Mathematics for Physicists

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

July 31, 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.

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.

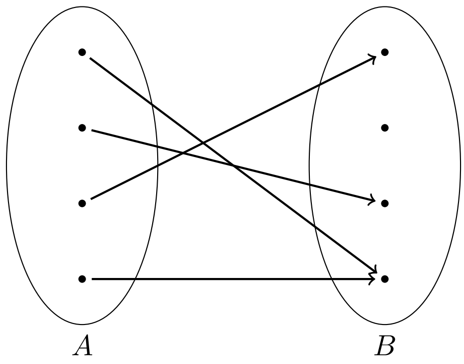


Figure 1.1: A mapping $f : A \to B$

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)

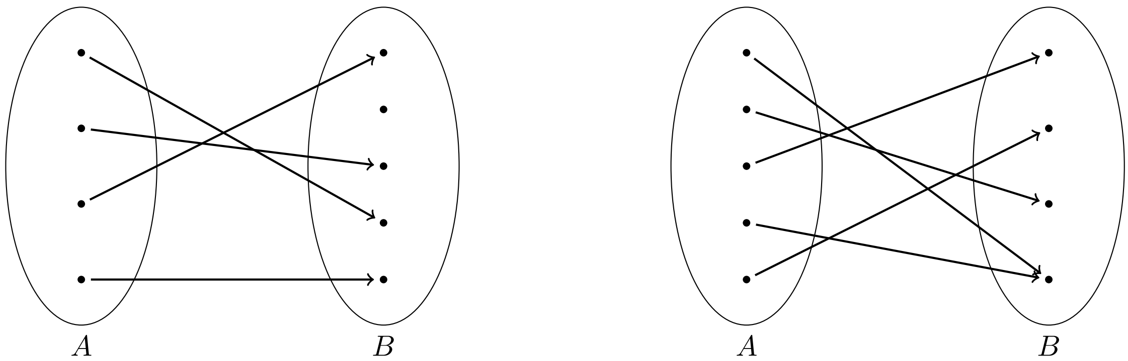
$$f \text{ is a mapping from } A \text{ to } B \iff f(x) \text{ is a value in } B$$

(iv)

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

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

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



(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

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

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

□

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 $b + a = b$.

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

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

□

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

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

O4: Totality

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

O5:

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

O6:

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

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

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

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

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

(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)

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

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

$$\overset{\text{O6}}{\implies} 0 \leq (y-x)x^{-1}y^{-1} \overset{\text{R1}}{=} yx^{-1}y^{-1} - xx^{-1}y^{-1} = x^{-1} - y^{-1} \tag{1.20}$$

$$\overset{\text{O5}}{\implies} y^{-1} \leq x^{-1} \tag{1.21}$$

□

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 $\forall y \in A: \ y \leq x$

(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 $x \leq y$ 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

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

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 $x \in A_i \ \ \forall i \in I$.

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

□

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.

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

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

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

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 $(x,y),(\tilde{x},\tilde{y})\in\mathbb{C}$.

$$(x,y)+(\tilde{x},\tilde{y}):=(x+\tilde{x},y+\tilde{y})$$
$$(x,y)\cdot(\tilde{x},\tilde{y}):=(x\tilde{x}-y\tilde{y},x\tilde{y}+\tilde{x}y)$$

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

$$\Re(z)=\mathrm{Re}(z)=x \quad \text{the real part}$$
$$\Im(z)=\mathrm{Im}(z)=y \quad \text{the imaginary part}$$

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.

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

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 $-|y| \leq y$.

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

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

□

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

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

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 $\{x_n \,|\, n \in \mathbb{N}\}$ 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 $\forall n \geq N$:

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

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

(ii)

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

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

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

This implies that

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

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

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

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

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

□

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*

$$\lim_{n\to\infty} \frac{a_0+a_1n+a_2n^2+\cdots+a_{k-1}n^{k-1}+a_kn^k}{b_0+b_1n+b_2n^2+\cdots+b_{k-1}n^{k-1}+b_kn^k}=\frac{a_k}{b_k}$$

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

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

□

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).

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

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

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

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

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.

$$\begin{aligned} (x_n) \text{ is monotonic increasing } &\iff \frac{x_{n+1}}{x_n} \geq 1 \quad \forall n \in \mathbb{N} \\ (x_n) \text{ is monotonic decreasing } &\iff \frac{x_{n+1}}{x_n} \leq 1 \quad \forall n \in \mathbb{N} \end{aligned}$$

Example 2.27. Consider the following sequence

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

Notice that $0 < x_n \ \ \forall n \in \mathbb{N}$. 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 $x_n^2 \geq a \ \ \forall n \geq 2$, 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 $\frac{1}{a} < 1$

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

$\square$

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

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

$\square$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned} \Longrightarrow |x_n| = |x_n - x_N + x_N| &\leq |x_n - x_N| + |x_N| \\ &\leq |x_N| + 1 \quad \forall n \geq N \end{aligned} \tag{2.62}$$

We will now introduce the two auxiliary sequences

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

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

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

$$\implies y_n = \sup\{x_k | k \geq n\} \leq \sup\{x_k | k \geq \tilde{n}\} = y_{\tilde{n}} \tag{2.65}$$

$$\implies (x_n) \text{ monotonic decreasing and therefore converging to } y \tag{2.66}$$

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. $\square$

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)

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

(ii)

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

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

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

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

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

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

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

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

(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 $\forall n \in \mathbb{N}$

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

$\square$

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 $\cos\left(\frac{\pi}{2}\right) = 0$.

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

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
$+ : V \times V \longrightarrow V$	$\cdot : \mathbb{K} \times V \longrightarrow V$
$(x, y) \longmapsto x + y$	$(\alpha, y) \longmapsto \alpha x$

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:

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

(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 $m = n$. □

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

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

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

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

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

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*

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

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*

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

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*

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

and

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

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

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

□

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 $\operatorname{span}\{z_1, \cdots, z_n\}$, 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: $\forall v_1, v_2, w_1, w_2 \in V, \quad \lambda \in \mathbb{K}$

$$\begin{array}{rl} \text{Linearity} & \langle v_1 | w_1 + \lambda w_2 \rangle = \langle w_1 | w_1 \rangle + \lambda \langle v_1 | w_2 \rangle \\ \text{Conjugated symmetry} & \langle v_1 | w_1 \rangle = \overline{\langle w_1 | v_1 \rangle} \\ \text{Positivity} & \langle v_1 | v_1 \rangle \geq 0 \\ \text{Definedness} & \langle v_1 | v_2 \rangle = 0 \implies v_1 = 0 \\ \text{Conjugated linearity} & \langle v_1 + \lambda v_2 | w_1 \rangle = \langle v_1 | w_1 \rangle + \overline{\lambda} \langle v_2 | w_1 \rangle \end{array}$$

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

$$\left\langle (x_1, x_2, \cdots, x_n)^T \middle| (y_1, y_2, \cdots, y_n)^T \right\rangle = \sum_{k=1}^{n} x_k y_k$$

