Mathematics for Physicists

https://www.github.com/Lauchmelder23/Mathematics
Alma Mater Lipsiensis

March 29, 2021

Contents

Chapter 1

Fundamentals and Notation

1.1 Logic

Definition 1.1 (Statements)**.** A statement is a sentence (mathematically or colloquially) which can be either true or false.

Example 1.2. Statements are

- Tomorrow is Monday
- $x > 1$ where x is a natural number
- Green rabbits grow at full moon

No statements are

- What is a statement?
- $x + 20y$ where x, y are natural numbers
- This sentence is false

Definition 1.3 (Connectives)**.** When Φ, Ψ are statements, then

(i) $\neg\Phi$ (not Φ)

(ii) $\Phi \wedge \Psi$ (Φ and Ψ)

(iii) $\Phi \vee \Psi$ (Φ or Ψ)

(iv) $\Phi \implies \Psi$ (if Φ then Ψ)

(v) $\Phi \iff \Psi$ (Φ if and only if (iff.) Ψ)

are also statements. We can represent connectives with truth tables

Φ	Ψ	$\neg\Phi$	$\Phi \wedge \Psi$	$\Phi \vee \Psi$	$\Phi \implies \Psi$	$\Phi \iff \Psi$
t	t	f	t	t	t	t
t	f	f	f	t	f	f
f	t	t	f	t	t	f
f	f	t	f	f	t	t

Remark 1.4.

(i) $\vee$ is inclusive

(ii) $\Phi \implies \Psi$, $\Phi \impliedby \Psi$, $\Phi \iff \Psi$ are NOT the same

(iii) $\Phi \implies \Psi$ is always true if Φ is false (ex falso quodlibet)

Definition 1.5 (Hierarchy of logical operators)**.** $\neg$ is stronger than $\wedge$ and $\vee$, which are stronger than $\implies$ and $\iff$.

Example 1.6.

$$\begin{aligned}
\neg\Phi \wedge \Psi &\cong (\neg\Phi) \wedge \Psi \\
\neg\Phi \implies \Psi &\cong (\neg\Phi) \wedge \Psi \\
\Phi \wedge \Psi \iff \Psi &\cong (\Phi \wedge \Psi) \iff \Psi \\
\neg\Phi \vee \neg\Psi \implies \neg\Psi \wedge \Psi &\cong ((\neg\Phi) \vee (\neg\Psi)) \implies ((\neg\Psi) \wedge \Psi)
\end{aligned}$$

We avoid writing statements like $\Phi \wedge \Psi \vee \Theta$. A statement that is always true is called a tautology. Some important equivalencies are

$$\begin{aligned}
\Phi &\text{ equiv. } \neg(\neg\Phi)) \\
\Phi \implies \Psi &\text{ equiv. } \neg\Psi \implies \neg\Phi \\
\Phi \iff \Psi &\text{ equiv. } (\Phi \implies \Psi) \wedge (\Psi \implies \Phi) \\
\Phi \vee \Psi &\text{ equiv. } \neg(\neg\Phi \wedge \neg\Psi)
\end{aligned}$$

Logical operators are commutative, associative and distributive.

Definition 1.7 (Quantifiers)**.** Let $\Phi(x)$ be a statement depending on x. Then $\forall x\ \Phi(x)$ and $\exists x\ \Phi(x)$ are also statements. The interpretation of these statements is

- $\forall x\ \Phi(x)$: "For all x, $\Phi(x)$ holds."
- $\exists x\ \Phi(x)$: "There is (at least one) x s.t. $\Phi(x)$ holds."

Remark 1.8.

(i) $\forall x\ x \geq 1$ is true for natural numbers, but not for integers. We must specify a domain.

(ii) If the domain is infinite the truth value of $\forall x\ \Phi(x)$ cannot be algorithmically determined.

(iii) $\forall x\ \Phi(x)$ and $\forall y\ \Phi(y)$ are equivalent.

(iv) Same operators can be exchanged, different ones cannot.

(v) $\forall x\ \Phi(x)$ is equivalent to $\neg\exists x\ \neg\Phi(x)$.

1.2 Sets and Functions

Definition 1.9. A set is an imaginary "container" for mathematical objects. If A is a set we write

- $x \in A$ for "x is an element of A"
- $x \notin A$ for $\neg x \in A$

There are some specific types of sets

(i) $\varnothing$ is the empty set which contains no elements. Formally: $\exists x \forall y\ y \notin x$

(ii) Finite sets: $\{1, 3, 7, 20\}$

(iii) Let $\Phi(x)$ be a statement and A a set. Then $\{x \in A \mid \Phi(x)\}$ is the set of all elements from A such that $\Phi(x)$ holds.

There are relation operators between sets. Let A, B be sets

(i) $A \subset B$ means "A is a subset of B".

(ii) $A = B$ means "A and B are the same"

Each element can appear only once in a set, and there is no specific ordering to these elements. This means that $\{1, 3, 3, 7\} = \{3, 1, 7\}$. There are also operators between sets

(i) $A \cup B$ is the union of A and B.

$$x \in A \cup B \iff x \in A \vee x \in B$$

(ii) $A \cap B$ is the intersection of A and B.

$$x \in A \cap B \iff x \in A \wedge x \in B$$

This can be expanded to more than two sets ($A \cup B \cup C$). We can also use the following notation. Let A be a set of sets. Then

$$\bigcup_{C \in A} C$$

is the union of all sets contained in A.

(iii) $A \setminus B$ is the difference of A and B.

$$x \in A \setminus B \iff x \in A \wedge x \notin B$$

(iv) The power set of a set A is the set of all subsets of A. Example:

$$\mathcal{P}(\{1,2\}) = \{\varnothing, \{1\}, \{2\}, \{1,2\}\}$$

Theorem 1.10. *Let A, B, C be sets. Then*

$$\begin{aligned} A \setminus (B \cup C) &= (A \setminus B) \cap (A \setminus C) \\ A \setminus (B \cap C) &= (A \setminus B) \cup (A \setminus C) \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \\ A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \end{aligned}$$

Proof. Let A, B, C be sets.

$$\begin{aligned} x \in A \cap (B \cup C) &\iff x \in A \wedge x \in B \cup C \\ &\iff x \in A \wedge (x \in B \vee x \in C) \\ &\iff (x \in A \wedge x \in B) \vee (x \in A \wedge x \in C) \qquad (1.1) \\ &\iff x \in A \cap B \vee x\ inA \cap C \\ &\iff x \in (A \cap B) \cup (A \cap C) \end{aligned}$$

The other equations are left as an exercise to the reader. □

Definition 1.11. Let A, B be sets. For $x \in A$, $y \in B$ we call (x, y) the ordered pair from x, y. The Cartesian product is defined as

$$A \times B = \{(x, y) \,|\, x \in A \wedge y \in B\}$$

Remark 1.12.

(i) (x, y) is NOT equivalent to $\{x, y\}$. The former is an ordered pair, the latter a set. It is important to note that

$$(x, y) = (a, b) \iff x = a \wedge y = b$$

(ii) This can be extended to triplets, quadruplets, ...

$$A \times B \times C = \{(x, y, z) \,|\, x \in A \wedge y \in B \wedge z \in C\}$$

We use the notation $A \times A = A^2$

(iii) For $\mathbb{R}^2$ ($\mathbb{R}$ are the real numbers) we can view (x, y) as coordinates of a point in the plane.

Definition 1.13. Let A, B be sets. A mapping f from A to B assigns each $x \in A$ exactly one element $f(x) \in B$. A is called the domain and B the codomain.

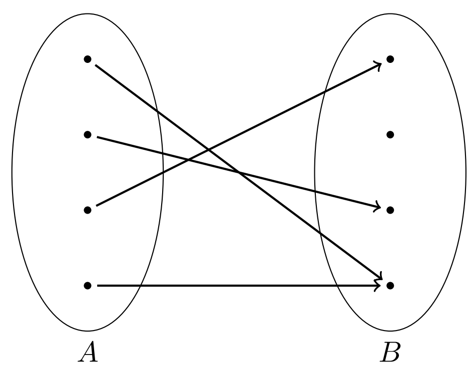


Figure 1.1: A mapping $f : A \to B$

As shown in figure 1.1, every element from A is assigned exactly one element from B, but not every element from B must be assigned to an element from A, and elements from B can be assigned more than one element from A. The notation for such mappings is

$$f : A \longrightarrow B$$

A mapping that has numbers ($\mathbb{N}, \mathbb{R}, \cdots$) as the codomain is called a function.

Example 1.14.

(i)

$$\begin{aligned} f : \mathbb{N} &\longrightarrow \mathbb{N} \\ n &\longmapsto 2n+1 \end{aligned}$$

(ii)

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \begin{cases} 0 & x \text{ rational} \\ 1 & x \text{ irrational} \end{cases} \end{aligned}$$

(iii) Addition on $\mathbb{N}$

$$f : \mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{N}$$

Instead of $f(x, y)$ we typically write $x + y$ for addition.

(iv) The identity mapping is defined as

$$\begin{aligned} \mathrm{id}_A : A &\longrightarrow A \\ x &\longmapsto x \end{aligned}$$

Remark 1.15 (Mappings as sets).

(i) A mapping $f : A \to B$ corresponds to a subset of $F = A \times B$, such that

$$\begin{aligned} &\forall x \in A\ \forall y, z \in B \quad (x, y) \in F \wedge (x, z) \in F \implies y = z \\ &\forall x \in A\ \exists y \in B \quad (x, y) \in F \end{aligned}$$

(ii) Simply writing "Let the function $f(x) = x^2$..." is NOT mathematically rigorous.

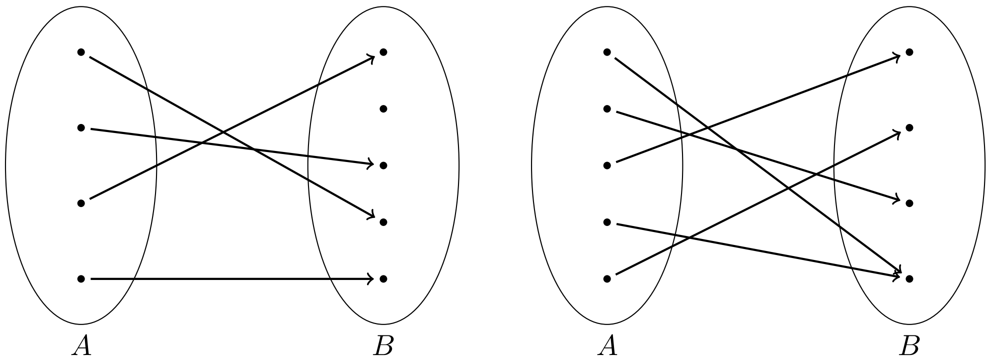
(iii)

$$f \text{ is a mapping from } A \text{ to } B \iff f(x) \text{ is a value in } B$$

(iv)

$$f, g : A \longrightarrow B \text{ are the same mapping} \iff \forall x \in A \quad f(x) = g(x)$$

Definition 1.16. We call $f : A \to B$



(a) Injective mapping. There is at most one arrow per point in B

(b) Surjective mapping. There is at least one arrow per point in B

Figure 1.2: Visualizations of injective and surjective mappings

- injective if $\forall x, \tilde{x} \in A \quad f(x) = f(\tilde{x}) \implies x = \tilde{x}$
- surjective if $\forall y \in B, \exists x \in A \quad f(x) = y$
- bijective if f is injective and surjective

Example 1.17.

(i)

$$\begin{aligned} f : \mathbb{N} &\longrightarrow \mathbb{N} \\ n &\longmapsto n^2 \end{aligned}$$

is not surjective (e.g. $n^2 \neq 3$), but injective.

(ii)

$$\begin{aligned} f : \mathbb{Z} &\longrightarrow \mathbb{N} \\ n &\longmapsto n^2 \end{aligned}$$

is neither surjective nor injective.

(iii)

$$\begin{aligned} f : \mathbb{N} &\longrightarrow \mathbb{N} \\ n &\longmapsto \begin{cases} \frac{n}{2} & n \text{even} \\ \frac{n+1}{2} & n \text{odd} \end{cases} \end{aligned}$$

is surjective but not injective.

Definition 1.18 (Function compositing)**.** Let $A,\ B,\ C$ be sets, and let $f : A \to B,\ g : B \to C$. Then the composition of f and g is the mapping

$$\begin{aligned} g \circ f : A &\longrightarrow C \\ x &\longmapsto g(f(x)) \end{aligned}$$

Remark 1.19. Compositing is associative (why?), but not commutative. For example let

$$\begin{aligned} f : \mathbb{N} &\longrightarrow \mathbb{N} \\ n &\longmapsto 2n \end{aligned} \qquad\qquad \begin{aligned} g : \mathbb{N} &\longrightarrow \mathbb{N} \\ n &\longmapsto n+3 \end{aligned}$$

Then

$$\begin{aligned} f \circ g(n) &= 2(n+3) = 2n+6 \\ g \circ f(n) &= 2n+3 \end{aligned}$$

Theorem 1.20. *Let $f : A \to B$ be a bijective mapping. Then there exists a mapping $f^{-1} : B \to A$ such that $f \circ f^{-1} = \mathrm{id}_B$ and $f^{-1} \circ f = \mathrm{id}_A$. f^{-1} is called the inverse function of f.*

Proof. Let $y \in B$ and f bijective. That means $\exists x \in A$ such that $f(x) = y$. Due to f being injective, this x must be unique, since if $\exists \tilde{x} \in A$ s.t. $f(\tilde{x}) = f(x) = y$, then $x = \tilde{x}$. We define $f(x) = y$ and $f^{-1}(y) = x$, therefore

$$f \circ f^{-1}(y) = f(f^{-1}(y)) = f(x) = y = \mathrm{id}_B(y) \implies f \circ f^{-1} = \mathrm{id}_B \tag{1.2}$$

and equivalently

$$f^{-1} \circ f(x) = \mathrm{id}_A(x) \implies f^{-1} \circ f = \mathrm{id}_A \tag{1.3}$$

$\square$

1.3 Numbers

Definition 1.21. The real numbers are a set $\mathbb{R}$ with the following structure

(i) Addition

$$+ : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$$

(ii) Multiplication

$$\cdot : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$$

Instead of $+(x, y)$ and $\cdot(x, y)$ we write $x + y$ and $x \cdot y$.

(iii) Order relations

$\leq$ is a relation on $\mathbb{R}$, i.e. $x \leq y$ is a statement.

Definition 1.22 (Axioms of Addition)**.**

A1: Associativity

$$\forall a, b, c \in \mathbb{R}: \quad (a+b)+c = a+(b+c)$$

A2: Existence of a neutral element

$$\exists 0 \in \mathbb{R}\ \forall x \in \mathbb{R}: \quad x+0 = x$$

A3: Existence of an inverse element

$$\forall x \in \mathbb{R}\ \exists (-x) \in \mathbb{R}: \quad x+(-x) = 0$$

A4: Commutativity

$$\forall x, y \in \mathbb{R}: \quad x+y = y+x$$

Theorem 1.23. $x, y \in \mathbb{R}$

(i) *The neutral element is unique*

(ii) $\forall x \in \mathbb{R}$ *the inverse is unique*

(iii) $-(-x) = x$

(iv) $-(x+y) = (-x)+(-y)$

Proof.

(i) Assume $a, b \in \mathbb{R}$ are both neutral elements, i.e.

$$\forall x \in \mathbb{R}: x+a = x = x+b \tag{1.4}$$

This also implies that $a+b = a$ and $b+a = b$.

$$\implies b = b+a \stackrel{\text{A4}}{=} a+b = a \tag{1.5}$$

Therefore $a = b$.

(ii) Assume $c, d \in \mathbb{R}$ are both inverse elements of $x \in \mathbb{R}$, i.e.

$$x + c = 0 = x + d \tag{1.6}$$

$$c = 0 + c = x + d + c \overset{\text{A4}}{=} x + c + d = 0 + d = d \tag{1.7}$$

Therefore $c = d$.

(iii) Left as an exercise for the reader.

(iv)

$$\begin{aligned} x + y + ((-x) + (-y)) &= x + y + (-x) + (-y) \\ &\overset{\text{A4}}{=} x + (-x) + y + (-y) = 0 \end{aligned} \tag{1.8}$$

Therefore $(-x) + (-y)$ is the inverse element of $(x + y)$, i.e. $-(x + y) = (-x) + (-y)$.

$\square$

Definition 1.24 (Axioms of Multiplication)**.**

M1: $\forall x, y, z \in \mathbb{R} : \quad (xy)z = x(yz)$

M2: $\exists 1 \in \mathbb{R}\ \forall x \in \mathbb{R} : \quad x1 = x$

M3: $\forall x \in \mathbb{R} \setminus \{0\}\ \exists x^{-1} \in \mathbb{R} : \quad xx^{-1} = 1$

M4: $\forall x, y \in \mathbb{R} : \quad xy = yx$

Definition 1.25 (Compatibility of Addition and Multiplication)**.**

R1: Distributivity

$$\forall x, y, z \in \mathbb{R} : \quad x \cdot (y + z) = (x \cdot y) + (x \cdot z)$$

R2: $0 \neq 1$

Theorem 1.26. $x, y \in \mathbb{R}$

(i) $x \cdot 0 = 0$

(ii) $-(x \cdot y) = x \cdot (-y) = (-x) \cdot y$

(iii) $(-x) \cdot (-y) = x \cdot y$

(iv) $(-x)^{-1} = -(x^{-1})$ *(only for* $x \neq 0$*)*

(v) $xy = 0 \implies x = 0 \vee y = 0$

Proof.

(i) $x \in \mathbb{R}$

$$x \cdot 0 \overset{\text{A2}}{=} x \cdot (0+0) \overset{\text{R1}}{=} x \cdot 0 + x \cdot 0 \tag{1.9}$$

$$\overset{\text{A3}}{\Longrightarrow} 0 = x \cdot 0 \tag{1.10}$$

(ii) $x, y \in \mathbb{R}$

$$xy + (-(xy)) \overset{\text{A3}}{=} 0 \overset{\text{(i)}}{=} x \cdot 0 = x(y + (-y)) \overset{\text{R1}}{=} xy + x(-y) \tag{1.11}$$

$$\overset{\text{A3}}{\Longrightarrow} -(xy) = x \cdot (-y) \tag{1.12}$$

(iii) Left as an exercise for the reader.

(iv) $x \in \mathbb{R}$

$$x \cdot (-(-x)^{-1}) \overset{\text{(ii)}}{=} -(x \cdot (-x)^{-1}) \overset{\text{(ii)}}{=} (-x) \cdot (-x)^{-1} \overset{\text{M3}}{=} 1 \overset{\text{M3}}{=} x \cdot x^{-1} \tag{1.13}$$

$$\overset{\text{M3}}{\Longrightarrow} -(-x)^{-1} = x^{-1} \overset{1.23(iii)}{\Longrightarrow} (-x)^{-1} = -(x^{-1}) \tag{1.14}$$

(v) $x, y \in \mathbb{R}$ and $y \neq 0$. Then $\exists y^{-1} \in \mathbb{R}$:

$$xy = 0 \implies xyy^{-1} \overset{\text{M3}}{=} x \cdot 1 \overset{\text{M2}}{=} x = 0 = 0 \cdot y^{-1} \tag{1.15}$$

□

Remark 1.27. A structure that fulfils all the previous axioms is called a field. We introduce the following notation for $x, y \in \mathbb{R},\ y \neq 0$

$$\frac{x}{y} = xy^{-1}$$

Definition 1.28 (Order relations)**.**

O1: Reflexivity

$$\forall x \in \mathbb{R}: \quad x \leq x$$

O2: Transitivity

$$\forall x, y, z \in \mathbb{R}: \quad x \leq y \wedge y \leq z \implies x \leq z$$

O3: Anti-Symmetry

$$\forall x, y \in \mathbb{R}: \quad x \leq y \land y \leq x \implies x = y$$

O4: Totality

$$\forall x, y \in \mathbb{R}: \quad x \leq y \lor y \leq x$$

O5:

$$\forall x, y, z \in \mathbb{R}: \quad x \leq y \implies x + z \leq y + z$$

O6:

$$\forall x, y \in \mathbb{R}: \quad 0 \leq x \land 0 \leq y \implies 0 \leq x \cdot y$$

We write $x < y$ for $x \leq y \land x \neq y$

Theorem 1.29. $x, y \in \mathbb{R}$

(i) $x \leq y \implies -y \leq -x$

(ii) $x \leq 0 \land y \leq 0 \implies 0 \leq xy$

(iii) $0 \leq 1$

(iv) $0 \leq x \implies 0 \leq x^{-1}$

(v) $0 < x \leq y \implies y^{-1} \leq x^{-1}$

Proof.

(i)

$$\begin{aligned} x \leq y &\overset{\text{O5}}{\implies} x + (-x) + (-y) \leq y + (-x) + (-y) \\ &\iff -y \leq -x \end{aligned} \tag{1.16}$$

(ii) With $y \leq 0 \overset{\text{(i)}}{\implies} 0 \leq -y$ and $x \leq 0 \overset{\text{(i)}}{\implies} 0 \leq -x$ follows from O6:

$$0 \leq (-x)(-y) = xy \tag{1.17}$$

(iii) Assume $0 \leq 1$ is not true. From O4 we know that

$$1 \leq 0 \overset{\text{(ii)}}{\implies} 0 \leq 1 \cdot 1 = 1 \tag{1.18}$$

(iv) Left as an exercise for the reader.

(v)

$$0 \le x^{-1} \wedge 0 \le y^{-1} \overset{\text{O6}}{\Longrightarrow} 0 \le x^{-1}y^{-1} \tag{1.19}$$

From $x \le y$ follows $0 \le y - x$

$$\overset{\text{O6}}{\Longrightarrow} 0 \le (y-x)x^{-1}y^{-1} \overset{\text{R1}}{=} yx^{-1}y^{-1} - xx^{-1}y^{-1} = x^{-1} - y^{-1} \tag{1.20}$$
$$\overset{\text{O5}}{\Longrightarrow} y^{-1} \le x^{-1} \tag{1.21}$$

□

Remark 1.30. A structure that fulfils all the previous axioms is called an ordered field.

Definition 1.31. Let $A \subset \mathbb{R}$, $x \in \mathbb{R}$.

(i) x is called an upper bound of A if $\forall y \in A: \; y \le x$

(ii) x is called a maximum of A if x is an upper bound of A and $x \in A$

(iii) x is called supremum of A is x is an upper bound of A and if for every other upper bound $y \in \mathbb{R}$ the statement $x \le y$ holds. In other words, x is the smallest upper bound of A.

A is called bounded above if it has an upper bound. Analogously, there exists a lower bound, a minimum and an infimum. We introduce the notation $\sup A$ for the supremum and $\inf A$ for the infimum.

Definition 1.32. $a, b \in \mathbb{R}$, $a < b$. We define

- $(a, b) := \{x \in \mathbb{R} \,|\, a < x \wedge x < b\}$
- $[a, b] := \{x \in \mathbb{R} \,|\, a \le x \wedge x \le b\}$
- $(a, \infty) := \{x \in \mathbb{R} \,|\, a < x\}$

Example 1.33. $(-\infty, 1)$ is bounded above ($1, 2, 1000, \cdots$ are upper bounds), but has no maximum. 1 is the supremum.

Definition 1.34 (Completeness of the real numbers)**.** Every non-empty subset of $\mathbb{R}$ with an upper bound has a supremum.

Definition 1.35. A set $A \subset \mathbb{R}$ is called inductive if $1 \in A$ and

$$x \in A \implies x + 1 \in A$$

Lemma 1.36. *Let I be an index set, and let A_i be inductive sets for every $i \in I$. Then $\bigcap_{i\in I} A_i$ is also inductive.*

Proof. Since A_i is inductive $\forall i \in I$, we know that $1 \in A_i$. Therefore

$$1 \in \bigcap_{i\in I} A_i \tag{1.22}$$

Now let $x \in \bigcap_{i\in I} A_i$, this means that $x \in A_i \quad \forall i \in I$.

$$\implies x+1 \in A_i \quad \forall i \in I \implies x+1 \in \bigcap_{i\in I} A_i \tag{1.23}$$

□

Definition 1.37. The natural numbers are the smallest inductive subset of $\mathbb{R}$. I.e.

$$\bigcap_{A \text{ inductive}} A =: \mathbb{N}$$

Theorem 1.38 (The principle of induction)**.** *Let $\Phi(x)$ be a statement with a free variable x. If $\Phi(1)$ is true, and if $\Phi(x) \implies \Phi(x+1)$, then $\Phi(x)$ holds for all $x \in \mathbb{N}$.*

Proof. Define $A = \{x \in \mathbb{R} \,|\, \Phi(x)\}$. According to the assumptions, A is inductive and therefore $\mathbb{N} \subset A$. This means that $\forall n \in \mathbb{N} : \quad \Phi(n)$. □

Corollary 1.39. $m, n \in \mathbb{N}$

(i) $m + n \in \mathbb{N}$

(ii) $mn \in \mathbb{N}$

(iii) $1 \leq n \quad \forall n \in \mathbb{N}$

Proof. We will only proof (i). (ii) and (iii) are left as an exercise for the reader. Let $n \in \mathbb{N}$. Define $A = \{m \in \mathbb{N} \,|\, m + n \in \mathbb{N}\}$. Then $1 \in A$, since $\mathbb{N}$ is inductive. Now let $m \in A$, therefore $n + m \in \mathbb{N}$.

$$\implies n + m + 1 \in \mathbb{N} \tag{1.24}$$

$$\iff m + 1 \in A \tag{1.25}$$

Hence A is inductive, so $\mathbb{N} \subset A$. From $A \subset N$ follows that $\mathbb{N} = A$. □

Theorem 1.40. $n \in \mathbb{N}$. *There are no natural numbers between n and $n+1$.*

Heuristic Proof. Show that $x \in \mathbb{N} \cap (1, 2)$ implies that $\mathbb{N} \setminus \{x\}$ is inductive. Now show that if $\mathbb{N} \cap (n, n+1) = \varnothing$ and $x \in \mathbb{N} \cap (n+1, n+2)$ then $\mathbb{N} \setminus \{x\}$ is inductive. ☐

Theorem 1.41 (Archimedian property)**.**

$$\forall x \in \mathbb{R}\ \exists n \in \mathbb{N} : \quad x < n$$

Proof. If $x < 1$ there is nothing to prove, so let $x \geq 1$. Define the set

$$A = \{n \in \mathbb{N} \,|\, n \leq x\} \tag{1.26}$$

A is bounded above by definition. There exists the supremum $s = \sup A$. By definition, $s - 1$ is not an upper bound of A, i.e. $\exists m \in A : \quad s - 1 < m$. Therefore $s \leq m + 1$.

$$m \in A \subset \mathbb{N} \implies m + 1 \in \mathbb{N} \tag{1.27}$$

Since s is an upper bound of A, this implies that $m + 1 \not\subset A$, so therefore $m + 1 > x$. ☐

Corollary 1.42. *Every non-empty subset of* $\mathbb{N}$ *has a minimum, and every non-empty subset of* $\mathbb{N}$ *that is bounded above has a maximum.*

Proof. Let $A \subset \mathbb{N}$. Propose that A has no minimum. Define the set

$$\tilde{A} := \{n \in \mathbb{N} \,|\, \forall m \in A : \ n < m\} \tag{1.28}$$

1 is a lower bound of A, but according to the proposition A has no minimum, so therefore $1 \notin A$. This implies that $1 \in \tilde{A}$.

$$n \in \tilde{A} \implies n < m \ \forall m \in A \tag{1.29}$$

But since there exists no natural number between n and $n + 1$, this means that $n + 1$ is also a lower bound of A, and therefore

$$n + 1 \leq m \ \forall m \in A \implies n + 1 \in \tilde{A} \tag{1.30}$$

So $\tilde{A}$ is an inductive set, hence $\tilde{A} = \mathbb{N}$. Therefore $A = \varnothing$. ☐

Definition 1.43. We define the following new sets:

$$\mathbb{Z} := \{x \in \mathbb{R} \,|\, x \in \mathbb{N}_0 \vee (-x) \in \mathbb{N}_0\}$$
$$\mathbb{Q} := \left\{\frac{p}{q} \,|\, p, q \in \mathbb{Z} \wedge q \neq 0\right\}$$

$\mathbb{Z}$ are called integers, and $\mathbb{Q}$ are called the rational numbers. $\mathbb{N}_0$ are the natural numbers with the 0 ($\mathbb{N}_0 = \mathbb{N} \cap \{0\}$).

Remark 1.44.

$$\begin{aligned} x,y \in \mathbb{Z} &\implies x+y, x\cdot y, (-x) \in \mathbb{Z} \\ x,y \in \mathbb{Q} &\implies x+y, x\cdot y, (-x) \in \mathbb{Q} \text{ and } x^{-1} \in \mathbb{Q} \text{ if } x \neq 0 \end{aligned}$$

The second statement implies that $\mathbb{Q}$ is a field.

Corollary 1.45 (Density of the rationals)**.** $x, y \in \mathbb{R},\ x < y$. *Then*

$$\exists r \in \mathbb{Q}: \quad x < r < y$$

Proof. This proof relies on the Archimedean property.

$$\exists q \in \mathbb{N}: \quad \frac{1}{y-x} < q \left(\iff \frac{1}{q} < y - x \right) \tag{1.31}$$

Let $p \in \mathbb{Z}$ be the greatest integer that is smaller than $y \cdot q$. The existence of p is ensured by corollary Corollary 1.42. Then $\frac{p}{q} < y$ and

$$p + 1 \geq y \cdot q \implies y \leq \frac{p}{q} + \frac{1}{q} < \frac{p}{q} + (y - x) \tag{1.32}$$

$$\implies x < \frac{p}{q} < y \tag{1.33}$$

□

Definition 1.46 (Absolute values)**.** We define the following function

$$\begin{aligned} |\cdot| : \mathbb{R} &\longrightarrow [0, \infty) \\ x &\longmapsto \begin{cases} x & , x \geq 0 \\ -x & , x < 0 \end{cases} \end{aligned}$$

Theorem 1.47.

$$x, y \in \mathbb{R} \implies |xy| = |x||y|$$

Proof. Left as an exercise for the reader. □

Definition 1.48 (Complex numbers)**.** Complex numbers are defined as the set $\mathbb{C} = \mathbb{R}^2$. Addition and multiplication are defined as mappings $\mathbb{C} \times \mathbb{C} \to \mathbb{C}$. Let $(x, y), (\tilde{x}, \tilde{y}) \in \mathbb{C}$.

$$\begin{aligned} (x, y) + (\tilde{x}, \tilde{y}) &:= (x + \tilde{x}, y + \tilde{y}) \\ (x, y) \cdot (\tilde{x}, \tilde{y}) &:= (x\tilde{x} - y\tilde{y}, x\tilde{y} + \tilde{x}y) \end{aligned}$$

$\mathbb{C}$ is a field. Let $z = (x, y) \in \mathbb{C}$. We define

$$\begin{aligned} \Re(z) = \operatorname{Re}(z) = x &\quad \text{the real part} \\ \Im(z) = \operatorname{Im}(z) = y &\quad \text{the imaginary part} \end{aligned}$$

Remark 1.49.

(i) We will not prove that $\mathbb{C}$ fulfils the field axioms here, this can be left as an exercise to the reader. However, we will note the following statements

- Additive neutral element: $(0, 0)$
- Additive inverse of (x, y): $(-x, -y)$
- Multiplicative neutral element: $(1, 0)$
- Multiplicative inverse of $(x, y) \neq (0, 0)$: $\left(\frac{x}{x^2+y^2}, -\frac{y}{x^2+y^2}\right)$

(ii) Numbers with $y = 0$ are called real.

(iii) The imaginary unit is defined as $i = (0, 1)$

$$(0, 1) \cdot (x, y) = (-y, x)$$

Especially

$$i^2 = (0, 1)^2 = (-1, 0) = -(1, 0) = -1$$

We also introduce the following notation

$$(x, y) = (x, 0) + i \cdot (y, 0) = x + iy$$

Theorem 1.50 (Fundamental theorem of algebra)**.** *Every non-constant, complex polynomial has a complex root. I.e. for $n \in \mathbb{N}$, $\alpha_0, \cdots, \alpha_n \in \mathbb{C}$, $\alpha_n \neq 0$ there is some $x \in \mathbb{C}$ such that*

$$\sum_{i=0}^{n} \alpha_i x^i = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_n x^n = 0$$

Proof. Not here. □

Chapter 2

Real Analysis: Part I

2.1 Elementary Inequalities

Example 2.1.

- $x \in \mathbb{R} \implies x^2 \geq 0$
- $x^2 - 2xy + y^2 = (x-y)^2 \geq 0 \;\; \forall x, y \in \mathbb{R}$
- $x^2 + y^2 \geq 2xy$

Theorem 2.2 (Absolute inequalities)**.** *Let $x \in \mathbb{R}$, $c \in [0, \infty)$. Then*

(i) $-|x| \leq x \leq |x|$

(ii) $|x| \leq c \iff -c \leq x \leq c$

(iii) $|x| \geq c \iff x \leq -c \vee c \leq x$

(iv) $|x| = 0 \iff x = 0$

Theorem 2.3 (Triangle inequality)**.** *Let $x, y \in \mathbb{R}$. Then*

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

Proof. From Theorem 2.2 follows $x \leq |x|$ and $y \leq |y|$.

$$\implies x + y \leq |x| + |y| \tag{2.1}$$

However, from the same theorem follows $-|x| \leq x$ and $-|y| \leq y$.

$$\implies -|x| - |y| = x + y \tag{2.2}$$

$$\implies |x + y| \leq |x| + |y| \tag{2.3}$$

$\square$

Corollary 2.4. *$n \in \mathbb{N}$, $x_1, \cdots, x_n \in \mathbb{R}$. Then*

$$\left| \sum_{i=1}^{n} x_i \right| \leq \sum_{i=1}^{n} |x_i|$$

Proof. Proof by induction. Let $n = 1$:

$$|x_1| \leq |x_1| \tag{2.4}$$

This statement is trivially true. Now assume the corollary holds for $n \in \mathbb{N}$. Then

$$\begin{aligned} \left| \sum_{i=1}^{n+1} x_i \right| = \left| \sum_{i=1}^{n} x_i + x_{n+1} \right| &\leq \left| \sum_{i=1}^{n} x_n \right| + |x_{n+1}| \\ &\leq \sum_{i=1}^{n} |x_i| + |x_{n+1}| \\ &= \sum_{i=1}^{n+1} |x_i| \end{aligned} \tag{2.5}$$

□

Theorem 2.5 (Bernoulli inequality)**.** *Let $x \in [-1, \infty)$ and $n \in \mathbb{N}$. Then*

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

Proof. Proof by induction. Let $n = 1$:

$$1 + x \geq 1 + 1 \cdot x \tag{2.6}$$

This is trivial. Now assume the theorem holds for $n \in \mathbb{N}$. Then

$$\begin{aligned} (1+x)^{n+1} = (1+x)^n(1+x) &\geq (1+nx)(1+x) \\ &= 1 + (n+1)x + nx^2 \\ &\geq 1 + (n+1)x \end{aligned} \tag{2.7}$$

□

2.2 Sequences and Limits

Definition 2.6. Let M be a set (usually M is $\mathbb{R}$ or $\mathbb{C}$). A sequence in M is a mapping from $\mathbb{N}$ to M. The notation is $(x_n)_{n \in \mathbb{N}} \subset M$ or $(x_n) \subset M$. x_n is called element of the sequence at n.

