}$$

is not continuous in any point, since each interval of positive length contains both irrational and rational numbers.

Definition 3.39. (continuously extended)
Let $f : I \to \mathbb{R}$ be a function over an interval I. If f is continuous on $I \setminus \{x_0\}$ and $\lim_{x \to x_0} f(x) = y$ exists, then f may be *continuously extended in* x_0 with $f(x_0) = y$.

Example 3.40.
Let $f : \mathbb{R} \setminus \{-2\} \to \mathbb{R} : x \mapsto \frac{x^2-4}{x+2}$ be a function. Then f is continuous as a composition of continuous functions (polynomials are continuous). Furthermore f can be continuously extended in -2 with

$$\lim_{x \to -2} \frac{x^2 - 4}{x + 2} = \lim_{x \to -2} x - 2 = -4.$$

Then we have

$$\widehat{f} : \mathbb{R} \to \mathbb{R} : x \mapsto \begin{cases} \frac{x^2-4}{x+2}, & , x \in \mathbb{R} \setminus \{-2\} \\ -4 & , x = -2 \end{cases}$$

as the continuous extension of f.

There are some basic rules for continuous functions.
Theorem 3.41.
Let $I \subseteq \mathbb{R}^n$, $x_0 \in I$, $I_1, I_2 \subseteq \mathbb{R}$ and $\widetilde{x}_0 \in I_1$.

i) If $f, g : I \to \mathbb{R}$ are continuous at x_0, then so are

$$f \pm g,\ f \cdot g \text{ and } \frac{f}{g}$$

with $0 \notin g(I)$ for the last.

ii) If $f : I_1 \to I_2$ is continuous in $\widetilde{x}_0$ and $g : I_2 \to \mathbb{R}$ is continuous in $f(\widetilde{x}_0)$ then $g \circ f : I_1 \to \mathbb{R}$ is also continuous in $\widetilde{x}_0$.

The following theorem gives for each continuous function f defined on an interval, that f will map to all points between the image of two different points.

Theorem 3.42. (Intermediate Value Theorem)
Let $a < b$ be real numbers and $f : [a, b] \to \mathbb{R}$ a continuous function. Then f takes all values in $[f(a), f(b)]$.

The intermediate value theorem yields the existence of zeroes for continuous functions if $f(a) \leq 0 \leq f(b)$.

Corollary 3.43.
Let $a < b$ be real numbers and $f : [a, b] \to \mathbb{R}$ a continuous function with $f(a) \leq 0 \leq f(b)$. Then f has at least one zero in $[a, b]$, that is there exists a $x \in [a, b]$ such that $f(x) = 0$.

Theorem 3.44. (Theorem about Maximum and Minimum)
Let $I = [a, b] \subset \mathbb{R}$ be an interval with $a < b$ and $f : I \to \mathbb{R}$ continuous. Then

$$f(I) = f([a, b]) =: [c, d]$$

is an interval for some $c, d \in \mathbb{R}$. c is the *minimum value* and d is the *maximum value* that f assumes on $I = [a, b]$.

We may rephrase the previous theorem in the well known theorem of Weierstraß.

Corollary 3.45. (Theorem of Weierstraß)
Each continuous function on a compact space has a maximum and minimum value.

A *compact space* is a space, that is both closed (bounds are part of the space) and bounded.

Remark 3.46.
It is crucial, that in the previous theorem both bounds are contained in the interval. If they are not, the maximum and minimum value does not need to exists. For example $f : (0, 1) \to \mathbb{R} : x \mapsto x$ has neither a maximum nor a minimum.

3.4 Power Series

In this chapter we want to talk about power series. A lot of functions, like *sin*, *cos* or *exp* may be defined as a power series. They serve as a to compute the value of a function in a numerical way (e.g. the *sin* may not be easily evaluated by geometric means).

Definition 3.47. (power series)
Let $a_k \in \mathbb{R}$ for all k, $x_0 \in \mathbb{R}$ be a point and x be a variable. The the series

$$\sum_{k=0}^{\infty} a_k (x - x_0)^k$$

is called a *power series* (*centered at* x_0) with coefficients $a_0, a_1, a_2, \ldots$.

So one may describe the power series as generalization of polynomials, like a polynomial of infinite degree.

Example 3.48.
Some of the more useful functions in mathematics may be defined by power series:

i)

$$\sin x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}.$$

ii)

$$\cos x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}.$$

iii)

$$\exp x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

Now the question arises, if a power series converges as a series or not.

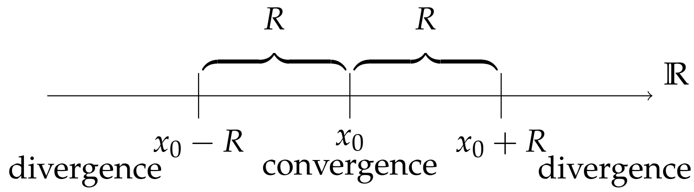
Theorem 3.49.
For any power series

$$\sum_{k=0}^{\infty} a_k (x - x_0)^k$$

there is a unique number $R \geq 0$ (or $R = \infty$) such that the sollowing holds:

The power series $\begin{cases} \text{converges (absolutely) for all } x \in (x_0 - R, x_0 + R), \\ \text{diverges for all } x \in (-\infty, x_0 - R) \cup (x_0 + R, \infty). \end{cases}$

Definition 3.50. (radius of convergence)
The number R is called the *radius of convergence*.



Remark 3.51.
The definition of the radius of convergence does not tell us if the power series converges or diverges for R.

Theorem 3.52.
Let $\sum_{k=0}^{\infty} a_k (x - x_0)^k$ be a power series and $\lim_{k \to \infty} (sup\{ \sqrt[m]{|a_m|} \mid m \geq k\}) = r$. Then the radius of convergence is

$$R = \begin{cases} \infty & , r = 0, \\ \frac{1}{r} & , r > 0, \\ 0 & , r = \infty. \end{cases}$$

In particular $r = \lim_{k \to \infty} \sqrt[k]{|a_k|}$ holds, if the limit exists.

Remark 3.53.
For numerous power series the radius of convergence may also be determined with the ratio test, that is

$$R = \lim_{k\to\infty} \left| \frac{a_k}{a_{k+1}} \right|.$$

Example 3.54.
Consider the power series

$$\sum_{k=0}^{\infty} \frac{2k^2}{3} x^k.$$

Using the ratio test yields

$$\left| \frac{a_k}{a_{k+1}} \right| = \left| \frac{\frac{2k^2}{3}}{\frac{2(k+1)^2}{3}} \right| = \frac{2k^2}{3} \cdot \frac{3}{2(k+1)^2} = \frac{k^2}{k^2+2k+1} \overset{k\to\infty}{\longrightarrow} 1$$

So the radius of convergence is $R = 1$.

By the radius of convergence we can define an interval on which the power series is an actual function.
Definition 3.55. (convergence interval)
Let

$$f : (x_0 - R, x_0 + R) \to \mathbb{R} : x \mapsto \sum_{k=0}^{\infty} a_k (x - x_0)^k$$

and R the radius of convergence of the power series. Then f is a function and the domain $(x_0 - R, x_0 + R)$ is called the *convergence interval.*

Example 3.56.
The following series are power series that can be defined as a function on their respected convergence intervals:

i) **Polynomials:** A polynomial p is a power series with $a_k = 0$ for all $k > deg(p)$. The the radius of convergence is $R = \infty$.

ii) **Exponential function:**

$$\exp : \mathbb{R} \to \mathbb{R} : x \mapsto \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

The radius of convergence of the exponential function is $R = \infty$, that is the exponential function converges for all $x \in \mathbb{R}$ and the center is $x_0 = 0$. Furthermore the following holds:

$$\exp(x + y) = \exp(x) + \exp(y),\ \forall x, y \in \mathbb{C}.$$

iii) **Sine and Cosine:**

$$\sin : \mathbb{R} \to \mathbb{R} : x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}.$$

$$\cos : \mathbb{R} \to \mathbb{R} : x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}.$$

The sine and cosine are centered at $x_0 = 0$ and have a radius of convergence of $R = \infty$.

iv) **Geometric series:**

$$f : (-1, 1) \to \mathbb{R} : x \mapsto \sum_{k=0}^{\infty} x^k$$

The convergence interval is $(-1, 1)$ since the power series is centered at $x_0 = 0$ with a radius of convergence of $R = 1$. So as we have seen before, the function converges to $f(x) = \frac{1}{1-x}$.

v) **Logarithm:**

$$f : (0, 2) \to \mathbb{R} : x \mapsto \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(x-1)^k}{k}$$

The power series is centered at $x_0 = 1$ with a radius of convergence of $R = 1$, which yield the convergence interval of $(0, 2)$.

Remark 3.57.
We can also consider x as a complex variable. Then there is still a radius of convergence in the complex plane to the complex power series

$$\sum_{k=0}^{\infty} a_k (x - x_0)^k, \ a_k \in \mathbb{C}.$$

The measure of distance in regards to the radius of convergence is the absolut value, that is:
The power series $\begin{cases} \text{converges for all } x \in \mathbb{C} \text{ with } |x - x_0| < R, \\ \text{diverges for all } x \in \mathbb{C} \text{ with } |x - x_0| > R. \end{cases}$ The radius of convergence is then the radius R of a circle around the center x_0 of the complex power series. Extending the power series to complex numbers will also yield

$$\exp(ix) = \cos x + i \sin x.$$

4 Calculus

4.1 Differentiation

In this chapter we are interested in continuous functions and their rate of change in a specific point. Therefore we will define the derivative of a function.

Definition 4.1. (differentiable, derivative)
Let $I = (a, b) \subseteq \mathbb{R}$ be an open interval and $x_0 \in I$. Let $f : I \to \mathbb{R}$ be a function. The limit

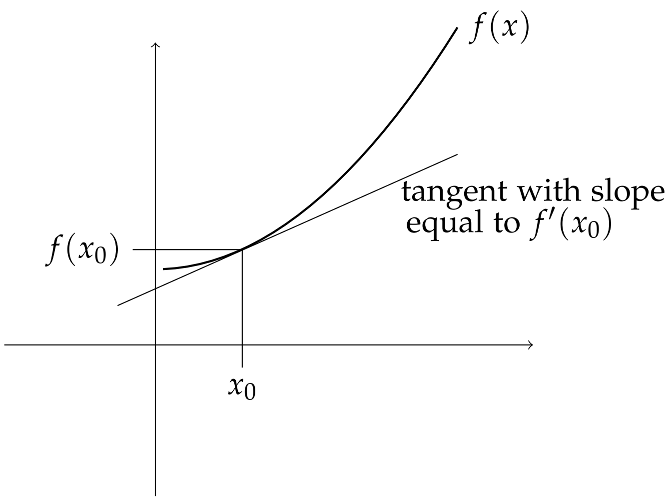
$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} =: f'(x_0) =: \frac{df}{dx}(x_0)$$

is called the *derivative* of f at x_0, if it exists. In case of its existence f is called *differentiable at* x_0. If f is differentiable at all $x_0 \in I$, then f is called differentiable. In this case the function

$$f' : I \to \mathbb{R} : x \mapsto f'(x)$$

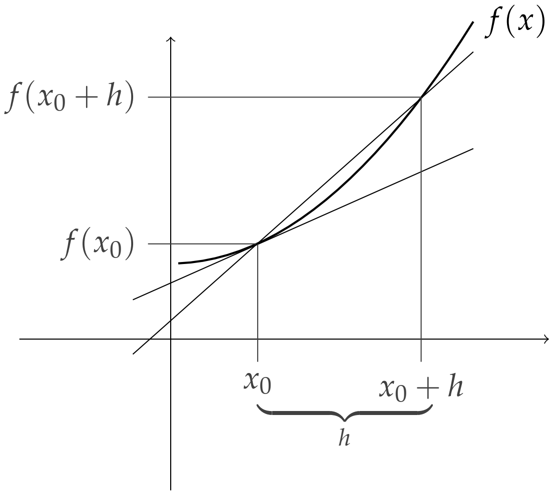
may be defined, which is called the *derivative* of f.