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 $\mathbb{C}^n$

$$\left\langle (u_1, u_2, \cdots, u_n)^T \middle| (v_1, v_2, \cdots, v_n)^T \right\rangle = \sum_{k=1}^{n} \overline{u_k} v_k$$

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*

$$\begin{array}{rl} \textit{Cauchy-Schwarz-Inequality} & |\langle v | w \rangle| \leq \|v\| \|w\| \\ \textit{Triangle Inequality} & \|v + w\| \leq \|v\| + \|w\| \end{array}$$

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

$$v_1 = (\cos\phi, \sin\phi)^T \qquad\qquad v_2 = (-\sin\phi, \cos\phi)^T$$

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*

$$\langle v|Aw\rangle = \left\langle A^H v\middle|w\right\rangle$$

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

$$\begin{aligned}
\langle v | Aw \rangle &= \sum_{i=1}^{n} \overline{v_i} (Aw)_i = \sum_{i=1}^{n} \left(\overline{v_i} \cdot \sum_{j=1}^{n} A_{ij} w_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} \overline{v_i} w_j \\
&= \sum_{j=1}^{n} \left(\sum_{i=1}^{n} Aij \overline{v_i} \right) w_j = \sum_{j=1}^{n} \left(\overline{\sum_{i=1}^{n} \overline{A_{ij}} v_i} \right) w_j \\
&= \sum_{j=1}^{n} \overline{(A^H v)_j} \cdot w_j \\
&= \left\langle A^H v \middle| w \right\rangle
\end{aligned} \tag{3.55}$$

□

Definition 3.60. A matrix $A \in \mathbb{R}^{n \times n}$ is called orthogonal if

$$A^T A = A A^T = I$$

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 $x, y \in \mathbb{R}^n$

$$\begin{aligned} \langle Ax | Ay \rangle &= \left\langle A^T Ax \middle| y \right\rangle = \langle x | y \rangle \\ \|Ax\| &= \sqrt{\langle Ax | Ax \rangle} = \sqrt{\langle x | x \rangle} = \|x\| \end{aligned}$$

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

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

The eigenspaces are

$$\begin{aligned} E_2 &= \left\{ \kappa \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \middle| \kappa \in \mathbb{R} \right\} \\ E_1 &= \left\{ \kappa \cdot \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + \rho \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \middle| \kappa, \rho \in \mathbb{R} \right\} = \operatorname{span} \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\} \end{aligned}$$

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

$$D^k = (T^{-1}AT)^k = \underbrace{T^{-1}AT \cdot T^{-1}AT \cdot \cdots}_{k \text{ times}} = T^{-1}A^kT$$

$$\implies A^k = TD^kT^{-1}$$

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

$$\lambda \langle v|v\rangle = \langle v|\lambda v\rangle = \langle v|Av\rangle = \left\langle A^H v\middle|v\right\rangle = \langle Av|v\rangle = \langle \lambda v|v\rangle = \overline{\lambda} \langle v|v\rangle \tag{3.63}$$

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

$$\underbrace{(\rho - \lambda)}_{\neq 0} \underbrace{\langle v|w\rangle}_{=0} = 0 \implies v \perp w \tag{3.67}$$

□

75

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 $\overline{D}$.

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. $\overline{\mathbb{Q}} = \mathbb{R}$.

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.

$$x \in \mathbb{R} \textit{ is a boundary point of } D \subset \mathbb{R} \iff \exists\, (x_n) \subset D \textit{ such that } x_n \to x$$

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

$$2 = |f(x) - f(-x)| \leq \underbrace{|f(x) - y|}_{<1} + \underbrace{|y - f(-x)|}_{<1} < 2$$

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.

$$\exists \epsilon > 0\ \forall \delta > 0\ \exists x \in D : \quad |x - x_0| < \delta \land |f(x) - y| \geq \epsilon \tag{4.8}$$

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

$$\frac{1}{1 - x_n} \xrightarrow{n \to \infty} 1$$

$$\implies \lim_{x \to 0} \frac{1}{1 - x} = 1$$

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*

$$\lim_{x \to x_0} f(x) = y \qquad \lim_{x \to x_0} g(x) = z$$

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

$$\lim_{n \to \infty} f(x_n) = y \quad (4.12a) \qquad \lim_{n \to \infty} g(x_n) = z \neq 0 \quad (4.12b)$$

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

$$\lim_{x \to 0}(1+x) = 1 \qquad\qquad \lim_{x \to 0}\left(\frac{1}{1-x}\right) = 1$$

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 $f : D \to \mathbb{K}$.

$$f \text{ is continuous in } x_0 \in D$$
$$\Longleftrightarrow$$
$$\forall\, (x_n) \subset D \text{ with } x_n \to x_0 : \quad \lim_{n\to\infty} f(x_n) = f(x_0)$$

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*

$$f : D \longrightarrow \mathbb{K} \textit{ continuous in } x_0 \in D \tag{4.16a}$$
$$g : f(D) \longrightarrow \mathbb{K} \textit{ continuous in } f(x_0) \tag{4.16b}$$

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

$$f : \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto |x|$$
$$g : \mathbb{R} \longrightarrow \mathbb{R} \setminus \{-1\}, \quad y \longmapsto \frac{1-y}{1+y}$$
$$h : \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \frac{1-|x|}{1+|x|}$$

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 $x_0 \in \mathbb{R}$

$$\begin{aligned} |\underbrace{\exp(i(x_0+h))}_{\exp(ix_0)\dot{\exp}(ih)} - \exp(ih_0)| &= \underbrace{|\exp(ix_0)|}_{1} |\exp(ih) - 1| \\ &\le 1 \cdot \left| \sum_{k=0}^{\infty} \frac{(ih)^k}{k!} - 1 \right| = \left| \sum_{k=1}^{\infty} \frac{(ih)^k}{k!} \right| \\ &\le \sum_{k=1}^{\infty} \left| \frac{(ih)^k}{k!} \right| \\ &= \sum_{k=1}^{\infty} \frac{|h|^k}{k!} = \sum_{k=0}^{\infty} \frac{|h|^k}{k!} - 1 = \exp(|h|) - 1 \end{aligned}$$

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

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

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. $\square$

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. $\square$

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

$$f(x_o - \tilde{epsilon}) \leq x_0 - \delta < y < y_0 + \delta \leq f(x_0 + \tilde{\epsilon}) \tag{4.29}$$

From the strict monotony of g we can conclude

$$x_0 - \tilde{epsilon} < g(y) < x_0 + \tilde{\epsilon} \tag{4.30}$$

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. □

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

$$\begin{aligned} \frac{f(g(x+h)) - f(g(x))}{h} &= \frac{f(g(x) + (g(x+h) - g(x))) - f(g(x))}{h} \\ &= \frac{\phi(g(x+h) - g(x))(g(x+h) - g(x))}{h} \\ &= \underbrace{\phi(g(x+h) - g(x))}_{\xrightarrow{h \to 0} 0} \cdot \underbrace{\frac{g(x+h) - g(x)}{h}}_{\xrightarrow{h \to 0} g'(x)} \\ &\xrightarrow{h \to 0} g'(x) \cdot f'(g(x)) \end{aligned} \tag{4.41}$$

□

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.