Example 2.7. Some real sequences are

- $x_n = \frac{1}{n} \quad (1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \cdots)$
- $x_n = \sum_{k=1}^{n} k \quad (1, 3, 6, 10, 15, \cdots)$
- $x_n =$ "smallest prime factor of n" $\quad (*, 2, 3, 2, 5, 2, 7, 2, 3, 2, \cdots)$

Definition 2.8 (Convergence)**.** Let $(x_n) \subset \mathbb{R}$ be a sequence, and $x \in \mathbb{R}$. Then

$$(x_n) \text{ converges to } x \iff \forall \epsilon > 0 \; \exists N \in \mathbb{N} : \quad |x_n - x| < \epsilon \;\; \forall n \geq N$$

A complex sequence $(z_n) \subset \mathbb{C}$ converges to $z \in \mathbb{C}$ if the real and imaginary parts of (z_n) converge to the real and imaginary parts of z. x (or z) is called the limit of the sequence. Common notation:

$$x_n \longrightarrow x \qquad\qquad x_n \xrightarrow{n\to\infty} x \qquad\qquad \lim_{n\to\infty} x_n = x$$

If a sequence converges to 0 it is called a null sequence.

Example 2.9.

(i) $x \in \mathbb{R}$, $x_n = x$ (constant sequence). This sequence converges to x. To show this, let $\epsilon > 0$. Then for $N = 1$:

$$|x_n - x| = |x - x| = 0 < \epsilon$$

(ii) $x_n = \frac{1}{n}$ is a null sequence. Let $\epsilon > 0$. By the Archimedean property:

$$\exists N \in \mathbb{N} : \quad \frac{1}{\epsilon} < N$$

Then for $n \geq N$:

$$|x_n - 0| = |x_n| = \frac{1}{n} \leq \frac{1}{N} < \epsilon$$

(iii) The sequence

$$x_n = \begin{cases} 1 & , n \text{ even} \\ -1 & , n \text{ odd} \end{cases}$$

does not converge.

Remark 2.10. A property holds for almost every (a.e.) $n \in \mathbb{N}$ if it doesn't hold for only finitely many n. (e.g. $n < 10$ is true for a.e. $n \in \mathbb{N}$)

Theorem 2.11. *A sequence $(x_n) \subset \mathbb{R}$ (or $\mathbb{C}$) has at most one limit.*

Proof. Propose that $x, \tilde{x}$ are different limits of (x_n). Without loss of generality (w.l.o.g.) we can write $x < \tilde{x}$. Now define $\epsilon = \frac{1}{2}(\tilde{x} - x) > 0$.

$$x_n \longrightarrow x \iff \exists N_1 : \quad x_n \in (x - \epsilon, x + \epsilon) = \left(x - \epsilon, \frac{x + \tilde{x}}{2}\right) \tag{2.8}$$

$$x_n \longrightarrow \tilde{x} \iff \exists N_2 : \quad x_n \in (\tilde{x} - \epsilon, \tilde{x} + \epsilon) = \left(\frac{x + \tilde{x}}{2}, x + \epsilon\right) \tag{2.9}$$

Since these intervals are disjoint, the proposition led to a contradiction. □

Theorem 2.12. *Let $(x_n) \subset \mathbb{R}$ (or $\mathbb{C}$) be sequence with limit $x \in \mathbb{R}$. Then for $m \in \mathbb{N}$*

$$\lim_{n \to \infty} x_{n+m} = x$$

Proof. Left as an exercise for the reader. □

Definition 2.13. The sequence $(x_n) \subset \mathbb{R}$ is bounded above if $\{x_n \mid n \in \mathbb{N}\}$ is bounded above. A number $K \in \mathbb{R}$ is an upper bound if $\forall n \in \mathbb{N} : \ x_n \leq K$.

Theorem 2.14. *Every convergent sequence is bounded.*

Proof. Let $(x_n) \subset \mathbb{R}$ converge to $x \in \mathbb{R}$. For $\epsilon = 1$ we trivially know that

$$\exists N \in \mathbb{N} \ \forall n \geq N : \quad |x_n - x| < \epsilon = 1 \tag{2.10}$$

Let

$$K = \max\{x_1, x_2, \cdots, x_N, |x| + 1\} \tag{2.11}$$

Then

$$|x_n| \leq K \quad \forall n \in \mathbb{N} \tag{2.12}$$

This is trivial for $n \leq N$. For $n > N$ we can use the triangle inequality:

$$|x_n| = |(x_n - x) + x| \leq |x_n - x| + |x| \leq |x| + 1 \tag{2.13}$$

□

Theorem 2.15. *If $(x_n) \subset \mathbb{R}$ bounded and $(y_n) \subset \mathbb{R}$ null sequence, then $(x_n) \cdot (y_n)$ is also a null sequence.*

Proof. If (x_n) is bounded, this means that $\exists K \in (0, \infty)$ such that

$$|x_n| \leq K \quad \forall n \in \mathbb{N} \tag{2.14}$$

Since (y_n) is a null sequence we know that

$$\forall \epsilon > 0 \; \exists N \in \mathbb{N} \; \forall n \geq N : \quad |y_n| < \epsilon \tag{2.15}$$

Now let $\epsilon > 0$, then $\exists N \in \mathbb{N}$ such that

$$\forall n \geq N : \quad |y_n| < \frac{\epsilon}{K} \tag{2.16}$$

$$|x_n \cdot y_n| = |x_n||y_n| \leq K\frac{\epsilon}{K} = \epsilon \tag{2.17}$$

Therefore $(x_n)(y_n)$ is a null sequence. □

Theorem 2.16 (Squeeze theorem)**.** *Let $(x_n), (y_n), (z_n) \subset \mathbb{R}$ be sequences such that*

$$x_n \leq y_n \leq z_n$$

for a.e. $n \in \mathbb{N}$, and let $x_n \to x$, $z_n \to x$. Then

$$\lim_{n\to\infty} y_n = x$$

Proof. Let $\epsilon > 0$. Then $\exists N_1, N_2, N_3 \in \mathbb{N}$ such that

$$\forall n \geq N_1 : \quad x_n \leq y_n \leq z_n \tag{2.18}$$

$$\forall n \geq N_2 : \quad |x_n - x| < \epsilon \tag{2.19}$$

$$\forall n \geq N_3 : \quad |z_n - x| < \epsilon \tag{2.20}$$

Choose $N = \max\{N_1, N_2, N_3\}$. Then

$$\forall n \geq N : \quad -\epsilon < x_n - x \leq y_n - x \leq z_n - x < \epsilon \tag{2.21}$$

Therefore $|y_n - x| < \epsilon$ □

Example 2.17. $\forall n \in \mathbb{N} : \quad n \leq n^2$ (why?).

$$\implies 0 \leq \frac{1}{n^2} \leq \frac{1}{n} \implies \lim_{n\to\infty} \frac{1}{n^2} = 0$$

Theorem 2.18. *Let $(x_n), (y_n) \subset \mathbb{R}$ and $x_n \to x$, $y_n \to y$. Then $x \leq y$.*

Proof. Left as an exercise for the reader. □

Remark 2.19. If $x_n < y_n \quad \forall n \in \mathbb{N}$, then $x = y$ can still be true.

Lemma 2.20. *Let $(x_n) \in \mathbb{R}$ and $x \in \mathbb{R}$.*

$$(x_n) \longrightarrow x \iff (|x_n - x|) \text{ is null sequence}$$

Especially:

$$(x_n) \text{ null sequence} \iff |x_n| \text{ null sequence}$$

Proof.

$$||x_n - x| - 0| = |x_n - x| \tag{2.22}$$

□

Theorem 2.21. *Let $(x_n), (x_n) \subset \mathbb{R}$ (or $\mathbb{C}$) with $x_n \to x$, $y_n \to y$ $(x, y \in \mathbb{R})$. Then all of the following are true:*

(i)

$$\lim_{n\to\infty} x_n + y_n = x + y = \lim_{n\to\infty} x_n + \lim_{n\to\infty} y_n$$

(ii)

$$\lim_{n\to\infty} x_n y_n = xy = \lim_{n\to\infty} x_n \cdot \lim_{n\to\infty} y_n$$

(iii) If $y \neq 0$:

$$\lim_{n\to\infty} \frac{x_n}{y_n} = \frac{x}{y} = \frac{\lim_{n\to\infty} x_n}{\lim_{n\to\infty} y_n}$$

Proof.

(i) Let $\epsilon > 0$. Then $\exists N_1, N_2 \in \mathbb{N}$ such that

$$\forall n \geq N_1 : \quad |x_n - x| < \frac{\epsilon}{2} \tag{2.23}$$

$$\forall n \geq N_2 : \quad |y_n - y| < \frac{\epsilon}{2} \tag{2.24}$$

Now choose $N = \max\{N_1, N_2\}$. Then $\forall n \geq N$:

$$\begin{aligned} |x_n + y_n - (x + y)| &= |(x_n - x) + (y_n - y)| \\ &\leq |x_n - x| + |y_n - y| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \end{aligned} \tag{2.25}$$

$$\Longrightarrow x_n + y_n \longrightarrow x + y \tag{2.26}$$

(ii)

$$\begin{aligned} 0 \le |x_n y_n - xy| &= |(x_n y_n - x_n y) + (x_n y - xy)| \\ &\le |x_n(y_n - y)| + |(x_n - x)y| \\ &= |x_n||y_n - y| + |x_n - x||y| \longrightarrow 0 \end{aligned} \tag{2.27}$$

Therefore $|x_n y_n - xy|$ is a null sequence and

$$x_n y_n \longrightarrow xy \tag{2.28}$$

(iii) Now we need to show that if $y \neq 0$ then $\frac{1}{y_n} \to \frac{1}{y}$. We know that $|y| > 0$. So $\exists N \in \mathbb{N}$ such that

$$\forall n \ge N: \quad |y_n - y| < \frac{|y|}{2} \tag{2.29}$$

This implies that

$$\forall n \ge N: \quad 0 < \frac{|y|}{2} \le |y_n| \tag{2.30}$$

From this we now know that $\frac{1}{y_n}$ is defined and bounded

$$\left|\frac{1}{y_n}\right| = \frac{1}{|y_n|} \le \frac{2}{|y|} \tag{2.31}$$

So finally

$$\left|\frac{1}{y_n} - \frac{1}{y}\right| = \left|\frac{1}{y_n}\left(1 - y_n\frac{1}{y}\right)\right| = \left|\frac{1}{y_n}\right|\left|1 - y_n\frac{1}{y}\right| \longrightarrow 0 \tag{2.32}$$

And therefore

$$\begin{aligned} y_n \longrightarrow y \implies & \frac{y_n}{y} \longrightarrow 1 \\ \stackrel{\text{Thm. 2.15}}{\implies} & \left|1 - \frac{y_n}{y}\right| \text{ is a null sequence} \\ \stackrel{\text{Lem. 2.20}}{\implies} & \frac{1}{y_n} \longrightarrow \frac{1}{y} \end{aligned} \tag{2.33}$$

□

Corollary 2.22. *Let $k \in \mathbb{N}$, $a_0, \cdots, a_k, b_0, \cdots, b_k \in \mathbb{R}$ and $b_k \neq 0$. Then*

$$\lim_{n\to\infty} \frac{a_0 + a_1 n + a_2 n^2 + \cdots + a_{k-1}n^{k-1} + a_k n^k}{b_0 + b_1 n + b_2 n^2 + \cdots + b_{k-1}n^{k-1} + b_k n^k} = \frac{a_k}{b_k}$$

Proof. Multiply the numerator and the denominator with $\frac{1}{n^k}$

$$\frac{\frac{a_0}{n^k} + \frac{a_1}{n^{k-1}} + \frac{a_2}{n^{k-2}} + \cdots + \frac{a_{k-1}}{n} + a_k}{\frac{b_0}{n^k} + \frac{b_1}{n^{k-1}} + \frac{b_2}{n^{k-2}} + \cdots + \frac{b_{k-1}}{n} + b_k} \xrightarrow[n\to\infty]{} 0 \tag{2.34}$$

□

Example 2.23. Let $x \in (-1, 1)$. Then $\lim_{n\to\infty} x^n = 0$

Proof. For $x = 0$ this is trivial. For $x \neq 0$ it follows that $|x| \in (0, 1)$ and $\frac{1}{|x|} \in (1, \infty)$. Choose $s = \frac{1}{|x|} - 1 > 0$ and apply the Bernoulli inequality (Theorem 2.5).

$$(1 + s)^n \geq 1 + n \cdot s \tag{2.35}$$

$$0 \leq |x|^n = \left(\frac{1}{1+s}\right)^n = \frac{1}{(1+s)^n} \leq \frac{1}{1 + n \cdot s} = \frac{1 + n \cdot 0}{1 + n \cdot s} \xrightarrow{2.22} 0 \tag{2.36}$$

The squeeze theorem now tells us that $|x^n| = |x|^n \to 0$ and therefore $x^n \to 0$. □

Definition 2.24. A sequence $(x_n) \subset \mathbb{R}$ is called monotonic increasing (decreasing) if $x_{n+1} \geq x_n$ $(x_{n+1} \leq x_n)$ $\forall n \in \mathbb{N}$.

Theorem 2.25 (Monotone convergence theorem)**.** *Let $(x_n) \subset \mathbb{R}$ be a monotonic increasing (or decreasing) sequence that is bounded above (or below). Then (x_n) converges.*

Proof. Let (x_n) be monotonic increasing and bounded above. Define

$$x = \sup \underbrace{\{x_n \mid n \in \mathbb{N}\}}_{A} \tag{2.37}$$

Now let $\epsilon > 0$, then $x - \epsilon$ is not an upper bound of A, this means $\exists N \in \mathbb{N}$ such that $x_N > x - \epsilon$. The monotony of (x_n) implies that

$$\forall n \geq N : \quad x_n > x - \epsilon \tag{2.38}$$

So therefore

$$x - \epsilon < x_n < x + \epsilon \implies |x_n - x| < \epsilon \tag{2.39}$$

□

Remark 2.26.

$$(x_n) \text{ is monotonic increasing} \iff \frac{x_{n+1}}{x_n} \geq 1 \quad \forall n \in \mathbb{N}$$

$$(x_n) \text{ is monotonic decreasing} \iff \frac{x_{n+1}}{x_n} \leq 1 \quad \forall n \in \mathbb{N}$$

Example 2.27. Consider the following sequence

$$x_1 = 1$$
$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right), \quad a \in [0, \infty)$$

Notice that $0 < x_n \quad \forall n \in \mathbb{N}$. For $n \in \mathbb{N}$ one can show that

$$\begin{aligned} x_{n+1}^2 = \frac{1}{4}\left(x_n^2 + 2a + \frac{a^2}{x_n^2}\right) &= \frac{1}{4}\left(x_n^2 - 2a + \frac{a^2}{x_n^2}\right) + a \\ &= \frac{1}{4}\left(x_n - \frac{a}{x_n}\right)^2 + a \geq a \end{aligned}$$

So $x_n^2 \geq a \quad \forall n \geq 2$, and therefore $\frac{a}{x_n} \leq x_n$. Finally

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right) \leq \frac{1}{2}(x_n + x_n) = x_n \quad \forall n \geq 2$$

This proves that (x_n) is monotonic decreasing and bounded below.

Theorem 2.28 (Square root)**.** *This theorem doubles as the definition of the square root. Let $a \in [0, \infty)$. Then $\exists! x \in [0, \infty)$ such that $x^2 = a$. Such an x is called the square root of a, and is notated as $x = \sqrt{a}$.*

Proof. First we want to prove the uniqueness of such an x. Assume that $x^2 = y^2 = a$ with $x, y \in [0, \infty)$. Then $0 = x^2 - y^2 = (x - y)(x + y)$.

$$\implies x + y = 0 \implies x = y = 0 \tag{2.40}$$
$$\implies x - y = 0 \implies x = y \tag{2.41}$$

Now to prove the existence, review the previous example.

$$x_n \longrightarrow x \text{ for some } x \in [0, \infty) \tag{2.42}$$

By using the recursive definition we can write

$$2x_n \cdot x_{n+1} = x_n^2 + a \longrightarrow x^2 + a \tag{2.43}$$

$$\implies 2x^2 = x^2 + a \implies x^2 = a \tag{2.44}$$

□

Remark 2.29. Analogously $\exists! x \in [0,\infty)\ \forall a \in [0,\infty)$ such that $x^n = a$. (Notation: $\sqrt[n]{a}$ or $x = a^{\frac{1}{n}}$). We will also introduce the power rules for rational exponents. Let $x, y \in \mathbb{R}$, $u, v \in \mathbb{Q}$.

$$(x \cdot y)^u = x^u y^u \qquad x^u \cdot x^v = x^{u+v} \qquad (x^u)^v = x^{u \cdot v}$$

Theorem 2.30. *Let $x, y \in \mathbb{R}$, $n \in \mathbb{N}$. Then*

$$0 \leq x < y \implies \sqrt[n]{x} < \sqrt[n]{y}$$

Let $n, m \in \mathbb{N}$, $n < m$, $x \in (1,\infty)$, $y \in (0,1)$. Then

$$\sqrt[n]{x} > \sqrt[m]{x} \qquad \sqrt[n]{y} < \sqrt[m]{y}$$

Proof. Left as an exercise for the reader. □

Theorem 2.31. *Let $a \in (0,\infty)$. Then*

$$\lim_{n\to\infty} \sqrt[n]{n} = 1 \qquad \lim_{n\to\infty} \sqrt[n]{a} = 1$$

Proof. Let $\epsilon > 0$. Then

$$\frac{n}{(n+\epsilon)^n} \xrightarrow{n\to\infty} 0 \tag{2.45}$$

This means that

$$\exists N \in \mathbb{N}\ \forall n \geq N : \quad \frac{n}{(n+\epsilon)^n} < 1 \tag{2.46}$$

Therefore

$$n < (1+\epsilon)^n \implies 1-\epsilon < 1 \leq \sqrt[n]{n} < 1+\epsilon \iff \left|\sqrt[n]{n} - 1\right| < \epsilon \tag{2.47}$$

This proves the first statement. The second statement is trivially true for $a = 1$, so let $a > 1$. Then $\exists n \in \mathbb{N}$ such that $a < n$:

$$\implies 1 < \sqrt[n]{a} < \sqrt[n]{n} \longrightarrow 1 \tag{2.48}$$

$$\overset{\text{Squeeze}}{\implies} \sqrt[n]{a} \xrightarrow{n\to\infty} 1 \tag{2.49}$$

Now let $a < 1$. Then $\frac{1}{a} < 1$

$$\lim_{n\to\infty} \sqrt[n]{a} = \lim_{n\to\infty} \frac{1}{\sqrt[n]{\frac{1}{a}}} \xrightarrow{n\to\infty} \frac{1}{1} = 1 \tag{2.50}$$

□

Definition 2.32. Let $z \in \mathbb{C}$, $x, y \in \mathbb{R}$ such that $z = x + iy$.

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

Theorem 2.33. *Let $u, v \in \mathbb{C}$. Then*

$$|u \cdot v| = |u||v| \qquad \left|\frac{1}{u}\right| = \frac{1}{|u|} \qquad |u+v| \leq |u| + |v|$$

Proof.

$$|uv| = \sqrt{uv \cdot \bar{u}\bar{v}} = \sqrt{u\bar{u} \cdot v\bar{v}} = \sqrt{u\bar{u}} \cdot \sqrt{v\bar{v}} = |u||v| \tag{2.51}$$

$$\left|\frac{1}{u}\right| |u| = \left|\frac{1}{u} u\right| = |1| \implies \left|\frac{1}{u}\right| = \frac{1}{|u|} \tag{2.52}$$

For the final statement, remember that complex numbers can be represented as $z = x + iy$, and then

$$\operatorname{Re}(z) \leq |\operatorname{Re}(z)| \leq |z| \tag{2.53}$$

$$\operatorname{Im}(z) \leq |\operatorname{Im}(z)| \leq |z| \tag{2.54}$$

So therefore

$$\begin{aligned} |u+v|^2 &= (u+v) \cdot (\bar{u} + \bar{v}) \\ &= u\bar{u} + v\bar{u} + u\bar{v} + v\bar{v} \\ &= |u|^2 + 2\operatorname{Re}(\bar{u}v) + |v|^2 \\ &\leq |u|^2 + 2|\bar{u}v| + |v|^2 \\ &= |u|^2 + 2|u||v| + |v|^2 \\ &= (|u| + |v|)^2 \end{aligned} \tag{2.55}$$

□

Lemma 2.34. *Let $(z_n) \subset \mathbb{C}$, $z \in \mathbb{C}$.*

$$(z_n) \longrightarrow z \iff (|z_n - z|) \text{ null sequence}$$

Proof. Let $x_n = \operatorname{Re}(z_n)$ and $y_n = \operatorname{Im}(z_n)$. Then $x = \operatorname{Re}(z)$ and $y = \operatorname{Im}(z)$. First we prove the " $\Longleftarrow$ " direction. Let $(|z_n - z|)$ be a null sequence.

$$0 \leq |x_n| - |x| = |\operatorname{Re}(z_n - z)| \leq |z_n - z| \longrightarrow 0 \tag{2.56}$$

Analogously, this holds for y_n and y. We know that $(|x_n - x|)$ is a null sequence if $x_n \longrightarrow x$ (same for y_n and y), therefore

$$\implies z_n \longrightarrow z \tag{2.57}$$

To prove the " $\implies$ " direction we use the triangle inequality:

$$\begin{aligned} 0 \le |z_n - z| &= |(x_n - x) + i(y_n - y)| \\ &\le |x_n - x| + \underbrace{|i(y_n - y)|}_{|y_n - y|} \longrightarrow 0 \end{aligned} \tag{2.58}$$

By the squeeze theorem, $|z_n - z|$ is a null sequence. □

Remark 2.35. Lemma 2.34 allows us to generalize Theorem 2.21 and Corollary 2.22 for complex sequences.

Definition 2.36 (Cauchy sequence)**.** A sequence $(x_n) \subset \mathbb{R}$ (or $\mathbb{C}$) is called Cauchy sequence if

$$\forall \epsilon > 0\ \exists N \in \mathbb{N}\ \forall n, m \ge N: \quad |x_n - x_m| < \epsilon$$

Theorem 2.37 (Cauchy convergence test)**.** *A sequence $(x_n) \subset \mathbb{R}$ (or $\mathbb{C}$) converges if and only if it is a Cauchy sequence.*

Proof. Firstly, let (x_n) converge to x, and let $\epsilon > 0$. Then

$$\exists N \in \mathbb{N}\ \forall n \ge N: \quad |x_n - x| < \frac{\epsilon}{2} \tag{2.59}$$

So therefore $\forall n, m \ge N$:

$$|x_n - x_m| = |x_n - x + x - x_m| \le |x_n - x| + |x - x_m| < \epsilon \tag{2.60}$$

This proves the " $\implies$ " direction of the theorem. To prove the inverse let (x_n) be a Cauchy sequence. That means

$$\exists N \in \mathbb{N}\ \forall n, m \ge N: \quad |x_n - x_m| \le 1 \tag{2.61}$$

$$\begin{aligned} \implies |x_n| = |x_n - x_N + x_N| &\le |x_n - x_N| + |x_N| \\ &\le |x_N| + 1 \quad \forall n \ge N \end{aligned} \tag{2.62}$$

We will now introduce the two auxiliary sequences

$$y_n = \sup\{x_k \,|\, k \ge n\} \qquad\qquad z_n = \inf\{x_k \,|\, k \ge n\} \tag{2.63}$$

(y_n) and (z_n) are bounded, and for $\tilde{n} \le n$

$$\{x_k \,|\, k \ge \tilde{n}\} \supset \{x_k \,|\, k \ge n\} \tag{2.64}$$

$$\implies y_n = \sup\{x_k | k \geq n\} \leq \sup\{x_k | k \geq \tilde{n}\} = y_{\tilde{n}} \tag{2.65}$$
$$\implies (x_n) \text{ monotonic decreasing and therefore converging to } y \tag{2.66}$$

Analogously, this holds true for (z_n) as well. Trivially,

$$z_n \leq x_n \leq y_n \tag{2.67}$$

If $y = z$, then (x_n) converges according to the squeeze theorem. Assume $z < y$. Choose $\epsilon > \frac{y-z}{2} > 0$. If N is big enough, then

$$\sup\{x_k \,|\, k \geq N\} = y_N > y - \epsilon \tag{2.68}$$
$$\inf\{x_k \,|\, k \geq N\} = z_N < z + \epsilon \tag{2.69}$$

So for every $N \in \mathbb{N}$, we know that

$$\exists k \geq N: \quad x_k > y - 2\epsilon \tag{2.70}$$
$$\exists l \geq N: \quad x_l < z + 2\epsilon \tag{2.71}$$

For these elements the following holds

$$|x_k - x_l| \geq \epsilon = \frac{y-z}{2} \tag{2.72}$$

This is a contradiction to our assumption that (x_n) is a Cauchy sequence, so $y = z$ and therefore (x_n) converges. □

Remark 2.38.

(i) $x_n = (-1)^n$. For this sequence the following holds

$$\forall n \in \mathbb{N}: \quad |x_n - x_{n+1}| = 2$$

So this sequence isn't a Cauchy sequence-

(ii) It is NOT enough to show that $|x_n - x_{n+1}|$ tends to 0! Example: $(x_n) = \sqrt{n}$

$$\begin{aligned}\sqrt{n+1} - \sqrt{n} &= (\sqrt{n+1} - \sqrt{n})\frac{\sqrt{n+1} + \sqrt{n}}{\sqrt{n+1} + \sqrt{n}} \\ &= \frac{\cancel{n} + 1 - \cancel{n}}{\sqrt{n+1} + \sqrt{n}} \\ &= \frac{1}{\sqrt{n+1} + \sqrt{n}} \xrightarrow{n\to\infty} 0\end{aligned}$$

However $(\sqrt{n})$ doesn't converge.

(iii) We introduce the following

$$\text{Limes superior} \qquad \limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup\{x_k \mid k \geq n\}$$
$$\text{Limes inferior} \qquad \liminf_{n\to\infty} x_n = \lim_{n\to\infty} \inf\{x_k \mid k \geq n\}$$

$\limsup_{n\to\infty} x_n \geq \liminf_{n\to\infty} x_n$ always holds, and if (x_n) converges then

$$x_n \xrightarrow{n\to\infty} x \iff \limsup_{n\to\infty} x_n = \liminf_{n\to\infty} x_n$$

Definition 2.39. A sequence $(x_n) \subset \mathbb{R}$ is said to be properly divergent to ∞ if

$$\forall k \in (0,\infty)\ \exists N \in \mathbb{N}\ \forall n \geq N : \quad x_n > k$$

We notate this as

$$\lim_{n\to\infty} x_n = \infty$$

Theorem 2.40. *Let $(x_n) \subset \mathbb{R}$ be a sequence that diverges properly to ∞. Then*

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

Conversely, if $(y_n) \subset (0,\infty)$ is a null sequence, then

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

Proof. Let $\epsilon > 0$. By condition

$$\exists N \in \mathbb{N}\ \forall n \geq N : \quad |x_n| > \frac{1}{\epsilon} \quad \left(\iff \frac{1}{|x_n|} < \epsilon \right) \tag{2.73}$$

Therefore $\frac{1}{x_n}$ is a null sequence. The second part of the proof is left as an exercise for the reader. □

Remark 2.41 (Rules for computing). In this remark we will introduce some basic "rules" for working with infinities. These rules are exclusive to this topic, and are in no way universal! This should become obvious with our first two rules:

$$\frac{1}{\pm\infty} = 0 \qquad\qquad \frac{1}{0} = \infty$$

Obviously, division by 0 is still a taboo, however it works in this case since we are working with limits, and not with absolutes. Let $a \in \mathbb{R}$, $b \in (0, \infty)$, $c \in (1, \infty)$, $d \in (0, 1)$. The remaining rules are:

$$\begin{aligned} a + \infty &= \infty & a - \infty &= -\infty \\ \infty + \infty &= \infty & -\infty - \infty &= -\infty \\ b \cdot \infty &= \infty & b \cdot (-\infty) &= -\infty \\ \infty \cdot \infty &= \infty & \infty \cdot (-\infty) &= -\infty \\ c^{\infty} &= \infty & c^{-\infty} &= 0 \\ d^{\infty} &= 0 & d^{-\infty} &= \infty \end{aligned}$$

There are no general rules for the following:

$$\infty - \infty \qquad \frac{\infty}{\infty} \qquad 0 \cdot \infty \qquad 1^{\infty}$$

Theorem 2.42. *Let $(x_n) \subset \mathbb{R}$ be a sequence converging to x, and let $(k_n) \subset \mathbb{N}$ be a sequence such that*

$$\lim_{n \to \infty} k_n = \infty$$

Then

$$\lim_{n \to \infty} x_{k_n} = x$$

Proof. Let $\epsilon > 0$. Then

$$\exists N \in \mathbb{N}\ \forall n \geq N : \quad |x_n - x| < \epsilon \tag{2.74}$$

Furthermore

$$\exists \tilde{N} \in \mathbb{N}\ \forall n \geq \tilde{N} : \quad k_n > N \tag{2.75}$$

Therefore

$$\forall n \geq \tilde{N} : \quad |x_{k_n} - x| < \epsilon \tag{2.76}$$

□

Example 2.43. Consider the following sequence

$$x_n = \frac{n^{2n} + 2n^n}{n^{3n} - n^n}$$

This can be rewritten as

$$\frac{n^{2n} + 2n^n}{n^{3n} - n^n} = \frac{(n^n)^2 + 2(n^n)}{(n^n)^3 - (n^n)}$$

Introduce the subsequence $k_n = n^n$:

$$\lim_{k \to \infty} \frac{k^2 + 2k}{k^3 - k} = 0 \implies \lim_{n \to \infty} \frac{n^{2n} + 2n^n}{n^{3n} - n^n} = 0$$

2.3 Convergence of Series

Definition 2.44. Let $(x_n) \subset \mathbb{R}$ (or $\mathbb{C}$). Then the series

$$\sum_{k=1}^{\infty} x_k$$

is the sequence of partial sums (s_n):

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

If the series converges, then $\sum_{k=1}^{\infty}$ denotes the limit.

Theorem 2.45. *Let $(x_n) \subset \mathbb{R}$ (or $\mathbb{C}$). Then*

$$\sum_{n=1}^{\infty} x_n \text{ converges} \implies (x_n) \text{ null sequence}$$

Proof. Let $s_n = \sum_{n=1}^{\infty} x_n$. This is a Cauchy series. Let $\epsilon > 0$. Then

$$\exists N \in \mathbb{N}\ \forall n \geq N : \quad |s_{n+1} - s_n| = |x_{n+1}| < \epsilon \tag{2.77}$$

□

Example 2.46 (Geometric series). Let $x \in \mathbb{R}$ (or $\mathbb{C}$). Then

$$\sum_{k=1}^{\infty} x^k$$

converges if $|x| < 1$. (Why?)

Example 2.47 (Harmonic series). This is a good example of why the inverse of Theorem 2.45 does not hold. Consider

$$x_n = \frac{1}{n}$$

This is a null sequence, but $\sum_{k=1}^{\infty} \frac{1}{k}$ does not converge. (Why?)

Lemma 2.48. *Let $(x_n) \subset \mathbb{R}$ (or $\mathbb{C}$). Then*

$$\sum_{k=1}^{\infty} x_n \text{ converges} \iff \sum_{k=N}^{\infty} x_n \text{ converges for some } N \in \mathbb{N}$$

Proof. Left as an exercise for the reader. □

Theorem 2.49 (Alternating series test)**.** *Let $(x_n) \subset [0, \infty)$ be a monotonic decreasing null sequence. Then*

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

converges, and

$$\left| \sum_{k=1}^{\infty} (-1)^k x_k - \sum_{k=1}^{N} (-1)^k x_k \right| \leq x_{N+1}$$

Proof. Let $s_n = \sum_{k=1}^{n} (-1)^k x_n$, and define the sub sequences $a_n = s_{2n}$, $b_n = s_{2n+1}$. Then

$$a_{n+1} = s_{2n} - \underbrace{(x_{2n+1} - x_{2n+2})}_{\geq 0} \leq s_{2n} = a_n \tag{2.78}$$

Hence, (a_n) is monotonic decreasing. By the same argument, (b_n) is monotonic decreasing. Let $m, n \in \mathbb{N}$ such that $m \leq n$. Then

$$b_m \leq b_n = a_n - x_{2n+1} \leq a_n \leq a_m \tag{2.79}$$

Therefore (a_n), (b_n) are bounded. By Theorem 2.25, these sequence converge

$$(a_n) \xrightarrow{n \to \infty} a \qquad\qquad (b_n) \xrightarrow{n \to \infty} b \tag{2.80}$$

Furthermore

$$b_n - a_n = -x_{2n+1} \xrightarrow{n \to \infty} 0 \implies a = b \tag{2.81}$$

From eq. (2.79) we know that

$$b_m \leq b = a \leq a_m \tag{2.82}$$

So therefore

$$|s_{2n} - a| = a_n - a \leq a_n - b_n = x_{2n+1} \tag{2.83}$$

$$|s_{2n+1} - a| = b - b_n \leq a_{m+1} - b_n = x_{2n+2} \tag{2.84}$$

□

Example 2.50 (Alternating harmonic series).

$$\begin{aligned}
s = \sum_{k=1}^{\infty}(-1)^{k+1}\frac{1}{k} &= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots \\
&= \left(1 - \frac{1}{2}\right) - \frac{1}{4} + \left(\frac{1}{3} - \frac{1}{6}\right) - \frac{1}{8} + \left(\frac{1}{5} - \frac{1}{10}\right) - \frac{1}{12} + \cdots \\
&= \frac{1}{2} - \frac{1}{4} + \frac{1}{6} - \frac{1}{8} + \frac{1}{10} - \frac{1}{12} + \cdots \\
&= \frac{1}{2}\left(1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots\right) \\
&= \frac{1}{2}s
\end{aligned}$$

But $s \in \left[\frac{1}{2}, 1\right]$, this is an example on why rearranging infinite sums can lead to weird results.