The idea behind the derivative is to approximate any function f in a certain point by a linear function, passing through the same point and having the exact same rate of change as the original function.



The limit given in the definition is actually a short version for the definition of the coefficient of the linear factor of the tangent. The coefficient is created by taking two points on the function f, one in the point we want to estiamte and the other close by, and moving the second point close to the first. If we move the two points close enough to each other, we will have the limit:

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

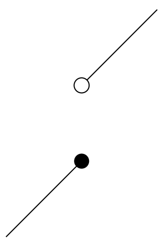


So we approximate the tangent slope with the slope of secants. When in the limit $h \to 0$ holds, then the secants approach the tangent.

We said earlier, that we want to look at continuous functions. The following theorem states, why we only try to approximate the tangent for a function that is continuous in x_0.

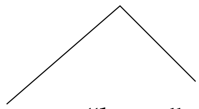
Theorem 4.2.
If a function $f : I \to \mathbb{R}$ is differentiable in $x_0 \in I$, then it is continuous in x_0.

So for differential functions there are two shapes that are not allowed to occur. By continuity at some point x_0 there is no "leap" allowed in x_0.



a "leap"

And by differentiability at some point x_0 it is forbidden to have a "knee".



a “knee”

We will add the rules for working with derivatives. This will allow us to derive more difficult functions.

Theorem 4.3. (rules for derivatives, product rule, quotient rule)
Let $f, g : I \to \mathbb{R}$ be two functions, both differentiable in $x_0 \in I$. Let $\lambda \in \mathbb{R}$ be a real number. Then we have

i) $\lambda \cdot f$ is differentiable in x_0 with

$$(\lambda \cdot f)'(x_0) = \lambda \cdot f'(x_0).$$

ii) $f + g$ is differentiable in x_0 with

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

iii) $f \cdot g$ is differentiable in x_0 with

$$(f \cdot g)'(x_0) = f'(x_0) \cdot g(x_0) + f(x_0) \cdot g'(x_0),$$

which is called the *product rule*.

iv) If $g(x_0) \neq 0$ then $\frac{f}{g}$ is differentiable in x_0 with

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

which is called the *quotient rule*.

Proof:
We will prove the cases *ii*) and *iii*):

ii)

$$\begin{aligned}(f + g)'(x_0) &= \lim_{h \to 0} \frac{(f + g)(x_0 + h) - (f + g)(x_0)}{h} \\ &= \lim_{h \to 0} \frac{f(x_0 + h) + g(x_0 + h) - f(x_0) - g(x_0)}{h} \\ &= \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} + \lim_{h \to 0} \frac{g(x_0 + h) - g(x_0)}{h} \\ &= f'(x_0) + g'(x_0)\end{aligned}$$

iii)

$$
\begin{aligned}
(f \cdot g)'(x_0) &= \lim_{h\to 0} \frac{(f\cdot g)(x_0+h) - (f\cdot g)(x_0)}{h} \\
&= \lim_{h\to 0} \frac{f(x_0+h)\cdot g(x_0+h) - f(x_0)\cdot g(x_0)}{h} \\
&= \lim_{h\to 0} \frac{f(x_0+h)\cdot g(x_0+h) - f(x_0)\cdot g(x_0) \overbrace{- f(x_0)\cdot g(x_0+h) + f(x_0)\cdot g(x_0+h)}^{=0}}{h} \\
&= \lim_{h\to 0} \frac{f(x_0+h) - f(x_0)}{h} \cdot g(x_0+h) + f(x_0)\cdot \lim_{h\to 0} \frac{g(x_0+h) - g(x_0)}{h} \\
&= \lim_{h\to 0} \frac{f(x_0+h) - f(x_0)}{h} \cdot \underbrace{\lim_{h\to 0} g(x_0+h)}_{=g(x_0)} + f(x_0)\cdot \lim_{h\to 0} \frac{g(x_0+h) - g(x_0)}{h} \\
&= f'(x_0)\cdot g(x_0) + f(x_0)\cdot g'(x_0).
\end{aligned}
$$

We have $\lim_{h\to 0} g(x_0+h) = g(x_0)$ since g is continuous. □

Example 4.4.

i) Let $f : \mathbb{R} \to \mathbb{R} : x \mapsto c$ with $c \in \mathbb{R}$ be a constant function. Then f is differentiable with derivative

$$f'(x_0) = \lim_{h\to 0} \frac{f(x_0+h) - f(x_0)}{h} = \lim_{h\to 0} \frac{c-c}{h} = \lim_{h\to 0} \frac{0}{h} = 0$$

for all $x_0 \in \mathbb{R}$.

ii) Let $f : \mathbb{R} \to \mathbb{R} : x \mapsto x$ be the identity function. Then f is differentiable with derivative

$$f'(x_0) = \lim_{h\to 0} \frac{f(x_0+h) - f(x_0)}{h} = \lim_{h\to 0} \frac{x_0+h-x_0}{h} = \lim_{h\to 0} \frac{h}{h} = \lim_{h\to 0} 1 = 1$$

for all $x_0 \in \mathbb{R}$.

iii) Let $n \in \mathbb{N}$ and $f : \mathbb{R} \to \mathbb{R} : x \to x^n$ be a power function. Then f is differentiable and with derivative

$$f'(x_0) = n \cdot x_0^{n-1}$$

for all $x_0 \in \mathbb{R}$.

Proof:
We proof this by induction.
base case:

$$(x)' = 1 = x^{1-1}$$

which was treated in the previous example.
inductive step: Assume that the claim is true for some $n \in \mathbb{N}$. Then we have

$$\begin{aligned}(x^{n+1})' &= (x^n \cdot x)' \underset{rule}{\overset{power}{=}} (x^n)' \cdot x + x^n \cdot (x)' \\ &\overset{I.H.}{=} n \cdot x^{n-1} \cdot x + x^n \cdot 1 = n \cdot x^n + x^n = (n+1) \cdot x^n.\end{aligned}$$

Thus by induction the claim is ture for all $n \in \mathbb{N}$. □

iv) The absolute value function $|\cdot| : \mathbb{R} \to \mathbb{R} : x \mapsto |x|$ is not differentiable at $x_0 = 0$ because the derivative limit does not exist. We have

$$\lim_{h \searrow 0} \frac{|0+h| - |0|}{h} = \lim_{h \searrow 0} \frac{h}{h} = \lim_{h \searrow 0} 1 = 1$$

and

$$\lim_{h \nearrow 0} \frac{|0+h| - |0|}{h} = \lim_{h \nearrow 0} \frac{-h}{h} = \lim_{h \nearrow 0} -1 = -1.$$

For the limit to exist, we need to have the same limit for both the right side and the left side. The absolute value is an example for having a “knee” in a function.

In the last chapter we have seen, that there are a lot of functions, which can be defined as power series. We can use the power series to determine the derivative of functions like the exponential function, the sine or the cosine.

Theorem 4.5.
Let $f : (x_0 - R, x_0 + R) \to \mathbb{R} : x \mapsto \sum_{k=0}^{\infty} a_k (x - x_0)^k$ be a power series function with radius of convergence R. Then f is differentiable with derivative

$$f' : (x_0 - R, x_0 + R) \to \mathbb{R} : x \mapsto \sum_{k=1}^{\infty} k \cdot a_k (x - x_0)^{k-1}.$$

With the help of the last theorem we can prove the derivative of the exponential and other functions.

Example 4.6.
i) We have $\exp'(x) = \exp(x)$ since

$$\exp'(x) = \left(\sum_{k=0}^{\infty} \frac{x^k}{k!} \right)' = \sum_{k=1}^{\infty} \frac{k}{k!} x^{k-1} = \sum_{k=1}^{\infty} \frac{1}{(k-1)!} x^{k-1} \underset{shift}{\overset{index}{=}} \sum_{k=0}^{\infty} \frac{1}{k!} x^k = \exp(x)$$

holds.

ii)

$$\sin'(x) = \cos(x)$$

and

$$\cos'(x) = -\sin(x)$$

may be shown via the last theorem, too.

iii) With the help of the quotient rule and the last theorem, we are able to determine the derivative of the tangent function:

$$\begin{aligned}
\tan'(x) &= \left(\frac{\sin}{\cos}\right)(x) \\
&= \frac{\sin'(x)\cdot\cos(x) - \sin(x)\cdot\cos'(x)}{\cos(x)^2} \\
&= \frac{\cos(x)\cdot\cos(x) - \sin(x)\cdot(-\sin(x))}{\cos(x)^2} \\
&= \frac{\cos(x)^2 + \sin(x)^2}{\cos(x)^2} \\
&\left(= \frac{1}{\cos(x)^2}\right) \\
&= 1 + \frac{\sin(x)^2}{\cos(x)^2} \\
&= 1 + \tan(x)^2
\end{aligned}$$

for all $x \in \mathbb{R} \setminus (\frac{\pi}{2} + \pi\cdot\mathbb{Z})$.

Besides the product and quotient rule there is another important rule for derivatives.

Theorem 4.7. (chain rule)
Let $f : I \to \mathbb{R}$ be differentiable at $x_0 \in I$ and $g : \widetilde{I} \to \mathbb{R}$ with $f(I) \subseteq \widetilde{I}$ be differentiable at $f(x_0) \in \widetilde{I}$. Then the composition $g \circ f : I \to \mathbb{R}$ is differentiable at x_0 with the derivative

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

We have seen already, that we can derive x^a with $a \in \mathbb{R}$. With the chain rule we are able to determine the derivative if the base and exponent are switched, that is a^x with $a \in \mathbb{R}_+$.

Example 4.8.
For $a > 0$ we have

$$(a^x)' = \exp(\log(a^x))' = \exp(x \cdot \log(a))' = \underbrace{\exp(x \cdot \log(a))}_{a^x} \cdot \underbrace{(x \cdot \log(a))'}_{\log(a)} = a^x \cdot \log(a).$$

Another application of the chain rule is the following theorem:

Theorem 4.9.
Let $f, g : I \to \mathbb{R}$ be differentiable at $x_0 \in I$ and $f(x) > 0$ for all $x \in I$. Then $f^g : I \to \mathbb{R} : x \mapsto f(x)^{g(x)}$ is differentiable at x_0 with derivative

$$(f^g)'(x_0) = f(x_0)^{g(x_0)-1} \cdot [f(x_0) \cdot g'(x_0) \cdot \log(f(x_0)) + f'(x_0) \cdot g(x_0)].$$

After we have seen, what the derivative of the exponential function is, we state the derivative of its inverse:

Example 4.10.
The following holds for $x > 0$:

$$(\log(x))' = \frac{1}{x}.$$

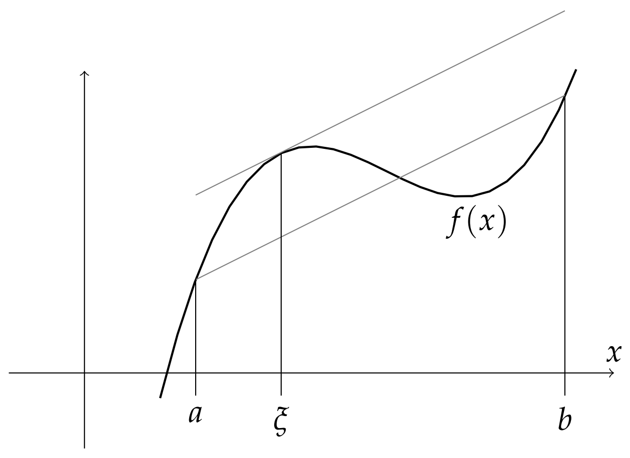
Another important theorem for differentiable functions is the so called mean value theorem:

Theorem 4.11. (Mean Value Theorem)
Let $a < b$ be real numbers and $f : [a, b] \to \mathbb{R}$ continuous. Furthermore f is differentiable on (a, b). Then there exists some $\xi \in (a, b)$ such that

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

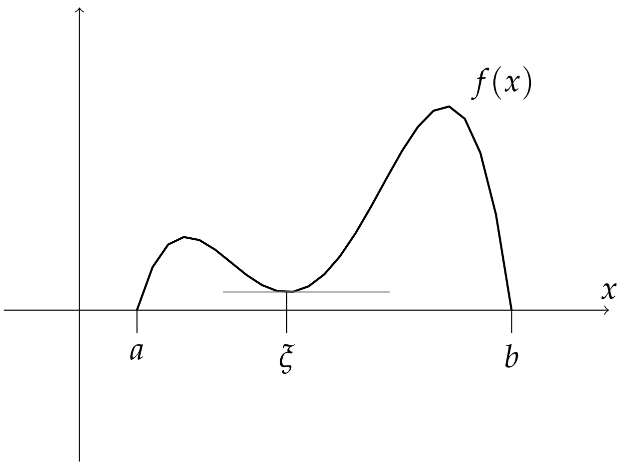
holds.