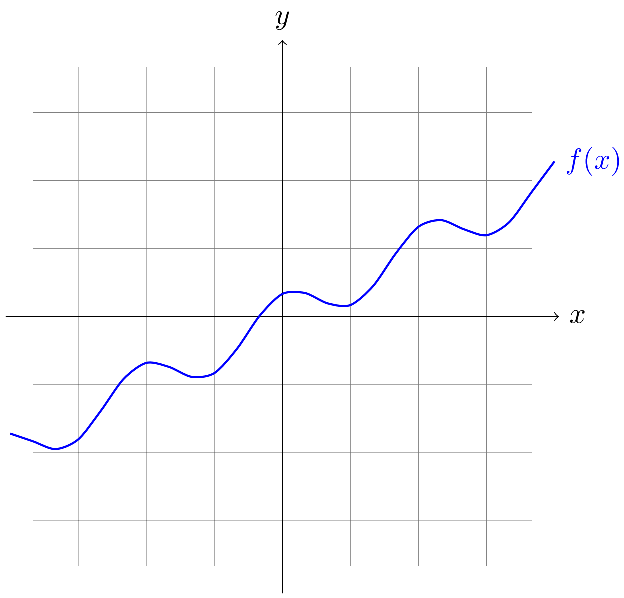
Example 4.39. (i) Let $f : [-1, 1] \to \mathbb{R}$, $f(x) = x^2$.

- $x_0 = 0$ 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*

$$f \textit{ differentiable in } x_0 \implies f'(x_0) = 0$$

Proof. Assume $f'(x_0) \neq 0$ (w.l.o.g. $f'(x_0) > 0$, otherwise consider $-f$). Then

$$\exists \delta > 0 : \quad \left| \frac{f(x_0 + h) - f(x)}{h} - f'(x_0) \right| < f'(x_0) \quad \forall h \in (-\delta, \delta) \tag{4.42}$$

Especially

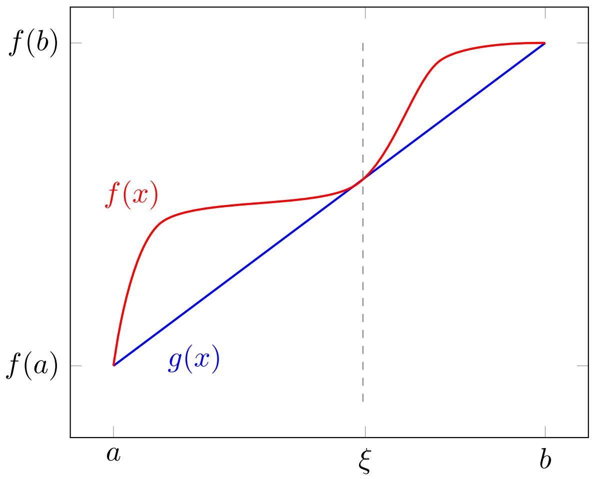
$$0 < \frac{f(x_0 + h) - f(x_0)}{h} \quad \forall h \in (-\delta, \delta) \tag{4.43}$$

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 $g'(x) \neq 0 \quad \forall x \in I \setminus \{x_0\}$ 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. $g(x) \neq 0 \quad \forall x \in I \setminus \{x_0\}$. 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

$$\lim_{x\to 0} \frac{x^2}{1-\cos x} = \frac{“0”}{0}$$

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

$$\lim_{x\to 0} \frac{2x}{\sin x} = \frac{“0”}{0}$$

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

$$\begin{aligned} \frac{f^{-1}(y+h) - f^{-1}(y)}{h} &= \frac{\xi(h)}{f(x+\xi(h)) - f(x)} \\ &= \left(\frac{f(x+\xi(h)) - f(x)}{\xi(h)}\right)^{-1} \\ &\xrightarrow{h \to 0} \left(f'(x)\right)^{-1} = \frac{1}{f'(f^{-1}(y))} > 0 \end{aligned} \tag{4.53}$$

□

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

$$f^{(n)}(x) > 0 \implies x \text{ is a local maximum}$$
$$f^{(n)}(x) < 0 \implies x \text{ is a local minimum}$$

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. □

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)

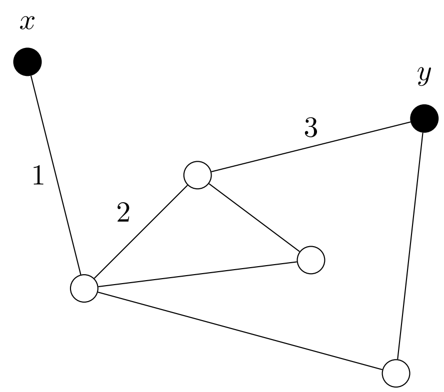
Example 5.2. (i) Let M be a set. Then

$$d(x, y) = \begin{cases} 1, & x \neq y \\ 0, & \text{else} \end{cases}$$

is called the discrete metric.

(ii) Let X be the set of edges of a graph.

$$d(x, y) := \text{Minimum amount of edges that have to be passed to get from } x \text{ to } y$$



(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

$$\{x \in X \mid x \text{ is inner point of } A\}$$

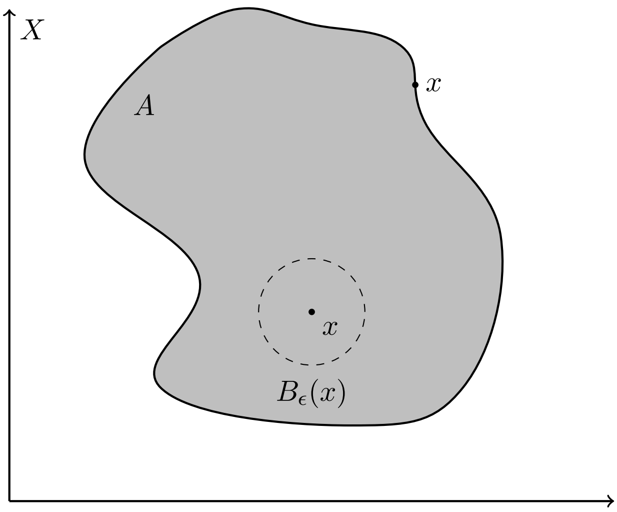
is called the interior of A, and is denoted as $\mathring{A}$.

(iv) The set

$$\{x \in X \mid x \text{ is boundary point of} A\}$$

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 $\overline{A}$.