Remark 2.51.

(i) The convergence behaviour does not change if we rearrange finitely many terms.

(ii) Associativity holds without restrictions

$$\sum_{k=1}^{\infty} x_k = \sum_{k=1}^{\infty}(x_{2k} + x_{2k-1})$$

(iii) Let I be a set, and define

$$\begin{aligned} I &\longrightarrow \mathbb{R} \\ i &\longmapsto a_i \end{aligned}$$

Consider the sum

$$\sum_{i \in I} a_i$$

If I is finite, there are no problems. However if I is infinite then the solution of that sum can depend on the order of summation!

Definition 2.52. Let $(x_n) \subset \mathbb{R}$ (or $\mathbb{C}$). The series $\sum_{k=1}^{\infty} x_k$ is said to converge absolutely if $\sum_{k=1}^{\infty} |x_k|$ converges.

Remark 2.53. Let $(x_n) \subset [0, \infty)$. Then the sequence

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

is monotonic increasing. If (s_n) is bounded it converges, if it is unbounded it diverges properly. The notation for absolute convergence is

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

Lemma 2.54. *Let $\sum_{k=1}^{\infty} x_k$ be a series. Then the following are all equivalent*

(i)

$$\sum_{k=1}^{\infty} x_k \textit{ converges absolutely}$$

(ii)

$$\left\{ \sum_{k \in I} |x_k| \,\middle|\, I \subset \mathbb{N} \textit{ finite} \right\} \textit{ is bounded}$$

(iii)

$$\forall \epsilon > 0 \; \exists I \subset \mathbb{N} \textit{ finite } \forall J \subset \mathbb{N} \textit{ finite}: \quad \sum_{k \in J \setminus I} |x_k| < \epsilon$$

Proof. To prove the equivalence of all of these statements, we will show that (i) $\implies$ (ii) $\implies$ (iii) $\implies$ (i). This is sufficient. First we prove (i) $\implies$ (ii). Let

$$\sum_{n=1}^{\infty} |x_n| = k \in [0, \infty) \tag{2.85}$$

Let $I \subset \mathbb{N}$ be a finite set, and let $N = \max I$. Then

$$\sum_{n \in I} |x_n| \leq \sum_{n=1}^{N} |x_n| \underset{\substack{\uparrow \\ \text{Monotony of the partial sums}}}{\leq} \sum_{n=1}^{\infty} |x_n| \tag{2.86}$$

Now to prove (ii) $\implies$ (iii), set

$$K := \left\{ \sum_{k \in I} |x_k| \,\middle|\, I \subset \mathbb{N} \text{ finite} \right\} \tag{2.87}$$

Let $\epsilon > 0$. Then by definition of sup

$$\exists I \subset \mathbb{N} \text{ finite}: \quad \sum_{k \in I} |x_k| > k - \epsilon \tag{2.88}$$

Let $J \subset \mathbb{N}$ finite. Then

$$k - \epsilon < \sum_{k \in I} |x_k| \leq \sum_{k \in I \cup J} |x_k| \leq K \tag{2.89}$$

Hence

$$\sum_{k \in J \setminus I} |x_k| = \sum_{k \in I \cup J} |x_k| - \sum_{k \in I} |x_k| \leq \epsilon \tag{2.90}$$

Finally we show that (iii) $\implies$ (i). Choose $I \subset \mathbb{N}$ finite such that

$$\forall J \subset \mathbb{N} \text{ finite}: \quad \sum_{k \in J \setminus I} |x_k| < 1 \tag{2.91}$$

Then $\forall J \subset \mathbb{N}$ finite

$$\sum_{k \in J} |x_k| \leq \sum_{k \in J \setminus I} |x_k| + \sum_{k \in I} |x_k| \leq \sum_{k \in I} |x_k| + 1 \tag{2.92}$$

Therefore $\sum_{k=1}^{n} |x_k|$ is bounded and monotonic increasing, and hence it is converging. So $\sum_{k=1}^{\infty} |x_k| < \infty$. □

Theorem 2.55. *Every absolutely convergent series converges and the limit does not depend on the order of summation.*

Proof. Let $\sum_{k=1}^{\infty} x_k$ be absolutely convergent and let $\epsilon > 0$. Choose $I \subset \mathbb{N}$ finite such that

$$\forall J \subset \mathbb{N}: \quad \sum_{k \in I} |x_k| < \epsilon \tag{2.93}$$

Choose $N = \max I$. Define the series

$$s_n = \sum_{k=1}^{n} x_k \tag{2.94}$$

Then for $n \leq m \leq N$

$$|s_n - s_m| \leq \sum_{k=m+1}^{n} |x_k| \leq \sum_{k \in \{1, \cdots, n\} \setminus I} |x_k| < \epsilon \tag{2.95}$$

Hence s_n is a Cauchy sequence, so it converges. Let $\phi : \mathbb{N} \to \mathbb{N}$ be a bijective mapping. According to Lemma 2.54 the series $\sum_{k=1}^{\infty} x_{\phi(n)}$ converges absolutely. Let $\epsilon > 0$. According to the same Lemma

$$\exists I \subset \mathbb{N} \text{ finite } \forall J \subset \mathbb{N} \text{ finite}: \quad \sum_{k \in J \setminus I} |x_k| < \frac{\epsilon}{2} \tag{2.96}$$

Choose $N \in \mathbb{N}$ such that

$$I \subset \{1, \cdots, N\} \cap \{\phi(1), \phi(2), \cdots, \phi(n)\} \tag{2.97}$$

Then for $n \geq N$

$$\begin{aligned} \left| \sum_{k=1}^{\infty} x_k - \sum_{k=1}^{n} x_{\phi(k)} \right| &= \left| \sum_{k \in \{1, \cdots, N\} \setminus I} x_k - \sum_{k \in \{\phi(1), \cdots, \phi(n)\} \setminus I} x_k \right| \\ &\leq \sum_{k \in \{1, \cdots, N\} \setminus I} |x_k| + \sum_{k \in \{\phi(1), \cdots, \phi(n)\} \setminus I} |x_k| < \epsilon \end{aligned} \tag{2.98}$$

Therefore

$$\lim_{n \to \infty} \left(\sum_{k=1}^{n} x_k - \sum_{k=1}^{n} x_{\phi(k)} \right) = 0 \tag{2.99}$$

□

Theorem 2.56. *Let $\sum_{k=1}^{\infty} x_k$ be a converging series. Then*

$$\left| \sum_{k=1}^{\infty} x_k \right| \leq \sum_{k=1}^{\infty} |x_k|$$

Proof. Left as an exercise for the reader. □

Theorem 2.57 (Direct comparison test)**.** *Let $\sum_{k=1}^{\infty} x_k$ be a series. If a converging series $\sum_{k=1}^{\infty} y_k$ exists with $|x_k| \leq y_k$ for all sufficiently large k, then $\sum_{k=1}^{\infty} x_k$ converges absolutely. If a series $\sum_{k=1}^{\infty} z_k$ diverges with $0 \leq z_k \leq x_k$ for all sufficiently large k, then $\sum_{k=1}^{\infty} x_k$ diverges.*

Proof.

$$\sum_{k=1}^{n} |x_k| \leq \sum_{k=1}^{n} y_k \implies \sum_{k=1}^{n} x_k \text{ bounded} \overset{\text{Lem. 2.54}}{\implies} \sum_{k=1}^{\infty} |x_k| < \infty \tag{2.100}$$

$$\sum_{k=1}^{n} z_k \leq \sum_{k=1}^{n} x_k \implies \sum_{k=1}^{\infty} x_k \text{ unbounded} \tag{2.101}$$

□

Corollary 2.58 (Ratio test)**.** *Let (x_n) be a sequence. If $\exists q \in (0,1)$ such that*

$$\left|\frac{x_{n+1}}{x_n}\right| \leq q$$

for a.e. $n \in \mathbb{N}$, then $\sum_{k=1}^{\infty} x_k$ converges absolutely. If

$$\left|\frac{x_{n+1}}{x_n}\right| \geq 1$$

then the series diverges.

Proof. Let $q \in (0,1)$ and choose $N \in \mathbb{N}$ such that

$$\forall n \geq N : \quad \left|\frac{x_{n+1}}{x_n}\right| \leq q \tag{2.102}$$

Then

$$|x_{N+1}| \leq q|x_N|, \; |x_{N+2}| \leq q|x_{N+1}| \leq q^2|x_N|, \; \cdots \tag{2.103}$$

This means that

$$\sum_{k=1}^{\infty} |x_k| \leq \sum_{k=1}^{N} |x_k| + \sum_{k=N+1}^{\infty} q^{k-N} \cdot |x_N| < \infty \tag{2.104}$$

Hence, $\sum_{k=1}^{\infty} x_k$ converges absolutely. Now choose $N \in \mathbb{N}$ such that

$$\forall n \geq N : \quad \left|\frac{x_{n+1}}{x_n}\right| > 1 \tag{2.105}$$

However this means that

$$|x_{n+1}| \geq |x_n| \quad \forall n \geq N \tag{2.106}$$

So (x_n) is monotonic increasing and therefore not a null sequence. Hence $\sum_{k=1}^{\infty} x_k$ diverges. □

Corollary 2.59 (Root test)**.** *Let (x_n) be a sequence. If $\exists q \in (0,1)$ such that*

$$\sqrt[n]{|x_n|} \leq q$$

for a.e. $n \in \mathbb{N}$, then $\sum_{k=1}^{\infty} x_k$ converges absolutely. If

$$\sqrt[n]{|x_n|} \geq 1$$

for all $n \in \mathbb{N}$ then $\sum_{k=1}^{\infty} x_k$ diverges.

Proof. Left as an exercise for the reader. □

Remark 2.60. The previous tests can be summed up by the formulas

$$\lim_{n\to\infty}\left|\frac{x_{n+1}}{x_n}\right| < 1 \qquad\qquad \lim_{n\to\infty}\sqrt[n]{|x_n|} < 1$$

$$\lim_{n\to\infty}\left|\frac{x_{n+1}}{x_n}\right| > 1 \qquad\qquad \lim_{n\to\infty}\sqrt[n]{|x_n|} > 1$$

for convergence and divergence respectively. If any of these limits is equal to 1 then the test is inconclusive.

Example 2.61. Let $z \in \mathbb{C}$. Then

$$\exp(z) := \sum_{k=0}^{\infty} \frac{z^k}{k!}$$

converges. To prove this, apply the ratio test:

$$\frac{|z|^{k+1}k!}{(k+1)!|z|^k} = \frac{|z|}{k+1} \longrightarrow 0$$

The function $\exp : \mathbb{C} \to \mathbb{C}$ is called the exponential function.

Remark 2.62 (Binomial coefficient). The binomial coefficient is defined as

$$\binom{n}{0} := 1 \qquad\qquad \binom{n}{k+1} = \binom{n}{k} \cdot \frac{n-k}{k+1}$$

and represents the number of ways one can choose k objects from a set of n objects. Some rules are

(i)
$$\binom{n}{k} = 0 \quad \text{if } k > n$$

(ii)
$$k \le n : \quad \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

(iii)
$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$$

(iv)

$$\forall x, y \in \mathbb{C} : \quad (x+y)^n = \sum_{k=1}^{n} \binom{n}{k} x^k y^{n-k}$$

Theorem 2.63.

$$\forall u, v \in \mathbb{C} : \quad \exp(u+v) = \exp(u) \cdot \exp(v)$$

Proof.

$$\begin{aligned} \exp(u) \cdot \exp(v) = \left(\sum_{n=0}^{\infty} \frac{u^n}{n!} \right) \cdot \left(\sum_{m=0}^{\infty} \frac{v^m}{m!} \right) &= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{u^n v^m}{n! m!} \\ &= \sum_{l=0}^{\infty} \sum_{k=0}^{l} \frac{u^k v^{l-k}}{k!(l-k)!} \qquad (2.107) \\ &= \sum_{l=0}^{\infty} \frac{(u+v)^l}{l!} \\ &= \exp(u+v) \end{aligned}$$

□

Remark 2.64. We define Euler's number as

$$e := \exp(1)$$

We will also take note of the following rules $\forall x \in \mathbb{C}, n \in \mathbb{N}$

$$\exp(0) = \exp(x)\exp(-x) = 1 \implies \exp(-x) = \frac{1}{\exp(x)}$$

$$\exp(nx) = \exp(x + x + x + \cdots + x) = \exp(x)^n$$

$$\exp(x)^{\frac{1}{n}} = \exp\left(\frac{x}{n}\right)$$

Alternatively we can write

$$\exp(z) = e^z$$

Theorem 2.65. *Let $x, y \in \mathbb{R}$.*

(i)

$$x < y \implies \exp(x) < \exp(y)$$

(ii)

$$\exp(x) > 0 \quad \forall x \in \mathbb{R}$$

(iii)

$$\exp(x) \geq 1 + x \quad \forall x \in \mathbb{R}$$

(iv)

$$\lim_{n\to\infty} \frac{n^d}{\exp(n)} = 0 \quad \forall d \in \mathbb{N}$$

Proof.

(i) Left as an exercise for the reader.

(ii) For $x \geq 0$ this is trivial. For $x < 0$

$$\exp(x) = \frac{1}{\exp(-x)} > 0 \tag{2.108}$$

(iii) For $x \geq 0$ this is trivial. For $x < 0$

$$\sum_{k=0}^{\infty} \frac{x^k}{k!} \tag{2.109}$$

is an alternating series, and therefore the statement follows from Theorem 2.49.

(iv) Let $d \in \mathbb{N}$. Then $\forall n \in \mathbb{N}$

$$0 < \frac{n^d}{\exp(n)} < \frac{n^d}{\sum_{k=0}^{d+1} \frac{n^k}{k!}} \xrightarrow{n\to\infty} 0 \tag{2.110}$$

□

Definition 2.66. Define

$$\sin, \cos : \mathbb{R} \longrightarrow \mathbb{R}$$

as

$$\sin(x) := \mathrm{Im}(\exp(ix))$$
$$\cos(x) := \mathrm{Re}(\exp(ix))$$

Remark 2.67.

(i) Euler's formula

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

(ii) $\forall z \in \mathbb{C}: \quad \overline{\exp(z)} = \exp(\bar{z})$

$$|\exp(ix)|^2 = \exp(ix) \cdot \overline{\exp(ix)} = \exp(ix) \cdot \exp(-ix) = 1$$

Also:

$$1 = \cos^2(x) + \sin^2(x)$$

On the symmetry of cos and sin:

$$\cos(-x) + i\sin(-x) = \exp(-ix) = \overline{\exp(ix)} = \cos(x) - i\sin(x)$$

(iii) From

$$\exp(ix) = \sum_{k=0}^{\infty} \frac{(ix)^k}{k!} \quad (i^0 = 1, i^1 = i, i^2 = -1, i^3 = -i, i^4 = 1, \cdots)$$

follow the following series

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

(iv) For $x \in \mathbb{R}$

$$\begin{aligned} \exp(i2x) &= \cos(2x) + i\sin(2x) \\ &= (\cos(x) + i\sin(x))^2 \\ &= \cos^2(x) - \sin^2(x) + 2i\sin(x)\cos(x) \end{aligned}$$

By comparing the real and imaginary parts we get the following identities

$$\begin{aligned} \cos(2x) &= \cos^2(x) - \sin^2(x) \\ \sin(2x) &= 2\sin(x)\cos(x) \end{aligned}$$

(v) Later we will show that cos as exactly one root in the interval $[0, 2]$. We define π as the number in the interval $[0, 4]$ such that $\cos\left(\frac{\pi}{2}\right) = 0$.

$$\implies \sin\left(\frac{\pi}{2}\right) = \pm 1$$

cos and sin are 2π-periodic.

Theorem 2.68. $\forall z \in \mathbb{C}$

$$\lim_{n\to\infty}\left(1+\frac{z}{n}\right)^n = \lim_{n\to\infty}\left(1-\frac{z}{n}\right)^{-n} = \exp(z)$$

Proof. Without proof. $\square$

Chapter 3

Linear Algebra

3.1 Vector Spaces

We introduce the new field $\mathbb{K}$ which will stand for any field. It can be either $\mathbb{R}$, $\mathbb{C}$ or any other set that fulfils the field axioms.

Definition 3.1. A vector space is a set V with the operations

Addition	Scalar Multiplication
$\begin{aligned} + : V \times V &\longrightarrow V \\ (x, y) &\longmapsto x + y \end{aligned}$	$\begin{aligned} \cdot : \mathbb{K} \times V &\longrightarrow V \\ (\alpha, y) &\longmapsto \alpha x \end{aligned}$

We require the following conditions for these operations

(i) $\exists 0 \in V \ \forall x \in V : \quad x + 0 = x$

(ii) $\forall x \in V \ \exists (-x) \in V : \quad x + (-x) = 0$

(iii) $\forall x, y \in V : \quad x + y = y + x$

(iv) $\forall x, y, z \in V : \quad (x + y) + z = x + (y + z)$

(v) $\forall \alpha \in \mathbb{K} \ \forall x, y \in V : \quad \alpha(x + y) = \alpha x + \alpha y$

(vi) $\forall \alpha, \beta \in \mathbb{K} \ \forall x \in V : \quad (\alpha + \beta)x = \alpha x + \beta x$

(vii) $\forall \alpha, \beta \in \mathbb{K} \ \forall x \in V : \quad (\alpha\beta)x = \alpha(\beta x)$

(viii) $\forall x \in V : \quad 1 \cdot x = x$

Elements from V are called vectors, elements from $\mathbb{K}$ are called scalars.

Remark 3.2. We now have two different addition operations that are denoted the same way:

(i) $+ : V \times V \to V$

(ii) $+ : \mathbb{K} \times \mathbb{K} \to \mathbb{K}$

Analogously there are two neutral elements and two multiplication operations.

Example 3.3.

(i) $\mathbb{K}$ is already a vector space

(ii) $V = \mathbb{K}^2$. In the case that $\mathbb{K} = \mathbb{R}$ this vector space is the two-dimensional Euclidean space. The neutral element is $(0,0)$, and the inverse is $(\chi_1, \chi_2) \to (-\chi_1, -\chi_2)$. This can be extended to $\mathbb{K}^n$.

(iii) $\mathbb{K}$-valued sequences:

$$V = \left\{ (\chi_n)_{n\in\mathbb{N}} \,\middle|\, \chi \in \mathbb{K} \quad \forall n \in \mathbb{N} \right\}$$

(iv) Let M be a set. Then the set of all $\mathbb{K}$-valued functions on M is a vector space

$$V = \{ f \mid f : M \to \mathbb{K} \}$$

Definition 3.4. Let V be a vector space, let $x, x_1, \cdots, x_n \in V$ and let $M \subset V$.

(i) x is said to be a linear combination of $x_1, \cdots, x_n$ if $\exists \alpha_1, \cdots, \alpha_n \in \mathbb{K}$ such that

$$x = \sum_{k=1}^{n} \alpha_k x_k$$

(ii) The set of all linear combinations of elements from M is called the *span*, or the *linear hull* of M

$$\operatorname{span} M := \left\{ \sum_{k=1}^{n} \alpha_k x_k \,\middle|\, n \in \mathbb{N},\ \alpha_1, \cdots, \alpha_n \in \mathbb{K},\ x_1, \cdots, x_n \in V \right\}$$

(iii) M (or the elements of M) are said to be linearly independent if $\forall \alpha_1, \cdots, \alpha_n \in \mathbb{K},\ x_1, \cdots, x_n \in V$

$$\sum_{k=1}^{n} \alpha_k x_k = 0 \implies \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$$

(iv) M is said to be a generator (of V) if

$$\operatorname{span} M = V$$

(v) M is said to be a basis of V if it is a generator and linearly independent.

(vi) V is said to be finite-dimensional if there is a finite generator.

Example 3.5.

(i) For $V = \mathbb{R}^2$ consider the vectors $x = (1, 0)$, $y = (1, 1)$. These vectors are linearly independent, since

$$\alpha x + \beta y = \alpha(1, 0) + \beta(1, 1) = (0, 0) \implies \alpha + \beta = 0 \wedge \beta = 0$$

So therefore $\alpha = \beta = 0$. We can show that $\operatorname{span}\{x, y\} = \mathbb{R}^2$ because

$$(\alpha, \beta) = (\alpha - \beta)x + \beta y$$

So $\{x, y\}$ is a generator, hence $\mathbb{R}^2$ is finite-dimensional.

(ii) For $V = \mathbb{R}^3$ consider $x = (1, -1, 2)$, $y = (2, -1, 0)$, $z = (4, -3, 3)$. These vectors are linearly dependent because

$$2x + y - z = (0, 0, 0)$$

(iii) Let $V = \{f \mid f : \mathbb{R} \to \mathbb{R}\}$. Consider the vectors

$$\begin{aligned} f_n : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto x^n \end{aligned}$$

The $f_0, f_1, \cdots, f_n, \cdots$ are linearly independent, because

$$0 = \sum_{=1}^{\infty} k = 0^n \alpha_k f_k = \sum_{=1}^{\infty} k = 0^n \alpha_k x^k$$

implies $\alpha_0 = \alpha_1 = \cdots = \alpha_n = 0$. The span of the f_k is the set of all polynomials of $(\leq n)$-th degree. The function $x \mapsto (x-1)^3$ is a linear combination of $f_0, \cdots, f_3$:

$$(x - 1)^3 = x^3 - 3x^2 + 3x - 1$$

Remark 3.6. Let V be a vector space, $y \in V$ a linear combination of $y_1, \cdots, y_n$, and each of those a linear combination of $x_1, \cdots, x_n$. I.e.

$$\exists \alpha_1, \cdots, \alpha_n \in \mathbb{K}: \quad y = \sum_{k=1}^{n} \alpha_k y_k$$

and

$$\exists \beta_{k,l} \in \mathbb{K}: \quad y_k = \sum_{l=1}^{n} \beta_{k,l} x_l$$

Then

$$y = \sum_{k=1}^{n} \alpha_k y_k = \sum_{k=1}^{n} \alpha_k \sum_{l=1}^{n} \beta_{k,l} x_l = \sum_{l=1}^{n} \underbrace{\left(\sum_{k=1}^{n} \alpha_k \beta_{k,l} \right)}_{\in \mathbb{K}} x_l$$

So therefore

$$\operatorname{span}(\operatorname{span}(M)) = \operatorname{span}(M)$$

Theorem 3.7. *Let V be a finite-dimensional vector space, and let $x_1, \cdots, x_n \in V$. Then the following are equivalent*

(i) $x_1, \cdots, x_n$ *is a basis.*

(ii) $x_1, \cdots, x_n$ *is a minimal generator (Minimal means that no subset is a generator).*

(iii) $x_1, \cdots, x_n$ *is a maximal linearly independent system (Maximal means that* $x_1, \cdots, x_n, y$ *is not linearly independent).*

(iv) $\forall x \in V$ *there exists a unique* $\alpha_1, \cdots, \alpha_n \in \mathbb{K}$

$$x = \sum_{k=1}^{n} \alpha_k x_k$$

Proof. First we prove "(i) $\implies$ (ii)". Let $x_1, \cdots, x_n$ be a basis of V. By definition $x_1, \cdots, x_n$ is a generator. Assume that $x_2, \cdots, x_n$ is still a generator, then

$$\exists \alpha_2, \cdots, \alpha_n \in \mathbb{K}: \quad x_1 = \sum_{k=1}^{n} \alpha_k x_k \tag{3.1}$$

However this contradicts the linear independence of the basis. Next, to prove "(ii) $\implies$ (iii)" let $x_1, \cdots, x_n$ be a minimal generator. Let $\alpha_1, \cdots, \alpha_n \in \mathbb{K}$ such that

$$0 = \sum_{k=1}^{n} \alpha_k x_k \tag{3.2}$$

Assume that one coefficient is $\neq 0$ (w.l.o.g. $\alpha_1 = 0$). Then

$$x_1 = \sum_{k=2}^{n} -\frac{\alpha_k}{\alpha_1} x_k \tag{3.3}$$

$x_1, \cdots, x_n$ is a generator, i.e. for $x \in V$

$$\exists \beta_1, \cdots, \beta_n \in \mathbb{K}: \quad x = \sum_{k=1}^{n} \beta_k x_k = \sum_{k=2}^{n} \left(\beta_k - \frac{\alpha_k}{\alpha_1} \right) x_k \tag{3.4}$$

But this implies that $x_2, \cdots, x_n$ is a generator. That contradicts the assumption that $x_1, \cdots, x_n$ was minimal.

$$\implies \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0 \tag{3.5}$$

Now let $y \in V$. Then

$$\exists \gamma_1, \cdots, \gamma_n \in \mathbb{K}: \quad y = \sum_{k=1}^{n} \gamma_k x_k \tag{3.6}$$

So $x_1, \cdots, x_n, y$ is linearly dependent, and therefore $x_1, \cdots, x_n$ is maximal. To prove "(iii) $\implies$ (iv)" let $x_1, \cdots, x_n$ be a maximal linearly independent system. If $y \in V$, then

$$\exists \alpha_1, \cdots, \alpha_k, \beta \in \mathbb{K}: \quad \sum_{k=1}^{n} \alpha_k x_k + \beta y = 0 \tag{3.7}$$

Assume $\beta = 0$, then consequently

$$x_1, \cdots, x_n \text{ linearly independent} \implies \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0 \tag{3.8}$$

This is a contradiction, so therefore $\beta \neq 0$:

$$y = \sum_{k=1}^{n} -\frac{\alpha_k}{\beta} x_k \tag{3.9}$$

The uniqueness of these coefficients are left as an exercise for the reader. Finally, to finish the proof we need to show "(iv) $\implies$ (i)". By definition

$$V = \operatorname{span}\{x_1, \cdots, x_n\} \tag{3.10}$$

Hence, $\{x_1, \cdots, x_n\}$ is a generator. In case

$$0 = \sum_{k=1}^{n} \alpha_k x_k \tag{3.11}$$

holds, then $\alpha_1 = \cdots = \alpha_n = 0$ follows from the uniqueness. □

Corollary 3.8. *Every finite-dimensional vector space has a basis.*

Proof. By condition, there is a generator $x_1, \cdots, x_n$. Either this generator is minimal (then it would be a basis), or we remove elements until it is minimal. □

Lemma 3.9. *Let V be a vector space and $x_1, \cdots, x_k \in V$ a linearly independent set of elements. Let $y \in V$, then*

$$x_1, \cdots, x_k, y \textit{ linearly independent } \iff y \notin \operatorname{span}\{x_1, \cdots, x_k\}$$

Proof. To prove " $\Longleftarrow$ ", assume $y \neq \operatorname{span}\{x_1, \cdots, x_k\}$. Therefore $x_1, \cdots, x_k, y$ must be linearly independent. To see this, consider

$$0 = \sum_{k=1}^{n} \alpha_k x_k + \beta y \quad \alpha_1, \cdots, \alpha_n \in \mathbb{K} \tag{3.12}$$

Then $\beta = 0$, otherwise we could solve the above for y, and that would contradict our assumption. The argument works in the other direction as well. □

Theorem 3.10 (Steinitz exchange lemma)**.** *Let V be a finite-dimensional vector space. If $x_1, \cdots, x_m$ is a generator and $y_1, \cdots, y_n$ a linear independent set of vectors, then $n \leq m$. In case $x_1, \cdots, x_m$ and $y_1, \cdots, y_n$ are both bases, then $n = m$.*

Heuristic Proof. Let $K \in \{0, \cdots, \min\{m, n\} - 1\}$ and let

$$x_1, \cdots, x_K, y_{K+1}, \cdots, y_n \tag{3.13}$$

be linearly independent. Assume that

$$x_{K+1}, \cdots, x_m \in \operatorname{span}\{x_1, \cdots, x_k, y_{K+2}, \cdots, y_n\} \tag{3.14}$$

Then

$$y_{K+1} \in \operatorname{span}\{x_1, \cdots, x_m\} \subset \operatorname{span}\{x_1, \cdots, x_K, y_{K+2}, \cdots, y_m\} \tag{3.15}$$

This contradicts with the linear independence of $x_1, \cdots, x_K, y_{K+2}, \cdots y_n$. Furthermore,

$$\exists x_i \in V: \quad x_i \notin \operatorname{span}\{x_1, \cdots, x_K, y_{K+2}, \cdots, y_n\} \tag{3.16}$$

W.l.o.g. $x : i = x_{K+1}$. By Lemma 3.9, $x_1, \cdots, x_{K+1}, y_{K+2}, \cdots y_n$ is linearly independent. We can now sequentially replace y_i with x_i without losing the linear independence. Assume $n > m$, then this process leads to a linear independent system $x_1, \cdots, x_m, y_{m+1}, \cdots, y_n$. But since $x_1, \cdots, x_m$ is a generator, y_{m+1} is a linear combination of $x_1, \cdots, x_m$. If $x_1, \cdots, x_m$ and $y_1, \cdots, y_n$ are both bases, then we cannot change the roles and therefore $m = n$. $\square$

Definition 3.11. The amount of elements in a basis is said to be the dimension of V, and is denoted as $\dim V$.

Example 3.12.

(i) Let $V = \mathbb{R}^n$ (or $\mathbb{C}^n$). Define

$$e_k = (0, 0, \cdots, 0, \underset{\substack{\uparrow \\ \text{k-th position}}}{1}, 0, \cdots, 0)$$

Then $e_1, \cdots, e_n$ is a basis, in fact, it is the standard basis of $\mathbb{R}^n$ ($\mathbb{C}^n$).

(ii) Let V be the vector space of polynomials

$$V = \left\{ f : \mathbb{R} \longrightarrow \mathbb{R} \,\middle|\, n \in \mathbb{N},\ \alpha_1, \cdots, \alpha_n \in \mathbb{R}, \quad f(x) = \sum_{k=1}^{n} \alpha_k x^k \quad \forall x \in \mathbb{R} \right\}$$

This space has the basis

$$\{x \longmapsto x^n \mid n \in \mathbb{N}_0\}$$

Corollary 3.13. *In an n-dimensional vector space, every generator has at least n elements, and every linearly independent system has at most n elements.*

Proof. Let $M \subset \operatorname{span}\{x_1, \cdots, x_n\}$. Then

$$V = \operatorname{span} M \subset \operatorname{span} x_1, \cdots, x_n \tag{3.17}$$

Hence, $x_1, \cdots, x_n$ is a generator. On the other hand, assume

$$\exists y \in M \setminus \operatorname{span}\{x_1, \cdots, x_n\} \tag{3.18}$$

Then $x_1, \cdots, x_n, y$ is linearly independent (Lemma 3.9), and we can sequentially add elements from M until $x_1, \cdots, x_n, y_{n+1}, \cdots, y_{n+m}$ is a generator. □

Definition 3.14 (Vector subspace)**.** Let V be a vector space. A non-empty set $W \subset V$ is called a vector subspace if

$$\forall x, y \in W \ \forall \alpha \in \mathbb{K} : \quad x + \alpha y \in W$$

Example 3.15. Consider

$$W = \left\{(\chi, \chi) \in \mathbb{R}^2 \,\middle|\, \chi \in \mathbb{R}\right\}$$

This is a subspace, because

$$(\chi, \chi) + \alpha(\eta, \eta) = (\chi + \alpha\eta, \chi + \alpha\eta)$$

However,

$$A = \left\{(\chi, \eta) \in \mathbb{R}^2 \,\middle|\, \chi^2 + \eta^2 = 1\right\}$$

is not a subspace, because $(1, 0), (0, 1) \in A$, but $(1, 1) \notin A$.

Remark 3.16.

(i) Every subspace $W \subset V$ contains the 0 and the inverse elements.

(ii) Let $W \subset V$ be a subspace. Then

$$\forall x_1, \cdots, x_n \in W, \ \alpha_1, \cdots, \alpha_n \in \mathbb{K} : \quad \sum_{k=1}^{n} \alpha_k x_k \in W$$

Furthermore, $M \subset W \implies \operatorname{span} M \subset W$.

(iii) $M \subset V$ is a subspace if and only of $\operatorname{span} M = M$.

(iv) Let I be an index set, and $W_i \subset V$ subspaces. Then

$$\bigcap_{i \in I} W_i$$

is also a subspace

(v) The previous doesn't hold for unions.

(vi) Let $M \subset V$:

$$\operatorname{span} M = \bigcap_{W \supset M \text{ subspace of } V} W$$

3.2 Matrices and Gaussian elimination

Definition 3.17. Let $a_{ij} \in \mathbb{K}$, with $i \in \{1, \cdots, n\}$, $j \in \{1, \cdots, m\}$. Then

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}$$

is called an $n \times m$-matrix. (n, m) is said to be the dimension of the matrix. An alternative notation is

$$A = (a_{ij}) \in \mathbb{K}^{n \times m}$$

$\mathbb{K}^{n \times m}$ is the space of all $n \times m$-matrices. The following operations are defined for $A, B \in \mathbb{K}^{n \times m}$, $C \in \mathbb{K}^{m \times l}$:

(i) Addition

$$A + B = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1m} + b_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} + b_{n1} & \cdots & a_{nm} + b_{nm} \end{pmatrix}$$

(ii) Scalar multiplication

$$\alpha \cdot A = \begin{pmatrix} \alpha a_{11} & \cdots & \alpha a_{1m} \\ \vdots & \ddots & \vdots \\ \alpha a_{n1} & \cdots & \alpha a_{nm} \end{pmatrix}$$

(iii) Matrix multiplication

$$A \cdot C = \begin{pmatrix} a_{11}c_{11} + a_{12}c_{21} + \cdots + a_{1m}c_{m1} & \cdots & a_{11}c_{1l} + a_{12}c_{2l} + \cdots + a_{1m}c_{ml} \\ \vdots & \ddots & \vdots \\ a_{n1}c_{11} + a_{n2}c_{21} + \cdots + a_{nm}c_{m1} & \cdots & a_{n1}c_{1l} + a_{n2}c_{2l} + \cdots + a_{nm}c_{ml} \end{pmatrix}$$

or in shorthand notation

$$(AC)_{ij} = \sum_{k=1}^{m} a_{ik}c_{kj}$$

(iv) Transposition

The transposed matrix $A^T \in \mathbb{K}^{m\times n}$ is created by writing the rows of A as the columns of A^T (and vice versa).