It states, that for each differentiable function on (a, b) and continuous function on $[a, b]$ there is a derivative in some point $\xi \in (a, b)$ such that the derivative has the same value as the slope of the line through a and b.



From the mean value theorem we can deduce the following corollary:

Corollary 4.12. (Rolle's theorem)
Let $a < b$ be real numbers and $f : [a,b] \to \mathbb{R}$ continuous with $f(a) = f(b) = 0$. Furthermore f is differentiable on (a,b). Then the derivative f' has a zero in (a,b), that is there exists some $\xi \in (a,b)$ such that $f'(\xi) = 0$ holds.



So somehow Rolle's theorem states, that the derivative will vanish for our differentiable function $f(x)$. In the last image we also get some kind of a low point for f in ξ. We will call those points *local minimum*.

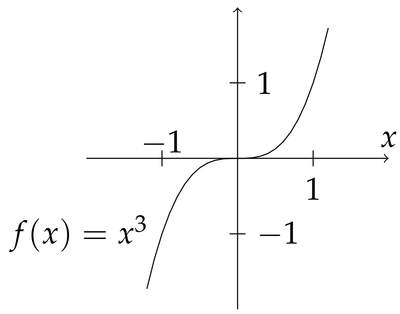
Definition 4.13. (local maximum, local minimum)
Let $I \subseteq \mathbb{R}$ be a subset, $x_0 \in I$ and $f : I \to \mathbb{R}$ a function. Then f has a *local maximum* (*local minimum*) at x_0 if there is some $\delta > 0$ such that $f(x) \leq f(x_0)$ ($f(x) \geq f(x_0)$) for all $x \in I$ with $|x - x_0| < \delta$.

If we look at the relationship between derivatives and maxima and minima, then we can deduce the following necessary condition.

Theorem 4.14.
Let $f : I \to \mathbb{R}$ be differentiable with derivative $f' : I \to \mathbb{R}$. Let f have a local maximum or a local minimum at $x_0 \in I$. Then $f'(x_0) = 0$ holds.

So local maxima or local minima will yield a derivative of zero, as we have seen already in the last image. But we can not deduce that a vanishing derivative will yield a local maximum or a local minimum. This implication is **false**.

Example 4.15.
Let $f : (-1,1) \to \mathbb{R} : x \mapsto x^3$ be a function. As a polynomial we can derive f with derivative $f'(x) = 3x^2$. So the polynomial f' has a zero in $x_0 = 0 \in (-1,1)$. But f does not have a local minimum or local maximum in $x_0 = 0$ because $f(x) < 0$ for all $x < 0$ and $f(x) > 0$ for all $x > 0$ holds.



Definition 4.16. (n-th derivative)
f' is called the *first derivative* of f, while f'' is called the *second derivative* of f. In general

$$f^{(n)} := \left(f^{(n-1)}\right)'$$

is called the *n-th derivative* of f if it exists.

Example 4.17.

i) The sine can be differentiated infinitely often. So with

$$\sin' x = \cos x,\ \sin'' x = -\sin x,\ \sin''' x = -\cos x,\ \sin^{(4)} x = \sin x,\ ...$$

we get

$$\sin^{(4n+k)} = \begin{cases} \sin x\ ,\ k = 0 \\ \cos x\ ,\ k = 1 \\ -\sin x\ ,\ k = 2 \\ -\cos x\ ,\ k = 3. \end{cases}$$

ii) The function

$$f : \mathbb{R} \to \mathbb{R} : x \mapsto |x| \cdot x$$

can be differentiated only once with

$$f' : \mathbb{R} \to \mathbb{R} : x \mapsto 2|x|.$$

The second derivative does not exist at $x_0 = 0$.

We have seen a couple of ways to determine the limit of functions in specific points. Yet there are many cases in which we are not able to calculate the limit. *L'Hospital's Rule* will extend the cases, for which we are able to derive the limits for functions. It is based upon differentiation.

Theorem 4.18. (L'Hospital's Rule)
Let $f, g : I \to \mathbb{R}$ be two differentiable functions with $0 \notin g'(I)$. Let x_0 be one of the two boundary points of I (including $\pm\infty$). If either

$$\lim_{x\to x_0} f(x) = \quad 0 \quad = \lim_{x\to x_0} g(x)$$

or

$$\lim_{x\to x_0} f(x) = \pm\infty = \lim_{x\to x_0} g(x)$$

holds and the limit

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

exists, then so does the limit

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} = a.$$

So in the cases, that we have a fraction and the nominator and denominator converge to zero or both diverge to infinity, we can calculate the limit by differentiating the nominator and denominator, if it exists.

Example 4.19.
Determine the limit

$$\lim_{x\to 0} \frac{\exp(x) - 1}{\sin(x)}.$$

Now we have, that $\lim\limits_{x\to 0} \exp(x) - 1 = 0$ and $\lim\limits_{x\to 0} \sin(x) = 0$ holds. So we have a case of $[\frac{0}{0}]$. Define $f(x) := \exp(x) - 1$ and $g(x) := \sin(x)$. Now we try to determine the limit of $\lim\limits_{x\to 0} \frac{f'(x)}{g'(x)}$ and get

$$\lim_{x\to 0} \frac{f'(x)}{g'(x)} = \lim_{x\to 0} \frac{\exp(x)}{\cos(x)} = \frac{\exp(0)}{\cos(0)} = \frac{1}{1} = 1.$$

Since the limit exists, we get by L'Hospital's Rule

$$\lim_{x\to 0} \frac{f(x)}{g(x)} = \lim_{x\to 0} \frac{f'(x)}{g'(x)} = 1.$$

We want to conclude the chapter with the *Taylor polynomial* and some of its applications. When we talked about derivatives $f'(x)$ to some function $f(x)$, we constructed the tangent as a linear approximation to our function f in a specific point. The tangent for a function $f(x)$ can be expressed in $\begin{pmatrix} x_0 \\ f(x_0) \end{pmatrix}$ as

$$y = f'(x_0) \cdot x + f(x_0).$$

The Taylor polynomials are approximations of higher degrees of a function $f(x)$ in a specific point. In Particular, the tangent, stated a few sentences ago, is the Taylor polynomial of f of first degree at x_0.

Definition 4.20. (Taylor polynomial)
Let $f : I \to \mathbb{R}$ be a function whose n derivatives $f', ..., f^{(n)} : I \to \mathbb{R}$ exist. Let $x_0 \in I$. Then

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

is called the *Taylor polynomial of f of degree n at x_0*, where $f^{(0)} = f$ is.

Example 4.21.
Let $f(x) = \sin x$ be the sine function. Determine the Taylor polynomial of f of degree 3 at $x_0 = 0$. We have

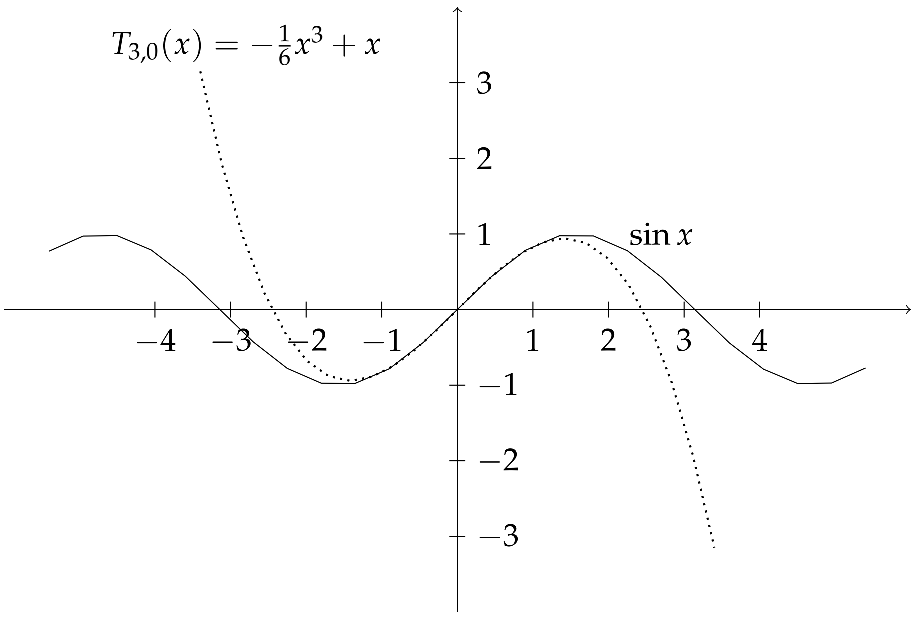
$$f(x) = \sin x,\ f'(x) = \cos x,\ f''(x) = -\sin x,\ f'''(x) = -\cos x$$

and for the point $x_0 = 0$ we calculate

$$f(0) = 0,\ f'(0) = 1,\ f''(0) = 0,\ f'''(0) = -1.$$

So the Taylor polynomial is

$$\begin{aligned} T_{3,0}(x) &= f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(x)}{3!}x^3 \\ &= 0 + 1x + \frac{0}{2}x^2 + \frac{-1}{6}x^3 \\ &= -\frac{1}{6}x^3 + x. \end{aligned}$$



The approximation of a function $f(x)$ gets better, if the degree of the Taylor polynomial increases. The gap between the original function $f(x)$ and the Taylor polynomial $T_{n,x_0}(x)$ in some point $\widetilde{x}$ can be approximated by the *Taylor formula*. Yet still the Taylor polynomial might only approximate the original function in a specific area around its origin x_0.

Theorem 4.22. (Taylor formula)
Let $f : I \to \mathbb{R}$ be a function whose $n+1$ derivatives $f', ..., f^{(n+1)} : I \to \mathbb{R}$ exist and are continuous, that is f is $n+1$ times *continuously differentiable*. Furthermore R_{n,x_0} is defined by the equation

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + ... + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + R_{n,x_0}.$$

Then the following holds:

i) $R_{n,x_0} = \frac{f^{(n+1)}(t)}{n!}(x-t)^n(x-x_0)$ with t between x and x_0,

ii) $R_{n,x_0} = \frac{f^{(n+1)}(t)}{(n+1)!}(x-x_0)^{n+1}$ with t between x and x_0.

R_{n,x_0} is called the *remainder* of the Taylor formula. *i*) is called the *Cauchy form* and *ii*) is called the *Lagrange form of the remainder*.