Example 5.22. Consider $X = \mathbb{R}^2$. Then

$$\begin{aligned} A &= \{(x, y) \in \mathbb{R} \mid 0 \leq y < 1\} \\ \mathring{A} &= \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, 0 \leq y < 1\right\} \\ \partial A &= \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, y = 1 \vee y = 0\right\} \\ \overline{A} &= \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, 0 \leq y \leq 1\right\} \end{aligned}$$

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 $A \subset X$.*

$$A \textit{ open} \iff X \setminus A \textit{ closed}$$

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

$$\begin{aligned} B_r(x) &= \{y \in X \mid d(x,y) < r\} \text{ is open} \\ K_r(x) &= \{y \in X \mid d(x,y) < r\} \text{ is closed} \end{aligned}$$

Remark 5.29. Consider the special case $a, b \in \mathbb{R}$ with $a < b$

$$\begin{aligned} (a,b) &= B_{\frac{b-a}{2}}\left(\frac{a+b}{2}\right) \text{ open} \\ [a,b] &= K_{\frac{b-a}{2}}\left(\frac{a+b}{2}\right) \text{ closed} \end{aligned}$$

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

$$\forall \epsilon > 0\ \exists N \in \mathbb{N}: \quad x_n \in B_\epsilon(x) \quad \forall n \geq N \tag{5.21}$$

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*

$$\mathring{A} \text{ open} \implies \partial A, \overline{A} \text{ closed}$$

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

$$\overline{A} = A \cup \partial A = \mathring{A} \cup \partial A = X \setminus \mathring{B} \text{ closed} \tag{5.28}$$

□

Theorem 5.34. *Let (X, d) be a metric space and $A \subset X$*

$$\bigcup_{\substack{O \text{ open} \\ O \subset A}} O = \mathring{A} \qquad \textit{and} \qquad \bigcap_{\substack{C \text{ closed} \\ A \subset C}} C = \overline{A}$$

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. $\square$

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.

$$f \text{ converges to } y \in Y \text{ in } x \iff \forall\, (x_n) \subset X : \quad f(x_n) \xrightarrow{x_n \to x} y$$

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 $d(f(x_n), y) \geq \epsilon$. □

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*

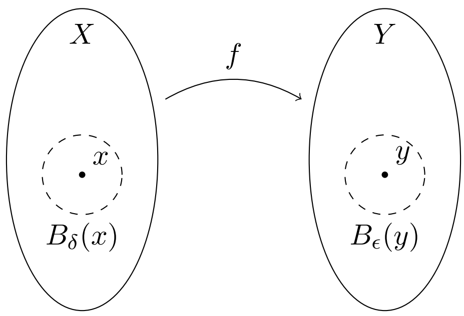
$$\lim_{\tilde{x} \to x} f(\tilde{x}) = y_1 \qquad\qquad \lim_{\tilde{x} \to x} g(\tilde{x}) = y_2$$

Then

$$\lim_{\tilde{x} \to x} (f + g)(\tilde{x}) = y_1 + y_2 \qquad\qquad \lim_{\tilde{x} \to x} (f \cdot g)(\tilde{x}) = y_1 \cdot y_2$$
$$\lim_{\tilde{x} \to x} \left(\frac{f}{g} \right)(\tilde{x}) = \frac{y_1}{y_2}$$

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 $(0, 0)$.

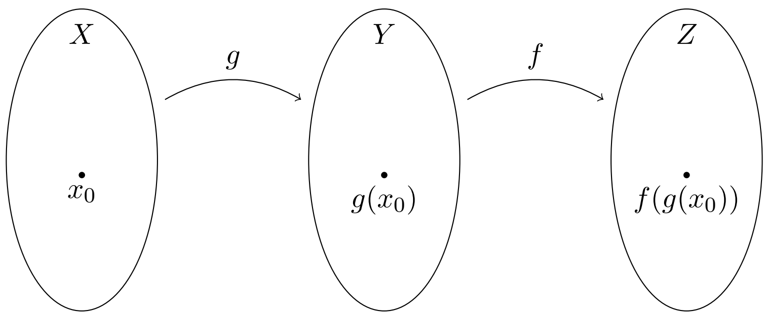
Remark 5.45. (i)

$$f \text{ continuous in } x \iff \forall \epsilon > 0\ \exists \delta > 0 : \quad f(B_\delta(x)) \subset B_\epsilon(f(x))$$

(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 $f : X \to Y$.*

$$f \text{ is continuous in } x \in X \iff \substack{x \text{ is an isolated point in } X \\ \text{or } \lim_{\tilde{x}\to x} f(\tilde{x})=f(x)}$$

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

$$f(B_\delta) \subset B_\epsilon(f(x)) \tag{5.42}$$

□

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 $c_\eta \in \mathbb{K} \quad \forall \eta$ with $|\eta| \leq N \quad N \in \mathbb{N}$. 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)

$$f \text{ is continuous in } x \iff \begin{matrix} f^{-1}(U) \text{ is a neighbourhood of} \\ x \quad \forall U \text{neighbourhood of } f(x) \end{matrix}$$

(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. $A \setminus B_n(0) \neq \varnothing \quad \forall n \in \mathbb{N}$. 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

$$\begin{aligned} f(C) &= \{f(x) \in B \mid x \in C\} \\ &= \{f(g(y)) \in B \mid g(y) \in C\} \\ &= \{y \in B \mid g(y) \in C\} = g^{-1}(C) \end{aligned} \tag{5.54}$$

So $g^{-1}(C)$ is bounded, and since C was an arbitrary closed set, g is also continuous. □

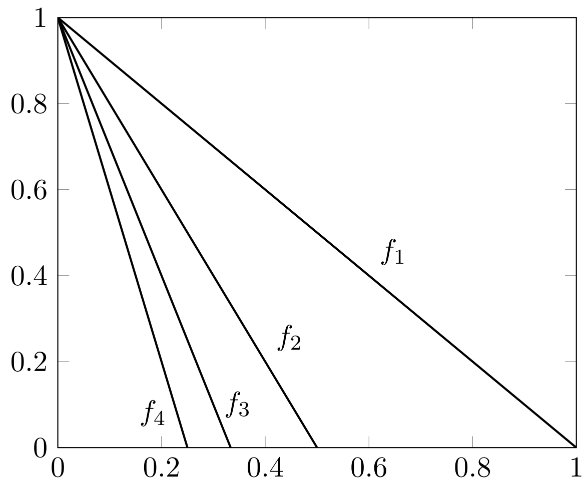
5.5 Convergence of Function sequences

Definition 5.66 (Pointwise convergence)**.** Let M be a set, $f_n : M \to \mathbb{K} \ \ \forall n \in \mathbb{N}$ and $f : M \to \mathbb{K}$. The sequence (f_n) is said to be pointwise convergent to f if

$$\lim_{n\to\infty} f_n(x) = f(x) \ \ \forall x \in M$$

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

$$\|f\|_\infty = 0 \implies |f(x)| = 0 \quad \forall x \in M \tag{5.55}$$

This implies

$$f(x) = 0 \quad \forall x \in M \implies f = 0 \tag{5.56}$$

The triangle inequality is proven by first considering

$$|f(x)| \leq \|f\|_\infty \quad \forall f \in B(M) \; \forall x \in M \tag{5.57}$$