(v) Conjugate transposition

$$A^H = \left(\overline{A}\right)^T$$

Remark 3.18.

(i) $\mathbb{K}^{n\times m}$ (for $n, m \in \mathbb{N}$) is a vector space.

(ii) $A \cdot B$ is only defined if A has as many columns as B has rows.

(iii) $\mathbb{K}^{n\times 1}$ and $\mathbb{K}^{1\times n}$ can be trivially identified with $\mathbb{K}^n$.

(iv) Let A, B, C, D, E matrices of fitting dimensions and $\alpha \in \mathbb{K}$. Then

$$\begin{aligned}(A+B)C &= AC + BC\\ A(B+C) &= AB + AC\\ A(CE) &= (AC)E\\ \alpha(AC) &= (\alpha A)C = A(\alpha C)\end{aligned}$$

$$\begin{aligned}(A+B)^T &= A^T + B^T & \qquad \overline{(A+B)} &= \overline{A} + \overline{B}\\ (\alpha A)^T &= \alpha (A)^T & \overline{(\alpha A)} &= \overline{A}\overline{A}\\ (AC)^T &= C^T \cdot A^T & \overline{(AC)} &= \overline{C}\overline{A}\end{aligned}$$

Proof of associativity. Let $A \in \mathbb{K}^{n\times m}, C \in \mathbb{K}^{m\times l}, E \in \mathbb{K}^{l\times p}$. Furthermore let $i \in \{1, \cdots, n\}, j \in \{1, \cdots, p\}$.

$$\begin{aligned}((AC)E)_{ij} &= \sum_{k=1}^{l} (AC)_{ik} E_{kj} = \sum_{k=1}^{l} \left(\sum_{\tilde{k}=1}^{m} a_{i\tilde{k}} c_{\tilde{k}k} \right) \cdot e_{kj}\\ &= \sum_{k=1}^{l} \sum_{\tilde{k}=1}^{m} a_{i\tilde{k}} \cdot c_{\tilde{k}k} \cdot e_{kj}\\ &= \sum_{\tilde{k}=1}^{m} a_{i\tilde{k}} \left(\sum_{k=1}^{l} c_{\tilde{k}k} e_{kj} \right)\\ &= \sum_{\tilde{k}=1}^{m} a_{i\tilde{k}} \cdot (CE)_{\tilde{k}j}\\ &= (A(CE))_{ij}\end{aligned} \tag{3.19}$$

$$\implies A(CE) = A(CE) \tag{3.20}$$

□

(v) Matrix multiplication is NOT commutative. First off, AB and BA are only well defined when $A \in \mathbb{K}^{n\times m}$ and $B \in \mathbb{K}^{m\times n}$. Example:

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

(vi) Let $n, m \in \mathbb{N}$. There exists exactly one neutral additive element in $\mathbb{K}^{n\times m}$, which is the zero matrix. Multiplication with the zero matrix yields a zero matrix.

(vii) We define

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0 & \text{else} \end{cases}$$

The respective matrix $I = (\delta_{ij}) \in \mathbb{K}^{n\times m}$ is called the identity matrix.

(viii) $A \neq 0$ and $B \neq 0$ can still result in $AB = 0$:

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

Example 3.19 (Linear equation system)*.* Consider the following linear equation system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m &= b_n \end{aligned}$$

This can be rewritten using matrices

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix} \qquad x = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \qquad b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

Which results in

$$Ax = B, \quad A \in \mathbb{K}^{m\times n}, x \in \mathbb{K}^{m\times 1}, b \in \mathbb{K}^{n\times 1}$$

Such an equation system is called homogeneous if $b = 0$.

Theorem 3.20. *Let $A \in \mathbb{K}^{n\times m}, b \in \mathbb{K}^n$. The solution set of the homogeneous equation system $Ax = 0$, (that means $\{x \in \mathbb{K}^m \mid Ax = 0\} \subset \mathbb{K}^m$) is a linear subspace. If x and $\tilde{x}$ are solutions of the inhomogeneous system $Ax = b$, then $x - \tilde{x}$ solves the corresponding homogeneous problem.*

Proof. $A \cdot 0 = 0$ shows that $Ax = 0$ has a solution. Let x, y be solutions, i.e. $Ax = 0$ and $Ay = 0$. Then $\forall \alpha \in \mathbb{K}$:

$$A(x + \alpha y) = Ax + A(\alpha y) = \underbrace{Ax}_{0} + \alpha(\underbrace{Ay}_{0}) = 0 \tag{3.21}$$

$$\implies x + \alpha y \in \{x \in \mathbb{K}^m \mid Ax = 0\} \tag{3.22}$$

Next, let $x, \tilde{x}$ be solutions of $Ax = b$, i.e.

$$Ax = b, \ A\tilde{x} = b \tag{3.23}$$

Then

$$A(x - \tilde{x}) = Ax - A\tilde{x} = b - b = 0 \tag{3.24}$$

Therefore, $x - \tilde{x}$ is the solution of the homogeneous equation system □

Remark 3.21 (Finding all solutions). First find a basis $e_1, \cdots, e_k$ of

$$\{x \in \mathbb{K}^m \mid Ax = 0\}$$

Next find some $x_0 \in \mathbb{K}^m$ such that $Ax_0 = b$. Then every solution of $Ax = b$ can be written as

$$x = x_0 + \alpha_1 e_1 + \cdots + \alpha_k e_k$$

Example 3.22. Let

$$A = \begin{pmatrix} 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad b = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \qquad c = \begin{pmatrix} 3 \\ 2 \\ 1 \\ 0 \end{pmatrix}$$

Then $Ax = b$ has no solution, since the fourth row would state $0 = 4$. However, $Ax = c$ has the particular solution

$$x = \begin{pmatrix} 3 \\ 0 \\ 2 \\ 1 \\ 0 \end{pmatrix}$$

If we consider the homogeneous problem $Ay = 0$, we can come up with the solution

$$y = \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} y_2 + \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} y_5$$

and in turn find the set of solutions

$$\left\{y \in \mathbb{K}^5 \,\middle|\, Ay = 0\right\} = \operatorname{span}\left\{(-2,1,0,0,0)^T, (-1,0,0,1,1)^T\right\}$$
$$\left\{x \in \mathbb{K}^5 \,\middle|\, Ax = c\right\} = \left\{(3,0,2,1,0)^T + \alpha(-2,1,0,0,0)^T + \beta(-1,0,0,1,1)^T\right\}$$

Definition 3.23 (Row Echelon Form)**.** A zero row is a row in a matrix containing only zeros. The first element of a row that isn't zero is called the pivot.

A matrix in row echelon form must meet the following conditions

(i) Every zero row is at the bottom

(ii) The pivot of a row is always strictly to the right of the pivot of the row above it

A matrix in reduced row echelon form must additionally meet the following conditions

(i) All pivots are 1

(ii) The pivot is the only non-zero element of its column

Remark 3.24. Let $A \in \mathbb{K}^{n\times m}$ and $b \in \mathbb{K}^n$. If A is in reduced row echelon form, then $Ax = b$ can be solved through trivial rearranging.

Definition 3.25 (Matrix row operations)**.** Let A be a matrix. Then the following are row operations

(i) Swapping of rows i and j

(ii) Addition of row i to row j

(iii) Multiplication of a row by $\lambda \neq 0$

(iv) Addition of row i multiplied by $lambda$ to row j

Theorem 3.26 (Gaussian Elimination)**.** *Every matrix can be converted into reduced row echelon form in finitely many row operations.*

Heuristic Proof. If A is a zero matrix the proof is trivial. But if it isn't:

- Find the first column containing a non-zero element.
  - Swap rows such that this element is in the first row
- Multiply every other row with multiples of the first row, such that all other entries in that column disappear.
- Repeat, but ignore the first row this time

At the end of this the matrix will be in reduced row echelon form. □

Definition 3.27. $A \in \mathbb{K}^{n\times n}$ is called invertible if there exists a multiplicative inverse. I.e.

$$\exists B \in \mathbb{K}^{n\times n}: \quad AB = BA = I$$

We denote the multiplicative inverse as A^{-1}

Remark 3.28. We have seen matrices $A \neq 0$ such that $A^2 = 0$. Such a matrix is not invertible.

Theorem 3.29. *Let $A, B, C \in \mathbb{K}^{n\times n}$, B invertible and $A = BC$. Then*

$$A \textit{ invertible} \iff C \textit{ invertible}$$

Especially, the product of invertible matrices is invertible.

Proof. Without proof. □

Remark 3.30. Matrix multiplication with A from the left doesn't "mix" the columns of matrix B

Theorem 3.31. *Let A be a matrix, and let $\tilde{A}$ be the result of row operations applied to A. Then*

$$\exists T \textit{ invertible}: \quad \tilde{A} = TA$$

We say: The left multiplication with T applies the row operations.

Heuristic proof. You can find invertible matrices $T_1, \cdots, T_n$ that each apply one row operation. Then we can see that

$$\tilde{A} = \underbrace{T_n T_{n-1} \cdots T_1}_{T} A \tag{3.25}$$

Since T is the product of invertible matrices, it must itself be invertible. □

Corollary 3.32. *Let $A \in \mathbb{K}^{n\times m}$, $b \in \mathbb{K}^n$, $T \in \mathbb{K}^{n\times m}$. Then $Ax = b$ and $TAx = Tb$ have the same solution sets.*

Proof. If $Ax = b$ it is trivial that

$$Ax = b \implies TAx = Tb \tag{3.26}$$

If $TAx = Tb$, then

$$Ax = T^{-1}TAx = T^{-1}Tb = b \tag{3.27}$$

□

Lemma 3.33. *Let $A \in field^{n\times m}$ be in row echelon form. Then*

$$A \textit{ invertible} \iff \textit{The last row is not a zero row}$$

and

$$A \textit{ invertible} \iff \textit{All diagonal entries are non-zero}$$

Proof. Let A be invertible with a zero-row as its last row. Then

$$(0, \cdots, 0, 1) \cdot A = (0, \cdots, 0, 0) \tag{3.28}$$

Multiplying with A^{-1} from the right would result in a contradiction. Therefore the last row of A can't be a zero row.

Now let the diagonal entries of A be non-zero. This means we can use row operations to transform A into the identity matrix, i.e.

$$\exists T \text{ invertible}: \quad TA = I \implies A = T^{-1} \tag{3.29}$$

□

Corollary 3.34. *Let $A \in \mathbb{K}^{n\times n}$. Then*

$$A \textit{ invertible} \iff \textit{Every row echelon form has non-zero diagonal entries}$$

and

$$A \textit{ invertible} \iff \textit{The reduced row echelon form is the identity matrix}$$

Proof. Every row echelon form of A has the form TA with T an invertible matrix. Especially, $\exists S$ invertible such that SA is in reduced row echelon form. Then

$$TA \text{ invertible} \iff A \text{ invertible} \tag{3.30}$$

$\square$

Remark 3.35. Let $A \in \mathbb{K}^{n\times n}$ be invertible, $B \in \mathbb{K}^{n\times m}$. Our goal is to compute $A^{-1}B$. First, write $(A \,|\, B)$. Now apply row operations until we reach the form $(I \,|\, \tilde{B})$. Let S be the matrix realising these operations, i.e. $SA = I$. Then $\tilde{B} = SB = A^{-1}B$. If $B = I$ this can be used to compute A^{-1}.

Example 3.36. Let

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

Rewrite this as

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right)$$

Turn this into

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right)$$

And finally

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right)$$

The right part of the above matrix is A^{-1}.

Definition 3.37. Let $A \in \mathbb{K}^{n\times m}$ and let $z_1, \cdots, z_n \in \mathbb{K}^{1\times m}$ be the rows of A. The row space of A is defined as

$$\operatorname{span}\{z_1, \cdots, z_n\}$$

The dimension of the row space is the row rank of the matrix. Analogously this works for the column space and the column rank. Later we will be able to show that row rank and column rank are always equal. They're therefore simply called rank of the matrix.

Theorem 3.38. *The row operations don't effect the row space.*

Proof. It is obvious that multiplication with λ and swapping of rows don't change the row space. Furthermore it is clear that every linear combination of $z_1 + z_2, z_2, \cdots, z_n$ is also a linear combination of $z_1, z_2, \cdots, z_n$, and vice versa. □

Theorem 3.39. *Let A be in row echelon form. Then the non-zero rows of the matrix are a basis of the row space of the matrix.*

Proof. Let $z_1, \cdots, z_k \in \mathbb{K}^{1\times n}$ be the non-zero rows of A. They create the space span $\{z_1, \cdots, z_n\}$, since $z_k, \cdots z_n$ are only zero rows. Analogously,

$$\alpha_1 z_1 + \alpha_2 z_2 + \cdots + \alpha_k z_k = 0 \tag{3.31}$$

Let j be the index of the column of the pivot of z_1. Then $z_2, \cdots, z_k$ have zero entries in the j-th column. Therefore

$$\alpha_1 \underbrace{z_{ij}}_{\neq 0} = 0 \implies \alpha_1 = 0 \tag{3.32}$$

By inductivity, this holds for every row. □

Remark 3.40. (i) To compute the rank of A, bring A into row echelon form and count the non-zero rows.

(ii) Let $v_1, \cdots, v_m \in \mathbb{K}^n$. To find a basis for

$$\operatorname{span}\{v_1, \cdots v_m\}$$

write $v_1, \cdots, v_m$ as rows of a matrix and bring it into row echelon form.

3.3 The Determinant

In this section we always define $A \in \mathbb{K}^{n\times n}$ and $z_1, \cdots, z_n$ the row vectors of A. We declare the mapping

$$\det : \mathbb{K}^{n\times n} \longrightarrow \mathbb{K}$$

and define

$$\det(A) := \det(z_1, z_2, \ldots, z_n)$$

Definition 3.41. There exists exactly one mapping det such that

(i) It is linear in the first row, i.e.

$$\det(z_1 + \lambda\tilde{z_1}, z_2, \cdots, z_n) = \det(z_1, z_2, \cdots, z_n) + \lambda \det(\tilde{z_1}, z_2, \cdots, z_n)$$

(ii) If $\tilde{A}$ is obtained from A by swapping two rows

$$\det(A) = -\det\left(\tilde{A}\right)$$

(iii) $\det(I) = 1$

This mapping is called the determinant, and we write

$$\det A = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

Example 3.42.

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}$$

$$\begin{aligned} \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} =& a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ & - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12} \end{aligned}$$

Remark 3.43. (i) Every determinant is linear in every row

(ii) If two rows are equal then $\det(A) = 0$

(iii) If one row (w.l.o.g. z_1) is a linear combination of the others, so

$$z_1 = \alpha_2 z_2 + \alpha_3 z_3 + \cdots + \alpha_n z_n, \quad \alpha_1, \cdots, \alpha_n \in \mathbb{K}$$

then

$$\begin{aligned} \det(z_1, z_2, \cdots, z_n) =& \alpha_2 \underbrace{\det(z_2, z_2, z_3, \cdots, z_n)}_{0} + \\ & \alpha_3 \underbrace{\det(z_3, z_2, z_3, \cdots, z_n)}_{0} + \\ & \cdots + \\ & \alpha_n \underbrace{\det(z_n, z_2, z_3, \cdots, z_n)}_{0} \\ =& 0 \end{aligned}$$

(iv) Adding a multiple of a row to another doesn't change the determinant

(v) Define

T_{ij}	swaps rows i and j
$M_i(\lambda)$	multiplies row i with $\lambda \neq 0$
$L_{ij}(\lambda)$	adds λ-times row j to row i

Then

$$\begin{aligned} \det(T_{ij}A) &= -\det(A) \\ \det(L_{ij}(\lambda)A) &= \det(A) \\ \det(M_i(\lambda)A) &= \lambda \det(A) \end{aligned}$$

Lemma 3.44. *Let* det *be the determinent, and $A, B \in \mathbb{K}^{n\times n}$. Let A be in row echelon form, then*

$$\det(AB) = a_{11} \cdot a_{22} \cdot \dots \cdot a_{nn} \cdot \det(B)$$

Proof. First consider the case of A not being invertible. This means that the last row of A is a zero row, which in turn means that $\det(A) = 0$. This also means that the last row of AB is a zero row and therefore $\det(AB) = 0$.

Now let A be invertible. This means that all the diagonal entries are non-zero. It is possible to bring A into diagonal form without changing the diagonal entries themselves. So, w.l.o.g. let A be in diagonal form:

$$A = M_n(a_{nn}) \cdot \dots \cdot M_2(a_{22}) M_1(a_{11}) I \tag{3.33}$$

and thus

$$\begin{aligned} \det(AB) &= \det(M_n(a_{nn}) \cdot \dots \cdot M_2(a_{22}) M_1(a_{11}) B) \\ &= a_{nn} \cdot \dots \cdot a_{22} \cdot a_{11} \det(B) \end{aligned} \tag{3.34}$$

□

Remark 3.45. For $B = I$ this results in

$$\det(A) = a_{11} a_{22} \cdots a_{nn}$$

Theorem 3.46. *Let $A, B \in \mathbb{K}^{n\times n}$. Then*

$$\det AB = \det A \cdot \det B$$

Proof. Let $i, j \in \{1, \cdots, n\}$ and $\lambda \neq 0$. Then

$$\det(T_{ij}AB) = -\det(AB) \tag{3.35a}$$

$$\det(L_{ij}(\lambda)AB) = \det(AB) \tag{3.35b}$$

Bring A with T_{ij} and $L_{ij}(\lambda)$ operations into row echelon form. Then

$$\det(AB) = a_{11}a_{22}\cdots a_{nn} \cdot \det(B) \tag{3.36}$$

and therefore

$$\det(AB) = \det A \cdot \det B \tag{3.37}$$

□

Corollary 3.47.

$$A \in \mathbb{K}^{n\times n} \textit{ invertible} \iff \det A \neq 0$$

Proof. Row operations don't effect the invertibility or the determinant (except for the sign) of a matrix. Therefore we can limit ourselves to matrices in row echelon form (w.l.o.g.). Let A be in row echelon form, then

$$\begin{aligned} \det A \neq 0 &\iff a_{11}a_{22}\cdots a_{nn} \neq 0 \\ &\iff a_{11} \neq 0, a_{22} \neq 0, \cdots, a_{nn} \neq 0 \\ &\iff A \text{ invertible since diagonal entries are non-zero} \end{aligned} \tag{3.38}$$

□

Theorem 3.48.

$$\det A = \det A^T$$

Proof. First consider the explicit representation of row operations:

$$T_{ij} = \begin{array}{cc} & \begin{array}{ccccc} & j & & i & \end{array} \\ \begin{array}{c} \\ i \\ \\ j \\ \\ \end{array} & \begin{pmatrix} 1 & & & & \\ & 0 & & 1 & \\ & & 1 & & \\ & 1 & & 0 & \\ & & & & 1 \end{pmatrix} \end{array} \tag{3.39a}$$

$$L_{ij}(\lambda) = \begin{array}{cc} & \begin{array}{ccccc} & & & j & \end{array} \\ \begin{array}{c} \\ i \\ \\ \\ \\ \end{array} & \begin{pmatrix} 1 & & & & \\ & 1 & & \lambda & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{pmatrix} \end{array} \tag{3.39b}$$

Thus we can see

$$\det(T_{ij}) = \det\big(T_{ij}^T\big) = -1 \tag{3.40a}$$

$$\det(L_{ij}(\lambda)) = \det\big(L_{ij}(\lambda)^T\big) = 1 \tag{3.40b}$$

Let T be one of those matrices. Then

$$\begin{aligned}\det\big((TA)^T\big) &= \det\big(A^T \cdot T^T\big) \\ &= \det A^T \cdot \det T^T \\ &= \det A^T \cdot \det T\end{aligned} \tag{3.41}$$

and

$$\det TA = \det A \cdot \det T \tag{3.42}$$

And therefore

$$\det\big((TA)^T\big) = \det(TA) \iff \det A^T = \det A \tag{3.43}$$

Now w.l.o.g. let A be in row echelon form. Let A be non-invertible, i.e. the last row is a zero row. Thus $\det A = 0$. This implies that A^T has a zero column. Row operations that bring A^T into row echelon form (w.l.o.g.) perserve this zero column. Therefore the resulting matrix must also have a zero column, and thus $\det\big(A^T\big) = 0$.

Now assume A is invertible, and use row operations to bring A into a diagonalised form (w.l.o.g.). For diagonalised matrices we know that

$$A = A^T \implies \det A = \det A^T \tag{3.44}$$

□

Remark 3.49. Let A_{ij} be the matrix you get by removing the i-th row and the j-th column from A.

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} \cdot a_{ij} \cdot \det(A_{ij}), \quad j \in \{1, \cdots, n\}$$

Remark 3.50 (Leibniz formula). Let $n \in \mathbb{N}$, and let there be a bijective mapping

$$\sigma : \{1, \cdots, n\} \longrightarrow \{1, \cdots, n\}$$

σ is a permutation. The set of all permutations is labeled S_n, and it contains $n!$ elements. Then

$$\det A = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)}$$

A permutation that swaps exactly two elements is called elementary permutation. Every permutation can be written as a number of consecutively executed elementary permutations.

$$\operatorname{sgn}(\sigma) = (-1)^k$$

where σ is the permutation in question and k is the number of elementary permutations it consists of.

3.4 Scalar Product

In this section V will always denote a vector space and $\mathbb{K}$ a field (either $\mathbb{R}$ or $\mathbb{C}$).

Definition 3.51. A scalar product is a mapping

$$\langle \cdot, \cdot \rangle : V \times V \longrightarrow \mathbb{K}$$

that fulfils the following conditions: $\forall v_1, v_2, w_1, w_2 \in V, \quad \lambda \in \mathbb{K}$

$$\begin{aligned}
\text{Linearity} \qquad & \langle v_1, w_1 + \lambda w_2 \rangle = \langle w_1, w_1 \rangle + \lambda \langle v_1, w_2 \rangle \\
\text{Conjugated symmetry} \qquad & \langle v_1, w_1 \rangle = \overline{\langle w_1, v_1 \rangle} \\
\text{Positivity} \qquad & \langle v_1, v_1 \rangle \geq 0 \\
\text{Definedness} \qquad & \langle v_1, v_2 \rangle = 0 \implies v_1 = 0 \\
\text{Conjugated linearity} \qquad & \langle v_1 + \lambda v_2, w_1 \rangle = \langle v_1, w_1 \rangle + \overline{\lambda} \langle v_2, w_1 \rangle
\end{aligned}$$

The mapping

$$\begin{aligned}
\|\cdot\| : V &\longrightarrow \mathbb{K} \\
v &\longmapsto \sqrt{\langle v, v \rangle}
\end{aligned}$$

Example 3.52. On $\mathbb{R}^n$ the following is a scalar product

$$\langle (x_1, x_2, \cdots, x_n)^T, (y_1, y_2, \cdots, y_n)^T \rangle = \sum_{k=1}^{n} x_k y_k$$

The norm is then equivalent to the Pythagorean theorem

$$\|v\| = \sqrt{\langle v, v \rangle} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

Analogously for $\mathbb{C}^n$

$$\langle (u_1, u_2, \cdots, u_n)^T, (v_1, v_2, \cdots, v_n)^T \rangle = \sum_{k=1}^{n} \overline{u_k} v_k$$

Remark 3.53.
- The length of $v \in V$ is $\|v\|$
- The distance between elements $v, w \in V$ is $\|v - w\|$
- The angle ϕ between $v, w \in V$ is $\cos\phi = \frac{\langle v,w\rangle}{\|v\| \cdot \|w\|}$

Theorem 3.54. *Let $v, w \in V$. Then*

Cauchy-Schwarz-Inequality	$\lvert\langle v, w\rangle\rvert \leq \|v\|\|w\|$
Triangle Inequality	$\|v + w\| \leq \|v\| + \|w\|$

Proof. For $\lambda \in \mathbb{K}$ we know that

$$\begin{aligned} 0 \leq \langle v - \lambda w, v - \lambda w\rangle &= \langle v - \lambda w, v\rangle - \lambda\langle v - \lambda w, w\rangle \\ &= \langle v, v\rangle - \overline{\lambda}\langle w, v\rangle - \lambda\langle v, w\rangle + \underbrace{\lambda\overline{\lambda}}_{|\lambda|^2}\langle w, w\rangle \end{aligned} \tag{3.45}$$

Let $\lambda = \frac{\langle w,v\rangle}{\|w\|^2}$. Then

$$\begin{aligned} 0 &\leq \|v\|^2 - \frac{\overline{\langle w, v\rangle}}{\|w\|^2} \cdot \langle w, v\rangle - \frac{\langle w, v\rangle}{\|w\|^2} \cdot \langle v, w\rangle + \frac{|\langle w, v\rangle|^2}{\|w\|^4}\|w\|^2 \\ &= \|v\|^2 - \frac{|\langle w, v\rangle|^2}{\|w\|^2} - \cancel{\frac{|\langle w, v\rangle|^2}{\|w\|^2}} + \cancel{\frac{|\langle w, v\rangle|^2}{\|w\|^2}} \\ &= \|v\|^2 - \frac{|\langle w, v\rangle|^2}{\|w\|^2} \end{aligned} \tag{3.46}$$

Through the monotony of the square root this implies that

$$|\langle w, v\rangle| \leq \|v\|\|w\| \tag{3.47}$$

To prove the triangle inequality, consider

$$\begin{aligned} ||v + w||^2 &= \langle v + w, v + w\rangle \\ &= \underbrace{\langle v, v\rangle}_{\|v\|^2} + \langle v, w\rangle + \underbrace{\langle w, v\rangle}_{\overline{\langle v,w\rangle}} + \underbrace{\langle w, w\rangle}_{\|w\|^2} \\ &\leq \|v\|^2 + 2 \cdot \mathrm{Re}\langle v, w\rangle + \|w\|^2 \\ &\leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 \\ &= (\|v\| + \|w\|)^2 \end{aligned} \tag{3.48}$$

Using the same argument as above, this implies

$$\|v + w\| \leq \|v\| + \|w\| \tag{3.49}$$

□

Definition 3.55. $v, w \in V$ are called orthogonal if

$$\langle v, w \rangle = 0$$

The elements $v_1, \cdots, v_m \in V$ are called an orthogonal set if they are non-zero and they are pairwise orthogonal. I.e.

$$\forall i, j \in \{1, \cdots, m\} : \langle v_i, v_j \rangle = 0$$

If $\|v_i\| = 1$, then the v_i are called an orthonormal set. If their span is V they are an orthonormal basis.

Theorem 3.56. *If $v_1, \cdots, v_n$ are an orthonormal set, they are linearly independent.*

Proof. Let $\alpha_1, \cdots, \alpha_n \in \mathbb{K}$, such that

$$0 = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n \tag{3.50}$$

Then

$$\begin{aligned} 0 &= \langle v_i, 0 \rangle = \langle v_i, \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n \rangle \\ &= \alpha_1 \langle v_i, v_1 \rangle + \alpha_2 \langle v_i, v_2 \rangle + \cdots + \alpha_n \langle v_i, v_n \rangle \\ &= \alpha_i \langle v_i, v_i \rangle \quad i \in \{1, \cdots, n\} \end{aligned} \tag{3.51}$$

Since v_i is not a zero vector, $\langle v_i, v_i \rangle \neq 0$, and thus $\alpha_i = 0$. Since i is arbitrary, the v_i are linearly independent. □

Example 3.57. (i) The canonical basis in $\mathbb{R}^n$ is an orthonormal basis regarding the canonical scalar product.

(ii) Let $\phi \in \mathbb{R}$. Then

$$v_1 = (\cos\phi, \sin\phi)^T \qquad v_2 = (-\sin\phi, \cos\phi)^T$$

are an orthonormal basis for $\mathbb{R}^2$

Theorem 3.58. *Let $v_1, \cdots, v_n$ be an orthonormal basis of V. Then for $v \in V$:*

$$v = \sum_{i=1}^{n} \langle v_i, v \rangle v_i$$

Proof. Since $v_1, \cdots, v_n$ is a basis,

$$\exists \alpha_1, \cdots, \alpha_n \in \mathbb{K}: \quad v = \sum_{i=1}^{n} \alpha_i v_i \tag{3.52}$$

And therefore, for $j \in \{1, \cdots, n\}$

$$\langle v_j, v \rangle = \sum_{i=1}^{n} \alpha_i \langle v_j, v_i \rangle = \alpha_j \underbrace{\langle v_j, v_j \rangle}_{\|v_j\|^2 = 1} \tag{3.53}$$

□

Theorem 3.59. *Let $A \in \mathbb{K}^{m \times n}$ and $\langle \cdot, \cdot \rangle$ the canonical scalar product on $\mathbb{K}^n$. Then*

$$\langle v, Aw \rangle = \langle A^H v, w \rangle$$

Proof. First consider

$$(Aw)_i = \sum_{j=1}^{n} A_{ij} w_i \tag{3.54a}$$

$$(A^H w)_j = \sum_{i=1}^{n} A_{ji} v_i \tag{3.54b}$$

Now we can compute

$$\begin{aligned}
\langle v, Aw \rangle &= \sum_{i=1}^{n} \overline{v_i} (Aw)_i = \sum_{i=1}^{n} \left(\overline{v_i} \cdot \sum_{j=1}^{n} A_{ij} w_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} \overline{v_i} w_j \\
&= \sum_{j=1}^{n} \left(\sum_{i=1}^{n} Aij \overline{v_i} \right) w_j = \sum_{j=1}^{n} \left(\overline{\sum_{i=1}^{n} \overline{A_{ij}} v_i} \right) w_j \\
&= \sum_{j=1}^{n} \overline{(A^H v)_j} \cdot w_j \\
&= \langle A^H v, w \rangle
\end{aligned} \tag{3.55}$$

□

Definition 3.60. A matrix $A \in \mathbb{R}^{n\times n}$ is called orthogonal if

$$A^T A = AA^T = I$$

or

$$A^T = A^{-1}$$

The set of all orthogonal matrices

$$O(n) := \left\{ A \in \mathbb{R}n \times n \,\middle|\, A^T A = I \right\}$$

is called the orthogonal group.

$$SO(n) = \left\{ A = \mathbb{R}n \times n \,\middle|\, A^T A = I \wedge \det A = 1 \right\} \subset O(n)$$

is called the special orthogonal group.6

Example 3.61. Let $\phi \in [0, 2\pi]$, then

$$A = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$$

is orthogonal.

Remark 3.62. (i) Let $A, B \in \mathbb{K}^{n\times n}$, then

$$AB = I \implies BA = I$$

(ii)
$$1 = \det I = \det A^T A = \det A^T \cdot \det A = \det^2 A$$

(iii) The i-j-component of $A^T A$ is equal to the canonical scalar product of the i-th row of A^T and the j-th column of A. Since the rows of A^T are the columns of A, we can conclude that

$$A \text{ orthogonal } \iff \langle r_i, r_j \rangle = \delta_{ij}$$

where the r_i are the columns of A. In this case, the r_i are an orthonormal basis on $\mathbb{R}^n$. This works analogously for the rows.

(iv) Let A be orthogonal, and $x, y \in \mathbb{R}^n$

$$\begin{aligned} \langle Ax, Ay \rangle &= \langle A^T Ax, y \rangle = \langle x, y \rangle \\ \|Ax\| &= \sqrt{\langle Ax, Ax \rangle} = \sqrt{\langle x, x \rangle} = \|x\| \end{aligned}$$

A perserves scalar products, lengths, distances and angles. These kinds of operations are called mirroring and rotation.

(v) Let $A, B \in O(n)$

$$(AB)^T \cdot (AB) = B^T A^T AB = B^T IB = I$$

This implies $(AB) \in O(n)$. It also implies $I \in O(n)$. Now consider $A \in O(n)$. Then

$$(A^{-1})^T A^{-1} = (A^T)^T \cdot A^T = AA^T = I$$

This implies $A^{-1} \in O(T)$. Such a structure (a set with a multiplication operation, neutral element and multiplicative inverse) is called a group.

Example 3.63. $O(n)$, $SO(n)$, $\mathbb{R} \setminus \{0\}$, $\mathbb{C} \setminus \{0\}$, $Gl(n)$ (set of invertible matrices) and S_n are all groups.

Definition 3.64. A matrix $U \in \mathbb{C}^{n\times n}$ is called unitary if

$$U^H U = I = UU^H$$

We also introduce

$$\left\{U \in \mathbb{C}n \times n \,\middle|\, U^H U = I\right\}$$

the unitary group, and

$$\left\{U \in \mathbb{C}n \times n \,\middle|\, U^H U = I \wedge \det U = 1\right\}$$

the special unitary group.

3.5 Eigenvalue problems

Definition 3.65. Let $A \in \mathbb{K}^{n\times n}$. Then $\lambda \in \mathbb{K}$ is called an eigenvalue of A, if

$$\exists v \in \mathbb{K}^n,\ v \neq 0 : \quad Av = \lambda v$$

Such a vector v is called eigenvector. We call

$$\{v \in \mathbb{K}^n \mid Av = \lambda v\} =: E_\lambda$$

eigenspace belonging to λ.

Example 3.66. Let

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

Then

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

The eigenspaces are

$$E_2 = \left\{ \kappa \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \middle| \kappa \in \mathbb{R} \right\}$$
$$E_1 = \left\{ \kappa \cdot \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + \rho \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \middle| \kappa, \rho \in \mathbb{R} \right\} = \operatorname{span} \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

Remark 3.67. The eigenspace to an eigenvalue λ is a linear subspace.

Remark 3.68. We want to find $\lambda \in \mathbb{K}$, $v \in \mathbb{K}^n$ such that

$$Av = \lambda v \iff (\underbrace{A - \lambda I}_{\in \mathbb{K}^{n \times n}})v = 0$$

If $(A - \lambda I)$ is invertible, then $v = 0$. So the interesting case is when $(A - \lambda I)$ not invertible.

$$(A - \lambda I) \text{ not invertible} \iff \det(A - \lambda I) = 0$$

This determinant is called the characteristic polynomial. This polynomial has degree n, and the eigenvalues are the roots of that polynomial. So let λ be an eigenvalue of A, then

$$(A - \lambda I)v = 0$$

is a linear equation system for the components of v.