Example 4.23.
Let $f(x) = \sin x$. As we have seen, the Taylor polynomial of $f(x)$ of third degree at $x_0 = 0$ is

$$T_{3,0}(x) = -\frac{1}{6}x^3 + x.$$

The remainder in Lagrange form is

$$R_{3,0} = \frac{f^{(4)}(t)}{4!}(x-0)^4 = \frac{\sin(t)}{24}x^4$$

with t between x and $x_0 = 0$. Let's say we want to know how good the approximation is for the domain $[-1, 1]$. By $|\sin \tilde{x}| \leq 1$ for all $\tilde{x} \in \mathbb{R}$ we have

$$|R_{3,0}| = \left| \frac{\sin(t)}{24}x^4 \right| \leq \frac{1}{24},$$

since $t \in [-x, x]$ and $x \in [-1, 1]$ holds. So the error is at most $\frac{1}{24} \approx 0.0416$. On the other hand if we get farther away from the origin of the Taylor polynomial, the error may increase proportionally. In fact we know that $|\sin x| \leq 1$ for all $x \in \mathbb{R}$, but $T_{3,0}(x) \overset{x\to\infty}{\longrightarrow} \infty$ holds. So in this case the Taylor polynomial is only a good approximation for x close to $x_0 = 0$.

From the Taylor formula we are able to deduce the next theorem. It states for any point x_0 for a function $f(x)$ with enough derivatives, wether that function $f(x)$ has a local minimum or local maximum or neither at x_0 by using the higher derivatives of the function $f(x)$.

Theorem 4.24.
Let $n \in \mathbb{N}$, $n \geq 2$, $x_0 \in I$ and $f : I \to \mathbb{R}$ be a function for which the first n derivatives exist. Further more

$$f'(x_0) = ... = f^{(n-1)}(x_0) = 0 \text{ and } f^{(n)}(x_0) \neq 0$$

holds. Then we have:

i) If n is even and $f^{(n)}(x_0) > 0$, then f has a local minimum at x_0.

ii) If n is even and $f^{(n)}(x_0) < 0$, then f has a local maximum at x_0.

iii) If n is odd, then f has neither a local maximum nor a local minimum at x_0.

Example 4.25.
We have already seen, that the function $f(x) : \mathbb{R} \to \mathbb{R} : x \mapsto x^3$ has no local maximum or local minimum in $x_0 = 0$, even though $f'(x_0) = f'(0) = 3 \cdot 0^2 = 0$ holds. With the last theorem we are able to prove it:

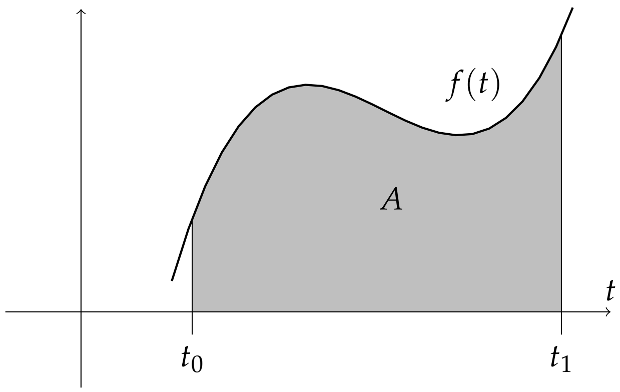
$$\begin{aligned} f'(x) &= 3x^2 &\Rightarrow\quad f'(0) &= 3 \cdot 0^2 = 0 \\ f''(x) &= 6x &\Rightarrow\quad f''(0) &= 6 \cdot 0 = 0 \\ f'''(x) &= 6 &\Rightarrow\quad f'''(0) &= 6 \neq 0 \end{aligned}$$

Since the third derivative is the first of our derivatives, that is not equal to zero at x_0, we have $n = 3$. Since n is odd, there is no local maximum or minimum at x_0.

4.2 Integration

We start with a little example:

Example 4.26.
Let a car and its velocity be given over a specific time interval. Let us say, we start at time t_0 and stop at t_1. The velocity can be modelled with a function $f(t)$ with $t \in [t_0, t_1]$. We are interested in the distance, that the car travels between the two points t_0 and t_1. This is actually the area A, that is enclosed by the function and its t-axis, since the distance travelled in a certain time interval is the velocity.



So we need to find a way, to calculate the area below a function. Later on we will call this area with a given start and end point an integral.

The basic idea to solve the problem in the example is to divide the given interval into small rectangles.

Definition 4.27. (lower/upper sum)
Let $[a, b] = [x_0, x_1] \cup [x_1, x_2] \cup ... \cup [x_{n-1}, x_n]$, with $a = x_0 < x_1 < ... < x_n = b$ be a partition P of the interval $[a, b]$. Let

$$m_i := \inf\{f(x)|x \in [x_{i-1}, x_i]\}$$

and

$$M_i := \sup\{f(x)|x \in [x_{i-1}, x_i]\}$$

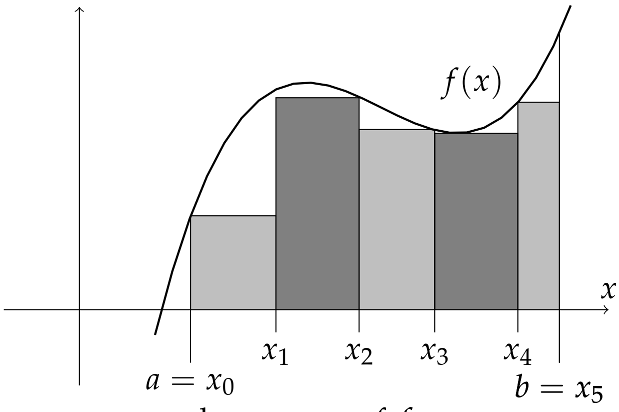
for all $i = 1, ..., n$.. Then

$$L(f, P) := \sum_{i=1}^{n} m_i \cdot (x_i - x_{i-1})$$

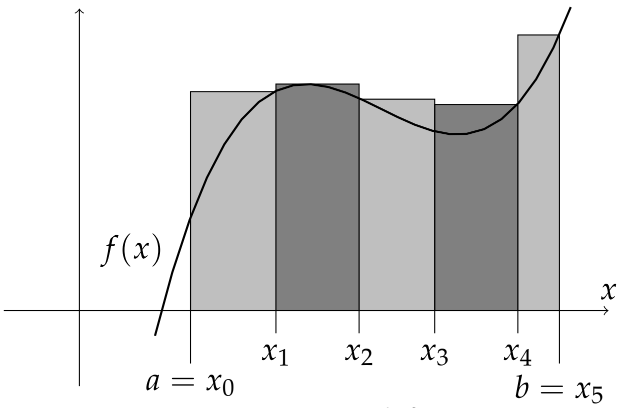
is called the *lower sum* of f with respect to the partition P and

$$U(f, P) := \sum_{i=1}^{n} M_i \cdot (x_i - x_{i-1})$$

is called the *upper sum* of f with respect to the partition P.



lower sum of f



upper sum of f

Remark 4.28.
Observe that for the precise area A below a function f we have

$$L(f, P) \leq A \leq U(f, P).$$

Now we want to refine our view on those sums, so we define the lower and upper integral for a bounded function. This means, that we use the best partition of the interval possible for our purposes. So for a continiuous function, which derivative is not constant, this means we refine the partition into an infinite set rectangles.

Definition 4.29. (lower/upper integral)
Let $f : I \to \mathbb{R}$ be a bounded function. Then

$$_*\int_a^b f(x)\mathrm{d}x := \sup\{L(f, P) | P \text{ is a partition of } [a, b]\} \tag{1}$$

is called the *lower integral* of f on $[a, b]$.

$$^*\int_a^b f(x)\mathrm{d}x := \inf\{U(f, P) | P \text{ is a partition of } [a, b]\} \tag{2}$$

is called the *upper integral* of f on $[a, b]$.

Since we work with the infimum and the supremum of a summed up function, we get the following remark.

Remark 4.30.
It always holds

$$ {}_*\int_a^b f(x)\mathrm{d}x \leq {}^*\int_a^b f(x)\mathrm{d}x. $$

By the last remark we want to know, what happens if the lower integral equals the upper integral, which should yield the precise area below a function.

Definition 4.31. (integrable, integral)
Let $f : I \to \mathbb{R}$ be a bounded function. If

$$ {}_*\int_a^b f(x)\mathrm{d}x = {}^*\int_a^b f(x)\mathrm{d}x. $$

holds, then f is called *integrable* on $[a, b]$ with *integral*

$$ \int_a^b f(x)\mathrm{d}x := {}_*\int_a^b f(x)\mathrm{d}x \left(= {}^*\int_a^b f(x)\mathrm{d}x \right). $$

We now have the definition of the precise area under a function. So first we state two huge classes of functions that are integrable.

Theorem 4.32.
Let $f : I \to \mathbb{R}$ be a monotonically decreasing or increasing funtion. Then f is integrable I.

Theorem 4.33.
Let $f : I \to \mathbb{R}$ be a continuous function. Then f is integrable on I.

After we defined some classes that are integrable, we want to give some rules about how we work with integrals.

Theorem 4.34. (rules for integrals)
Let $f, g : [a, b] \to \mathbb{R}$ be integrable functions and $\alpha, \beta \in \mathbb{R}$. Then:

i) $\alpha \cdot f + \beta \cdot g$ is integrable with

$$ \int_a^b \alpha \cdot f(x) + \beta \cdot g(x)\mathrm{d}x = \alpha \cdot \int_a^b f(x)\mathrm{d}x + \beta \cdot \int_a^b g(x)\mathrm{d}x. $$

ii) If $f(x) \leq g(x)$ for all $x \in [a, b]$, then

$$\int_a^b f(x)\mathrm{d}x \leq \int_a^b g(x)\mathrm{d}x$$

holds. Multiplying both sides of the equation with (-1) will yield the analogous theorem for $f(x) \geq g(x)$.

iii) $|f|$ is integrable with

$$\left|\int_a^b f(x)\mathrm{d}x\right| \leq \int_a^b |f(x)|\mathrm{d}x.$$

iv) If $c \in [a, b]$ holds, then $f : [a, c] \to \mathbb{R}$ and $f : [c, b] \to \mathbb{R}$ are integrable with

$$\int_a^c f(x)\mathrm{d}x + \int_c^b f(x)\mathrm{d}x = \int_a^b f(x)\mathrm{d}x.$$

Now we have stated some rules about how integrals may be transformed or calculated with. But we do not know how we actually are able to compute such an integral. The solution to this problem was the discovery, that integration is somehow the inverse operation of differentiation. So we state the *anti-derivative*.

Definition 4.35. (anti-derivative)
Let $F, f : I \to \mathbb{R}$ be functions and F be differentiable. If $F' = f$ holds, then F is called an *anti-derivative* of f.

Remark 4.36.
If F is an anti-derivative for f, then so is $F + c$ with $c \in \mathbb{R}$. Hence anti-derivatives are not unique with $F' = (F + c)' = f$. Conversely, if F and G are both anti-derivatives of f, then there exists a constant $c \in \mathbb{R}$ with $F = G + c$.

Example 4.37.
Since we already know the derivatives of some functions, we can also easily guess an anti-derivative of them.

i) Let $f(x) = x^n$ with $n \in \mathbb{Z} \setminus \{-1\}$. Then $F(x) = \frac{1}{n+1} \cdot x^{n+1}$ is an anti-derivative of f.

Proof:

$$F'(x) = \left(\frac{1}{n+1} \cdot x^{n+1}\right)' = \frac{n+1}{n+1} \cdot x^{n+1-1} = x^n = f(x). \qquad \square$$

ii) Let $f(x) = \sin x$. Then $F(x) = -\cos x$ is an anti-derivative of f.

Proof:

$$F'(x) = (-\cos x)' = \sin x = f(x). \qquad \square$$

iii) Let $f(x) = \exp x$. Then $F(x) = \exp x$ is an anti-derivative of f.