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

$$\exists N \in \mathbb{N}: \quad \|f_n \cdot f_m\|_\infty < \epsilon \ \ \forall n, m \geq N \tag{5.62}$$

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

$$\exists N \in \mathbb{N}: \quad \|f - f_n\|_\infty y\frac{\epsilon}{3} \ \ \forall n \geq N \tag{5.68}$$

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. □

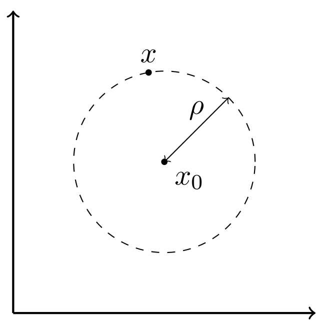
Definition 5.75. Let $x_0 \in \mathbb{K}$ and $(a_n) \subset \mathbb{K}$. Then

$$\sum_{n=1}^{\infty} a_n (x - x_0)^n$$

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 $\rho > 0$. □

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

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

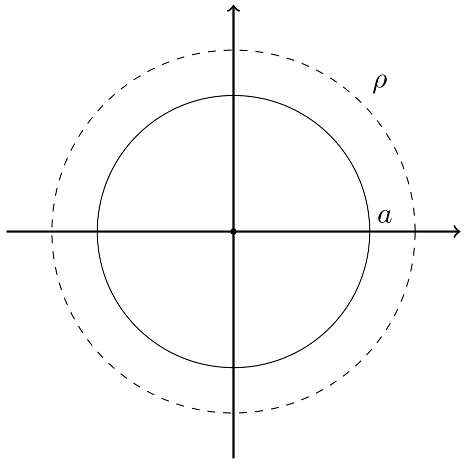
is convergent on $(-1, 1)$, so $\rho = 1$. The limit function is

$$x \longmapsto \frac{1}{1 - x}$$

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. □

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

$$\begin{aligned}\sum_{n=1}^{\infty}\left\|\frac{A^n}{n!}\right\| = \sum_{n=1}^{\infty}\frac{1}{n!}\|A^n\| \leq \sum_{n=1}^{\infty}\frac{1}{n!}\|A\|^n \\ = \exp(\|A\|) < \infty\end{aligned}$$

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

$$\lim_{y \to x} \frac{\phi(x)}{|y - x|} = \left(\frac{f(y) - f(x)}{y - x} - f'(x) \right) \cdot \frac{y - x}{|y - x|} = 0 \tag{6.3}$$

□

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 $\partial_i f(x) = Ae_i$. □

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

$$Df(x) =: \vec{\nabla} f(x)$$

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

$$\exists \phi_f : V \to \mathbb{R}^k :: \quad \lim_{z \to g(x)} \phi_f(z) \|z - g(x)\|^{-1} = 0 \tag{6.7}$$

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.

$$\Longrightarrow \psi(g(y)) \cdot \frac{\|g(y)-g(x)\|}{\|y-x\|} \longrightarrow 0 \tag{6.16}$$

□

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.

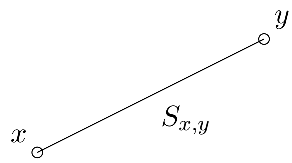
$$\begin{aligned}
\partial_i x^\eta &= \partial_i \left(x_1^{\eta_1}, x_2^{\eta_2}, \cdots, x_n^{\eta_n} \right) \\
&= \eta_i x_1^{\eta_1} \cdots x_{i-1}^{\eta_{i-1}} x_i^{\eta_{i-1}} x_{i+1}^{\eta_{i+1}} \cdots x_n^{\eta_n}
\end{aligned}$$

   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. □

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*

$$\exists \xi \in S_{x,y} : \quad \|f(y) - f(x)\| \le \|Df(\xi)(y - x)\|$$

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

$$\begin{aligned} \|f(y) - f(x)\|^2 &= |\langle f(y) - f(x) | Df(\xi)(y - x) \rangle| \\ &\le \|f(y) - f(x)\| \|Df(\xi)(y - x)\| \end{aligned} \tag{6.21}$$

□

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

$$\forall y \in B_\epsilon(x)\ \exists \xi \in S_{x,y} : \quad \|f(y) - f(x)\| \leq \|Df(\xi)(y-x)\| = 0 \tag{6.22}$$

This implies

$$\|f(y) - f(x)\| = 0 \implies f(y) = f(x) \quad \forall y \in B_\epsilon(x) \tag{6.23}$$

□

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

$$\begin{aligned} D^2f(x) : \mathbb{R}^n \times \mathbb{R}^n &\longrightarrow \mathbb{R}^m \\ (u, v) &\longmapsto D(Df(\cdot)u)(x)v \end{aligned}$$

$D^2f(x)$ is linear in v and u, and

$$\begin{aligned} D^2f(x)(u_1 + \lambda u_2, v) &= D(Df(\cdot)(u_1 + \lambda u_2))(x)v \\ &= D(Df(\cdot)u_1 + \lambda Df(\cdot)u_2)(x)v \\ &= D(Df(\cdot)u_1)(x)v + \lambda D(Df(\cdot)u_2)(x)v \\ &= D^2f(x)(u_1, v) + \lambda D^2f(x)(u_2, v) \end{aligned}$$

$D^2f(x)$ 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

$$\begin{aligned} D^{k+1}f(x) : (\mathbb{R}^n)^{k+1} &\longrightarrow \mathbb{R}^m \\ (u_1, \cdots, u_k, v) &\longmapsto D(D^kf(\cdot)(u_1, \cdots, u_k))(x)v \end{aligned}$$

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.

$$(D^kf(x)(u_1, \cdots, U_k))_j = D^kf_j(x)(u_1, \cdots, u_k)$$

Remark 6.21. $D^kf(x)$ 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! □

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

$$\begin{aligned}
&\lim_{t\to 0}\frac{f(x+tv+tu)-f(x+tv)-f(x+tu)+f(x)}{t^2}\\
&\quad=\lim_{t\to 0} D^2 f\,\underbrace{(x+\eta v+\xi u)}_{\longrightarrow x}(u,v)\\
&\quad= D^2 f(x)(u,v)
\end{aligned}\tag{6.31}$$

The left side is symmetric in terms of swapping u and v, so the right side must be as well. $\square$

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

$$\begin{aligned}
g:[0,1] &\longrightarrow \mathbb{R}\\
t &\longmapsto f(x+th)
\end{aligned}\tag{6.32}$$

$\square$

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 $l > 2$.

6.3 Function Sequences and Differentiability

Example 6.30. Consider (f_n):

$$\begin{aligned} f_n : \mathbb{R} &\longrightarrow \mathbb{C} \\ x &\longmapsto \frac{1}{n} e^{inx} \end{aligned}$$

Then

$$\|f_n\|_\infty = \frac{1}{n} \longrightarrow 0$$
$$\Longleftrightarrow$$
$$(f_n) \text{ converges uniformly to the zero function}$$

But

$$f_n'(x) = ie^{inx} = i(e^{ix})^n$$

converges (pointwise even) only for $x = 2k\pi, \quad k \in \mathbb{Z}$.