Example 3.69. Let

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in \mathbb{C}^{2 \times 2}$$

The characteristic polynomial is

$$\det(A - \lambda I) = \begin{vmatrix} -\lambda & 1 \\ -1 & -\lambda \end{vmatrix} = \lambda^2 + 1$$

Its roots are

$$\lambda_1 = i \qquad\qquad \lambda_2 = -i$$

To find the eigenvector belonging to λ_1, we declare $v_1 = (x, y) \in \mathbb{C}^2$ and solve the linear equation system

$$(A - \lambda_1 I)v_1 = 0 \qquad\qquad \begin{aligned} -ix + 1y &= 0 \\ -1x - iy &= 0 \end{aligned}$$

It has the solutions $x = -i$ and $y = 1$, so

$$v_1 = \begin{pmatrix} -i \\ 1 \end{pmatrix}$$

Doing the same for v_2 yields

$$v_2 = \begin{pmatrix} i \\ 1 \end{pmatrix}$$

It is to be noted that the eigenvectors aren't unique (multiples of eigenvectors are also eigenvectors).

Example 3.70. Let D be a diagonal matrix, with the diagonal entries λ_j. Then

$$\det(D - \lambda I) = \begin{vmatrix} \lambda_1 - \lambda & & & \\ & \lambda_2 - \lambda & & \\ & & \ddots & \\ & & & \lambda_n - \lambda \end{vmatrix}$$

The roots (eigenvalues) are $\lambda_1, \lambda_2, \cdots, \lambda_n$, and the eigenvectors are $De_i = \lambda_i e_i$.

Definition 3.71. $A \in \mathbb{K}^{n \times n}$ is called diagonalizable if there exists a basis of $\mathbb{K}^n$ that consists of eigenvectors.

Theorem 3.72. *A matrix $A \in \mathbb{K}^{n \times n}$ is diagonalizable, if and only if there exists a diagonal matrix D and a invertible matrix T such that*

$$D = T^{-1} AT$$

Proof. Let $e_1, e_2, \cdots, e_n$ be the canonical basis of $\mathbb{K}^n$. Define $TDT^{-1} = A$, and let $\lambda_1, \cdots, \lambda_n$ be the diagonal entries of D. Then we know that

$$De_i = \lambda_i e_i, \quad \forall i \in \{1, \cdots n\} \tag{3.56}$$

Since T is invertible, the $Te_1, \cdots Te_n$ form a basis.

$$A(Te_i) = T(T^{-1}AT)e_i = TDe_i = T\lambda_i e_i = \lambda_i(Te_i) \tag{3.57}$$

Therefore Te_i is an eigenvector of A to the eigenvalue λ_i. Now let $v_1, \cdots, v_n$ be a basis of $\mathbb{K}^n$ and

$$Av_i = \lambda_i v_i, \quad \lambda_1, \cdots, \lambda_n \in \mathbb{K}^n \tag{3.58}$$

Write write $v_1, \cdots, v_n$ as the columns of a matrix, therefore

$$T = (v_1, v_2, \cdots, v_n) \tag{3.59a}$$

$$D = \begin{pmatrix} \lambda_1 & & \\ & \vdots & \\ & & \lambda_n \end{pmatrix} \tag{3.59b}$$

So $Te_i = v_i$, and thus

$$A(Te_i) = Av_i = \lambda_i v_i = \lambda_i(Te_i) = T\lambda_i e_i = TDe_i \tag{3.60}$$

This means that $(AT - TD)e_i = 0$, $\forall i \in \{1, \cdots, n\}$.

$$\implies AT = TD \tag{3.61}$$

T is invertible (left as an exercise for the reader), and thus

$$\implies T^{-1}AT = D \tag{3.62}$$

$\square$

Example 3.73. (i) Let

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

The eigenvalues and eigenvectors are

$$A \cdot \begin{pmatrix} -i \\ 1 \end{pmatrix} = i \begin{pmatrix} -i \\ 1 \end{pmatrix} \qquad\qquad A \cdot \begin{pmatrix} i \\ 1 \end{pmatrix} = -i \begin{pmatrix} i \\ 1 \end{pmatrix}$$

Therefore

$$T = \begin{pmatrix} -i & i \\ 1 & 1 \end{pmatrix}$$

which has the inverse

$$T^{-1} = \frac{1}{2} \begin{pmatrix} i & 1 \\ -i & 1 \end{pmatrix}$$

Finally,

$$T^{-1}AT = \frac{1}{2} \begin{pmatrix} i & 1 \\ -i & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 2i & 0 \\ 0 & -2i \end{pmatrix} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

This is a diagonal matrix, therefore A is diagonalizable.

(ii) The matrix

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

is not diagonalizable since its only eigenvector is $(1,0)^T$.

Remark 3.74. For diagonal matrices the following is true

$$\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_3 \end{pmatrix}^k = \begin{pmatrix} \lambda_1^k & & & \\ & \lambda_2^k & & \\ & & \ddots & \\ & & & \lambda_3^k \end{pmatrix}$$

If $T^{-1}AT = D$ (where D is a diagonal matrix), then

$$D^k = (T^{-1}AT)^k = \underbrace{T^{-1}AT \cdot T^{-1}AT \cdot \dots \cdot}_{k \text{ times}} = T^{-1}A^kT$$

$$\implies A^k = TD^kT^{-1}$$

Theorem 3.75. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix, i.e. $A = A^T$. (Or if $A \in \mathbb{C}^{n\times n}$ a self-adjoint matrix $A = A^H$). Then A has an orthonormal basis consisting of eigenvectors and is diagonalizable.*

Proof. Let $\lambda \in \mathbb{C}$ be an eigenvalue of $A \in \mathbb{K}^{n\times n}$ with eigenvector $v \in \mathbb{K}^n$ and $A = A^H$. Then

$$\lambda\langle v, v\rangle = \langle v, \lambda v\rangle = \langle v, Av\rangle = \langle A^H v, v\rangle = \langle Av, v\rangle = \langle \lambda v, v\rangle = \overline{\lambda}\langle v, v\rangle \tag{3.63}$$

Therefore

$$(\lambda - \overline{\lambda})\underbrace{\langle v, v\rangle}_{0} = 0 \tag{3.64}$$

$$\implies (\lambda - \overline{\lambda}) = 0 \implies \lambda = \overline{\lambda} \implies \lambda \in \mathbb{R} \tag{3.65}$$

Now let $\lambda, \rho \in \mathbb{R}$ be eigenvalues to the eigenvectors v, w, and require $\lambda \neq \rho$. Then

$$\rho\langle v, w\rangle = \langle v, Aw\rangle = \langle Av, w\rangle = \overline{\lambda}\langle v, w\rangle = \lambda\langle v, w\rangle \tag{3.66}$$

And thus

$$\underbrace{(\rho - \lambda)}_{\neq 0}\underbrace{\langle v, w\rangle}_{=0} = 0 \implies v \perp w \tag{3.67}$$

$\square$

Chapter 4

Real Analysis: Part II

4.1 Limits and Functions

In this chapter we will introduce the notation

$$B_\epsilon(x) = (x - \epsilon, x + \epsilon)$$

Definition 4.1. Let $D \subset \mathbb{R}$ and $x \in \mathbb{R}$. x is called a boundary point of D if

$$\forall \epsilon > 0: \quad D \cap B_\epsilon(x) \neq 0$$

The set of all boundary points of D is called closure and is denoted as $\bar{D}$.

Example 4.2. (i) $x \in D$ is always a boundary point of D, because

$$x \in D \cap B_\epsilon(x)$$

(ii) Boundary points don't have to be elements of D. If $D = (0, 1)$, then 0 and 1 are boundary points, because

$$\frac{\epsilon}{2} \in (0, 1) \cap B_\epsilon(0) = (-\epsilon, \epsilon) \quad \forall \epsilon > 0$$

(iii) Let $D = \mathbb{Q}$. Every $x \in \mathbb{R}$ is a boundary point, because $\forall \epsilon > 0$, $B_\epsilon(x)$ contains at least one rational number. I.e. $\bar{\mathbb{Q}} = \mathbb{R}$.

Remark 4.3. If x is a boundary point, then

$$\forall \epsilon > 0 \; \exists y \in D: \quad |x - y| < \epsilon$$

If x is not a boundary point, then

$$\exists \epsilon > 0 \; \forall y \in D: \quad |x - y| \geq \epsilon$$

Theorem 4.4.

$$x \in \mathbb{R} \text{ is a boundary point of } D \subset \mathbb{R} \iff \exists\, (x_n) \subset D \text{ such that } x_n \to x$$

Proof. Let x be a boundary point of D. Then

$$\forall n \in \mathbb{N}\ \exists x_n \in D \cap \left(x - \frac{1}{n}, x + \frac{1}{n}\right) \tag{4.1}$$

The resulting sequence (x_n) is in D, and

$$|x - x_n| \leq \frac{1}{n} \tag{4.2}$$

holds. Therefore, x_n converges to x. Now let $(x_n) \subset D$, with $x_n \to x$. This means

$$\forall \epsilon > 0\ \exists N \in \mathbb{N} : \quad |x - x_N| < \epsilon \tag{4.3}$$

Then

$$x_N \in D \cap B_\epsilon(x) \tag{4.4}$$

Since ϵ is arbitrary, x is a boundary point of D. □

Definition 4.5. Let $D \subset \mathbb{R}$ and $f : D \to \mathbb{R}$. Let x_0 be a boundary point of D. We say that f converges to $y \in \mathbb{R}$ for $x \to x_0$ and write

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

if

$$\forall \epsilon > 0\ \exists \delta > 0 : \quad |x - x_0| < \delta \implies |f(x) - f(y)| < \epsilon$$

Remark 4.6. This definition only makes sense for boundary points x_0 of D. The most imoprtant case is

$$D = (x_0 - a, x_0 + a) \setminus \{x_0\}$$

Example 4.7. (i) Let $a \in \mathbb{R}$

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto ax \end{aligned}$$

Consider $a \neq 0$: Let $\epsilon > 0$. We want that

$$|f(x) - 0| = |a||x| \overset{!}{<} \epsilon$$

Choose $\delta = \frac{\epsilon}{|a|}$. Then we have

$$|x| = |x - 0| < \delta \implies |f(x) - 0| = |a||x| < |a|\delta = |a|\frac{\epsilon}{|a|} = \epsilon$$

Therefore

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

(ii) Consider

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \begin{cases} 1, & x > 0 \\ -1, & x < 0 \end{cases} \end{aligned}$$

f doesn't converge for $x \to 0$. Assume $y \in \mathbb{R}$ is the limit of x at 0. This means that there is a $\delta > 0$ such that

$$|f(x) - y| < 1 \text{ if } |x| = |x - 0| < \delta$$

Then, for any $x \in (0, \delta)$ we have

$$2 = |f(x) - f(-x)| \le \underbrace{|f(x) - y|}_{<1} + \underbrace{|y - f(-x)|}_{<1} < 2$$

which is a contradiction.

Theorem 4.8. *Let $f : D \to \mathbb{R}$, x_0 a boundary point of D and $y \in \mathbb{R}$. Then*

$$\lim_{x \to x_0} f(x) = y \iff \forall\, (x_n) \subset D \text{ with } x_n \longrightarrow x_0 : \quad \lim_{n \to \infty} f(x_n) = x_0$$

Proof. Assume that $\lim_{x \to x_0} f(x)$ and that there is $(x_n) \subset D$ converging to x. Let $\epsilon > 0$, then

$$\exists \delta > 0 : \quad |x - x_0| < \delta \implies |f(x) - y| < \epsilon \tag{4.5}$$

Since $x_n \to x_0$, we know that

$$\exists N \in \mathbb{N}\ \forall n > N : \quad |x_n - x_0| < \delta \tag{4.6}$$

For such n, the epsilon criterion $|f(x_n) - y| < \epsilon$ also holds, and thus

$$f(x_n) \xrightarrow{n \to \infty} y \tag{4.7}$$

Now to prove the " $\Longleftarrow$ " direction, assume that $\lim_{x\to x_0} f(x) \neq y$, i.e.

$$\exists\epsilon > 0\ \forall\delta > 0\ \exists x \in D: \quad |x - x_0| < \delta \wedge |f(x) - y| \geq \epsilon \tag{4.8}$$

Choose $\forall x \in \mathbb{N}$ one x_n such that

$$|x_n - x_0| < \frac{1}{n} \text{ but } |f(x_n) - y| \geq \epsilon \tag{4.9}$$

Then $x_n \to x_0$, but $|f(x_n) - y| \geq \epsilon\ \forall n \in \mathbb{N}$, so

$$\lim_{n\to\infty} f(x_n) \neq y \tag{4.10}$$

This indirectly proves " $\Longleftarrow$ ". □

Example 4.9. Consider $D = \mathbb{R} \subset \{0\}$, we want to prove

$$\lim_{x\to 0} \frac{1}{1-x} = 1$$

So let $(x_n) \subset D$ with $x_n \to 0$. Then

$$\frac{1}{1-x_n} \xrightarrow{n\to\infty} 1$$

$$\Longrightarrow \lim_{x\to 0} \frac{1}{1-x} = 1$$

However, the limit $\lim_{x\to 1}$ doesn't exist. Let $x_n = \frac{1}{n} + 1$ with $x_n \to 1$. Then

$$\frac{1}{1-(\frac{1}{n}+1)} = -n \xrightarrow{n\to\infty} -\infty$$

This doesn't converge, thus there is no limit.

Corollary 4.10. *Let $f, g : D \to \mathbb{R}$, x_0 a boundary point and $y, z \in \mathbb{R}$ such that*

$$\lim_{x\to x_0} f(x) = y \qquad\qquad \lim_{x\to x_0} g(x) = z$$

Then

$$\lim_{x\to x_0} (f(x) + g(x)) = y + z$$
$$\lim_{x\to x_0} (f(x) \cdot g(x)) = y \cdot z$$

If $z \neq 0$, then

$$\lim_{x\to x_0} (\frac{f(x)}{g(x)}) = \frac{y}{z}$$

Proof. Here we will only prove the last statement. Let $\lim_{x \to x_0} = z \neq 0$. Then

$$\exists \delta > 0 \; \forall x \in B_\delta(x_0) : \quad |g(x) - z| < |z| \tag{4.11}$$

g doesn't have any roots on $B_\delta(x_0)$. Let $(x_n) \subset D \cap B_\delta(x_0)$ converge to x_0. According to prerequisites, we have

$$\lim_{n \to \infty} f(x_n) = y \tag{4.12a}$$

$$\lim_{n \to \infty} g(x_n) = z \neq 0 \tag{4.12b}$$

Thus

$$\implies \lim_{n \to \infty} \frac{f(x_n)}{g(x_n)} = \frac{y}{z} \implies \lim_{x \to x_0} \frac{f(x)}{g(x)} = \frac{y}{z} \tag{4.13}$$

□

Corollary 4.11 (Squeeze Theorem)**.** *Let $f, g, h : D \to \mathbb{R}$ and x a boundary point of D. If for $y \in \mathbb{R}$*

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

and

$$f(x) \leq g(x) \leq h(x) \quad \forall x \in B_\epsilon(x_0)$$

then

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

Example 4.12. Consider $\exp(x)$. We already know that

$$1 + x \leq \exp(x) \quad \forall x \in \mathbb{R}$$

This also implies that

$$1 - x \leq \exp(-x) = \frac{1}{\exp(x)} \quad \forall x \in \mathbb{R}$$

So

$$1 + x \leq \exp(x) \leq \frac{1}{1 - x}$$

The limits of these terms are

$$\lim_{x \to 0} (1 + x) = 1 \qquad \lim_{x \to 0} \left(\frac{1}{1 - x} \right) = 1$$

And using the squeeze theorem this results in

$$\lim_{x \to 0} \exp(0) = 1$$

Definition 4.13. Let $f : D \to \mathbb{R}$ and x_0 a boundary point of D. We say f diverges to infinity for $x \to x_0$ and write

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

if

$$\forall K \in (0, \infty)\ \exists \delta > 0 : \quad |x - x_0| < \delta \implies f(x) \geq K$$

Definition 4.14. Let $D \subset \mathbb{R}$ be unbounded above. We say f converges for $x \to \infty$ to $y \in \mathbb{R}$ and write

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

if

$$\forall \epsilon > 0\ \exists K \in (0, \infty)\ \forall x > K : \quad |f(x) - y| < \epsilon$$

Remark 4.15. Let $f : D \to \mathbb{C}$ and x_0 a boundary point of D. Then

$$\begin{aligned} & \lim_{x \to x_0} f(x) = y \in \mathbb{C} \\ \iff & \lim_{x \to x_0} \operatorname{Re}(f(x)) = \operatorname{Re}(y) \wedge \lim_{x \to x_0} \operatorname{Im}(f(x)) = \operatorname{Im}(y) \\ \iff & \lim_{x \to x_0} |f(x) - y| = 0 \end{aligned}$$

Definition 4.16. Let $D \subset K$, $f : D \to K$ and $x_0 \in D$. f is called continuous in x_0 if

$$\forall \epsilon > 0\ \exists \delta > 0 : \quad |x - x_0| < \delta \implies |f(x) - f(x_0)| < \epsilon$$

If f is continuous in every point of D, we call f continuous.

f is called Lipschitz continuous if

$$\exists L \in (0, \infty)\ \forall x, y \in D : \quad |f(x) - f(y)| \leq L|x - y|$$

L is called Lipschitz constant

Remark 4.17. Let $f : D \to \mathbb{K}$

$$f \text{ is continuous in } x_0 \in D \iff \lim_{x \to x_0} f(x) = f(x_0)$$

Example 4.18. We want to show that

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto x^2 \end{aligned}$$

is continuous. To do that, let $x_0 \in \mathbb{R}$, $\epsilon > 0$. We want

$$|f(x) - f(x_0)| = |x^2 - x_0^2| = |x - x_0||x + x_0| \overset{!}{\leq} \epsilon$$

So we choose

$$\delta = \min\left\{1, \frac{\epsilon}{2|x_0| + 1}\right\} > 0$$

Then for every x with $|x - x_0| < \delta$

$$\begin{aligned}|f(x) - f(x_0)| &= |x - x_0||x + x_0| \leq \delta(|x| + |x_0|) \leq \delta(|x_0| + \delta + |x_0|) \\ &\leq \delta(2|x_0| + 1) \leq \frac{\epsilon}{2|x_0| + 1}(2|x_0| + 1) = \epsilon\end{aligned}$$

Theorem 4.19. *Every Lipschitz continuous function is continuous*

Proof. Let $f : D \to \mathbb{K}$ be a Lipschitz continuous function with Lipschitz constant $L > 0$. I.e.

$$\forall x, y \in D: \quad |f(x) - f(y)| \leq L|x - y| \tag{4.14}$$

Let $x_0 \in \mathbb{R}$ and $\epsilon > 0$. Choose $\delta = \frac{\epsilon}{L}$. Then $|x - x_0| < \delta$ implies

$$|f(x) - f(x_0)| \leq L|x - x_0| \leq L \cdot \delta = \epsilon \tag{4.15}$$

□

Example 4.20. (i) Consider the constant function $x \mapsto c$, $c \in \mathbb{K}$.

$$|f(x) - f(y)| = |c - c| = 0 \leq 1 \cdot |x - y|$$

(ii) Consider the linear function $x \mapsto cx$, $c \in \mathbb{K}$.

$$|f(x) - f(y)| = |cx - cy| = |c||x - y|$$

These two functions are Lipschitz continuous, and therefore continuous.

(iii) Consider $x \mapsto \operatorname{Re}(x)$. Then

$$|\operatorname{Re}(x) - \operatorname{Re}(y)| = |\operatorname{Re}(x - y)| \leq |x - y|$$

Analogously this works for $\operatorname{Im}(x)$. Both of those are Lipschitz continuous.

(iv) Lipschitz continuity depends on D. E.g.

$$\begin{aligned} f : [0,1] &\longrightarrow \mathbb{R} \\ x &\longmapsto x^2 \end{aligned}$$

is Lipschitz continuous:

$$|f(x) - f(y)| = |x-y||x+y| \leq 2 \cdot |x-y|$$

However,

$$\begin{aligned} g : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto x^2 \end{aligned}$$

is NOT Lipschitz continuous, because

$$\frac{|g(n+1) - g(n)|}{(n+1) - n} = 2n+1 \xrightarrow{n\to\infty} \infty$$

Remark 4.21. Let $f : D \to \mathbb{K}$.

$$\begin{gathered} f \text{ is continuous in } x_0 \in D \\ \Longleftrightarrow \\ \forall\, (x_n) \subset D \text{ with } x_n \to x_0 : \quad \lim_{n\to\infty} f(x_n) = f(x_0) \end{gathered}$$

If f, g are continuous in x_0, then $f+g$ and $f \cdot g$ are also continuous in x_0, and if $g(x_0) \neq 0$ then f/g is also continuous in x_0. Notably, polynomials are continuous. A rational function (the quotient of two polynomials) is continuous in all points that are not roots of the denominator.

Theorem 4.22. *Let $D \subset \mathbb{K}$, and let*

$$f : D \longrightarrow \mathbb{K} \text{ continuous in } x_0 \in D \tag{4.16a}$$

$$g : f(D) \longrightarrow \mathbb{K} \text{ continuous in } f(x_0) \tag{4.16b}$$

Then $g \circ f$ is also continuous in x_0.

Proof. Let $\epsilon > 0$. Since g is continuous in $f(x_0)$,

$$\exists \delta_1 > 0 : \quad |y - f(x_0)| < \delta_1 \implies |g(y) - g(f(x_0))| < \epsilon \tag{4.17}$$

Since f is continuous in x_0,

$$\exists \delta_2 > 0 : \quad |x - x_0| < \delta_2 \implies |f(x) - f(x_0)| < \delta_1 \tag{4.18}$$

For such x the following holds

$$|(g \circ f)(x) - (g \circ f)(x_0)| = |g(f(x)) - g(f(x_0))| < \epsilon \tag{4.19}$$

which implies that $g \circ f$ is continuous in x_0. $\square$

Example 4.23. Consider the following mappings

$$\begin{aligned} f &: \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto |x| \\ g &: \mathbb{R} \longrightarrow \mathbb{R} \setminus \{-1\}, \quad y \longmapsto \frac{1-y}{1+y} \\ h &: \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \frac{1-|x|}{1+|x|} \end{aligned}$$

It is clear that $h = g \circ f$. Since f, g are continuous, h must also be continuous.

Example 4.24. The functions exp, sin and cos are continuous. We know that

$$\lim_{h \to 0} \frac{\exp(k) - 1}{h} = 1$$

From this follows that

$$\lim_{h \to 0} \exp(k) = \exp(0) = 0$$

Thus, exp is continuous in 0. Let $x_0 \in \mathbb{R}$, then

$$\begin{aligned} \lim_{x \to x_0} \exp(x) &= \lim_{h \to 0} \exp(x_0 + h) = \lim_{h \to 0} \exp(x_0) \exp(h) \\ &= \exp(x_0) - \lim_{h \to 0} \exp(h) = \exp\{x_0\} \end{aligned}$$

Now, consider the function $x \mapsto \exp(ix)$. For $x_0 \in \mathbb{R}$

$$\begin{aligned} |\underbrace{\exp(i(x_0 + h))}_{\exp(ix_0)\exp(ih)} - \exp(ih_0)| &= \underbrace{|\exp(ix_0)|}_{1} |\exp(ih) - 1| \\ &\leq 1 \cdot \left| \sum_{k=0}^{\infty} \frac{(ih)^k}{k!} - 1 \right| = \left| \sum_{k=1}^{\infty} \frac{(ih)^k}{k!} \right| \\ &\leq \sum_{k=1}^{\infty} \left| \frac{(ih)^k}{k!} \right| \\ &= \sum_{k=1}^{\infty} \frac{|h|^k}{k!} = \sum_{k=0}^{\infty} \frac{|h|^k}{k!} - 1 = \exp(|h|) - 1 \end{aligned}$$

For $h \to 0$, the absolute function converges $|h| \to 0$, and therefore

$$\lim h0|\exp(i(x_0 + h)) - \exp(ix)| = 0$$

due to the squeeze theorem. I.e., $x \mapsto \exp(ix)$ is also continuous. Thus

$$\cos x = \operatorname{Re}(\exp(ix)) \qquad \sin x = \operatorname{Im}(\exp(ix))$$

are also continuous due to the concatination of continuous functions.

Lemma 4.25. *Let $a, b \in \mathbb{R}$ with $a < b$, and let*

$$f : [a, b] \longrightarrow \mathbb{R}$$

be a continuous function. Furthermore, let $y \in \mathbb{R}$. Now if the set

$$\{x \in [a, b] \,|\, f(x) \geq y\}$$

is non-empty, it has a smallest element.

Proof. Let M be non-empty. Set $x_0 = \inf \{M\}$. Then it is to be shown that $x_0 \in M$, or that $f(x_0) \geq y$. There exists a sequence $(x_n) \subset M$ such that $x_n \to x_0$. Because of the continuity of f,

$$f(x_0) = f(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} f(x_n) \geq y \tag{4.20}$$

holds, thus $x_0 \in M$. □

Theorem 4.26 (Extreme value theorem)**.** *Let $a, b \in \mathbb{R}$ with $a < b$, and let $f : [a, b] \to \mathbb{R}$ continuous. Then the function f attains a maximum, i.e.*

$$\exists x_0 \in [a, b] \; \forall x \in [a, b] : \quad f(x) \leq f(x_0)$$

Proof. First we show that f is bounded. Assume f is unbounded above, i.e.

$$\{x \in [a, b] \,|\, f(x) \geq n\} =: M_n, \quad n \in \mathbb{N} \tag{4.21}$$

According to the last lemma, every M_n has a smallest element x_n. The sequence $(x_n)_{n\in\mathbb{N}}$ is monotonically increasing ($M_{n+1} \subset M_n$) and bounded above by b. Thus, x_n converges to some $x_0 \in [a, b]$. Now consider the sequence $(f(x_n))_{n\in\mathbb{N}}$. By definition

$$\lim_{n\to\infty} f(x_n) \geq \lim_{n\to\infty} n = \infty \tag{4.22}$$

And since f is continuous, $\lim_{n\to\infty} f(x_n) = f(x_0)$ must hold. This contradicts the assumption, so f is bounded.

Now set

$$y = \sup\{f(x) \,|\, x \in [a,b]\} \tag{4.23}$$

In case f is equal to y everywhere, there is nothing to show. So assume that there are values for which $f \neq y$. According to the definition of the supremum, the sets

$$\left\{ x \in [a,b] \,\middle|\, f(x) \geq y - \frac{1}{n} \right\} \tag{4.24}$$

are non-empty for all $n \in \mathbb{N}$, and thus they have a smallest element x_n. The sequence $(x_n)_{n\in\mathbb{N}}$ is monotonically increasing and bounded, i.e. it converges to $x_0 \in [a,b]$. Therefore

$$y \geq f(x_0) = \lim_{n\to\infty} f(x_n) \geq \lim_{n\to\infty} y - \frac{1}{n} = y \tag{4.25}$$

From this follows

$$f(x_0) = y \implies f(x_0) \text{ upper bound of } \{f(x) \,|\, x \in [a,b]\} \tag{4.26}$$

□

Theorem 4.27 (Intermediate value theorem)**.** *Let $a, b \in \mathbb{R}$ with $a < b$, and $f : [a,b] \to \mathbb{R}$ a continuous function with $f(a) < f(b)$.*

$$y \in (f(a), f(b)) \implies \exists x_0 \in (a,b) : \quad f(x_0) = y$$

Proof. Without proof. □

Example 4.28. cos has in $[0,2]$ exactly one root. Consider the definition

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

We know that $\cos 0 = 1$. Furthermore we can show that

$$-1 = \underbrace{1 - \frac{2^2}{2!}}_{\text{2nd partial sum}} \leq \cos(2) \leq \underbrace{1 - \frac{2^2}{2!} + \frac{2^4}{4!}}_{\text{3rd partial sum}} < 0$$

The intermediate value theorem tells us that there exists a root in $[0,2]$. Now we need to show that cos is strictly monotonically decreasing on $[0,2]$. Choose $z \in [0,2]$. Then

$$z \leq \sin z \leq z - \frac{z^3}{3!}$$

The addition theorem tells us that

$$\cos(x) - \cos(y) = -2\sin\left(\frac{x+y}{2}\right)\sin\left(\frac{x-y}{2}\right) < 0$$

for $x, y \in (0,2]$ and $x > y$. Thus cos is strictly monotonically decreasing on $[0,2]$.

Corollary 4.29. *Let I be an interval and $f : I \to \mathbb{R}$ continuous. Then $f(I)$ is also an interval.*

Proof. Left as an exercise for the reader. □

Theorem 4.30. *Let I be an interval, $f : I \to \mathbb{R}$ continuous. If f is strictly monotonically increasing, then the inverse function $f^{-1} : f(I) \to I$ exists and is continuous.*

Heuristic Proof. $f(I)$ is an interval, and f is injective. This is because if $f(x) = f(\tilde{x})$, then $x = \tilde{x}$ or else f wouldn't be strictly monotonic. This means

$$\exists g : f(I) \longrightarrow \mathbb{R} : \quad f(x) = y \iff g(y) = x \tag{4.27}$$

Let $y_0 \in f(I)$ and $\epsilon > 0$. We require that x_0 is not a boundary point of I. Then choose $0 < \tilde{\epsilon} < \epsilon$ such that $(x_0 - \tilde{\epsilon}, x_0 + \widetilde{epsilon}) \in I$. Choose

$$\delta = \min\left\{\underbrace{f(x_0+\tilde{\epsilon}) - y_0}_{>0}, \underbrace{y_0 - f(x_0-\tilde{\epsilon})}_{>0}\right\} > 0 \tag{4.28}$$

If $y \in f(I)$ with $|y - y_0| < \delta$ then

$$f(x_o - \widetilde{epsilon}) \leq x_0 - \delta < y < y_0 + \delta \leq f(x_0 + \tilde{\epsilon}) \tag{4.29}$$

From the strict monotony of g we can conclude

$$x_0 - \widetilde{epsilon} < g(y) < x_0 + \tilde{\epsilon} \tag{4.30}$$

so

$$|g(y) - g(y_0)| = |g(y) - x_0| < \tilde{\epsilon} < \epsilon \tag{4.31}$$

Thus, g is continuous in y_0. Since $y_0 \in f(I)$ was chose arbitrarily, all of g is continuous. To prove the monotony of g, assume $y < \tilde{y}$ and $g(y) \geq g(\tilde{y})$ for $y, \tilde{y} \in f(I)$. From the monotony of f we know that

$$y \geq \tilde{y} \tag{4.32}$$

which is a contradiction, so g is strictly monotonic. □

Example 4.31. (i) Let $n \in \mathbb{N}$ and consider

$$\begin{aligned} f : [0, \infty) &\longrightarrow \mathbb{R} \\ x &\longmapsto x^n \end{aligned}$$

f is continuous and strictly monotonically increasing. Thus the inverse function

$$\sqrt[n]{\cdot} : [0, \infty) \longrightarrow \mathbb{R}^+$$

is also continuous.

(ii) Consider $\exp : \mathbb{R} \to \mathbb{R}$. It's clear that $\exp(\mathbb{R}) = (0, \infty)$, so the mapping

$$\ln : (0, \infty) \to \mathbb{R}$$

is continuous and strictly monotonically increasing.

(iii) Equal arguments can be made for the trigonometric functions.

4.2 Differential Calculus

Definition 4.32. Let I be an open interval ((a, b), $a < b$, $a, b = \infty$ possible). Let $f : I \to \mathbb{K}$ and $x \in I$. f is called differentiable in x if

$$f'(x) = \lim_{h \to 0} \underbrace{\frac{f(x+h) - f(x)}{h}}_{\text{Difference quotient}}$$

exists. $f'(x)$ is called the differential quotient, or derivative of f in x. f is called differentiable if it is differentiable in every x.

Example 4.33. (i) Let $f(x) = c$ with $c \in \mathbb{K}$ be a constant function

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

(ii) For $n \in \mathbb{N}$ consider $f : \mathbb{R} \to \mathbb{R} \quad x \mapsto x^n$

$$f'(x) = \lim_{h\to 0} \frac{(x+h)^n - x^n}{h} = \lim_{h\to 0} \sum_{k=0}^{n} \binom{n}{k} h^{k-1} x^{k-1} = nx^{n-1}$$

(iii) Consider the exponential function

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

Theorem 4.34. *Let $f : I \to \mathbb{K}$ be differentiable in x. Then f is also continuous in x.*

Proof. Let f be continuous in x. Then

$$\lim_{h\to 0} (f(x+h) - f(x)) = 0 \tag{4.33}$$

Assume f to be uncontinuous in x. This means that

$$\exists \epsilon > 0 \ \forall \delta > 0 \ \exists h \in (-\delta, \delta) : \quad |f(x+h) - f(x)| \geq \epsilon \tag{4.34}$$

In particular, for every n there exists an $h_n \in \left(\frac{-1}{n}, \frac{1}{n}\right) \subset \{0\}$, such that

$$|f(x+h_n) - f(x)| \geq \epsilon \tag{4.35}$$

h_n is a null sequence and

$$\left| \frac{f(x+h_n) - f(x)}{h_n} \right| \geq \frac{\epsilon}{\frac{1}{n}} = n \cdot \epsilon \longrightarrow \infty \tag{4.36}$$

So the above term doesn't converge, thus

$$\frac{f(x+h) - f(x)}{h} \longrightarrow \infty \tag{4.37}$$

Therefore, f isn't differentiable in x. □

Remark 4.35. The inverse is not true.