Proof:

$$F'(x) = (\exp x)' = \exp x = f(x). \qquad \square$$

Now we can state the *Fundamental Theorem of Calculus*, which connects the anti-derivative with integrals.

Theorem 4.38. (Fundamental Theorem of Calculus)
Let $f : [a, b] \to \mathbb{R}$ be continuous.

i) The function

$$F : [a, b] \to \mathbb{R} : x \mapsto \int_a^x f(t)\mathrm{d}t$$

is continuous and also differentiable on (a, b) with

$$F'(x) = f(x).$$

ii) Let $F : [a, b] \to \mathbb{R}$ be an anti-derivative of f, that is $F' = f$, then

$$\int_a^b f(x)\mathrm{d}x = F(b) - F(a) \left(=: [F(x)]_a^b \right).$$

holds.

Proof:

i) Let $h \geq 0$. By the definition of the derivative we have for $x \in [a, b]$

$$\begin{aligned} F'(x) &= \lim_{h \to 0} \frac{F(x+h) - F(x)}{h} \\ &= \lim_{h \to 0} \frac{1}{h} \left(\int_a^{x+h} f(t)\mathrm{d}t - \int_a^x f(t)\mathrm{d}t \right) \\ &= \lim_{h \to 0} \frac{1}{h} \left(\int_x^{x+h} f(t)\mathrm{d}t \right). \end{aligned}$$

Furthermore

$$m_h \cdot h \leq \int\limits_{x}^{x+h} f(t)\mathrm{d}t \leq M_h \cdot h$$

holds with

$$m_h := \inf\{f(x_0) | x \leq x_0 \leq x+h\}$$

and

$$M_h := \sup\{f(x_0) | x \leq x_0 \leq x+h\}.$$

Since f is continuous, we deduce

$$\lim_{h\to 0} m_h = \lim_{h\to 0} M_h = f(x)$$

and hence we get

$$\lim_{h\to 0} \frac{F(x+h) - F(x)}{h} = f(x).$$

So $F(x)$ is differentiable and continuous. The case $h < 0$ is anlaogous.

ii) Let F be an anti-derivative of f. According to $i)$ the function

$$\widetilde{F} : [a,b] \to \mathbb{R} : x \mapsto \int\limits_{a}^{x} f(t)\mathrm{d}t$$

is also an anti-derivative of f. Hence there exists a constant $c \in \mathbb{R}$ with

$$F(x) = \widetilde{F}(x) + c.$$

With $x = a$, we can deduce

$$F(a) = \underbrace{\widetilde{F}(a)}_{=0} + c$$
$$\Rightarrow F(a) = c.$$

Therefore we get

$$\int\limits_{a}^{b} f(x)\mathrm{d}x = \widetilde{F}(b) = F(b) - c = F(b) - F(a). \qquad \square$$

Example 4.39.

i) Let

$$f : [a, b] \to \mathbb{R} : x \mapsto \int\limits_a^x e^t \sin t \mathrm{d}t$$

be a function. Determine f'. By the fundamental theorem of calculus we get

$$f'(x) = e^x \sin x.$$

ii) Let

$$f : \mathbb{R} \to \mathbb{R} : x \mapsto \sin x$$

be a function. Determine

$$\int\limits_0^{\frac{5\pi}{2}} f(x)\mathrm{d}x.$$

By the fundamental theorem of calculus we get

$$\begin{aligned} \int\limits_0^{\frac{5\pi}{2}} f(x)\mathrm{d}x &= \int\limits_0^{\frac{5\pi}{2}} \sin x \mathrm{d}x \\ &= -\cos x\Big|_0^{\frac{5\pi}{2}} \\ &= -\left(\cos\frac{5\pi}{2} - \cos 0\right) \\ &= -(0-1) \\ &= 1 \end{aligned}$$

Definition 4.40. (indefinite integral)
Let f be a integrable function. We denote by

$$\int f(x)\mathrm{d}x$$

the *indefinite integral*.

As with the anti-derivatives, the indefinite integral is unique up to a constant $c \in \mathbb{R}$.

Example 4.41.
With the indefinite integral and our knowledge about derivatives we give the following often used anti-derivatives:

- $\int x^n \, \mathrm{d}x = \frac{1}{n+1} x^{n+1}$ with $n \in \mathbb{Z} \setminus \{-1\}$
- $\int \frac{1}{x} \, \mathrm{d}x = \log |x|$
- $\int \sin x \, \mathrm{d}x = -\cos x$
- $\int \cos x \, \mathrm{d}x = \sin x$
- $\int \tan x \, \mathrm{d}x = -\log |\cos(x)|$
- $\int \frac{1}{\cos^2 x} \, \mathrm{d}x = \tan x$
- $\int \exp x \, \mathrm{d}x = \exp x$
- $\int a^x \, \mathrm{d}x = \frac{1}{\log a} a^x$
- $\int \log x \, \mathrm{d}x = x \cdot \log x - x$

Not all of these indefinite integrals are defined on $\mathbb{R}$. So be careful by using them.

While the indefinite integrals of the last example can more or less be easily guessed, we want to state, that integration is by far more difficult than its "inverse" operation differentiation. There are lots of functions where the anti-derivatives are not known. Furthermore there are functions where we do not even know, if there exists an anti-derivative.

So we need some additional tools to tackle the problem of integration and to extend the amount of functions, that we are able to find indefinite integrals for. The first tool is called *integration by parts*.

Theorem 4.42. (integration by parts)
Let $f, g : [a, b] \to \mathbb{R}$ be differentiable with continiuous derivatives. Then

$$\int_a^b f(x) \cdot g'(x) \mathrm{d}x = [f(x) \cdot g(x)]_a^b - \int_a^b f'(x) \cdot g(x) \mathrm{d}x$$

holds.

Proof:
With the product rule of differentiation we get

$$(f(x) \cdot g(x))' = f'(x) \cdot g(x) + f(x) \cdot g'(x).$$

Integrating both sides yields

$$\int_a^b f'(x) \cdot g(x) + f(x) \cdot g'(x) \mathrm{d}x = \int_a^b (f(x) \cdot g(x))' \mathrm{d}x = [f(x) \cdot g(x)]_a^b.$$

Using the linearity of the integration o rearrange the parts, we deduce

$$\int_a^b f(x) \cdot g'(x) \mathrm{d}x = [f(x) \cdot g(x)]_a^b - \int_a^b f'(x) \cdot g(x) \mathrm{d}x \qquad \square$$

Example 4.43.
Determine

$$\int x \cdot \cos x \, \mathrm{d}x.$$

We put

$$f(x) := x \qquad \Rightarrow f'(x) = 1$$

and

$$g(x) := \sin x \Rightarrow g'(x) = \cos x.$$

Integration by parts yields

$$\begin{aligned} \int x \cdot \cos x \, \mathrm{d}x &= [x \cdot \sin x] - \int 1 \cdot \sin x \, \mathrm{d}x \\ &= x \cdot \sin x + \cos x \end{aligned}$$

Integration by parts is often used by functions, where $f^{(n)}(x)$ is constant for some $n \in \mathbb{N}$ and the integration of $g(x)$ will yield the same indefinite integral every few integration steps.

Another tool for integration is *integration by substitution*. Where integration by parts was a variation of the product rule, integration by substitution is a variation of the chain rule.

Theorem 4.44. (integration by substitution)
Let $f : I \to \mathbb{R}$ be continiuous and $g : [a, b] \to \mathbb{R}$ be differentiable with continiuous derivative and $g([a, b]) \subseteq I$. Then

$$\int_a^b f(g(x)) \cdot g'(x) \, \mathrm{d}x = \int_{g(a)}^{g(b)} f(\widetilde{x}) \, \mathrm{d}\widetilde{x}$$

holds.

Proof:
Let F be an anti-derivative of f on $g([a,b])$. So we have

$$\int_{g(a)}^{g(b)} f(y)\ \mathrm{d}y = F(g(b)) - F(g(a)).$$

Furthermore $(f \circ g) \cdot g'$ is continiuous as a composition of continiuous functions. With the chain rule of differentiation we also get

$$(F \circ g)' = (F' \circ g) \cdot g' = (f \circ g) \cdot g'.$$

Hence $F \circ g$ is an anti-derivative of $(f \circ g) \cdot g'$ and so

$$\int_a^b f(g(x)) \cdot g(x)'\ \mathrm{d}x = F(g(b)) - F(g(a)).$$

So

$$\int_a^b f(g(x)) \cdot g'(x)\ \mathrm{d}x = \int_{g(a)}^{g(b)} f(y)\ \mathrm{d}y$$ □

Example 4.45.

i) Determine

$$\int_0^1 (2x+1) \cdot \sin(x^2+x)\ \mathrm{d}x.$$

Define $f(x) := \sin x$ and $g(x) := x^2 + x$. Then we have

$$\begin{aligned}\int_0^1 \underbrace{2x+1}_{g'(x)} \cdot \overbrace{\sin\underbrace{(x^2+x)}_{g(x)}}^{f(g(x))}\ \mathrm{d}x &= \int_{g(0)}^{g(1)} \sin y\ \mathrm{d}y \\ &= \int_0^2 \sin y\ \mathrm{d}y \\ &= -\cos y\big|_0^2 \\ &= -\cos(2) + 1\end{aligned}$$

ii) Determine

$$\int \tan x \, \mathrm{d}x.$$

For any some intervall $[a,b]$ with $\tan x$ continiuous on $[a,b]$, we get

$$\begin{aligned}
\int_a^b \tan x \, \mathrm{d}x &= -\int_a^b \frac{-\sin x}{\cos x} \, \mathrm{d}x \\
&= -\int_a^b \frac{1}{\cos x} \cdot (-\sin x) \, \mathrm{d}x \\
&= -\int_a^b \frac{1}{\cos x} \cdot \cos' x \, \mathrm{d}x
\end{aligned}$$

Now integration by substitution yields with $f(x) := \frac{1}{y}$ and $g(x) := \cos x$

$$\begin{aligned}
-\int_a^b \frac{1}{\cos x} \cdot \cos' x \, \mathrm{d}x &= -\int_{\cos a}^{\cos b} \frac{1}{u} \, \mathrm{d}u \\
&= -\log(u)\big|_{\cos a}^{\cos b} \\
&\overset{res.}{=} -\log(\cos x)\big|_a^b \\
&= -(\log(\cos b) - \log(\cos a)).
\end{aligned}$$

The second last operation is called resubstitution. Hence we get as an indefinite integral

$$\int \tan x \, \mathrm{d}x = -\log|\cos x|.$$

Definition 4.46. (improper integral)
Let $\widetilde{b} \in \mathbb{R} \cup \{+\infty\}$ and $f : [a,b] \to \mathbb{R}$ be integrable for all real numbers b with $a < b < \widetilde{b}$. Then

$$\int_a^{\widetilde{b}} f(x)\mathrm{d}x := \lim_{b \nearrow \widetilde{b}} \int_a^b f(x)\mathrm{d}x$$

is called an *improper integral*. The improper integral converges, if the corresponding limit exists. Otherwise it diverges.

Remark 4.47.
There is an analogous definition for the improper integral for the lower bound with $\widetilde{a} \in \mathbb{R} \cup \{-\infty\}$and $\widetilde{a} < a < b$. Also both the upper and lower point may be chosen

for limitting at the same time. In this case the integration interval is split by means of some $c \in \mathbb{R}$ with $\widetilde{a} < c < \widetilde{b}$ and

$$\begin{aligned}\int_{\widetilde{a}}^{\widetilde{b}} f(x)\mathrm{d}x := &\int_{\widetilde{a}}^{c} f(x)\mathrm{d}x + \int_{c}^{\widetilde{b}} f(x)\mathrm{d}x \\ =&\lim_{a \searrow \widetilde{a}} \int_{a}^{c} f(x)\mathrm{d}x + \lim_{b \nearrow \widetilde{b}} \int_{c}^{b} f(x)\mathrm{d}x.\end{aligned}$$

Then this improper integral converges if both integrals

$$\lim_{a \searrow \widetilde{a}} \int_{a}^{c} f(x)\mathrm{d}x \text{ and } \lim_{b \nearrow \widetilde{b}} \int_{c}^{b} f(x)\mathrm{d}x$$

converge.