Remark 6.31. Let $f : X \to V$ where V is a normed vector space. Define

$$\|f\|_\infty = \sup \{\|f(x)\| \,|\, x \in X\}$$

the supermum norm. Also define

- $B(X, V)$ the space of bounded functions from $X \to V$
- $C_B(X, V)$ the space of continuous, bounded functions from $X \to V$

Theorem 6.32. *Let $U \subset \mathbb{R}^n$ be open and $f_n : U \to \mathbb{R}^m$ continuously differentiable $\forall n \in \mathbb{N}$. If (f_n) and (Df_n) converge uniformly to $f : U \to \mathbb{R}^m$ and $g : U \to \mathbb{R}^{m\times m}$, then f is differentiable and $Df = g$.*

Proof. First consider $m = 1$. We use the operator norm on $\mathbb{R}^{m\times m}$. First, let Df_n be continuous $\forall n$ and thus g is continuous. Choose $x \in U$ and $\epsilon > 0$, then

$$\exists \delta > 0 : \quad \|g(y) - g(x)\| < \frac{\epsilon}{3} \quad \text{if} \quad \|y - x\| < \delta \tag{6.33}$$

Furthermore

$$\exists N \in \mathbb{N} : \quad \|Df_n - g\|_\infty < \frac{\epsilon}{3} \quad \forall n > N \tag{6.34}$$

Let $y \in B_\delta(x)$. Then according to the intermediate value theorem,

$$\forall n \in \mathbb{N} \; \exists \xi_n \in S_{x,y} = \{x + t(y - x) \,|\, t \in [0, 1]\} \tag{6.35}$$

such that

$$f_n(y) - f_n(x) = Df_n(\xi_n)(y - x) \tag{6.36}$$

We have $\xi_m \in B_\delta(x)$. Then

$$\begin{aligned}
&\frac{1}{\|y-x\|}\,|f_n(y)-f_n(x)-Df_n(x)(y-x)| \\
=&\frac{1}{\|y-x\|}\,\underbrace{|(Df_n(\xi_n)-Df_n(x))(y-x)|}_{\|Df_n(\xi_n)\|-Df_n(x)\|y-x\|} \\
\leq&\|Df_n(\xi_n)-Df_n(x)\| \\
\leq&\|Df_n(\xi_n)-g(\xi_n)\|+\|g(\xi_n)-g(x)\|+\|g(x)-Df_n(x)\| \\
\leq&\|Df_n-g\|_\infty+\|g(\xi_n)-g(x)\|+\|g-Df_n\|_\infty \\
=&2\|Df_n-g\|_\infty+\|g(\xi_n)-g(x)\|<\epsilon
\end{aligned} \tag{6.37}$$

For $n\to\infty$ we have

$$\frac{1}{\|y-x\|}|f(y)-f(x)-g(x)(y-x)|<\epsilon \quad \forall y\in B_\delta(x) \tag{6.38}$$

Since $\epsilon>0$ is arbitrary, we get

$$\lim_{y\to x}\frac{1}{\|y-x\|}|f(y)-f(x)-g(x)(y-x)|=0 \tag{6.39}$$

This means that f is differentiable in x with $Df(x)=g(x)$. □

Remark 6.33. On $C^1_B(U,\mathbb{R}^m)$ (the space of continuous, differentiable and bounded functions with bounded derivative) we can define a norm:

$$\|f\|_{C_1}:=\|f\|_\infty+\|Df\|_\infty$$

Then the above theorem is equivalent to the statement that $C^1_B(U,\mathbb{R}^m)$ with $\|f\|_{C_1}$ is complete.

Theorem 6.34. *Let $f(x)=\sum_{n=0}^\infty a_nx^n$ be a power series with positive convergence radius ρ. Then f is differentiable on $B_\rho(0)$ and*

$$f'(x)=\sum_{n=0}^\infty na_nx^{n-1}$$

Proof. We need to inspect the convergence radius R of

$$\sum_{n=0}^\infty na_nx^{n-1}=\frac{1}{x}\sum_{n=0}^\infty na_nx^n \tag{6.40}$$

$(\sqrt[n]{n})$ converges to 1, so $\exists\epsilon>0$ such that for sufficiently big n we have

$$(1-\epsilon)\sqrt[n]{a_n}\leq\sqrt{na_n}\leq(1+\epsilon)^n\sqrt{a_n} \tag{6.41}$$

and thus

$$\frac{1-\epsilon}{\rho} = (1-\epsilon)\cdot \limsup_{n\to\infty} \sqrt[n]{|a_n|} \le \limsup_{n\to\infty} \sqrt[n]{|na_n|} = \frac{1}{R} \le \frac{1+\epsilon}{\rho} \tag{6.42}$$

So

$$\implies \frac{1-\epsilon}{\rho} \le \frac{1}{R} \le \frac{1+\epsilon}{\rho} \tag{6.43}$$

Since this holds for every ϵ, this implies $\rho = R$. Now for $x \in B_\rho(0)$ set

$$g(x) := \sum_{k=1}^{\infty} na_n x^{n-1} \tag{6.44}$$

Let $x \in B_\rho(0)$ be fixed and choose $a > 0$ such that $|x| < a < \rho$. This means that

$$f_N(x) := \sum_{n=0}^{N} a_n x^n \qquad \text{and} \qquad g_N(x) := \sum_{n=0}^{N} a_n x^{n-1}$$

converge uniformly on $B_a(0)$ to f and g. Obviously, $f_N' = g_N$, so f is differentiable and $f' = g$. Since differentiabiility is a local property, the desired statement follows $\forall x \in B_\rho(0)$. $\square$

Corollary 6.35. *Let $f(x) = \sum_{n=0}^{\infty} a_n x^n$ be a power series with convergence radius $\rho > 0$. Then $f \in C^\infty(B_r(0))$, and*

$$a_k = f^{(k)}(0) \cdot ({}^{-1}k!)$$

Furthermore, the series representation (if it exists) is unique.

Proof. The infinite Differentiability follows inductively from the previous theorem. Also inductively we have

$$f^{(k)}(x) = \sum_{n=0}^{\infty} n(n-1)\cdots(n-k+1)a_n x^{n-k} \tag{6.45}$$

Choose $x = 0$ and receive

$$f^{(k)}(0) = n(n-1)\cdots(n-k+1)a_n \tag{6.46}$$

$\square$

Example 6.36 (Derivative of the exponential function).

$$(e^x)' = \sum_{n=0}^{\infty} \left(\frac{x^n}{n!}\right)' = \sum_{n=1}^{\infty} \frac{nx^{n-1}}{n!} = \sum_{n=1}^{\infty} \frac{x^{n-1}}{(n-1)!} = \sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x$$

Remark 6.37 (Taylor Series). We can define the Taylor series for $f : \mathbb{K} \to \mathbb{K}$

$$\sum_{n=0}^{\infty} \frac{f^(n)(0)}{n!} x^n = f(x)$$

- In general, this doesn't hold true for all x, not even for $f \in C^\infty$.
- The convergence radius could be 0
- There are examples of convergent Taylor series that don't converge to the initial function, e.g.

$$f : x \mapsto \begin{cases} \exp\left(-\frac{1}{x}\right), & x > 0 \\ 0, & \text{else} \end{cases}$$

  f is infinitely continuously differentiable in 0, but the Taylor series would converge to 0.