Theorem 4.36. *Let I be an open interval and $f, g : I \to \mathbb{K}$ differentiable in $x \in I$. Then $f + g$ and $f \cdot g$ are differentiable too, and if $g(x) \neq 0$ then f/g is also differentiable.*

$$(f+g)'(x) = f'(x) + g'(x)$$
$$(f \cdot g)'(x) = f'(x)g(x) + f(x)g'(x)$$
$$\left(\frac{1}{g}\right)'(x) = \frac{-g'(x)}{g(x)^2}$$

Proof. Left as an exercise for the reader. $\square$

Theorem 4.37 (Chain rule)**.** *Let I, J be open intervals, and let*

$$g : J \longrightarrow I \qquad\qquad f : i \longrightarrow \mathbb{K}$$

g and f are to be differentiable in x and $f(x)$ respectively. Then $f \circ g$ is differentiable in x and

$$(f \circ g)' = g'(x) \cdot f'(g(x))$$

Proof. Consider the following function

$$\phi : J \longrightarrow \mathbb{K} \qquad\qquad \phi(\xi) = \begin{cases} \frac{f(g(x)+\xi)-f(g(x))}{\xi}, & \xi \neq 0 \\ f'(g(x)), & \xi = 0 \end{cases} \tag{4.38}$$

ξ is continuous, since f is continuous and

$$\lim_{\xi \to 0} \phi(\xi) = f'(g(x)) = \phi(0) \tag{4.39}$$

$\forall \xi \in J$ the following holds

$$f(g(x) + \xi) - f(g(x)) = \phi(\xi) \cdot \xi \tag{4.40}$$

With this we can now show that

$$\begin{aligned} \frac{f(g(x+h)) - f(g(x))}{h} &= \frac{f(g(x) + (g(x+h) - g(x))) - f(g(x))}{h} \\ &= \frac{\phi(g(x+h) - g(x))(g(x+h) - g(x))}{h} \\ &= \underbrace{\phi(g(x+h) - g(x))}_{\xrightarrow{h \to 0} 0} \cdot \underbrace{\frac{g(x+h) - g(x)}{h}}_{\xrightarrow{h \to 0} g'(x)} \\ &\xrightarrow{h \to 0} g'(x) \cdot f'(g(x)) \end{aligned} \tag{4.41}$$

$\square$

Definition 4.38. Let I be an interval and $f : I \to \mathbb{R}$. $x_0 \in I$ is called a global maximum if

$$f(x) \leq f(x_0) \quad \forall x \in I$$

$x_0 \in I$ is called a local maximum if

$$\exists \epsilon > 0 : \quad f(x) \leq f(x_0) \quad \forall x \in (x_0 - \epsilon, x_0 + \epsilon)$$

An extremum is either maximum or minimum.

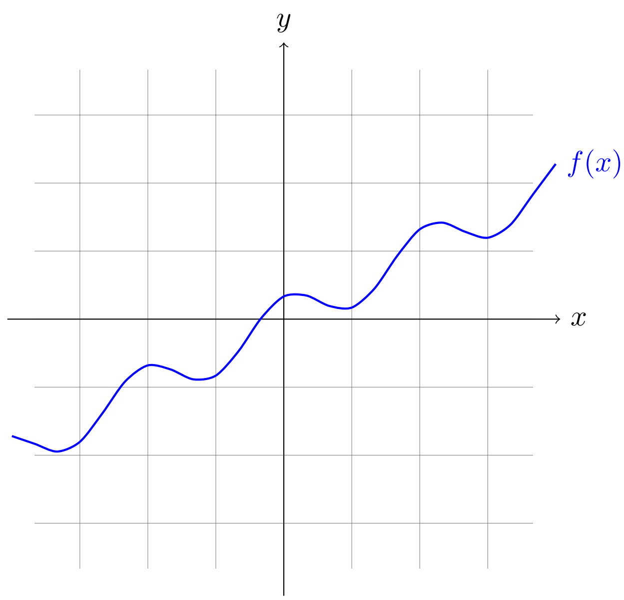
Example 4.39. (i) Let $f : [-1,1] \to \mathbb{R}$, $f(x) = x^2$.

- $x_0 = 0$ is a local and global minimum
- $x_0 = \pm 1$ is a local and global maximum

(ii) Consider

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \cos x + \frac{x}{2} \end{aligned}$$

f has infinitely many local extrema, but no global ones!



(iii) Consider

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \begin{cases} 1, & x \text{ rational} \\ 0, & x \text{ irrational} \end{cases} \end{aligned}$$

- x_0 rational is a global maximum
- x_0 irrational is a global minimum

Theorem 4.40. *Let I be an open interval, and $f : I\mathbb{R}\mathbb{R}$ a function with a local extremum at $x_0 \in I$. Then*

$$f \text{ differentiable in } x_0 \implies f'(x_0) = 0$$

Proof. Assume $f'(x_0) \neq 0$ (w.l.o.g. $f'(x_0) > 0$, otherwise consider $-f$). Then

$$\exists\delta > 0 : \quad \left|\frac{f(x_0+h) - f(x)}{h} - f'(x_0)\right| < f'(x_0) \ \ \forall h \in (-\delta, \delta) \tag{4.42}$$

Especially

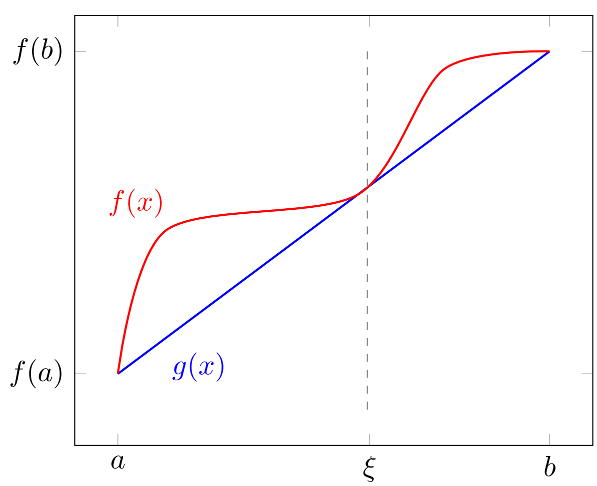
$$0 < \frac{f(x_0+h) - f(x_0)}{h} \ \ \forall h \in (-\delta, \delta) \tag{4.43}$$

For $h > 0$ this means $f(x_0 + h) > f(x_0)$. And for $h < 0$ this means that $f(x_0 + h) < f(x_0)$. Thus x_0 is not an extremum. □

Remark 4.41. Let $f : I \to \mathbb{R}$ be differentiable. To find the extrema of f, calculate f' and find its roots. However, the roots are to be insepcted more closely, as $f'(x_0) = 0$ is not a sufficient criterion (The function could have inflection points or behave badly at the boundaries of I).

Theorem 4.42 (Mean value theorem)**.** *Let $a, b \in \mathbb{R}$ with $a < b$, and let $f, g : [a, b] \to \mathbb{R}$ be differentiable. Then $\exists\xi \in (a, b)$ such that*

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



Proof. Consider all

$$h(x) = (f(b) - f(a))g(x) - f(x)(g(b) - f(a)) \tag{4.44}$$

h is differentiable, which means h is continuous on $[a, b]$:

$$h(a) = f(b)g(a) - f(a)g(b) = h(b) \tag{4.45}$$

We need to show that h' has a root in $[a, b]$. If h is constant, this is trivial. So we assume $\exists x \in (a, b)$ such that $h(x) > h(a)$. Since h is continuous on (a, b) there exists a global maximum $x_0 \in [a, b]$ with $x_0 \neq a$ and $x_0 \neq b$. This implies that $h'(x_0) = 0$. If $h(x) < h(a)$ the same argument can be made. □

Remark 4.43. This theorem is often written as

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

And if $g(x) = x$

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

Corollary 4.44. *Let I be an open interval and $f : I \to \mathbb{R}$ differentiable. Then*

(i) $f'(I) \subset [0, \infty) \iff$ *monotonically increasing*

(ii) $f'(I) \subset (0, \infty) \implies$ *strictly monotonically increasing*

(iii) $f'(I) \subset (-\infty, 0] \iff$ *monotonically decreasing*

(iv) $f'(I) \subset (-\infty, 0) \implies$ *striuctly monotonically decreasing*

Proof. We will only show the " $\implies$ " direction for (i). Assume f isn't monotonically increasing, then $\exists x, y \in I$ such that $x < y$ but $f(x) > f(y)$. The mean value theorem thus states, $\exists \xi \in (x, y)$ such that

$$f'(\xi) = \frac{f(y) - f(x)}{y - x} < 0 \tag{4.46}$$

All other statements are proven in the same fashion. □

Example 4.45. f strictly monotonically increasing does NOT imply that $f'(I) \subset (0, \infty)$. Consider $f(x) = x^3$.

Corollary 4.46 (L'Hôpital's rule)**.** *Let $a, b, x_0 \in \mathbb{R}$, with $a < x_0 < b$ and let $f, g : (a, b) \to \mathbb{R}$ be a differentiable function. We require $f(x_0) = g(x_0) = 0$. If $g'(x) \neq 0 \ \ \forall x \in I \setminus \{x_0\}$ and if*

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

exists, then

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

Proof. Between two roots of g there must be at least one root of g'. I.e. $g(x) \neq 0 \ \ \forall x \in I \setminus \{x_0\}$. This means, that

$$\forall x \in (a, x_0) \ \exists \xi_x : \quad \frac{f(x)}{g(x)} = \frac{f(x) - f(x_0)}{g(x) - g(x_0)} = \frac{f'(\xi_x)}{g'(\xi_x)} \implies \lim_{x \to x_0} \frac{f'(x)}{g'(x)} \tag{4.47}$$

Since $\xi_x \in (x, x_0)$

$$\xi_x \xrightarrow{x \to x_0} x_0 \tag{4.48}$$

For the limit from the left, this implies

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = \lim_{x \to x_0} \frac{f'(x)}{g'(x)} \tag{4.49}$$

This argument can be made for the limit from the right as well. □

Remark 4.47. (i) For the computation of the limit it is enough to consider f and g on $(x_0 - \delta, x_0 + \delta)$ with $\delta > 0$.

(ii) L'Hôpital's rule also works for one-sided limits

(iii) Let $f, g : (a, b) \setminus \{x_0\} \to \mathbb{R}$ be differentiable. Then it is enough to require

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

(iv) L'Hôpital's rule doesn't generally apply to complex valued functions.

(v) By substituring $\tilde{f}(x) = f\left(\frac{1}{x}\right)$ and $\tilde{g}(x) = g\left(\frac{1}{x}\right)$ we can also use

$$\lim_{x \to \infty} \frac{\tilde{f}(x)}{\tilde{g}(x)} = \lim_{x \to \infty} \frac{\tilde{f}'(x)}{\tilde{g}'(x)}$$

(vi) The inverse

$$L = \lim_{x\to x_0} \frac{f(x)}{g(x)} \implies \lim_{x\to 0} \frac{f'(x)}{g'(x)} = L$$

is NOT true.

Example 4.48. Consider

$$\lim_{x\to 0} \frac{x^2}{1-\cos x} = \frac{\text{“}0\text{”}}{0}$$

The functions here are

$$f(x) = x^2 \qquad\qquad g(x) = 1-\cos x$$

with the derivatives

$$f'(x) = 2x \qquad\qquad g'(x) = \sin x$$

However, the limit of the derivatives is still

$$\lim_{x\to 0} \frac{2x}{\sin x} = \frac{\text{“}0\text{”}}{0}$$

We can derive the functions again

$$f''(x) = 2 \qquad\qquad g''(x) = \cos x$$

And thus

$$\lim_{x\to 0} \frac{2}{\cos x} = 2 \implies \lim_{x\to 0} \frac{x^2}{1-\cos x} = 2$$

Theorem 4.49 (Derivative of inverse functions)**.** *Let I be an open inverval, and $f : I \to \mathbb{R}$ differentiable with $f'(I) \subset (0,\infty)$. Then f has a differentiable inverse function $f^{-1}(x) : f(I) \to \mathbb{R}$ and for $y \in f(I)$ we have*

$$\left(f^{-1}\right)'(y) = \frac{1}{f'\left(f^{-1}(y)\right)}$$

Proof. f is strictly monotonically increasing, thus f^{-1} exists and is continuous. Let $y \in f(I)$, $x := f^{-1}(y)$ and

$$\xi(h) = f^{-1}(y+h) - \underbrace{f^{-1}(y)}_{x} \tag{4.50}$$

Then

$$x + \xi(h) = f^{-1}(y+h) \implies f(x+\xi(h)) = y + h = f(x) + h \tag{4.51}$$

Which in turn implies

$$f(x+\xi(h)) - f(x) = h \tag{4.52}$$

Now we have

$$\begin{aligned}\frac{f^{-1}(y+h) - f^{-1}(y)}{h} =& \frac{\xi(h)}{f(x+\xi(h)) - f(x)} \\ =& \left(\frac{f(x+\xi(h)) - f(x)}{\xi(h)}\right)^{-1} \\ \xrightarrow{h\to 0}& \big(f'(x)\big)^{-1} = \frac{1}{f'(f^{-1}(y))} > 0\end{aligned} \tag{4.53}$$

□

Example 4.50. (i) Let $n \in \mathbb{N}$ and consider

$$\begin{aligned} f : (0,\infty) &\longrightarrow \mathbb{R} \\ x &\longmapsto x^n \end{aligned}$$

The derivative is $f'(x) = nx^{n-1}$. The inverse function is

$$g(y) = \sqrt[n]{y} \qquad g'(y) = \frac{1}{f'(g(y))} = \frac{1}{n\left(\sqrt[n]{y}\right)^{n-1}} = \frac{1}{n} \cdot y^{\left(\frac{1}{n}-1\right)}$$

(ii) The natural logarithm. Let $f(x)0\exp x$ and $g(y) = \ln y$. Then

$$(\ln y)' = \frac{1}{\exp(\ln(y))} = \frac{1}{y}$$

(iii) Let $f(x) = x^3$. Then

$$f^{-1}(y) = \begin{cases} \sqrt[3]{y}, & y \geq 0 \\ -\sqrt[3]{y}, & y < 0 \end{cases}$$

f^{-1} is not differentiable in $y = 0$.

Definition 4.51. Let I be an open interval. $f : I \to \mathbb{R}$ is said to be $(n+1)$-times differentiable if the n-th derivative of f $(f^{(n)})$ is differentiable.

f is said to be infinitely differentiable (or smooth) if f is n times differentiable for all $n \in \mathbb{N}$.

f is said to be n times continuously differentiable if the n-th derivative $f^{(n)}$ is continuous.

Definition 4.52. Let I be an open interval, and $f : I \to \mathbb{R}$ n times differentiable in $x \in I$. Then

$$T_n f(y) = \sum_{k=0}^{n} \frac{f^{(k)}(x)}{k!}(y-x)^k$$

is called the Taylor polynomial of n-th degree at x of f.

Theorem 4.53 (Taylor's theorem)**.** *Let I be an open interval and $f : I \to \mathbb{R}$ an $(n+1)$-times differentiable function. Let $x \in I$ and $h : I \to \mathbb{R}$ differentiable. For every $y \in I$, there exists a ξ between x and y such that*

$$(f(y) - T_n f(y)) \cdot h'(\xi) = \frac{f^{(n+1)}(\xi)}{n!}(y-\xi)^n(h(y) - h(x))$$

Proof. Let

$$\begin{aligned} g : I &\longrightarrow \mathbb{R} \\ t &\longmapsto \sum_{k=0}^{n} \frac{f^{(k)}(t)}{k!}(y-t)^k \end{aligned} \tag{4.54}$$

Apply the mean value theorem to g and h to get

$$g'(\xi)(h(y) - h(x)) = (g(y) - g(x))h'(\xi) = (f(y) - T_n f(y))h'(\xi) \tag{4.55}$$

and thus

$$\begin{aligned} g'(t) &= \sum_{k=0}^{n} \underbrace{\left(\frac{f^{(k+1)}(t)}{k!}(y-t)^k - \frac{f^{(k)}(t)}{k!}k(y-t)^{k-1} \right)}_{\text{Telescoping series}} \\ &= \frac{f^{n+1}(t)}{n!}(y-t)^n \end{aligned} \tag{4.56}$$

By inserting ξ we receive the desired equation. ☐

Remark 4.54. (i) This is useful for when $h'(\xi) \neq 0$

(ii) The choice of h can yield different errors

$$R_{n+1}(y, x) := f(y) - T_n f(y)$$

(iii) The Langrange error bound is for $h(t) = (y-t)^{n+1}$:

$$R_{n+1}(y, x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(y-x)^{n+1}$$

(iv) This theorem makes no statement about Taylor series.

Corollary 4.55. *Let $(a, b) \subset \mathbb{R}$ and $f : (a, b) \to \mathbb{R}$ a n-times continuously differentuable function with*

$$0 = f'(x) = f''(x) = \cdots = f^{(n-1)}(x)$$

and $f^{(n)} \neq 0$. If n is odd, then there is no local extremum in x. If n is even then

$$\begin{aligned} f^{(n)}(x) > 0 &\implies x \textit{ is a local maximum} \\ f^{(n)}(x) < 0 &\implies x \textit{ is a local minimum} \end{aligned}$$

Proof. W.l.o.g. $f^{(n)} > 0$. We will use the Taylor series with Lagrange error bound. According to prerequisites, $f^{(n)}$ is continuous, i.e. $\exists \epsilon > 0$ such that $f^{(n)}(\xi) > 0$ on $(x - \epsilon, x + \epsilon)$. The Taylor formula tells us, that $\forall y \in (x - \epsilon, x + \epsilon)$ $\exists \xi_y \in (x - \epsilon, x + \epsilon)$ such that

$$f(y) - T_{n-1}(f(y)) = f(y) - f(x) = \frac{f^{(n)}(\xi_y)}{n!}(y - x)^n \tag{4.57}$$

For n odd, $f(y) - f(x)$ assumes positive and negative values in every neighbourhood of x. If n is even then $f(y) - f(x)$ cannot be negative, thus x is a local minimum. $\square$

Chapter 5

Topology in Metric spaces

5.1 Metric and Normed spaces

Definition 5.1 (Metric space)**.** A metric space (X, d) is an ordered pair consisting of a set X and a mapping

$$d : X \times X \longrightarrow [0, \infty]$$

called metric. This mapping must fulfil the following conditions $\forall x, y, z \in X$:

- $d(x, y) \geq 0$ (Positivity)
- $d(x, y) = 0 \iff x = y$ (Definedness)
- $d(x, y) = d(y, x)$ (Symmetry)
- $d(x, y) \leq d(x, z) + d(z, y)$ (Triangle inequality)

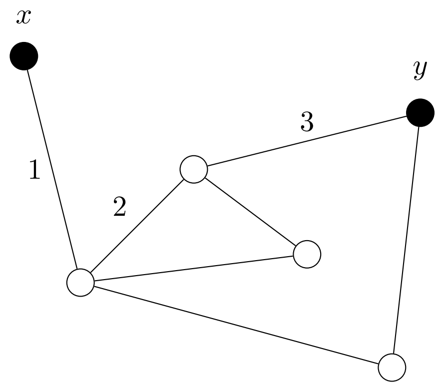
Example 5.2. (i) Let M be a set. Then

$$d(x, y) = \begin{cases} 1, & x \neq y \\ 0, & \text{else} \end{cases}$$

is called the discrete metric.

(ii) Let X be the set of edges of a graph.

$$\begin{aligned} d(x, y) := &\text{ Minimum amount of edges that have} \\ &\text{ to be passed to get from } x \text{ to } y \end{aligned}$$



(iii) Let X be the surface of a sphere.

$$d(x,y) := \text{"Bee line"}$$

(iv) Let X be the set of points of the European street network.

$$d(x,y) := \text{ Shortest route along this network}$$

(v) Let (X, d_X), (Y, d_Y) be metric spaces. Then

$$d_{X\times Y}((x_1,y_1),(x_2,y_2)) := d_X(x_1,x_2) + d_Y(y_1,y_2)$$

defines a metric on $X \times Y$.

Definition 5.3 (Normed space)**.** $(V, \|\cdot\|)$ is said to be a normed space if V is a vector space and

$$\|\cdot\| : V \longrightarrow [0,\infty)$$

is a mapping (called norm) with the following properties

- $\|x\| \geq 0$ (Positivity)
- $\|x\| = 0 \iff x = 0$ (Definedness)
- $\|\lambda x\| = |\lambda| \|x\|$
- $\|x + y\| \leq \|x\| + \|y\|$ (Triangle inequality)

To every norm belongs a unique induced metric

$$d(x,y) = \|x - y\|$$

Example 5.4 ($\mathbb{R}^n$ with Euclidian norm).

$$\begin{aligned} \|\cdot\| : \mathbb{R}^n &\longrightarrow [0, \infty) \\ (x_1, x_2, \cdots, x_n) &\longmapsto \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \end{aligned}$$

Then $(\mathbb{R}^n, \|\cdot\|)$ is a normed space.

Example 5.5. (i) $(x_1, x_2, \cdots, x_n) \mapsto |x_1| + |x_2| + \cdots + |x_n|$ is also a norm on $\mathbb{R}^n$.

(ii) On
$$V = \{f : [0,1] \longrightarrow \mathbb{R} \mid f \text{ continuous}\}$$
we can define the supremum norm
$$\|f\|_\infty = \sup\{|f(x)| \mid x \in [0,1]\}$$

(iii) We can define sequence spaces as
$$\ell^p = \left\{ (x_n) \subset \mathbb{C}^n \;\middle|\; \sum_{n=1}^\infty |x_n|^p < \infty \right\}$$
with the norm
$$\|(x_n)\|_p := \sqrt{\sum_{n=1}^\infty |x_n|^2}$$
A special space is ℓ^2, called Hilbert space

Remark 5.6. The Minkowski metric is not a metric in this sense.

Definition 5.7 (Balls and Boundedness)**.** Let (X, d) be a metric space, and $x \in X, r > 0$. We then define

$$\begin{aligned} B_r(x) &= \{y \in X \mid d(x,y) < r\} && \text{Open ball} \\ K_r(x) &= \{y \in X \mid d(x,y) \leq r\} && \text{Closed ball} \end{aligned}$$

A subset $M \subset X$ is called bounded if

$$\exists x \in X, r > 0 : \quad M \subset B_r(x)$$

5.2 Sequences, Series and Limits

Definition 5.8 (Sequences and Convergence)**.** Let (X, d) be a metric space. A sequence is a mapping $\mathbb{N} \to X$. We write $(x_n)_{n\in\mathbb{N}}$ or (x_n).

The sequence (x_n) is said to be convergent to $x \in X$ if

$$\forall \epsilon > 0 \; \exists N \in \mathbb{N} \; \forall n \geq N : \quad d(x_n, x) < \epsilon$$

x is said to be the limit, and sequences that aren't convergent are called divergent.

Remark 5.9. On $\mathbb{R}$ the metric is the Euclidian metric $|\cdot|$, therefore this new definition of convergence is merely a generalization of the old one.

Theorem 5.10. *Let (x_n) be a sequence in the metric space (X, d) and $x \in X$. Then the following statements are equivalent:*

(i) *(x_n) converges to x*

(ii) *$\forall \epsilon > 0$ $B_\epsilon(x)$ contains all but finitely many elements of the sequence (almost every (a.e.) element)*

(iii) *$(d(x, x_n))$ is a null sequence*

Proof. (ii) is merely a reformulation of (i), and $(ii) \iff (iii)$ follows from

$$d(x_n, x) = |d(x_n, x) - 0| \tag{5.1}$$

□

Theorem 5.11. *Let $\left(x^{(n)}\right) = (x_1^{(n)}, x_2^{(n)}, \cdots, x_d^{(n)}) \subset \mathbb{R}^d$ and*

$$x = (x_1, \cdots, x_d) \in \mathbb{R}^d$$

$\left(x^{(n)}\right)$ is said to converge to x if and only if $x_i^{(n)}$ converges to x_i for all i in $\{1, \cdots, d\}$

Proof. For $y = (y_1, \cdots, y_d) \in \mathbb{R}^d$ we have

$$\|y_i\| < \|y\| \quad \forall i \in \{1, \cdots, d\} \tag{5.2}$$

If $\left(x^{(n)}\right)$ converges to x, then

$$\left|x_i^{(n)} - x_i\right| \leq \left\|x^{(n)} - x\right\| \longrightarrow 0 \tag{5.3}$$

If $(x_i^{(n)})$ converges to $x_i \quad \forall i \in \{1, \cdots d\}$, then

$$\forall \epsilon > 0 \; \exists N \in \mathbb{N} \; \forall n > N : \quad \left| x_i^{(n)} - x_i \right| < \frac{\epsilon}{\sqrt{d}} \quad \forall i \in \{1, \cdots d\} \tag{5.4}$$

Thus

$$\begin{aligned} \left\| x^{(n)} - x \right\| &= \sqrt{(x_1^{(n)} - x_1)^2 + (x_2^{(n)} - x_2)^2 + \cdots + (x_d^{(n)} - x_d)^2} \\ &\leq \sqrt{\frac{\epsilon^2}{d} + \frac{\epsilon^2}{d} + \cdots + \frac{\epsilon}{2}} \\ &= \epsilon \end{aligned} \tag{5.5}$$

So $\left(x^{(n)}\right)$ converges to x. □

Theorem 5.12. *Every convergent sequence has exactly one limit and is bounded.*

Proof. Assume that x, y are limits of (x_n) with $x \neq y$. Then $d(x, y) > 0$. There exists $N_1, N_2 \in \mathbb{N}$, such that

$$d(x_n, x) < \frac{d(x,y)}{2} \quad \forall n \geq N_1 \tag{5.6a}$$

$$d(x_n, x) < \frac{d(x,y)}{2} \quad \forall n \geq N_2 \tag{5.6b}$$

From this follows that

$$d(x, y) \leq d(x, x_n) + d(x_n, y) < d(x, y) \quad \forall \max\{N_1, N_2\} \tag{5.7}$$

which is a contradiction, thus sequences can have only one limit.

Now if (x_n) converges to x, then

$$\exists N \in \mathbb{N} \; \forall n \geq N : \quad d(x_n, x) < 1 \tag{5.8}$$

Then

$$d(x_n, x) \leq \max\{d(x_1, x), d(x_2, x), \cdots, d(x_{N-1}, x), 1\} \tag{5.9}$$

□

Theorem 5.13. *Let $(V, \|\cdot\|)$ be a normed space over $\mathbb{K}$. Let $(x_n), (y_n) \subset V$ be sequences with limits $x, y \in V$ and $(\lambda_n) \subset \mathbb{K}$ a sequence with limit $\lambda \in \mathbb{K}$. Then*

$$x_n + y_n \longrightarrow x + y \qquad\qquad \lambda_n x_n \longrightarrow \lambda x$$

Proof. Left as an exercise for the reader. □

Definition 5.14 (Cauchy sequences and completeness)**.** A sequence (x_n) in a metric space (X, d) is called Cauchy sequence if

$$\forall \epsilon > 0 \; \exists N \in \mathbb{N} : \quad d(x_n, x_m) < \epsilon \;\; \forall m, n \geq N$$

A metric space is complete if every Cauchy sequence converges. A complete normed space is called Banach space.

Example 5.15.
$(\mathbb{R}, |\cdot|)$ and $(\mathbb{C}, |\cdot|)$ are complete
$(\mathbb{Q}, |\cdot|)$ is not complete

Theorem 5.16. *Every convering series is a Cauchy sequence*

Proof. Let $(x_n) \longrightarrow x$. This means that

$$\forall \epsilon > 0 \; \epsilon N \in \mathbb{N} : \quad d(x_n, x) < \frac{\epsilon}{2} \;\; \forall n \geq N \tag{5.10}$$

Then

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) < \epsilon \;\; \forall m, n \geq N \tag{5.11}$$

□

Theorem 5.17. $\mathbb{R}^n$ *with the Euclidian norm is complete.*

Proof. Let $\left(x^{(n)}\right) \subset \mathbb{R}^n$ be a Cauchy sequence. We know that

$$\forall y \in \mathbb{R}^n : \quad |y_i| \leq \|y\| \;\; \forall i \in \{1, \cdots, n\} \tag{5.12}$$

We also know that $(x_i^{(n)})$ are Cauchy sequences because

$$\left|(x_i^{(n)} - x_i^m)\right| \leq \left\|x^{(n)} - x^{(m)}\right\| \;\; \forall i \in \{1, \ldots, n\} \tag{5.13}$$

Thus $x_i^{(n)} \longrightarrow x_i$ and therefore $\left(x^{(n)}\right) \longrightarrow x$. □

Definition 5.18 (Series and (absolute) convergence)**.** Let $(V, \|\cdot\|)$ be a normed space and $(x_n) \subset V$. The series

$$\sum_{k=1}^{\infty} x_k$$

is the sequence of partial sums

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

If the series converges then $\sum_{k=1}^{\infty} x_k$ also denotes the limit. The series is said to absolutely convergent if

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

Theorem 5.19. *In Banach spaces every absolutely convergent series is convergent.*

Proof. Let $(V, \|\cdot\|)$, $(x_n) \subset V$ and require $\sum_{n=1}^{\infty} (V, \|\cdot\|) x_n < \infty$. We need to show that $s_n = \sum_{k=1}^{n} x_k$ is a Cauchy sequence. Let $\epsilon > 0$ and $t_n = \sum_{k=1}^{n} \|x_k\|$. (t_n) is convergent in $\mathbb{R}$, and thus a Cauchy sequence. I.e.

$$\exists N \in \mathbb{N} : \quad |t_n - t| < \epsilon \quad \forall m, n \geq N \tag{5.14}$$

For $n > m > N$:

$$\|s_n - s_m\| = \left\| \sum_{k=m+1}^{n} x_k \right\| \leq \sum_{k=m+1}^{n} \|x_k\| = t_n - t_m = |t_n - t_m| < \epsilon \tag{5.15}$$

□

Theorem 5.20. *Let $(V, \|\cdot\|)$ be a Banach space, $\sum_{k=1}^{\infty} x_k$ absolutely convergent and let $\sigma : \mathbb{N} \to \mathbb{N}$ be a bijective mapping. Then*

$$\sum_{k=1}^{\infty} x_k = \sum_{k=1}^{\infty} x_{\sigma(k)}$$

Proof. Analogous to *Theorem* 2.55 □

5.3 Open and Closed Sets

Definition 5.21 (Inner points and Boundary points)**.** Let (X, d) be a metric space, $A \subset X$ and $x \in X$.

(i) x is said to be an inner point of A, if

$$\exists \epsilon > 0 : \quad B_\epsilon(x) \subset A$$

(ii) x is said to be a boundary point of A if

$$\forall \epsilon > 0 : \quad \underbrace{B_\epsilon(x) \cap A \neq \emptyset}_{\substack{B_\epsilon(x) \text{ contains} \\ \text{points from } A}} \wedge \underbrace{B_\epsilon(x) \cap (X \setminus A) \neq \emptyset}_{\substack{B_\epsilon(x) \text{ contains points} \\ \text{from outside of } A}}$$

(iii) The set

$$\{x \in X \,|\, x \text{ is inner point of } A\}$$

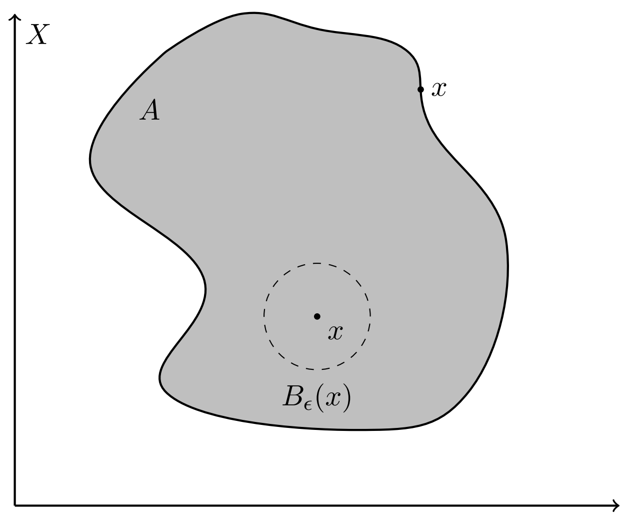
is called the interior of A, and is denoted as $\mathring{A}$.

(iv) The set

$$\{x \in X \,|\, x \text{ is boundary point of} A\}$$

is called the boundary of A, and is denoted as ∂A.

(v) $A \cup \partial A$ is said to be the closure of A, and is denoted as $\bar{A}$.



Example 5.22. Consider $X = \mathbb{R}^2$. Then

$$A = \{(x, y) \in \mathbb{R} \,|\, 0 \leq y < 1\}$$
$$\mathring{A} = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, 0 \leq y < 1\right\}$$
$$\partial A = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, y = 1 \vee y = 0\right\}$$
$$\bar{A} = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, 0 \leq y \leq 1\right\}$$

Remark 5.23. (i) $\mathring{A} \subset A$

(ii) Boundary points of A can be elements of A or not.

(iii) $A \subset \mathring{A} \cup \partial A, \quad \mathring{A} \cap \partial A = \varnothing$

(iv) $\partial A = \partial X \setminus A$

Theorem 5.24. *Let (X, d) be a metric space, $A \subset X$ and x an interior point or boundary point of A. Then*

$$\exists\, (x_n) \subset A : \quad x_n \longrightarrow x$$

Proof. If $x \in A$ then this is trivial, so let $x \notin A$. Then

$$\forall n \in \mathbb{N}\ \exists x_n \in \left(B_{\frac{1}{n}}(x) \cap A \neq \varnothing\right) \tag{5.16}$$

We need to show that (x_n) converges to x.

$$\forall \epsilon > 0\ \epsilon N \in \mathbb{N} : \quad \frac{1}{N} < \epsilon \tag{5.17}$$

For $n \geq N$ we have

$$\frac{1}{n} \leq \frac{1}{N} < \epsilon \tag{5.18}$$

and thus

$$d(x_n, x) < \frac{1}{n} < \epsilon \tag{5.19}$$

□

Definition 5.25 (Open and Closed sets)**.** Let (X, d) be a metric space. $A \subset X$ is said to be

(i) open, if every point in A is an interior point

(ii) closed, if A contains all its boundary point

(iii) neighbourhood of $x \in A$, if x is an interiot point of A

Theorem 5.26. *Let (X, d) be a metric space and $A \subset X$.*

$$A \text{ open} \iff X \setminus A \text{ closed}$$

Proof.

$$\begin{aligned} A \text{ open} &\iff \forall x \in A: \quad x \in \mathring{A} && (5.20a)\\ &\iff \forall x \in A: \quad x \in \partial A && (5.20b)\\ &\iff X \setminus A \text{ contains all boundary point of } A && (5.20c)\\ &\iff X \setminus A \text{ contains all boundary points of } X \setminus A && (5.20d)\\ &\iff X \setminus A \text{ closed} && (5.20e)\end{aligned}$$

□

Remark 5.27. That doesn't mean A has to be either open and closed.