Example 4.48.

i) Determine

$$\int_0^1 \frac{1}{\sqrt{x}}\, \mathrm{d}x.$$

We have

$$\int_0^1 \frac{1}{\sqrt{x}}\, \mathrm{d}x = \lim_{t \searrow 0} \int_t^0 \frac{1}{\sqrt{x}}\, \mathrm{d}x = \lim_{t \searrow 0} [2\sqrt{x}]_t^1 = \lim_{t \searrow 0} (2 - 2\sqrt{t}) = 2.$$

So the improper integral converges.

ii) Determine

$$\int_1^\infty \frac{1}{\sqrt{x}}\, \mathrm{d}x.$$

We have

$$\int_1^\infty \frac{1}{\sqrt{x}}\, \mathrm{d}x = \lim_{t \to \infty} \int_1^t \frac{1}{\sqrt{x}}\, \mathrm{d}x = \lim_{t \to \infty} [2\sqrt{x}]_1^t = \lim_{t \to \infty} (2\sqrt{t} - 2) = \infty.$$

So the improper integral diverges.

iii) Determine

$$\int_0^1 \frac{1}{x^2}\,\mathrm{d}x.$$

We have

$$\int_0^1 \frac{1}{x^2}\,\mathrm{d}x = \lim_{t\searrow 0}\int_t^1 \frac{1}{x^2}\,\mathrm{d}x = \lim_{t\searrow 0}\left[-\frac{1}{x}\right]_t^1 = \lim_{t\searrow 0}\left(-1+\frac{1}{t}\right) = \infty.$$

So the improper integral diverges.

iv) Determine

$$\int_1^\infty \frac{1}{x^2}\,\mathrm{d}x.$$

We have

$$\int_1^\infty \frac{1}{x^2}\,\mathrm{d}x = \lim_{t\to\infty}\int_1^t \frac{1}{x^2}\,\mathrm{d}x = \lim_{t\to\infty}\left[-\frac{1}{x}\right]_1^t = \lim_{t\to\infty}\left(-\frac{1}{t}+1\right) = 1.$$

So the improper integral converges.

v) Determine

$$\int_{-\infty}^\infty \frac{2x}{x^2+1}\,\mathrm{d}x.$$

We have

$$\begin{aligned}\int_{-\infty}^\infty \frac{2x}{x^2+1}\,\mathrm{d}x &= \lim_{t\to-\infty}\int_t^0 \frac{2x}{x^2+1}\,\mathrm{d}x + \lim_{t\to\infty}\int_0^t \frac{2x}{x^2+1}\,\mathrm{d}x \\ &= \lim_{t\to-\infty}\left[\log(x^2+1)\right]_t^0 + \lim_{t\to\infty}\left[\log(x^2+1)\right]_0^t \\ &= \underbrace{\lim_{t\to-\infty} -\log(t^2+1)}_{=-\infty} + \underbrace{\lim_{t\to\infty}\log(t^2+1)}_{=\infty}.\end{aligned}$$

So the improper integral diverges.

vi) The *Gaussian Integral*

$$\int_{-\infty}^\infty e^{-x^2}\,\mathrm{d}x = \sqrt{\pi}$$

converges.

5 Ordinary differential equations

5.1 Ordinary differential equations of first order

So far we have encountered equations where the solution was some interval, a set of some numbers or a single number. Now a differential equation is a function linked to its own derivatives. So solving a differential equation yields a function, that satisfies the given link to it's own derivatives. For example

$$y'(x) = x \cdot y(x)$$

is a differential equation, where the solution is some function $y(x)$. In this case the trivial function $y(x) = 0$ solves the equation as well as $y(x) = \exp\left(\frac{x^2}{2}\right)$.

Definition 5.1.
An equation relating a function

$$y : I \to \mathbb{R} : x \mapsto y(x)$$

to one or several of its derivatives $y', y'', ...$ as well as to its independent variable x is called an *ordinary differential equation*.

Remark 5.2.
Since we will only work with ordinary differential equations (ODEs) and not with partial differential equations (PDEs), we will skip the word ordinary.

If the solution of a differential equation is known, then it is easy to verify, that it is really a solution. The problem of finding that unknown solution is far more difficult. We will present some methods, that will help us to solve some differential equations.

The first method is called *seperation of variables*.

Theorem 5.3. (seperation of variables)
Differential equations of the form

$$y' = f(x) \cdot g(y)$$

with f and g are continuous functions and g has no zeroes may sometimes be solved by solving the integral equation

$$\int \frac{1}{g(y)} \, \mathrm{d}y - \int f(x) \, \mathrm{d}x = c$$

for y, where $c \in \mathbb{R}$ is a constant.

Example 5.4.
Solve the differential equation

$$y' = e^{2y}\cos(3x).$$

By seperation of variables we define $f(x) := \cos(3x)$ and $g(y) := e^{2y}$ and we get

$$\begin{aligned}
\int \frac{1}{g(y)}\,\mathrm{d}y &= \int f(x)\,\mathrm{d}x \\
\Leftrightarrow \int e^{-2y}\,\mathrm{d}y &= \int \cos(3x)\,\mathrm{d}x \\
\Leftrightarrow -\frac{1}{2}e^{-2y} + c_1 &= \frac{1}{3}\sin(3x) + c_2 \\
\Leftrightarrow e^{-2y} &= -\frac{2}{3}\sin(3x) + 2(c_1 - c_2) \\
\Leftrightarrow -2y &= \log\left(-\frac{2}{3}\sin(3x) + 2(c_1 - c_2)\right) \\
\Leftrightarrow y &= -\frac{1}{2}\log\left(-\frac{2}{3}\sin(3x) + 2(c_1 - c_2)\right)
\end{aligned}$$

Now there are some restrictions for this solution. For example, the logarithm may not be negative. Furthermore if a single point is known, like a initial value, then the constant of $c_3 := c_1 - c_2 \in \mathbb{R}$ can be determined. Now we have only a general solution.

The second method is called *variation of parameters*. It works on the special form of a differential equation called *linear differential equation of first order*.

Definition 5.5. (linear differential equation of first order)
A differential equation of form

$$y' + a(x)y = f(x)$$

with functions f, a defined on an interval I is called a *linear differential equation of first order*. The right side f is called *source term*. If $f = 0$ holds, then the differential equation

$$y' + a(x)y = 0$$

is called *homogeneous linear differential equation of first order*.

First we give the method to solve the homogeneous linear differential equation of first order.

Theorem 5.6.
Let $y' + a(x)y = 0$ be a homogeneous linear differential equation of first order and a be continuous on the interval I. Then the solution to the differential equation is given by

$$y(x) = Ce^{-A(x)},\ C \in \mathbb{R}$$

with an anti-derivative $A(x) = \int a(x)\mathrm{d}x$.

The solution to solve homogeneous linear differential equations of first order can be generalized by variation of its constants to general linear differential equations of first order.

Theorem 5.7. (variation of parameters)
Let $y' + a(x)y = f(x)$ be a linear differential equation of first order and a be continuous on the interval I and the initial value $y(x_0) = y_0$ is given. Then the solution to the initial value problem is given by

$$y(x) = e^{-A(x)} \left(y_0 + \int_{x_0}^{x} f(\xi)e^{A(\xi)}\, \mathrm{d}\xi \right)$$

with the anti-derivaitve $A(x) = \int_{x_0}^{x} a(\xi)\, \mathrm{d}\xi$.

Example 5.8.
Solve the linear differential equation

$$y' + 2xy = xe^{-x^2}$$

with initial value $(0, 1)$. Hence we have $a(x) = 2x$ and $f(x) = xe^{-x^2}$. Integrating $a(x)$ yields

$$\begin{aligned} A(x) &= \int_{x_0}^{x} a(\xi)\, \mathrm{d}\xi \\ &= \int_{0}^{x} 2\xi\, \mathrm{d}\xi \\ &= [\xi^2]_0^x \\ &= x^2. \end{aligned}$$

So the solution for the differential equation is

$$
\begin{aligned}
y(x) &= e^{-A(x)} \left(y_0 + \int_{x_0}^{x} f(\xi) e^{A(\xi)} \, \mathrm{d}\xi \right) \\
&= e^{-x^2} \cdot \left(1 + \int_{0}^{x} \xi e^{-\xi^2} \cdot e^{\xi^2} \, \mathrm{d}\xi \right) \\
&= e^{-x^2} \cdot \left(1 + \int_{0}^{x} \xi \, \mathrm{d}\xi \right) \\
&= e^{-x^2} \cdot \left(1 + \left[\frac{\xi^2}{2} \right]_0^x \right) \\
&= e^{-x^2} \cdot \left(1 + \frac{x^2}{2} \right).
\end{aligned}
$$

5.2 Non-linear ordinary differential equations of first order

We will continue with two special types of nonlinear differential equations. The next type is called a *Bernoulli differential equation*.

Definition 5.9. (Bernoulli differential equation)
Let $I \subseteq \mathbb{R}$ be an interval and $a, f : I \to \mathbb{R}$ be continuous functions. The differential equation

$$y' + a(x)y = f(x)y^{\alpha}, \ \alpha \in \mathbb{R}$$

is called *Bernoulli differential equation*.

Remark 5.10.
Observe that in the cases that $\alpha = 0$ or $\alpha = 1$ holds, that the differential equation is a linear differential equation.

Since we know how to solve Bernoulli differential equations if $\alpha = 0$ or $\alpha = 1$, we will only give a solution scheme for $0 \neq \alpha \neq 1$. The scheme is based on substituting the function y with another function z and to solve the differential equation in z before resubstituting z with y again.

Theorem 5.11.
Let a Bernoulli differential equation

$$y' + a(x)y = f(x)y^{\alpha}, \ \alpha \in \mathbb{R} \setminus \{0, 1\}$$

be given. Substitution of $z = y^{1-\alpha}$ yields the linear differential equation

$$z' + (1-\alpha)a(x)z = (1-\alpha)f(x)$$

in z. Then the solution to Bernoulli's differential equation is given by

$$y = z^{\frac{1}{1-\alpha}}.$$

Example 5.12.
Solve the Bernoulli differential equation

$$y' + 2xy = 2xy^2.$$

So we have $a(x) = 2x$, $f(x) = 2x$ and $\alpha = 2$. Hence we substitute $z = y^{1-\alpha} = \frac{1}{y}$ and we get the linear differential equation

$$\begin{aligned} & z' + (1-2)2xz = (1-2)2x \\ \Leftrightarrow \qquad & z' - 2xz = -2x. \end{aligned}$$

Solving the linear differential equation yields

$$z = 1 + Ce^{x^2}$$

which resubstitutes with

$$y = \frac{1}{z}$$

into

$$y = \frac{1}{1 + Ce^{x^2}}.$$

Before we move on to systems of linear differential equations, we want to add one last method. This method is based upon substitution, too, but one special solution is needed for the substitution.

Definition 5.13. (Riccati equation)
Let $I \subseteq \mathbb{R}$ be an interval and $a, b, f : I \to \mathbb{R}$ be continuous functions. The differential equation

$$y' + a(x)y + b(x)y^2 = f(x)$$

is called a *Riccati equation*.

Remark 5.14.
If $b(x) = 0$ the differential equation is a linear differential equation of first order. If $f(x) = 0$ holds, then the differential equation is a Bernoulli differential equation.