Definition 6.38. Let $a_\eta \in \mathbb{K}$ (Multiindex notation) be coefficients $\forall \eta \in \mathbb{N}_0^d$. Then

$$\sum_{\eta \in \mathbb{N}_0^d} a_\eta x^\eta$$

is said to be a (formal) power series with d variables.

A function $f : U \to \mathbb{K}$ with U neighbourhood around 0 is said to be analytic in 0, if and only if

$$\exists \epsilon > 0, a_\eta \in \mathbb{K} : \quad f(x) = \sum_{\eta \in \mathbb{N}_0^d} a_\eta x^\eta \quad \forall x \in B_\epsilon(0)$$

Remark 6.39. (i) The convergence of the series to $S(x)$ can be defined as follows: $\forall \epsilon > 0\ \exists A \subset \mathbb{N}_0^d$ finite such that $\forall B \supset A$ finite we have

$$\left| \sum_{\eta \in B} a_\eta x^\eta - S(x) \right| < \epsilon$$

(ii) If the series converges in $(y_1, \cdots, y_n)$, then it also absolutely converges in the open cuboid

$$\left\{ x \in \mathbb{R}^d \,\middle|\, |x_i| < |y_i| \quad \forall i \in \{1, \cdots, d\} \right\}$$

which means

$$\sum_{\eta \in \mathbb{N}_0^d} |a_\eta| (|x_1|, \cdots, |x_d|)^\eta < \infty$$

(iii) If the power series converges on a neighbourhood U around 0, then it is infinitely differentiable and

$$a_\eta = \frac{\partial^\eta f(0)}{\eta!}$$

with

$$\partial^\eta := \partial_1^{\eta_1} \partial_2^{\eta_2} \cdots \partial_d^{\eta_d} \qquad\qquad \eta! := \eta_1! \eta_2! \cdots \eta_d!$$

(iv) The formula above is only rarely useful to calculate the Taylor series. By inverting it we can calculate the derivative of a known series representation. E.g.

$$f(x) = xe^{x^2} = x \cdot \sum_{k=0}^{\infty} \frac{(x^2)^k}{k!} = \sum_{=1}^{\infty} k = 0^{\infty} \frac{x^{2k+1}}{k!} \quad \forall x \in \mathbb{K}$$

$f^{(k)}(0) = 0$ is k is even, and it is something else if k is odd.

(v) $C^{\omega}(U)$ is the space of all analytic functions.

$$C(U) \supset C^1(U) \supset C^2(U) \supset \cdots \supset C^k(U) \supset \cdots \supset C^{\infty}(U) \supset C^{\omega}(U)$$

(vi) The analytic functions are closed among sums, products and concatinations. A power series is analytic within its converges radius.

Example 6.40. Consider the power series

$$\sum_{n=0}^{\infty} (xy)^n = \sum_{\eta \in \mathbb{N}_0^2} (xy)^{\eta} \cdot a_{\eta}$$

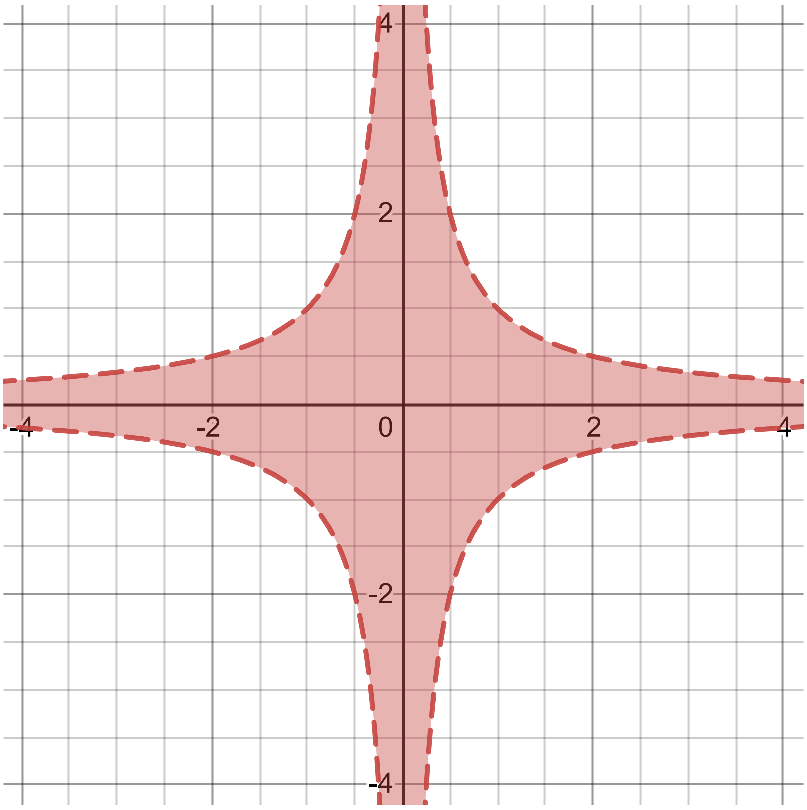
with

$$a_{\eta} = 1 \text{ if } \eta_1 = \eta_2$$
$$a_{\eta} = 0 \text{ else}$$

This series converges on

$$\{(x, y) \mid |xy| < 1\}$$

to $\frac{1}{1-xy}$.



So the convergence area must not necessarily be a sphere. The limit function is also defined outside of the convergence area.

6.4 The Banach Fixed-Point Theorem and the Implicit Function Theorem

Theorem 6.41 (Banach Fixed-Point Theorem)**.** *Let (X,d) be a complete metric space, and $\phi : X \to X$ strictly contractive, i.e.*

$$\exists C \in (0,1): \quad d(\phi(x),\phi(y)) \leq Cd(x,y) \quad \forall x,y \in X$$

Then there exists exactly one fixed point x of ϕ, i.e. $\phi(x) = x$.