Example 5.28. Let (X, d) be a metric space, $x \in X$ and $r > 0$. Then

$$B_r(x) = \{y \in X \mid d(x,y) < r\} \text{ is open}$$
$$K_r(x) = \{y \in X \mid d(x,y) < r\} \text{ is closed}$$

Remark 5.29. Consider the special case $a, b \in \mathbb{R}$ with $a < b$

$$(a,b) = B_{\frac{b-a}{2}}\left(\frac{a+b}{2}\right) \text{ open}$$
$$[a,b] = K_{\frac{b-a}{2}}\left(\frac{a+b}{2}\right) \text{ closed}$$

Theorem 5.30. *Let (X, d) be a metric space and $A \subset X$.*

$$A \textit{ closed} \iff \forall\, (x_n) \subset A \textit{ convergent}: \quad \lim_{n\to\infty} x_n \in A$$

Proof. Assume A is closed. Let $(x_n) \subset A$ be convergent to x. then

$$\forall \epsilon > 0\ \exists N \in \mathbb{N}: \quad x_n \in B_\epsilon(x) \ \ \forall n \geq N \tag{5.21}$$

This means that every ϵ-ball around x contains at least one point from A. I.e. x is always a point (or a boundary point) of A. From A closed follows $x \in A$.

Now assume $x \in \partial A$. Then

$$\exists\, (x_n) \subset A: \quad (x_n) \longrightarrow x \tag{5.22}$$

According to the prerequisites, $x \in A$. □

Theorem 5.31. *Let (X, d) be a metric space, and τ the set of all open subsets. Then* *(i)* $\varnothing \in \tau, \quad X \in \tau$

(ii) The union of any number of sets from τ is an open set

$$\bigcup_{t\in\tau} t \in \tau$$

(iii) The intersection of finitely many sets from τ is an open set

$$\bigcap_{t\in\tau} t \in \tau$$

Proof. Left as an exercise for the reader. □

Remark 5.32. (i) τ is said to be the topology induced by d

(ii)
- $\varnothing$, X are also closed
- The intersection of any number of closed sets is closed
- The union of finitely many closed sets is closed

(iii) Infinitely many intersections of open sets are not open in general.

Theorem 5.33. *Let (X, d) be a metric space and $A \subset X$. Then*

$$\mathring{A} \textit{ open} \implies \partial A, \bar{A} \textit{ closed}$$

Proof. Let $\mathring{A}$ be open and $x \in \mathring{A} \subset A$. This means

$$\exists \epsilon > 0 : \quad B_\epsilon(x) \subset A \tag{5.23}$$

We have to show that $B_\epsilon(x) \subset \mathring{A}$. Let $y \in B_\epsilon(x)$. Since $B_\epsilon(x)$ is open

$$\exists \delta > 0 : \quad B_\delta(y) \subset B_\epsilon(x) \subset A \tag{5.24}$$

This means that $y \in B_\epsilon(x)$ is interior point A. I.e. $\subset (x) \subset \mathring{A}$, and thus x is interior point of $\mathring{A}$.

Let $B = X \setminus A$. Then $\partial A = \partial B$

$$X = A \cup B = \mathring{A} \cup \partial A \cup \mathring{B} \cup \partial B = \mathring{A} \cup \partial A \cup \mathring{B} \tag{5.25}$$

Then

$$A \text{ and } B \text{ are disjoint} \implies \mathring{A}, \mathring{B} \text{ disjoint} \tag{5.26a}$$
$$\implies \partial A \text{ disjoint to } \mathring{A}, \mathring{B} \tag{5.26b}$$

This results in

$$\partial A = X \setminus (\underbrace{\mathring{A} \cup \mathring{B}}_{\text{open}}) \implies \partial A \text{ closed} \tag{5.27}$$

and

$$\bar{A} = A \cup \partial A = \mathring{A} \cup \partial A = X \setminus \mathring{B} \text{ closed} \tag{5.28}$$

□

Theorem 5.34. *Let (X, d) be a metric space and $A \subset X$*

$$\bigcup_{\substack{O \text{ open} \\ O \subset A}} O = \mathring{A} \qquad \text{and} \qquad \bigcap_{\substack{C \text{ closed} \\ A \subset C}} C = \bar{A}$$

Proof. Let $\mathring{A}$ is open and $\mathring{A} \subset A$

$$\implies \bigcup_{O \subset A \text{ open}} \supset \mathring{A} \tag{5.29}$$

Now let $O \subset A$ be open and $x \in O$, i.e.

$$\exists \epsilon > 0 : \quad B_\epsilon(x) \subset O \subset A \implies x \in \mathring{A} \tag{5.30}$$

This implies that $O \subset \mathring{A}$. Since this holds for all open $O \subset A$, this statement is proven. The other statement follows from the complement. □

Theorem 5.35. *Let (X, d) be a complete space and $A \subset X$ be closed. Then (A, d_A) is complete.*

Proof. Left as an exercise for the reader. □

Remark 5.36. Topological terms (open, closed, continuous, compact) don't just depend on A, but also on X.

Definition 5.37. Let (X, d) be a metric space and $x \in X$.

(i) x is said to be an isolated point if $\exists \epsilon > 0$ such that $B_\epsilon(x) = \{x\}$.

(ii) x is said to be a limit point if it's not an isolated point.

Definition 5.38 (Punctured neighbourhood, Punctured ball)**.** $\dot{U} \subset X$ is said to be a punctured neighbourhood, if there is a neighbourhood U of x with $\dot{U} = U \setminus \{x\}$

A punctured ball is $\dot{B}_\epsilon(x) = B_\epsilon \setminus \{x\}$.

Definition 5.39 (Limit of mappings)**.** Let $(X, d_X), (Y, d_Y)$ and x a limit point of X. Let $\dot{U}$ be a punctured neighbourhood of x and $f : \dot{U} \to Y$. Then f converges to $y \in Y$ in x (y is said to be the limit of f in x), if

$$\forall \epsilon > 0 \; \exists \delta > 0 : \quad f(\tilde{x}) \in B_\epsilon(y) \quad [d(f(\tilde{x}), y) < \epsilon]$$

if $\tilde{x} \in \dot{B}_\epsilon(x)$ $[d(\tilde{x}, x) < \delta]$

Example 5.40. Let $f, g : \mathbb{R}^2 \setminus \{0\} \to \mathbb{R}$.

$$f(x) := \|x\|^2 \qquad\qquad g(x) := \frac{1}{\|x\|}$$

Then $\lim_{x\to 0} f(x) = 0$, because for $\epsilon > 0$ and $\delta = \sqrt{\epsilon}$ we have

$$d(\tilde{x}, 0) = \|\tilde{x} - 0\| = \tilde{x} < \delta \implies d(f(\tilde{x}), 0) = \left| \|\tilde{x}\|^2 - 0 \right| = \|\tilde{x}\|^2 < \epsilon = \delta^2$$

Theorem 5.41.

$$f \textit{ converges to } y \in Y \textit{ in } x \iff \forall\, (x_n) \subset X : \quad f(x_n) \xrightarrow{x_n \to x} y$$

Proof. Let $(x_n) \subset X$ with $x_n \longrightarrow x$. Let $\epsilon > 0$, then

$$\exists \delta > 0 : \quad f(\tilde{x}) \in B_\epsilon(y) \text{ if } \tilde{x} \in B_\delta(x) \tag{5.31}$$

Furthermore

$$\exists N \in \mathbb{N} : \quad x_n \in B_\delta(x) \quad \forall n \geq N \tag{5.32}$$

Then

$$f(x_n) \in B_\epsilon(y) \quad \forall n \geq N \tag{5.33}$$

To prove the other direction, assume f doesn't converge to y in y. This means

$$\exists \epsilon > 0 : \quad \exists \tilde{x} \in B_\delta(x) \text{ but } f(\tilde{x}) \notin B_\epsilon(y) \quad \forall \delta > 0 \tag{5.34}$$

Therefore

$$\forall n \in \mathbb{N} : \quad \exists x_n \in B_{\frac{1}{n}}(x) \tag{5.35}$$

We know that $x_n \longrightarrow x$ since $d(x_n, x) < \frac{1}{n}$, but $f(x_n)$ doesn't converge to y since $d(f(x_n), y) \geq \epsilon$. $\square$

Corollary 5.42. *Let (X, d) be a metric space, $x \in X$ a limit point and $\dot{U}$ a punctured neighbourhood of x. Let $f, g : \dot{U} \to \mathbb{K}$ with*

$$\lim_{\tilde{x}\to x} f(\tilde{x}) = y_1 \qquad\qquad \lim_{\tilde{x}\to x} g(\tilde{x}) = y_2$$

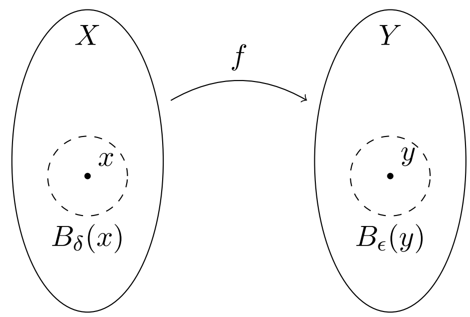
Then

$$\lim_{\tilde{x}\to x}(f+g)(\tilde{x}) = y_1 + y_2 \qquad \lim_{\tilde{x}\to x}(f\cdot g)(\tilde{x}) = y_1 \cdot y_2$$

$$\lim_{\tilde{x}\to x}\left(\frac{f}{g}\right)(\tilde{x}) = \frac{y_1}{y_2}$$

Heuristic Proof. Draw parallels back to number sequences □

5.4 Continuity



Definition 5.43. Let $(X, d_X), (Y, d_Y)$ be metric spaces. $f : x \to y$ is said to be continuous in $x \in X$ if

$$\forall \epsilon > 0\ \exists \delta > 0 : \quad \tilde{x} \in B_\delta(x) \implies f(\tilde{x}) \in B_\epsilon(f(x))$$

f is said to be continuous is it is continuous in every point.

Example 5.44. (i) Let (X, d) be a metric space.

$$\begin{aligned} \mathrm{id} : X &\longrightarrow X \\ x &\longmapsto x \end{aligned}$$

is continuous (choose $\delta = \epsilon$).

(ii) The function

$$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x, y) &\longmapsto (x, -y) \end{aligned}$$

is continuous. For $(\tilde{x}, \tilde{y}), (x, y) \in \mathbb{R}^2$ we have

$$\begin{aligned} \|f(\tilde{x}, \tilde{y}) - f(x, y)\|^2 &= \|(\tilde{x} - x, y - \tilde{y})\|^2 = (\tilde{x} - x)^2 + (y - \tilde{y})^2 \\ &= \|(\tilde{x}, \tilde{y}) - (x, y)\|^2 \end{aligned}$$

(iii) Consider

$$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x,y) &\longmapsto \begin{cases} 0, & x \cdot y = 0 \\ 1, & x \cdot y \neq 0 \end{cases} \end{aligned}$$

f is non continuous in $(0,0)$.

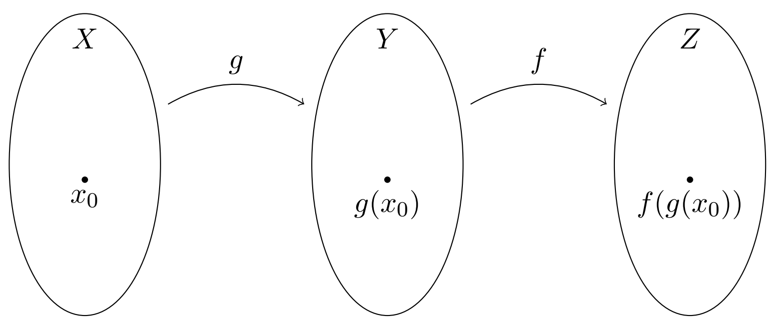
Remark 5.45. (i)

$$f \text{ continuous in } x \iff \forall \epsilon > 0\ \exists \delta > 0 : \quad f(B_\delta(x)) \subset B_\epsilon(f(x))$$

(ii) Continuity is a local property, this means if $x \in X$, U a neighbourhood of x and f, g functions with $f|_U = g|_U$, then

$$f \text{ continuous} \iff g \text{ continuous}$$

Theorem 5.46. *Let $x_0 \in X$, $g : X \to Y$ and $f : Y \to Z$. If g is continuous in x_0 and f is continuous in $g(x_0)$, then $f \circ g$ is continuous in x_0.*



Proof. Since f, g are continuous we know that

$$\forall \epsilon > 0\ \exists \delta > 0 : \quad y \in B_\delta(g(x_0)) \implies f(y) \in B_\epsilon(f(g(x_0))) \tag{5.36a}$$

$$\forall \delta > 0\ \exists \rho > 0 : \quad x \in B_\rho(x_0) \implies g(x) \in B_\delta(g(x_0)) \tag{5.36b}$$

Then $\forall x \in B_\rho(x_0)$ we have

$$(f \circ g)(x_0) = f(g(x_0)) \in B_\epsilon(f(g(x_0))) \tag{5.37}$$

□

Definition 5.47 (Lipschitz continuity)**.** A function $f : X \to Y$ is said to be Lipschitz continuous if

$$\exists L > 0 : \quad d_Y(f(x), f(y)) \leq L \cdot D_X(x, y)$$

L is called Lipschitz constant. If $L = 1$, f is called contraction.

Example 5.48. Let $f, g : [0, 1] \to \mathbb{R}$.

$$f(x) = x^2 \qquad\qquad g(x) = \sqrt{x}$$

f is Lipschitz continuous, g is not.

Theorem 5.49. *Every Lipschitz continuous function is continuous.*

Proof. Let $f : X \to Y$ be Lipschitz continuous, with Lipschitz constant L. Let $\epsilon > 0$, then for $x \in B_{\frac{\epsilon}{L}}(x_0)$

$$d(f(x), f(x_0)) \leq L \cdot d(x, x_0) < \epsilon \tag{5.38}$$

Thus, f is continuous in x_0, and since we chose an arbitrary x_0, f is continuous everywhere. ☐

Example 5.50. (i) Consider

$$\begin{aligned} \pi_i : \mathbb{K}^n &\longrightarrow \mathbb{K} \\ (x_1, x_2, \cdots, x_n) &\longmapsto x_i \end{aligned}$$

Then

$$|\pi_i(x) - \pi_i(y)| = |x_i - y_i| \leq \|x - y\|$$

So π_i is a contraction.

(ii) Let $(X, d), (X \times X, d_{X\times X})$ be metric spaces. Then

$$\begin{aligned} d : X \times X &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto d(x, y) \end{aligned}$$

is a contraction. Let $x_1, x_2, y_1, y_2 \in X$ and apply the triangle inequality

$$d(x_1, y_1) \leq d(x_1, x_2) + d(x_2, y_1) \leq d(x_1, x_2) + d(y_2, y_1) + d(x_2, y_2)$$

This implies

$$\begin{aligned} |d(x_1, y_1) - d(x_2, y_2)| &\leq d(x_1, x_2) + d(y_1, y_2) \\ &= d_{X\times X}((x_1, x_2), (y_1, y_2)) \end{aligned}$$

which means the metric is continuous.

(iii) Analogously, this works for $\|\cdot\|$.

Theorem 5.51. *Let $f : X \to Y$.*

$$f \text{ is continuous in } x \in X \iff {\scriptstyle x \text{ is an isolated point in } X \atop \text{or } \lim_{\tilde{x}\to x} f(\tilde{x})=f(x)}$$

Proof. Let f be continuous in $x \in X$. If x is an isolated point there is nothing to show, so let x be a limit point. Then

$$\forall \epsilon > 0\ \exists \delta > 0 : \quad f(\tilde{x}) \in B_\epsilon(f(x)) \ \ \forall \tilde{x} \in B_\delta(x) \tag{5.39}$$

Now let x be an isolated point, i.e. $\exists \delta > 0$ such that $B_\delta(x) = \{x\}$. Then

$$f(B_{delta}(x)) = \{f(x)\} \subset B_\epsilon(f(x)) \ \ \forall \epsilon > 0 \tag{5.40}$$

If x is a limit point and $\lim_{\tilde{x}\to x} f(\tilde{x}) = f(x)$, then let $\epsilon > 0$

$$\exists \delta > 0 : \quad f(\dot{B}_\delta(x)) \subset B_\epsilon(f(x)) \tag{5.41}$$

This then implies

$$f(B_\delta) \subset B_\epsilon(f(x)) \tag{5.42}$$

$\square$

Corollary 5.52.

$$f : X \to Y \text{ continuous in } x \in X \iff \forall\,(x_n) \subset X : \quad f(x_n) \xrightarrow{x_n \to x} f(x)$$

This means, for continuous f we have

$$\lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n)$$

Corollary 5.53. *Let $f_1, \cdots, f_n : \mathbb{R}^m \to \Re$. Then define*

$$\begin{aligned} f : \mathbb{R}^m &\longrightarrow \mathbb{R}^n \\ x &\longmapsto (f_1(x), f_2(x), \cdots, f_n(x)) \end{aligned}$$

f is continuous if and only if $f_1, \cdots, f_n$ are continuous.

Corollary 5.54. *Let $f, g : X \to \mathbb{R}$ be continuous in $x \in X$. Then*

$$f + g \qquad\qquad\qquad f \cdot g$$

are continuous in x, and if $g(x) \neq 0$ then

$$\frac{f}{g}$$

is also continuous in x.

Example 5.55. Let $\eta = (\eta_1, \cdots, \eta_n) \in \mathbb{N}_0^n$ and $x \in \mathbb{K}^n$. Define

$$x^\eta = x_1^{\eta_1} \cdot x_2^{\eta_2} \cdot x_3^{\eta_3} \cdot \cdots \cdot x_n^{\eta_n}$$

η is called multi index. We set

$$|\eta| := \eta_1 + \eta_2 + \eta_3 + \cdots + \eta_n$$

Let $c_\eta \in \mathbb{K} \ \ \forall \eta$ with $|\eta| \leq N \ \ N \in \mathbb{N}$. Then we call

$$\begin{aligned} f : \mathbb{K}^n &\longrightarrow \mathbb{K} \\ x &\longmapsto \sum_{|\eta| \leq N} c_\eta \cdot x^\eta \end{aligned}$$

a polynomial with n variables. Such polynomials are continuous. Example:

$$(x_1, x_2) \longmapsto x_1^2 + x_2^2 + x_1^9 + x_2^{17}$$

Remark 5.56. In the context of polynomials (and power series) we define

$$0^0 = 1$$

Reminder: If $f : X \to Y$ and $U \subset Y$ then $f^{-1}(U)$ is said to be the preimage of U under f. It's the set of all points of X that get mapped to U.

$$f^{-1}(U) = \{x \in X \mid f(x) \in U\}$$

Theorem 5.57. *Let $f : X \to Y$*

(i)

$$f \text{ is continuous in } x \iff \begin{array}{l} f^{-1}(U) \text{ is a neighbourhood of} \\ x \quad \forall U \text{neighbourhood of } f(x) \end{array}$$

(ii)

$$f \text{ is continuous} \iff f^{-1}(O) \text{ is open } \forall O \subset Y \text{ open}$$

(iii)

$$f \text{ is continuous} \iff f^{-1}(C) \text{ is closed } \forall C \subset Y \text{ closed}$$

Proof. We will prove (i). Let U be a neighbourhood of $f(x)$, i.e.

$$\exists \epsilon > 0 : \quad B_\epsilon(f(x)) \subset U \tag{5.43}$$

Since f is continuous

$$\exists \delta > 0 : \quad f(B_\delta(x)) \subset B_\epsilon(f(x)) \tag{5.44}$$

which in turn means

$$B_\delta(x) \subset f^{-1}(B_\epsilon(f(x))) \subset f^{-1}(U) \tag{5.45}$$

so $f^{-1}(U)$ is a neighbourhood of $f(x)$. Now let $\epsilon > 0$. Since $B_\epsilon(f(x))$ is a neighbourhood of $f(x)$, $f^{-1}(B_\epsilon(f(x)))$ is a neighbourhood of x. This means

$$\exists \delta > 0: \quad B_\delta(x) \subset f^{-1}(B_\epsilon(f(x))) \tag{5.46}$$

Thus $f(B_\delta(x)) \subset B_\epsilon(f(x))$ which means f is continuous in x.

(ii) and (iii) are left to the reader. □

Definition 5.58 (Subsequences and (sequential) compactness)**.** Let (X, d) be a metric space, and $(x_n) \subset X$, $(n_k) \subset \mathbb{N}$ are strictly monotonically increasing. Then (x_{n_k}) is said to be a subsequence of (x_n).

A subset $A \subset X$ is said to be (sequentially) compact, if every sequence $(x_n) \subset A$ has a subsequence convergent in A.

Remark 5.59. If (x_n) converges to $x \in X$, then every subsequence of (x_n) converges to x. However, consider

$$(x_n) = (-1)^n$$

This sequence doesn't converge, but the subsequences (x_{2n}) and (x_{2n+1}) converge to (different) values.

Example 5.60. Let $X = \mathbb{R}$, then $(0, 1)$ and $\mathbb{N}$ are not compact. Because

$$(x_n = \frac{1}{n}) \subset (0, 1) \qquad\qquad (x_n = n) \subset \mathbb{N}$$

have no convering subsequences.

Theorem 5.61.

$$A \subset \mathbb{R}^n \textit{ is compact } \iff A \textit{ closed and bounded}$$

Proof. Assume A is not closed, i.e. for $x \in \partial A \setminus A$

$$\exists\, (x_n) \subset A \text{ with } x_n \longrightarrow x \tag{5.47}$$

Every subequence of (x_n) converges to x, but $x \neq A$. From this follows that A is not compact. Assume A is not bounded, i.e. $A \setminus B_n(0) \neq \varnothing \ \ \forall n \in \mathbb{N}$. Now choose $(x_n) \subset A$ such that $\|(x_n)\| \geq n$. (x_n) cannot have a convergent subsequence, because on the one hand for (x_{n_k}) convergent to x we have

$\|x_{n_k}\| \to \|x\|$, but on the other hand $\|x_{n_k}\| \geq n_k \longrightarrow \infty$. This proves the " $\Longrightarrow$ " direction, to prove the inverse, consider the case $n = 1$: Let $A \subset \mathbb{R}$ be bounded and closed. Then

$$\exists K > 0 : \quad A \subset I_1 = [-K, K] \tag{5.48}$$

Let $(x_n) \subset A$ be a sequence. We recursively define more intervals. Let $I_k = [a, b)$ such that $x_n \in I_k$ for infinitely many $n \in \mathbb{N}$. Half the interval:

$$I_{k+1} = \left[a, \frac{b-a}{2}\right) \qquad \text{or} \qquad I_{k+1} = \left[\frac{b-a}{2}, b\right) \tag{5.49a}$$

such that $x_n \in I_{k+1}$ for infinitely many $n \in \mathbb{N}$. By doing this we are creating a sequence of nested intervals of length $K \cdot 2^{-k+2}$. Now set $n_1 = 1$, and then recursively define

$$n_{k+1} > \max\{n_1, \cdots, n_k\} \text{ and } x_{n_{k+1}} \in I_{k+1} \tag{5.50}$$

We now need to show that (x_{n_k}) is convergent. Apply the Cauchy criterion: For $l > k$ we know that x_{n_k} and $x_{n_l} \in I_k$, i.e.

$$|x_{n_k} - x_{n_l}| \leq K \cdot 2^{-k+2} \xrightarrow{k \to \infty} 0 \tag{5.51}$$

This means, x_{n_k} is a Cauchy sequence, so it converges to $x \in \mathbb{R}$. Since A is closed, we have $x \in A$. □

Theorem 5.62. *Continuous mappings map compact sets to compact sets.*

Proof. Let $f : X \to Y$ be continuous and $A \subset X$ compact. Let $(x_n) \subset f(A)$. We need to show that (x_n) has a convergent subsequence. We know that

$$\exists (y_n) \subset A : \quad x_n = f(y_n) \tag{5.52}$$

Since A is compact, there must be subsequences (y_{n_k}) with $y_{n_k} \xrightarrow{k \to \infty} y \in A$. Because of the continuity of f, we have

$$\underbrace{f(y_{n_k})}_{x_{n_k}} \longrightarrow f(y) \in f(A) \tag{5.53}$$

Thus, $f(A)$ is compact. □

Remark 5.63. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a continuous mapping. f maps closed, bounded sets to closed, bounded sets. In general, closed sets are NOT mapped to closed sets, and bounded sets are NOT mapped to bounded sets.

Example: $f : (0, \infty) \to \mathbb{R}, \quad x \mapsto x^{-1}$

$$f(\underbrace{(0,1)}_{\text{bounded}}) = \underbrace{(1,\infty)}_{\text{unbounded}} \qquad\qquad f(\underbrace{[1,\infty]}_{\text{closed}}) = \underbrace{(0,1]}_{\text{not closed}}$$

Corollary 5.64. *Let $A \subset \mathbb{R}^n$ be compact and $f : A \to \mathbb{R}$ continuous. Then f assumes its maximum on A. I.e.*

$$\exists x \in A : \quad f(y) \leq f(x) \quad \forall y \in A$$

Proof. $f(A)$ is compact, so it's closed and bounded. We want to show that compact subsets K of $\mathbb{R}$ have a maximum $M := \sup K$ such that $x_n \longrightarrow M$. Since K is closed we know that $M \in \mathbb{K}$, so M is a maximum. Especially, $\exists z \in f(A)$ maximum and $\exists x \in A$ with $f(x) = z$ □

Theorem 5.65. *Let $A \subset \mathbb{R}^n, B \subset \mathbb{R}^m$ be compact subsets and $f : A \to B$ a bijective, continuous mapping. Then f^{-1} is also continuous.*

Proof. Define $g := f^{-1}$. g is also bijective and maps $B \to A$. Let $C \subset A$ be closed. Since A is bounded, C is also bounded. Thus, $f(C)$ is also compact (i.e. bounded and closed), and we have

$$\begin{aligned} f(C) &= \{f(x) \in B \,|\, x \in C\} \\ &= \{f(g(y)) \in B \,|\, g(y) \in C\} \\ &= \{y \in B \,|\, g(y) \in C\} = g^{-1}(C) \end{aligned} \tag{5.54}$$

So $g^{-1}(C)$ is bounded, and since C was an arbitrary closed set, g is also continuous. □

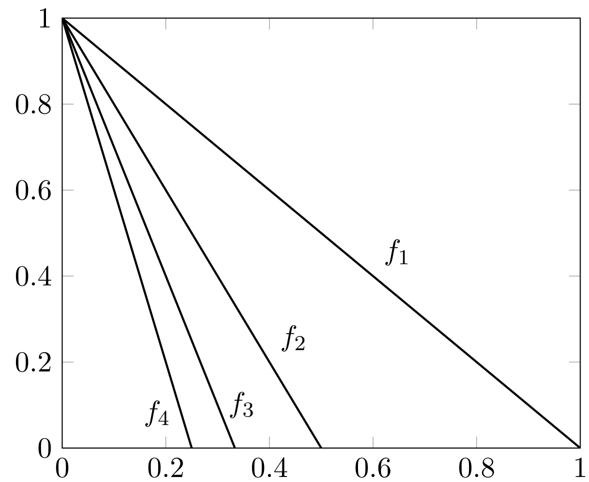
5.5 Convergence of Function sequences

Definition 5.66 (Pointwise convergence)**.** Let M be a set, $f_n : M \to \mathbb{K} \quad \forall n \in \mathbb{N}$ and $f : M \to \mathbb{K}$. The sequence (f_n) is said to be pointwise convergent to f if

$$\lim_{n\to\infty} f_n(x) = f(x) \quad \forall x \in M$$

Example 5.67. Consider

$$\begin{aligned} f_n : [0,1] &\longrightarrow \mathbb{R} \\ x &\longmapsto \begin{cases} 1 - nx, & x \in [0, \frac{1}{n}] \\ 0, & \text{else} \end{cases} \end{aligned}$$



The f_n are continuous for all $n \in \mathbb{N}$ and converge pointwise to

$$\begin{aligned} f : [0,1] &\longrightarrow \mathbb{R} \\ x &\longmapsto \begin{cases} 1, & x = 0 \\ 0, & x \neq 0 \end{cases} \end{aligned}$$

f is not continuous.

Remark 5.68. Let M be a set. Then

$$B(M) = \{f_n : M \longrightarrow \mathbb{K} \,|\, \exists K \in \mathbb{R} : \quad |f(x)| < K \quad \forall x \in M\}$$

is a linear subspace of the space of all functions $M \to \mathbb{K}$. We can define the supremum norm

$$\begin{aligned} \|\cdot\|_\infty : B(M) &\longrightarrow \mathbb{R} \\ f &\longmapsto \sup_{x \in M} \{|f(x)|\} \end{aligned}$$

Proof. We will now proof that $\|\cdot\|_\infty$ is a norm. It is defined, because

$$\|f\|_\infty = 0 \implies |f(x)| = 0 \;\; \forall x \in M \tag{5.55}$$

This implies

$$f(x) = 0 \;\; \forall x \in M \implies f = 0 \tag{5.56}$$

The triangle inequality is proven by first considering

$$|f(x)| \leq \|f\|_\infty \;\; \forall f \in B(M) \; \forall x \in M \tag{5.57}$$

Let $f, g \in B(M)$, then

$$|f(x) + g(x)| \leq |f(x)| + |g(x)| \leq \|f\|_\infty + \|g\|_\infty \quad \forall x \in M \tag{5.58}$$

Which implies

$$\|f + g\|_\infty = \sup_{x \in M} |f(x) + g(x)| \leq \|f\|_\infty + \|g\|_\infty \tag{5.59}$$

□

Definition 5.69 (Uniform convergence)**.** A sequence of bounded functions (f_n),

$$f_n : M \longrightarrow \mathbb{K}$$

is said to be uniformly convergent to $f : M \to \mathbb{K}$ if its norm converges.

$$\|f_n - f\|_\infty \xrightarrow{n \to \infty} 0$$

Remark 5.70. Formally, pointwise convergence means

$$\forall \epsilon > 0 \; \forall x \in M \; \exists N \in \mathbb{N} \; \forall n \geq N : \quad |f_n(x) - f(x)| < \epsilon$$

and uniform convergence means

$$\forall \epsilon > 0 \; \exists N \in \mathbb{N} \; \forall x \in M \; \forall n \geq N : \quad |f_n(x) - f(x)| < \epsilon$$

Theorem 5.71. *The function space $B(M)$ is complete.*

Proof. Let $(f_n) \subset B(M)$ be a Cauchy sequence in terms of $\|\cdot\|_\infty$. Firstly, we have for some fixed $x \in M$

$$|f_n(x) - f_m(x)| \leq \|f_n - f_m\|_\infty \tag{5.60}$$

Since (f_n) is a Cauchy sequence, $(f_n(x))$ is also a Cauchy sequence in $\mathbb{K}_0$. Because $\mathbb{K}$ is complete, $(f_n(x))$ converges, and we define

$$f(x) = \lim_{n \to \infty} f_n(x) \tag{5.61}$$

thus (f_n) converges pointwise to f. Let $\epsilon > 0$. Then

$$\exists N \in \mathbb{N} : \quad \|f_n \cdot f_m\|_\infty < \epsilon \quad \forall n, m \geq N \tag{5.62}$$

Then $\forall x \in M, \; \forall n, m \geq N$ we have

$$|f_n(x) - f_m(x)| \leq \|f_n - f_m\|_\infty < \epsilon \tag{5.63}$$

We can find the limit for $m \to \infty$

$$|f(x) - f_n(x)| \leq \epsilon \tag{5.64}$$

and

$$\|f\|_\infty = \sup_{x \in M} |f| \leq \sup_{x \in M} |f(x) - f_n(x)| + \sup_{x \in M} |f_n(x)| = \epsilon + \|f_n\|_\infty \tag{5.65}$$

Thus, f is bounded. Furthermore

$$\|f - f_n\|_\infty = \sup_{x \in M} |f(x) - f_n(x)| \leq \epsilon \tag{5.66}$$

which in turn implies

$$\|f - f_n\|_\infty \xrightarrow{n \to \infty} 0 \tag{5.67}$$

□

Definition 5.72. Let (X, d) be a metric space, then $C_b(X)$ is said to be the space of all continuous bounded functions.

Remark 5.73. If X is compact (e.g. a bounded, closed subset of $\mathbb{R}^n$) then all continuous functions are bounded. We then write $C(X)$ for $C_b(X)$.

Theorem 5.74. *Let (X, d) be a metric space. $C_b(X)$ is closed in $B(X)$. In other words, every uniformly convergent sequence of continuous functions converges to a continuous function.*

Proof. Let $(f_n) \subset C_b(X)$ be a sequence that uniformly converges to $f \in B(X)$. Let $x \in X$ and $\epsilon > 0$, then

$$\exists N \in \mathbb{N} : \quad \|f - f_n\|_\infty y \frac{\epsilon}{3} \quad \forall n \geq N \tag{5.68}$$

Choose a fixed $n \geq N$. Since f_n is continuous, this means that

$$\exists \delta > 0 : \quad |f_n(x) - f_n(y)| < \frac{\epsilon}{3} \quad \forall y \in B_\delta(x) \tag{5.69}$$

Then we have for all such y

$$\begin{aligned} |f(x) - f(y)| &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| \\ &\leq 2 \cdot \|f - f_n\|_\infty + f_n(x) - f_n(y) < \epsilon \end{aligned} \tag{5.70}$$

This proves the continuity of f in x. Since $x \in X$ was chosen arbitrarily, f is continuous everywhere. □

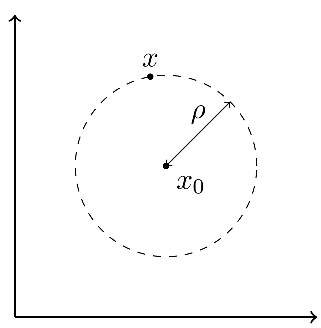
Definition 5.75. Let $x_0 \in \mathbb{K}$ and $(a_n) \subset \mathbb{K}$. Then

$$\sum_{n=1}^{\infty} a_n (x - x_0)^n$$

is called a power series around x_0. The number

$$\rho := \sup \left\{ |x - x_0| \,\middle|\, \sum_{n=1}^{\infty} a_n (x - x_0)^n \text{ converges} \right\}$$

is the convergence radius.



Remark 5.76. All results so far (including proofs) can be extended to $\mathbb{R}^n$-valued functions, or functions with values in a Banach space in general.