Theorem 5.15.
Let a Riccati equation

$$y' + a(x)y + b(x)y^2 = f(x)$$

be given. Let y_s be a special solution for the Riccati equation. Substitution by $z = \frac{1}{y-y_s}$ yields the linear differential equation

$$z' - [a(x) + 2b(x)y_s]z = b(x).$$

Then the solution to the Riccati equation is given by

$$y = y_s + \frac{1}{z}.$$

Example 5.16.
Solve the Riccati equation

$$y' - 3xy + xy^2 = -2x.$$

So we have $a(x) = -3x$, $b(x) = x$ and $f(x) = -2x$. First we guess a special solution y_s.

$$y_s(x) = 1$$

solves the Ricatti equation. So we substitute $z = \frac{1}{y-y_s} = \frac{1}{y-1}$ and we get the linear differential equation

$$\begin{aligned} & z' - [-3x + 2x \cdot 1]z = x \\ \Leftrightarrow \quad & z' + xz = x. \end{aligned}$$

Solving that linear differential equation yields

$$z = 1 + Ce^{-\frac{x^2}{2}}$$

which resubstitutes with

$$y = y_s + \frac{1}{z}$$

into

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

5.3 Systems of linear differential equations

Definition 5.17. (system of linear differential equations of first order)
Let

$$A(x) := \begin{pmatrix} a_{11}(x) & ... & a_{1n}(x) \\ \vdots & \ddots & \vdots \\ a_{n1}(x) & ... & a_{nn}(x) \end{pmatrix}, \; b(x) := \begin{pmatrix} b_1(x) \\ \vdots \\ b_n(x) \end{pmatrix}$$

be a matrix and a vector with functions $a_{ij}(x)$ and $b_i(x)$, $i, j \in \{1, ..., n\}$. Then

$$y' = A(x)y + b(x), \; y = (y_1, ..., y_n)$$

is called a *system of linear differential equations of first order*. If $b(x) = 0$ holds, then it is called *homogeneous*.

We will only look at the homogeneous systems of linear differential equations of first order, since the inhomogeneous system is too difficult for our means. If we have a homogeneous system, then the *fundamental system* forms some kind basis of the solutions for our differential equation.

Definition 5.18. (fundamental system)
A system of functions

$$\overline{y}_1, \overline{y}_2, ..., \overline{y}_n, \text{ with } \overline{y} \in \mathbb{R}^n$$

of n linearly independent solutions of the homogeneous system $y' = A(x)y$ is called a *fundamental system* or a *basis* of solutions.

Theorem 5.19.
If $\overline{y}_1, \overline{y}_2, ..., \overline{y}_n$ is a fundamental system of $y' = A(x)y$, then any solution y may be constructed in the form

$$y = c_1\overline{y}_1 + c_2\overline{y}_2 + ... + c_n\overline{y}_2, \text{ with } c_1, c_2, ..., c_n \in \mathbb{R}.$$

Since the general problem to solve this system is out of our league, we will solve an easier and much smaller problem. We will have, that the vector $b(x) = 0$ and therefore we have a homogeneous system. Furthermore the matrix A will satisfy, that all functions $a_{ij}(x)$ are constant, that is $A(x) \in \mathbb{R}^{n \times n}$ holds.

Theorem 5.20.
let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix, that is $A = A^T$. Let $v_1, v_2, ..., v_n$ be a basis of the eigenvectors and $\lambda_1, ..., \lambda_n$ the eigenvalues of A. Then

$$(v_1 e^{\lambda_1 x}, ..., v_n e^{\lambda_n x})$$

is a fundamental system of $y' = Ay$ and the solution to the differential equation is given by the linear combination

$$y = c_1 v_1 e^{\lambda_1 x} + ... + c_n v_n e^{\lambda_n x}, \text{ with } c_1, ..., c_n \in \mathbb{R}.$$

Example 5.21.
Solve the system

$$\begin{aligned} y_1' &= -2y_1 + y_2 \\ y_2' &= y_1 \\ y_3' &= -y_3. \end{aligned}$$

We deduce the system

$$y' = Ay$$

with

$$A = \begin{pmatrix} -2 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \ y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \text{ and } y' = \begin{pmatrix} y_1' \\ y_2' \\ y_3' \end{pmatrix}.$$

Hence the eigenvalues are $\lambda_1 = -1$, $\lambda_2 = \sqrt{2} - 1$ and $\lambda_3 = -1 - \sqrt{2}$ with their eigenvectors

$$v_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \ v_2 = \begin{pmatrix} \frac{1}{1+\sqrt{2}} \\ 1 \\ 0 \end{pmatrix}, \text{ and } v_3 = \begin{pmatrix} \frac{1}{1-\sqrt{2}} \\ 1 \\ 0 \end{pmatrix}.$$

So

$$(v_1 e^{\lambda_1 x}, v_2 e^{\lambda_2 x}, v_3 e^{\lambda_3 x})$$

is a fundamental system and the solution to the differential equation is given by

$$y(x) = c_1 \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} e^{-x} + c_2 \cdot \begin{pmatrix} \frac{1}{1+\sqrt{2}} \\ 1 \\ 0 \end{pmatrix} e^{(\sqrt{2}-1)x} + c_3 \cdot \begin{pmatrix} \frac{1}{1-\sqrt{2}} \\ 1 \\ 0 \end{pmatrix} e^{(-1-\sqrt{2})x}, \ c_1, c_2, c_3 \in \mathbb{R}.$$

Remark 5.22.
In the last theorem the condition was, that $A \in \mathbb{R}^{n \times n}$ is symmetric. That yields, that the multiplicity of any eigenvalue of A equals the geometric multiplicity of that eigenvalue. Furthermore all eigenvalues are real numbers. Hence we used a theorem that is only able to solve a few systems. But the last theorem may be extended to matrices where there are complex eigenvalues as well as to systems where the multiplicity of an eigenvalue does not match its geometric multiplicity. In that case, the way to solve those systems and the formula of the fundamental system itself will change slightly.

6 Appendix

6.1 Determinants for bigger sized matrices

6.2 Gram-Schmidt

6.3 Spectral Theorem

6.4 Cramer's Rule

6.5 Taylor Series

Definition 6.1. (smooth function)
A function is called *smooth*, if it has derivatives of all orders everywhere in its domain.

Definition 6.2. (Taylor series)
Let $f : I \to \mathbb{R}$ with $I \subset \mathbb{R}$ be a smooth function. Then the power series

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k$$

is called a *Taylor series* centered at x_0. $f^{(k)}$ denotes the k-th derivative of f.

Example 6.3.
We show that

$$\sin x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}$$

holds. For that we center the Taylor series at $x_0 = 0$. By the Taylor formula we have

$$\sin(x) = \sin(0) + \sin'(0)x + ... + \frac{\sin^{(n)}(0)}{n!}x^n + R_{n,0}(x)$$

with

$$R_{n,0}(x) = \frac{\sin^{(n+1)}(t)}{(n+1)!}x^{n+1}.$$

The higher derivatives of $\sin(x)$ are $\pm\cos(x)$ and $\pm\sin(x)$. Since the Taylor series is centered at 0, the $\sin(x)$ and its derivatives in 0 will yield $0, 1, 0, -1, 0, 1, 0, -1$. So $|\sin^{(n+1)}(t)| \leq 1$ and by that

$$\left|\frac{\sin^{(n+1)}(t)}{(n+1)!}x^{n+1}\right| \leq \left|\frac{x^{n+1}}{(n+1)!}\right|$$

holds. Hence for all fixed x the term $R_{n,0}$ converges to 0. Furthermore the series converges with the comparison test, since $\sum_{k=0}^{\infty} \frac{x^k}{k!}$ converges. So we have

$$\begin{aligned}
\sin x &= \sum_{n=0}^{\infty} \frac{\sin^{(n)}(0)}{n!} x^n = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + ... \\
&= \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}
\end{aligned}$$

Example 6.4.
Analoguesly one can determine, that

$$\cos x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}$$

or that

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

holds.

6.6 Fourier Series

6.7 Multivariable Calculus

Definition 6.5. (distance, absolute value)
Let $x, y \in \mathbb{R}^n$ be two points in the n-dimensional Euclidean space. Then

$$\|x - y\| = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{\frac{1}{2}}$$

is the *distance between the points x and y*.

Example 6.6.
Let $x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ and $y = \begin{pmatrix} -2 \\ 5 \\ -3 \end{pmatrix}$ be two points in $\mathbb{R}^3$. Then the distance between x and y is

$$\begin{aligned}
\|x - y\| &= \left(\sum_{i=1}^{3} (x_i - y_i)^2 \right)^{\frac{1}{2}} = \sqrt{(1-(-2))^2 + (2-5)^2 + (3-(-3))^2} \\
&= \sqrt{3^2 + (-3)^2 + 6^2} = \sqrt{9+9+36} = \sqrt{54}.
\end{aligned}$$

In the next definitions it is important to know that for x^m the value m states an index and not an exponent or power. It is used to differentiate x^m from x_m, where x_m denotes the m-th entry of the vector x. x^m on the other hand denotes the m-th element of the sequence $(x^m)_m$.

Definition 6.7. (convergent, limit)
Let $(x^m)_m$ be a sequence with points $x^m \in \mathbb{R}^n$. Then the sequence $(x^m)_m$ *converges to* a, if for all $\varepsilon > 0$ there exists an $n_0 := n_0(\varepsilon) \in \mathbb{N}$, such that $\|x^n - a\| < \varepsilon$ holds for all $n > n_0$. If the sequence converges, then a is called the *limit* denoted by

$$\lim_{m\to\infty} x^m = a.$$

Definition 6.8. (set closure)
Let $A \subseteq \mathbb{R}^n$ be subset of the real numbers. Then we denote by $\overline{A}$ the *closure* of A. The closure of A consists of all points of A and all boundary points of A. The set of all boundary points is denoted by ∂A.

Remark 6.9.
So $\overline{A} = A \cup \partial A$.

Example 6.10.

i) The closure of $A = (1,2)$ is $\overline{A} = [1,2]$.

ii) The closure of $A = (1,2]$ is $\overline{A} = [1,2]$.

iii) The closure of $A = (1,2] \times [1,2) \times (1,2)$ is $\overline{A} = [1,2] \times [1,2] \times [1,2]$.

Definition 6.11. (convergent, limit)
Let $f : I \to \mathbb{R}^m$, $I \subseteq \mathbb{R}^n$ be a function, $x_0 \in \overline{I}$ and $a \in \mathbb{R}^m$. Then $f(x)$ *converges* to a for x approaches x_0, that is $\lim\limits_{x\to x_0} f(x) = a$, if and only if for all $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\|f(x) - a\| < \varepsilon$$

for all $x \in I$ with $\|x - x_0\| < \delta$. Then

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

is called the *limit* of f at x_0.

Definition 6.12. (continuous)
Let $f : I \to \mathbb{R}$ be a function with $I \subseteq \mathbb{R}^n$.

i) The function f is called *continuous* at $x_0 \in I$, if

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

holds.

ii) The function f is called *continuous* on I, if f is continuous in each point $x_0 \in I$.