Theorem 5.77. *Let $\sum_{n=1}^{\infty} a_n (x - x_0)^n$ be a power series with convergence radius $\rho \in [0, \infty) \cup \{\infty\}$. If $|x - x_0| < \rho$ then the series converges absolutely, for $|x - x_0| > \rho$ it diverges.*

$$\frac{1}{\rho} = \limsup_{n \to \infty} \sqrt[n]{|a_n|}$$

Proof. W.l.o.g. choose $x_0 = 0$: For $|x| > \rho$ the series diverges by definition. If $|x| < \rho$ then there exists $y \in \mathbb{K}$ such that $|x| < |y| \leq \rho$ and $\sum_{n=1}^{\infty} a_n y^n$ convergent. Especially, $(a_n y^n)$ is a null sequence. This means $\exists C > 0$ such that $|a_n y^n| \leq C \quad \forall n \in \mathbb{N}$

$$\sum_{n=1}^{\infty} |a_n x^n| = \sum_{n=1}^{\infty} |a_n y^n| \left| \frac{x}{y} \right|^n \leq C \cdot \sum_{n=1}^{\infty} \left| \frac{x}{y} \right|^n < \infty \tag{5.71}$$

This statement only holds for $\rho > 0$. $\square$

Remark 5.78. (i) We have

$$\rho = \sup \left\{ a \in [0,\infty) \,\middle|\, \sum_{n=1}^{\infty} |a_n| a^n \text{ converges} \right\}$$

(ii) If the following limit exists, then

$$\rho = \lim_{n\to\infty} \frac{|a_n|}{|a_{n+1}|}$$

Example 5.79. The series

$$\sum_{n=1}^{\infty} x^n$$

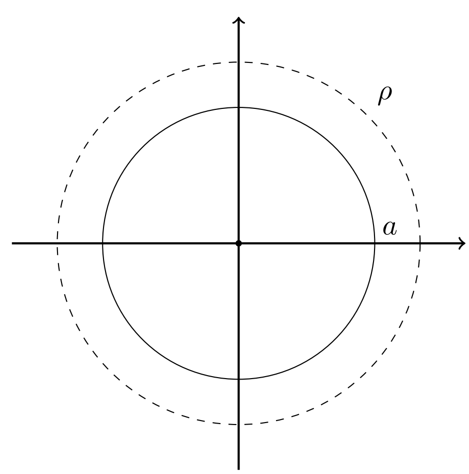
is convergent on $(-1,1)$, so $\rho = 1$. The limit function is

$$x \longmapsto \frac{1}{1-x}$$

Theorem 5.80. *Let* $\sum_{n=1}^{\infty} a_n(x-x_0)^n$ *be a power series with convergence radius* $\rho > 0$*. Let* $0 < a < \rho$*. Then this power series converges uniformly on* $K_a(x_0)$*. Especially*

$$\begin{aligned} f : B_\rho(x_0) &\longrightarrow \mathbb{R} \\ x &\longmapsto \sum_{n=1}^{\infty} a_n(x_n - x_0)^n \end{aligned}$$

Proof. W.l.o.g. choose $x_0 = 0$. Let $0 < a < \rho$. We know that $\sum_{n=1}^{\infty} a_n x^n$ converges on $K_a(0)$.



Define

$$\begin{aligned} f_n : K_a(0) &\longrightarrow \mathbb{K} \\ x &\longmapsto x^n \quad \forall n \in \mathbb{N} \end{aligned} \tag{5.72}$$

We can see that

$$\|f\|_\infty = \sup_{x \in K_a(0)} |f_n| = \sup_{x \in K_a(0)} = a^n \tag{5.73}$$

and thus

$$\sum_{n=1}^\infty a_n f_n \implies \sum_{n=1}^\infty \|a_n f_n\|_\infty = \sum_{n=1}^\infty |a_n|^n < \infty \tag{5.74}$$

because $a < \rho$. The series $\sum_{n=1}^\infty a_n f_n$ is absolutely convergent in $C(K_a(0))$. Since $C(K_a(0))$ is complete, $\sum_{n=1}^\infty a_n f_n$ is convergent because the partial sums $\sum_{n=1}^N a_n f_n$ are continuous $\forall N \in \mathbb{N}$. Therefore f is also continuous on $K_a(0)$. Let $x \in B_\rho(0)$. Then there exists some $a > 0$ such that $|x| < a < \rho$. Thus, f is continuous on $K_a(0)$. Since $K_a(0)$ contains a neighbourhood of x, and continuity is a local property, f is also continuous in x. Because $x \in B_\rho(0)$ was chosen arbitrarily, f is continuous. $\square$

Remark 5.81. exp, sin, cos are continuous.

Example 5.82. The statements above can be extended to Banach space-valued power series (e.g. matrix-valued functions). The norm on $\mathbb{R}^{n \times n}$ is

$$\|A\| = \sup\{\|Ax\| \mid \forall x \in B_1(0)\}$$

Define

$$\exp(A) := \sum_{0=1}^\infty \frac{A^n}{n!}$$

This converges $\forall A \in \mathbb{R}^{n \times n}$, because

$$\begin{aligned} \sum_{n=1}^\infty \left\| \frac{A^n}{n!} \right\| &= \sum_{n=1}^\infty \frac{1}{n!} \|A^n\| \le \sum_{n=1}^\infty \frac{1}{n!} \|A\|^n \\ &= \exp(\|A\|) < \infty \end{aligned}$$

Thus, $\sum_{n=1}^\infty \frac{A^n}{n!}$ converges absolutely. Now consider the function

$$\begin{aligned} \mathbb{R} &\longrightarrow \mathbb{R}^{n \times n} \\ t &\longmapsto \exp(At) \end{aligned}$$

This is a matrix-valued power series

$$\exp(At) = \sum_{n=1}^{\infty} \frac{(At)^n}{n!} = \sum_{n=1}^{\infty} \frac{A^n}{n!} t^n$$

with a convergence radius of $\rho = \infty$. In this case $\exp(A + B)$ doesn't necessarily have to equal $\exp(A) \cdot \exp(B)$.

Chapter 6

Multivariable Calculus

6.1 Partial and Total Differentiability

Definition 6.1. Let $U \subset \mathbb{R}^n$ be open, $x \in (x_1, \cdots, x_n) \in U$ and define the function $f : U \to \mathbb{R}^m$. The mapping f is said to be partially differentiable in x in terms of x_i if

$$t \longmapsto f(x_1, \cdots, x_{i-1}, t, x_{i+1}, \cdots, x_n)$$

is differentiable in x_i, i.e.

$$\partial_i f(x) = \lim_{h \to 0} \frac{f(x_1, \cdots, x_{i-1}, x_i + h, x_{i+1}, \cdots, x_n) - f(x_1, \cdots, x_n)}{h}$$

exists. $\partial_i f(x)$ is said to be the partial derivative of f in x in terms of x_i. Another notation is

$$\frac{\partial f}{\partial x_i}$$

This mapping is said to be partially differentiable in x if it is partially differentiable in terms of $x_i \ \ \forall i \in \{1, \cdots, n\}$.

Example 6.2. Consider

$$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto \begin{cases} 1, & x = 0 \vee y = 0 \\ 0, & \text{else} \end{cases} \end{aligned}$$

f is partially differentiable in $(0, 0)$, but not continuous.

Theorem 6.3. *Let $U \subset \mathbb{R}$ be open, $x \in U$ and $f : U \to \mathbb{K}$.*

$$f \text{ is differentiable in } x$$
$$\Longleftrightarrow$$
$$\exists a \in \mathbb{K}, \phi : U \to \mathbb{K} : \quad f(y) = f(x) + a(y-x) + \phi(y) \quad \forall y \in U$$

and

$$\lim_{y \to x} \frac{\phi(x)}{|y-x|} = 0$$

Proof. We will first prove the " $\Longleftarrow$ " direction. So let a, ϕ be as demanded in the theorem. Then

$$\frac{f(y)-f(x)}{y-x} = a + \frac{\phi(y)}{|y-x|} \cdot \frac{|y-x|}{y-x} \xrightarrow{y \to x} a \tag{6.1}$$

which means f is differentiable in x and $f'(x) = a$. Now let f be differentiable, and set

$$\phi(y) = f(y) - f(x) - f'(x)(y-x) \tag{6.2}$$

Which is equivalent to the equation in the theorem, with $a = f'(x)$. Then

$$\lim_{y \to x} \frac{\phi(x)}{|y-x|} = \left(\frac{f(y)-f(x)}{y-x} - f'(x) \right) \cdot \frac{y-x}{|y-x|} = 0 \tag{6.3}$$

$\square$

Definition 6.4. Let $U \subset \mathbb{R}^n$, $x \in U$ and $f : U \to \mathbb{R}^m$. f is said to be (totally) differentiable in x if a matrix $A \in \mathbb{R}^{m \times n}$ and a mapping $\phi : U \to \mathbb{R}^m$ exist, such that

$$f(y) = f(x) + A(y-x) + \phi(x) \quad \forall y \in U$$

and

$$\lim_{y \to x} \frac{\phi(y)}{\|y-x\|} = 0$$

f is said to be (totally) differentiable if it is (totally) differentiable in every point $x \in U$.

Theorem 6.5. *Let $U \subset \mathbb{R}^n$ be open, $x \in U$ and $f : U \to \mathbb{R}^m$ with*

$$f = (f_1, \cdots, f_m), \quad f_1, \cdots, f_m : U \longrightarrow \mathbb{R}$$

If f is totally differentiable in x, then it is partially differentiable as well, and the matrix A is given by

$$a_{ji} = \partial_i f_j(x)$$

Proof. Let A, ϕ be as demanded above. Let $e_1, \cdots, e_n$ be the canonical basis for $\mathbb{R}^n$. We insert $y = x + he_i$ and receive

$$f(x + he_i) = f(x) + h \cdot Ae_i + \phi(x + he_i) \tag{6.4}$$

By rearranging this yields

$$\frac{f(x + he_i) - f(x)}{h} = Ae_i + \frac{\phi(x + he_i)}{|h|} \cdot \frac{|h|}{h} \xrightarrow{h \to 0} Ae_i \tag{6.5}$$

Thus, f is partially differentiable in x in terms of x_i with $\partial_i f(x) = Ae_i$. $\square$

Definition 6.6. The matrix $(\partial_i f_j(x))_{ij}$ is called the Jacobian matrix of f in x. We write $Df(x)$. If f is totally differentiable, then $Df(x)$ is said to be the (total) derivative of f in x.

For $m = 1$ (so $f : \mathbb{R}^n \to \mathbb{R}$), the Jacobian matrix has one column, and we call it gradient

$$Df(x) =: \boldsymbol{\nabla} f(x)$$

Note: I will adhere to the physical notation of the gradient, using the Nabla operator ∇.

Example 6.7. Let $A \in \mathbb{R}^{m \times n}$ and define

$$\begin{aligned} f_A : \mathbb{R}^n &\longrightarrow \mathbb{R}^m \\ x &\longmapsto Ax \end{aligned}$$

Then we have

$$f_A(y) = Ay = Ax + A(y - x) = f_A(x) - f_A(y - x)$$

Thus, f_A is differentiable ($\phi = 0$) and the derivative is

$$Df_A(x) = A \quad \forall x \in \mathbb{R}^n$$

For another example, let

$$\begin{aligned} f : (0, \infty) \times (0, 2\pi) &\longrightarrow \mathbb{R}^2 \\ (r, \phi) &\longmapsto (r \cos \phi, r \sin \phi) \end{aligned}$$

Then f is partially differentiable.

$$Df(r, \phi) = \begin{pmatrix} \cos \phi & -r \sin \phi \\ \sin \phi & r \cos \phi \end{pmatrix}$$

So f is also totally differentiable (We'll get back to this later).

Remark 6.8. (i) Let $U \subset \mathbb{R}^n$ be open and $f : U \to \mathbb{R}^m$ differentiable, then the derivative Df is a function $U \to \mathbb{R}^{m \times n}$

(ii) Total differentiability is also called local linear approximation. Linearity is the property

$$A(x + \lambda y) = Ax + \lambda Ay \quad \forall x, y \in \mathbb{R}^n \ \lambda \in \mathbb{R}$$

(iii) For arbitrary vector spaces V, W, a mapping $V \to W$ is said to be linear if

$$A(x + \lambda y) = Ax + \lambda Ay \quad \forall x, y \in \mathbb{R}^n \ \lambda \in \mathbb{R}$$

So we can analogously define differentiability for mappings $f : V \to W$ between arbitrary normed vector spaces.

(iv) f is totally differentiable in x if and only if the Jacobian matrix exists and

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

(v) Let $f = (f_1, \cdots, f_m)$ with $f_1, \cdots, f_m : U \to \mathbb{R}$.

$$f \text{ totally differentiable} \iff f_i \text{ totally differentiable} \quad \forall i \in \{1, \cdots, n\}$$

The Jacobian matrix $Df_i(x)$ is the i-th row of $Df(x)$.

(vi) Total differentiability implies continuity.

(vii) Partial and total differentiability are local properties.

(viii) The mapping $h \mapsto Df(x) \cdot h$ is linear.

(ix) The derivative $x \mapsto Df(x)$ is not linear in general.

Theorem 6.9 (Chain rule)**.** *Let $U \subset \mathbb{R}^n$ be open, $V \subset \mathbb{R}^m$ open, $x \in U$, $g : U \to V$ differentiable in x, and $f : V \to \mathbb{R}^k$ differentiable in $g(x)$. Then $f \circ g$ is differentiable and*

$$D(f \circ g) = Df(g(x)) \cdot Dg(x)$$

Proof. Differentiability of g in x means

$$\exists \phi_g : U \longrightarrow \mathbb{R}^m : \quad g(y) - g(x) = D_g(x)(y - x) + \phi_g(y) \tag{6.6}$$

Differentiability of f in $g(x)$ means

$$\exists \phi_f : V \to \mathbb{R}^k :: \lim_{z \to g(x)} \phi_f(z) \|z - g(x)\|^{-1} = 0 \tag{6.7}$$

and

$$f(z) = f(g(x)) + D_f(g(x))(z - g(x)) + \phi_f(z) \tag{6.8}$$

Now set $z = g(y)$, then

$$\begin{aligned} \underbrace{f(g(y))}_{(f \circ g)(y)} = \underbrace{f(g(x))}_{(f \circ g)(x)} &+ D_f(g(x)) \cdot D_g(x)(y - x) \\ &+ (D_f(g(x))\phi_g(y) + \phi_f(g(y))) \end{aligned} \tag{6.9}$$

And we finally need to show

$$\frac{D_f(g(x))\phi_g(y) + \phi_f(g(y))}{\|y - x\|} \xrightarrow{y \to x} 0 \tag{6.10}$$

We know that

$$Df(g(x)) \frac{\phi_g(y)}{\|y - x\|} \longrightarrow 0 \tag{6.11}$$

because

$$z \longmapsto Df(g(x))z \text{ linear and thus continuous} \tag{6.12}$$

We define a new mapping

$$\begin{aligned} \psi : U &\longrightarrow \mathbb{R} \\ z &\longmapsto \begin{cases} \phi_f(z) - \|z - g(x)\|^{-1}, & z \neq g(x) \\ 0, & z = g(x) \end{cases} \end{aligned} \tag{6.13}$$

ψ is continuous in $g(x)$. Then $\forall y \in U$ we have

$$\frac{\phi_f(g(y))}{\|y - x\|} = \underbrace{\psi(g(y))}_{\xrightarrow{y \to x} 0} \cdot \frac{\|g(y) - g(x)\|}{\|y - x\|} \tag{6.14}$$

and

$$\begin{aligned} \frac{\|g(y) - g(x)\|}{\|y - x\|} &= \left\| Dg(x) \frac{y - x}{\|y - x\|} + \frac{\phi_g(y)}{\|y - x\|} \right\| \\ &\leq \underbrace{\left\| Dg(x) \frac{y - x}{\|y - x\|} \right\|}_{\leq \|Dg(x)\|} + \underbrace{\left\| \frac{\phi_g(y)}{\|y - x\|} \right\|}_{\xrightarrow{y \to x} 0} \end{aligned} \tag{6.15}$$

thus ψ is bounded.

$$\implies \psi(g(y)) \cdot \frac{\|g(y) - g(x)\|}{\|y - x\|} \longrightarrow 0 \tag{6.16}$$

□

Theorem 6.10. *Let $U \subset \mathbb{R}^n$ and $f : U \longrightarrow \mathbb{R}^m$. If $\forall x \in U$ the partial derivatives $\partial_i f(x)$ exist and are continuous $\forall i \in \{1, \cdots, n\}$. then f is totally differentiable.*

Proof. Without proof. □

Definition 6.11. Let $U \subset \mathbb{R}^n$ be open. $f : U \to \mathbb{R}^m$ is said to be continuously differentiable if all partial derivatives exist and are continuous. The vector space of all such functions is denoted as $C^1(U, \mathbb{R}^m)$, or in the special case $m = 1$ as $C^1(U)$.

Example 6.12.

1. Coming back to a previous example, we consider

$$Df(r, \phi) = \begin{pmatrix} \cos\phi & -r\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$$

Thus, f is continuously differentiable, and therefore totally differentiable.

2. Let $N \in \mathbb{N}$ and $c_\eta \in \mathbb{K}$ for every multiindex $\eta \in \mathbb{N}_0^n$ with $|\eta| \leq N$. Then the polynomial

$$\begin{aligned} P : \mathbb{R}^n &\longrightarrow \mathbb{K} \\ x &\longmapsto \sum_{\substack{\eta \\ |\eta| \leq N}} c_\eta x^\eta \end{aligned}$$

is continuously differentiable, and therefore totally differentiable.

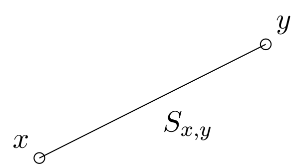
$$\begin{aligned} \partial_i x^\eta &= \partial_i \left(x_1^{\eta_1}, x_2^{\eta_2}, \cdots, x_n^{\eta_n} \right) \\ &= \eta_i x_1^{\eta_1} \cdots x_{i-1}^{\eta_{i-1}} x_i^{\eta_{i-1}} x_{i+1}^{\eta_{i+1}} \cdots x_n^{\eta_n} \end{aligned}$$

This is another polynomial, and therefore continuous.

We introduce the following new notation, for $x, y \in \mathbb{R}^n$:

$$\begin{aligned} S_{x,y} &:= \{x + t(y - x) \mid t \in (0, 1)\} \\ \overline{S_{x,y}} &:= \{x + t(y - x) \mid t \in [0, 1]\} \end{aligned}$$

They denote the connecting line between x and y.



Theorem 6.13 (Intermediate value theorem for $\mathbb{R}$-valued functions)**.** *Let $U \subset \mathbb{R}^n$ be open, $x, y \in U$ and $\overline{S_{x,y}} \subset U$. Now let $f : U \to \mathbb{R}$ differentiable on $S_{x,y}$ and continuous in x, y. Then*

$$\exists \xi \in \overline{S_{x,y}} : \quad f(y) - f(x) = Df(\xi)(y - x)$$

Proof. Consider

$$\begin{aligned} g : [0,1] &\longrightarrow \mathbb{R} \\ t &\longmapsto f(x + t(y - x)) \end{aligned} \tag{6.17}$$

Apply the one dimensional intermediate value theorem. Due to the chain rule, g fulfils the prerequisites. $\exists \theta \in (0, 1)$ such that

$$f(y) - f(x) = g(1) - g(0) = g(\theta) = Df(x + \theta(y - x))(y - x) \tag{6.18}$$

For $\xi = x + \theta(y - x)$ follows the initial statement. $\square$

Theorem 6.14 (Intermediate value theorem)**.** *Let $U \subset \mathbb{R}^n$ be open, $\overline{S_{x,y}} \subset U$ and $f : U \to \mathbb{R}^m$ differentiable on $S_{x,y}$ and continuous in x, y. Then*

$$\exists \xi \in S_{x,y} : \quad \|f(y) - f(x)\| \leq \|Df(\xi)(y - x)\|$$

Proof. For $a \in \mathbb{R}^m$, consider the (real) helper function

$$a^T f(x) = \langle a, f(x) \rangle \tag{6.19}$$

According to the previous theorem

$$\exists \xi \in B_\epsilon : \quad a^T f(y) - a^T f(x) = a^T Df(\xi)(y - x) \tag{6.20}$$

In this implication the chain rule has been applied. We can rewrite this using the scalar product

$$\begin{aligned} \|f(y) - f(x)\|^2 &= |\langle f(y) - f(x), Df(\xi)(y - x) \rangle| \\ &\leq \|f(y) - f(x)\| \|Df(\xi)(y - x)\| \end{aligned} \tag{6.21}$$

$\square$

Corollary 6.15. *Let $U \subset \mathbb{R}^n$ be open and $f : U \to \mathbb{R}^m$ a differentiable function.*

$$Df = 0 \text{ on } U \implies \exists V \subset U : f \text{ constant on } V$$

Proof. Let $x \in U$, choose $\epsilon > 0$ such that $B_\epsilon(x) \subset U$. Then

$$\forall y \in B_\epsilon(x)\ \exists \xi \in S_{x,y} : \quad \|f(y) - f(x)\| \le \|Df(\xi)(y-x)\| = 0 \tag{6.22}$$

This implies

$$\|f(y) - f(x)\| = 0 \implies f(y) = f(x) \quad \forall y \in B_\epsilon(x) \tag{6.23}$$

$\square$

Remark 6.16. Functions with vanishing derivatives must be constant. Consider

$$\begin{aligned} f : (-2,-1) \cup (1,2) &\longrightarrow \\ x &\longmapsto \begin{cases} -1, & x < 0 \\ 1, & x > 0 \end{cases} \end{aligned}$$

Local constancy implies constancy on connected sets.

6.2 Higher Derivatives

Definition 6.17. Let $U \subset \mathbb{R}^n$ and let f be (the only) partial derivative of order 0. Now define recursively

(i) f is said to be $(k+1)$-times partially differentiable if all partial derivatives of order k are partially differentiable.

(ii) The partial derivatives of order $(k+1)$ are the functions $\partial_i g \quad i \in \{1, \cdots, n\}$ where g is the partial derivative of order k of f.

The k-th partial derivative in terms of i of f is denoted as

$$\partial_i^k f$$

f is said to be k-times continuously differentiable if all partial derivatives of order k are continuous. $C^k(U, \mathbb{R}^m)$ is the vector space of all k-times continuously differentiable functions.

f is said to be infinitely differentiable (or smooth) is it is k-times differentiable $\forall k \in \mathbb{N}$, and the vector space of all infinitely differentiable functions is denoted as $C^\infty(U, \mathbb{R}^m)$.

For total differentiability we have

$$f : \mathbb{R}^n \longrightarrow \mathbb{R}^m \qquad\qquad Df : \mathbb{R}^m \longrightarrow \mathbb{R}^{m\times n}$$

Remark 6.18. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be sufficiently often differentiable. Consider for $u \in \mathbb{R}^n$

$$x \longmapsto Df(x)u = \underbrace{\lim_{k\to 0} \frac{f(x+hu) - f(x)}{h}}_{\text{Directional derivative along } u}$$

Now consider for fixed x

$$\begin{aligned} D^2 f(x) : \mathbb{R}^n \times \mathbb{R}^n &\longrightarrow \mathbb{R}^m \\ (u, v) &\longmapsto D(Df(\cdot)u)(x)v \end{aligned}$$

$D^2 f(x)$ is linear in v and u, and

$$\begin{aligned} D^2 f(x)(u_1 + \lambda u_2, v) &= D(Df(\cdot)(u_1 + \lambda u_2))(x)v \\ &= D(Df(\cdot)u_1 + \lambda Df(\cdot)u_2)(x)v \\ &= D(Df(\cdot)u_1)(x)v + \lambda D(Df(\cdot)u_2)(x)v \\ &= D^2 f(x)(u_1, v) + \lambda D^2 f(x)(u_2, v) \end{aligned}$$

$D^2 f(x)$ is a bi-linear mapping.

Definition 6.19. Let $U \subset \mathbb{R}^n$ and $f : U \to \mathbb{R}^m$. Define recursively for $k \geq 1$:

(i) f is said to be $(k+1)$ times (totally) differentiable on U, if the term $D^k(\cdot)(u_1, \cdots, u_k)$ is differentiable on $U \forall u_1, \cdots, u_k \in \mathbb{R}^n$.

(ii) The $(k+1)$-th derivative of f in $x \in U$ is the multi-linear mapping

$$\begin{aligned} D^{k+1} f(x) : (\mathbb{R}^n)^{k+1} &\longrightarrow \mathbb{R}^m \\ (u_1, \cdots, u_k, v) &\longmapsto D(D^k f(\cdot)(u_1, \cdots, u_k))(x)v \end{aligned}$$

Remark 6.20. Let $f_1, \cdots, f_m : U \to \mathbb{R}$, then the function

$$\begin{aligned} f : U &\longrightarrow \mathbb{R}^m \\ x &\longmapsto (f_1(x), \cdots, f_m(x)) \end{aligned}$$

is k-times totally differentiable if and only if the $f_1, \cdots, f_n$ are totally differentiable.

$$(D^k f(x)(u_1, \cdots, U_k))_j = D^k f_j(x)(u_1, \cdots, u_k)$$

Remark 6.21. $D^k f(x)$ really is multi-linear (linear in every point) $\forall k \in \mathbb{N}$. Other multi-linear mappings are

(i) The scalar product on $\mathbb{R}^n$

$$\mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$$

(ii) The determinant

$$\mathbb{R}^{n \times n} \longrightarrow \mathbb{R}$$

Remark 6.22. A matrix $A \in \mathbb{R}^{m \times n}$ is uniquely determined by its effect on the canonical basis $e_1, \cdots, e_n$. This means if $v \in \mathbb{R}$, then $\exists \alpha_1, \cdots, a_n \in \mathbb{R}$ that are uniquely determined such that

$$v = \alpha_1, e_1 + \cdots + \alpha_n e_n$$

Then

$$Av = \alpha_1 A e_1 + \cdots + \alpha_n A e_n$$

Ae_i is the i-th column of A. An analogous statement for multi-linear mappings would be, that

$$A : \mathbb{R}^{n \times k} \longrightarrow \mathbb{R}^m$$

is uniquely determined if $A(e_{i_1}, e_{i_2}, \cdots, e_{i_k})$ known $\forall i_1, \cdots, i_k \in \{1, \cdots, n\}$.

Theorem 6.23. *Let $U \subset \mathbb{R}^n$ be open, $f : U \to \mathbb{R}^m$ k-times differentiable in x and let $e_1, \cdots, e_n$ be the canonical basis of $\mathbb{R}^n$. Then*

$$D^k f(x)(e_{i_1}, \cdots, e_{i_k}) = \partial_{i_k} \cdots \partial_{i_1} f(x)$$

$\forall i_i, \cdots, i_k \in \{1, \cdots, n\}$.

Proof. For $k = 1$ this is already proven. So we can use proof by induction; assume the statement holds for a k, i.e. $\forall i_1, \cdots, i_k \in \{1, \cdots, k\}$

$$D^k f(x)(e_{i_1}, \cdots, e_{i_k}) = \partial_{i_k} \cdots \partial_{i_1} f(x)$$

Then for $i_1, \cdots, i_k, i_{k+1} \in \{1, \cdots, n\}$

$$\begin{aligned} D^{k+1} f(x)(e_{i_1, \cdots, e_{i_k}}) &= D(D^k f(\cdots)(e_{i_1}, \cdots, e_{i_k}))(x) \cdot e_{i_{k+1}} \\ &= D(\partial_{i_k}, \cdots \partial_{i_1} f(\cdot))(x) e_{i_{k+1}} \\ &= \partial_{i_{k+1}} \partial_{i_k} \cdots \partial_{i_1} f(x) \end{aligned} \tag{6.24}$$

The order in which partial derivatives are applied is important! $\square$

Example 6.24. Consider

$$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x_1, x_2) &\longmapsto x_1^2 \cos(x_2) \end{aligned}$$

Then we can calculate

$$D^2 f(x)(u, v) \quad u = u_1 e_1 + u_2 e_2, v = v_1 e_1 + v_2 e_2$$

As follows

$$\begin{aligned} D^2 f(x)(u, v) &= u_1 v_1 D^2 f(x)(e_1, e_1) + u_1 v_2 D^2 f(x)(e_1, e_2) \\ &\quad + u_2 v_1 D^2 f(x)(e^2, e^1) + u_2 v_2 D^2 f(x)(e^2, e^2) \\ &= u_1 v_1 \cdot 2 \cdot \cos(x_2) - 2x_1 \sin(x_2) u_1 v_2 \\ &\quad - 2x_1 \sin(x_2) v_1 u_2 - x_1^2 \cos(x_2) u_2 v_2 \end{aligned}$$

Theorem 6.25. *Let $U \subset \mathbb{R}^n$ be open, and $f : U \to \mathbb{R}^m$ k-times continuously differentiable. Then f is k-times totally differentiable.*

Proof. This is already proveb for $k = 1$. So we can use induction over k; assume the statement is correct for $k \in \mathbb{N}$. Let $u_1, \cdots, u_k \in \mathbb{R}^n$, then $D^k f(\cdot)(u_1, \cdots, u_k)$ is a linear combination of the partial derivative of f of order k, and is thus continuously differentiable once more. Therefore $D^2 f(\cdot)(u_1, \cdots, u_k)$ is totally differentiable, and thus f is $(k+1)$-times totally differentiable. □

Theorem 6.26 (Theorem of Schwarz)**.** *Let $U \subset \mathbb{R}^n$ be open, and also $f \in C^2(U, \mathbb{R}^m)$. Then*

$$\forall x \in U \ \forall u, v \in \mathbb{R}^n : \quad D^2 f(x)(u, v) = D^2 f(x)(v, u)$$

and

$$\forall x \in U \ \forall i_1, i_2 \in \{1, \cdots, n\} : \quad \partial_{i_1} \partial_{i_2} f(x) = \partial_{i_2} partial_{i_1} f(x)$$

Proof. Let $m = 1$, $x \in U$, $\epsilon > 0$ such that $B_\epsilon(x) \subset U$. If $u = 0$ or $v = 0$ then both sides of the equation vanish, so let $u, v \in \mathbb{R}^n \setminus \{0\}$ and

$$0 < t < c := \frac{\epsilon}{2 \cdot \max\{\|u\|, \|v\|\}} \tag{6.25}$$

Define the helper function

$$\begin{aligned} g_1 : [0, t] &\longrightarrow \mathbb{R} \\ s &\longmapsto f(x + tv + su) - f(x + su) \end{aligned} \tag{6.26}$$

And apply the one dimensional intermediate value theorem. $\exists \xi \in (0, t)$ such that

$$g_1(t) - g_1(0) = g_1'(\xi) \cdot t = (Df(x + tv + \xi u)u - Df(x + \xi u)u) \cdot t \tag{6.27}$$

Analogously, define and apply the intermediate value theorem to

$$\begin{aligned} g_2 : [0, t] &\longrightarrow \mathbb{R} \\ s &\longmapsto Df(x + sv + \xi u)u \end{aligned} \tag{6.28}$$

and get $\eta \in (0, t)$

$$\begin{aligned} g_2(t) - g_2(0) = g_2'(\eta)t &= D(Df(\cdot)u)(x + \eta v + \xi u)uvt \\ &= D^2 f(x + \eta v + \xi u)(u, v)t \end{aligned} \tag{6.29}$$

using these results, we can get $\xi, \eta \in (0, t)$ for all $t \in (0, c)$ such that

$$\begin{aligned} f(x{+}tv &+ tu) - f(x + tv) - f(x + tu) + f(x) \\ &= g_1(t) - g_1(0) = (Df(x + tv + \xi u)u - Df(x + \xi u)u)t \\ &= (g_2(t) - g_2(0))t = D^2 f(x + \eta v + \xi u)(u, v)t^2 \end{aligned} \tag{6.30}$$

So we can write

$$\begin{aligned} \lim_{t\to 0} &\frac{f(x + tv + tu) - f(x + tv) - f(x + tu) + f(x)}{t^2} \\ &= \lim_{t\to 0} D^2 f \underbrace{(x + \eta v + \xi u)}_{\longrightarrow x}(u, v) \\ &= D^2 f(x)(u, v) \end{aligned} \tag{6.31}$$

The left side is symmetric in terms of swapping u and v, so the right side must be as well. □

Note, that

$$D^2 f(x)(e_{i_1}, e_{i_2}) = \partial_{i_2}\partial_{i_1} f(x) = \partial_{i_1}\partial_{i_2} f(x) = D^2 f(x)(e_{i_2}, e_{i_1})$$

Remark 6.27. Via induction:

(i) $D^k f(x)(u_1, \cdots, u_k)$ is independent from the order of the u_i, if $D^k f$ is continuous.

(ii) The limit of the second derivaative is useful in the numerical discussion of differential equations.

Theorem 6.28 (Taylor's Theorem)**.** *Let $U \subset \mathbb{R}^n$ be open, $f : U \to \mathbb{R}$ be $(l+1)$-times differentiable and $h \in \mathbb{R}^n$ such that $x+th \in U$ $\forall t \in [0,1]$. Then $\exists \theta \in [0,1]$ such that*

$$f(x+h) = \sum_{k=1}^{l} \frac{1}{k!} D^k f(x)(h, \cdots, h) + \frac{1}{(l+1)!} D^{l+1} f(x+\theta h)(h, \cdots, h)$$

Heuristic Proof. Apply the one dimensional Taylor theorem with Lagrange error bound onto a helper function

$$\begin{aligned} g : [0,1] &\longrightarrow \mathbb{R} \\ t &\longmapsto f(x+th) \end{aligned} \tag{6.32}$$

□

Remark 6.29. (i) Consider $h = \sum_{i=1}^n h_i e_i$. Then

$$D^2 f(x)(h,h) = \sum_{i,j=1}^{n} h_i h_j D^2 f(x)(e_i, e_j) = \sum_{i,j=1}^{n} \partial_i \partial_j f(x) h_i h_j$$

(ii) Analogously to one dimension, we can formulate criteria for local extrema:

$$Df(x) = 0, \cdots, D^{l-1} f(x) = 0 \text{ and } D^l f(x) \neq 0$$

- x is a local minimum if l is even and $D^l f(x)$ is positive.
- x is a local maximum if l is even and $D^l f(x)$ is negative.
- x is no local extremum of l is odd or if $D^l f(x)$ is undefined.

Definedness is complicated to determine for $l > 2$.