Example 6.13.

i) Let $f : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}$ be a function with

$$f(x,y) = \frac{x^2 y}{x^2 + y^2}$$

Show that f is *continuously extendable* in $x_0 = (0,0)$ with $f(0,0) = 0$.
An often used way to show such a condition is to use polar coordinates. The argument is free to take any angle but we try to decrease the radius to zero to shrink the cricle spanned by the argument to a point. So we substitute (x,y) by $(r\cos\varphi, r\sin\varphi)$ and approach the origin, that is

$$(x,y) = (r\cos\varphi, r\sin\varphi) \to (0,0).$$

So for f we have

$$\begin{aligned} f(r\cos\varphi, r\sin\varphi) &= \frac{r^2\cos^2\varphi \cdot r\sin\varphi}{r^2\cos^2\varphi + r^2\sin^2\varphi} \\ &= \frac{r^3}{r^2}\cos^2\varphi \cdot \sin\varphi \\ &= r\cos^2\varphi \cdot \sin\varphi \\ &\xrightarrow{r\to 0} 0, \end{aligned}$$

since $|\cos^2\varphi \cdot \sin\varphi|$ is bounded. So $\lim\limits_{(x,y)\to(0,0)} f(x,y) = f(0,0) = 0$ holds. Hence f is continuously extendable at $(0,0)$ with $f(0,0) = 0$.

ii) Let $f : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}$ be a function with

$$f(x,y) = \frac{x^2}{x^2 + y}$$

Show that f is not continuously extendable in $x_0 = (0,0)$. We show this by using two different sequences, that approach $(0,0)$ but have a different limit.
We choose $(\frac{1}{n}, \frac{1}{n})_n$ as our first sequence with $(\frac{1}{n}, \frac{1}{n}) \xrightarrow{n\to\infty} (0,0)$ and

$$\begin{aligned} f\left(\frac{1}{n}, \frac{1}{n}\right) &= \frac{\frac{1}{n^2}}{\frac{1}{n^2} + \frac{1}{n}} \\ &= \frac{\frac{1}{n^2}}{\frac{n+1}{n^2}} \\ &= \frac{n^2}{n^2(n+1)} \\ &\xrightarrow{n\to\infty} 0. \end{aligned}$$

We choose $(\frac{1}{n},0)_n$ as our second sequence with $(\frac{1}{n},0) \overset{n\to\infty}{\longrightarrow} (0,0)$ and

$$\begin{aligned} f\left(\frac{1}{n},0\right) &= \frac{\frac{1}{n^2}}{\frac{1}{n^2}+0} \\ &= \frac{\frac{1}{n^2}}{\frac{1}{n^2}} \\ &= 1 \\ &\overset{n\to\infty}{\longrightarrow} 1. \end{aligned}$$

Hence the function is not continuously extendable in $(0,0)$, since two different sequences, that approach the origin, yield different limits in the function.

Definition 6.14. ((total) differentiable, (total) derivative)
Let $f : I \to \mathbb{R}^m$ be a function with $I \subseteq \mathbb{R}^n$ and $x_0 \in I$ is an inner point of I. Then f is *(total) differentiable* at x_0, if there exists a linear map $h : \mathbb{R}^n \to \mathbb{R}^m$, such that

$$\lim_{x\to x_0} \frac{R(x)}{\|x - x_0\|} = 0$$

holds with $R(x) := f(x) - f(x_0) - h(x - x_0)$. In this case the linear function will be denoted by $Df(a)$ and is called the *(total) derivative* or *(total) differential* of f at x_0.

Remark 6.15.
The (total) derivative is uniquely determined.

Example 6.16.
Let

$$f : \mathbb{R}^2 \to \mathbb{R} : x \mapsto \begin{cases} xy\frac{x^2-y^2}{x^2+y^2} & , (x,y) \neq (0,0) \\ 0 & , (x,y) = (0,0). \end{cases}$$

be a function. f is continuous in $(0,0)$. We want to show, that f is (total) differentiable in $(0,0)$. So it holds:

$$\begin{aligned} \lim_{(x,y)\to(0,0)} \frac{|f(x,y) - f(0,0)| - 0}{\|(x,y) - (0,0)\|} &= \lim_{(x,y)\to(0,0)} \frac{|xy| \cdot \frac{|x^2-y^2|}{x^2+y^2}}{\sqrt{x^2+y^2}} \\ &= \lim_{(x,y)\to(0,0)} \frac{|xy| \cdot |x^2 - y^2|}{(x^2+y^2)^{\frac{3}{2}}} \\ &= \lim_{r\to 0} \frac{r^2|\cos\varphi \sin\varphi| \cdot |r^2\cos^2\varphi - r^2\sin^2\varphi|}{r^3} \\ &= \lim_{r\to 0} \frac{r^4|\cos\varphi \sin\varphi| \cdot |\cos^2\varphi - \sin^2\varphi|}{r^3} \\ &= \lim_{r\to 0} r|\cos\varphi \sin\varphi| \cdot |\cos^2\varphi - \sin^2\varphi| \\ &= 0. \end{aligned}$$

Hence f is (total) differentiable in $(0,0)$.

Definition 6.17. (*i*-th partial derivative)
Let $f : I \to \mathbb{R}$ be a function with $I \subset \mathbb{R}^n$ and x_0 is an inner point of I. Then the *i-th partial derivative* of f at x_0 is defined as

$$D_i f(x_0) = \lim_{t \to 0} \frac{f(x_0 + t \cdot e_i) - f(x_0)}{t},$$

if it exists.

Remark 6.18.
For derivatives in one variable the differentiablility implied that the function is continuous. Now the partial derivatives do not imply that the original function is continuous.

Remark 6.19.
$D_i f$ is also often refered to as $\frac{\partial f}{\partial x_i}$ or for a function $f(x,y,z)$ we could have $\frac{\partial f(x,y,z)}{\partial y}$.

Remark 6.20.
The last definition can be extended for directions, that are not along the unit vectors.

Example 6.21.
Let

$$f : \mathbb{R}^3 \to \mathbb{R} : (x,y,z) \mapsto xyz + x^2yz + z^3$$

be a function. then the partial derivatives are

$$\begin{aligned}
\frac{\partial f}{\partial x}(x,y,z) &= yz + 2xyz, \\
\frac{\partial f}{\partial y}(x,y,z) &= xz + x^2z, \\
\frac{\partial f}{\partial z}(x,y,z) &= xy + x^2y + 3z^2.
\end{aligned}$$

Only because the partial derivatives exist, we do not necessarily have a total differentiable function. The next theorem connects partial derivatives to the total derivative.

Theorem 6.22.
Let $f : I \to \mathbb{R}^m$ be a function with $I \subset \mathbb{R}^n$ and x_0 is an inner point of I. If all partial derivatives $D_j f_i(a)$ exists in an area around x_0 and are also continuous at x_0, then f is (total) differentiable in x_0.

Theorem 6.23. (Mean value theorem in $\mathbb{R}^n$)
Let $f : I \to \mathbb{R}$ with $I \subseteq \mathbb{R}^n$ be a function that is continuously differentiable. Furthermore let $a, b \in I$ be points which link is a subset of I. Then there exists an $\xi \in (0,1)$ with

$$f(b) - f(a) = Df[a + \xi(b-a)](b-a).$$

Definition 6.24. (Jacobian matrix)
Let $f : I \to \mathbb{R}^m$ with $I \subseteq \mathbb{R}^n$ be a function which is differentiable in an inner point x of I. Then the $m \times n$ matrix

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

is called *Jacobian matrix* of f at x.

Example 6.25.
Let

$$f : \mathbb{R}^2 \to \mathbb{R}^3 : (x, y) \mapsto (x^2 y,\ 3xy^4 + x^2 y^2,\ 3 + xy^2)$$

be a function. Then the Jacobian matrix of f is

$$Df(x) = \begin{pmatrix} 2xy & x^2 \\ 3y^4 + 2xy^2 & 12xy^3 + 2x^2 y \\ y^2 & 2xy \end{pmatrix}.$$

Theorem 6.26. (Schwarz)
Let $f : I \to \mathbb{R}$ be a function on the open interval I. If f is two times continuously differentiable, then

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

holds for all i, j.

Example 6.27.
Let

$$f : \mathbb{R}^2_+ \to \mathbb{R} : (x, y) \mapsto \frac{xy^2 \exp y}{\log x}$$

be a function. Then we have

$$D_x f(x, y) = \frac{\partial f}{\partial x}(x, y) = \frac{\log x - 1}{\log^2 x} y^2 \exp y$$

and

$$\begin{aligned} D_y f(x, y) = \frac{\partial f}{\partial y}(x, y) &= \frac{x}{\log x}(2y \exp y + y^2 \exp y) \\ &= \frac{x}{\log x}(2 + y) y \exp y. \end{aligned}$$

Both will now yield

$$\frac{\partial^2 f}{\partial x \partial y}(x,y) = \frac{\log x - 1}{\log^2 x}(2+y)y \exp y = \frac{\partial^2 f}{\partial y \partial x}(x,y).$$

Definition 6.28. (Hessian matrix)
Let $f : I \to \mathbb{R}$ with $I \subseteq \mathbb{R}^n$ be a function which is two times differentiable in an inner point x of I. Then the $n \times n$ matrix

$$H_f(x) := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(x) & & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(x) \end{pmatrix}$$

is called *Hessian matrix* of f at x.

Remark 6.29.
The Hessian matrix is symmetric by the theorem of Schwarz, if it is not only two times differentiable but two times continuously differentiable. In general the Hessian matrix does not need to be symmetric.

Example 6.30.
Let

$$f : \mathbb{R}^2 \to \mathbb{R} : (x,y) \mapsto 3x + x^2y^2 + \exp y$$

be a function. Determine the Hessian matrix at $(1,0)$.
The partial derivatives are

$$\frac{\partial f}{\partial x}(x,y) = 3 + 2xy^2$$
$$\frac{\partial f}{\partial y}(x,y) = 2x^2y + \exp y$$

and we get

$$\frac{\partial^2 f}{\partial x \partial y}(x,y) = 4xy = \frac{\partial^2 f}{\partial x \partial y}(x,y).$$

Then the Hessian matrix at $(1,0)$ is

$$H_f(1,0) = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}.$$

Definition 6.31. (Taylor polynomial of second degree in $\mathbb{R}^n$)
Let $f : I \to \mathbb{R}$ be a two times continuously differentiable function with an open subset $I \subseteq \mathbb{R}^n$. Then the *Taylor polynomial of second degree* of f at $x_0 \in I$ is

$$T_{2,x_0} = f(x_0) + Df(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^T H_f(x_0)(x - x_0).$$

Example 6.32.
Let

$$f(x, y) = x \sin y + y \sin x$$

be a function. Determine $T_{2,(\frac{\pi}{2},0)}$. We have

$$f(x_0) = f\left(\frac{\pi}{2}, 0\right) = 0$$

and

$$\begin{aligned} Df(x, y) &= \begin{pmatrix} \sin y + y \cos x & x \cos y + \sin x \end{pmatrix} \\ \Rightarrow Df\left(\frac{\pi}{2}, 0\right) &= \begin{pmatrix} 0 & \frac{\pi}{2} + 1 \end{pmatrix}. \end{aligned}$$

Hence the Hessian matrix is

$$\begin{aligned} H_f(x) &= \begin{pmatrix} -y \sin x & \cos y + \cos x \\ \cos y + \cos x & -x \sin y \end{pmatrix} \\ \Rightarrow H_f\left(\frac{\pi}{2}, 0\right) &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \end{aligned}$$

So for the Taylor polynomial of second degree we get

$$\begin{aligned} T_{2,(\frac{\pi}{2},0)} &= 0 + \begin{pmatrix} 0 & \frac{\pi}{2} + 1 \end{pmatrix} \cdot \begin{pmatrix} x - \frac{\pi}{2} \\ y \end{pmatrix} + \frac{1}{2} \begin{pmatrix} x - \frac{\pi}{2} & y \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} x - \frac{\pi}{2} \\ y \end{pmatrix} \\ &= y + xy. \end{aligned}$$

Theorem 6.33. (Weierstraß)
Let $f : I \to \mathbb{R}$ be a continuous function with $I \subseteq \mathbb{R}^n$ is a compact space. Then there are $a, b \in I$ such that $f(a) \leq f(x) \leq f(b)$ holds for all $x \in I$.

In other words the theorem of Weierstraß states, that a continuous function on a compact space has a maximum and minimum value.

Definition 6.34. (minimum, maximum)

Theorem 6.35. (necessary condition)

Theorem 6.36. (sufficient condition)

Theorem 6.37. (Lagrange)