Proof. First, ϕ is Lipschitz continuous, and thus continuous. Let $x_0 \in X$, and recursively define $x_{n+1} = \phi(x_n)$. Then

$$d(x_{n+1},x_n) = d(\phi(x_n),\phi(x_{n-1})) \leq Cd(x_n,x_{n-1}) \tag{6.47}$$

and via induction

$$d(x_{n+k},x_{n+k-1}) \leq C^k d(x_n,x_{n-1}) \quad \forall k,n \in \mathbb{N} \tag{6.48}$$

Especially,

$$d(x_n, x_{n-1}) \leq C^{n-1} d(x_1, x_0) \tag{6.49}$$

Using the triangle inequality we can compute

$$\begin{aligned}
d(x_{n+k}, x_{n-1}) &\leq d(x_{n+k}, x_{n+k-1}) + d(x_{n+k-1}, x_{n+k-2}) + \cdots + d(x_n, x_{n-1}) \\
&\leq (C^k + C^{k-1} + C^{k-2} + \cdots + 1) d(x_n, x_{n-1}) \\
&\leq \frac{1 - C^{k+1}}{1 - C} \cdot d(x_n, x_{n-1}) \\
&\leq \frac{1 - C^{k+1}}{1 - C} C^{n-1} d(x_1, x_0) \\
&\leq \frac{C^{n-1}}{1 - C} d(x_1, x_0) \xrightarrow{n \to \infty} 0
\end{aligned} \tag{6.50}$$

This means

$$\forall \epsilon > 0 \; \exists N \in \mathbb{N} : \quad d(x_{n+k}, x_{n-1}) < \epsilon \;\; \forall n > N \; \forall k \in \mathbb{N} \tag{6.51}$$

Which in turn means that (x_n) is a Cauchy sequence, and thus convergent. (x_n) converges to $x \in X$

$$x = \lim_{n \to \infty} x_n = \lim_{n \to \infty} \phi(x_{n-1}) = \phi(\lim_{n \to \infty} x_{n-1}) = \phi(x) \tag{6.52}$$

To prove the uniqueness, let x, y both be fixed points. Then

$$d(x, y) = d(\phi(x), \phi(y)) \leq C d(x, y) \tag{6.53}$$

Since $C < 1$, we have

$$d(x, y) \implies x = y \tag{6.54}$$

□

Remark 6.42. The Banach fixed-point theorem implies that every map that is within the area it is mapping, will have a point on the map that lies directly on top of the point in the real world that it maps.

Example 6.43. Consider the equation

$$x - y^2 = 0$$

with the solutions

$$y = \sqrt{x} \qquad\qquad y = -\sqrt{x}$$

on $(0, \infty)$. For a point (ξ, η) that solves the equation, there exists a neighbourhood U and a function f such that all solutions of the equation on U are of the form $(x, f(x))$.

Remark 6.44. Let $F : \mathbb{R}^P \times \mathbb{R}^Q \to \mathbb{R}^Q$, and consider $x_1, \cdots, x_P \in \mathbb{R}$ as independent variables, and $y_1, \cdots, y_Q \in \mathbb{R}$ as dependent variables of the equation system

$$F(x, y) = 0, \quad x = (x_1, \cdots, x_P), y = (y_1, \ldots, y_Q)$$

Let (ξ, η) be a solution. The question is wether a $f : \mathbb{R}^P \to \mathbb{R}^Q$ exists, such that $(x, f(x))$ are solutions $\forall x \in U$, where U is a neighbourhood of ξ.

$$x \longmapsto F(x, f(x))$$

If F is differentiable, then let $D_y F(x, \eta) \in \mathbb{R}^{Q \times Q}$ denote the total derivative of the function. Analogously this works for y as the variable. We approximately have

$$F(x, y) \approx F(x, \eta) + D_y F(x, \eta)(y - \eta) = 0$$

Theorem 6.45 (Implicit Function Theorem)**.** *Let $U \subset \mathbb{R}^P, V \subset \mathbb{R}^Q$ be open, and*

$$F : U \times V \to \mathbb{R}^Q$$

continuously differentiable. Choose $\xi \in U, \eta \in V$ such that $F(\xi, \eta) = 0$, and $D_y F(\xi, \eta)$ invertible. Then there exists a neighbourhood $\tilde{U} \subset U$ of ξ, a neighbourhood $\tilde{V} \subset V$ of η and a continuous function $f : \tilde{U} \to \tilde{V}$ such that $f(\xi) = \eta$ and

$$F(x, f(x)) = 0 \quad \forall x \in \tilde{U}$$

.

Proof. Set $D = D_y F(\xi, \eta)$. Then consider

$$\begin{aligned} \phi : \text{function} &\longrightarrow \text{function} \\ \phi(g)(x) &\longmapsto g(x) - D^{-1} F(x, g(x)) \end{aligned} \tag{6.55}$$

where $g : \mathbb{R}^P \to \mathbb{R}^Q$. Then we have

$$\phi(g) = g \iff D^{-1} F(x, g(x)) = 0 \iff F(x, g(x)) = 0 \tag{6.56}$$

Since this is a fixed point problem, our goal is to apply the Banach fixed-point theorem. Let $I : \mathbb{R}^Q \to \mathbb{R}^Q$ be the identity mapping. Then the function

$$(x, y) \longmapsto \left\| I - D^{-1} D_y F(x, y) \right\| \tag{6.57}$$

is continuous and vanishes in (ξ, η). $\exists \delta, \epsilon > 0$ such that $B_\delta(\xi) \subset U$, and $B_\epsilon(\eta) \subset V$ and

$$\left\| I - D^{-1} D_y F(x, y) \right\| \leq \frac{1}{2} \quad \forall x \in B_\delta(\xi), y \in B_\epsilon(\eta) \tag{6.58}$$

Because of the continuity of

$$x \longmapsto \left\| D^{-1} F(x, \eta) \right\| \tag{6.59}$$

we can choose a (possibly smaller) $\delta > 0$, such that

$$\left\|D^{-1}F(x,\eta)\right\| \leq \frac{\epsilon}{4} \quad \forall x \in B_\delta(\xi) = \tilde{U} \tag{6.60}$$

Now let X denote the set of all continuous functions $g : \tilde{U} \to \mathbb{R}^Q$

$$g(\xi) = \eta \tag{6.61a}$$

$$\|g(x) - \eta\| \leq \frac{\epsilon}{2} \quad \forall x \in \tilde{U} \tag{6.61b}$$

Equation (6.61b) implies that $g(x) \in B_\epsilon(\eta) \subset V$. Furthermore X is a subset of $C_B(\tilde{U}, \mathbb{R}^Q)$, which is a complete set with the norm

$$\|g\|_\infty = \sup\left\{\|g(x)\| \,\middle|\, x \in \tilde{U}\right\} \tag{6.62}$$

X is non-empty (for example, it contains $g(\xi) = \eta$) and bounded, which means X is also complete. Now, for a fixed $x \in \tilde{U}$ and $\tilde{V} \subset B_\epsilon(\eta)$ consider the mapping

$$\begin{aligned} \Phi : \tilde{V} &\longrightarrow \mathbb{R}^Q \\ y &\longmapsto y - D^{-1}F(x,y) \end{aligned} \tag{6.63}$$

From the intermediate value theorem we can conclude

$$\begin{aligned} \|\Phi(y) - \Phi(z)\| &\leq \sup_{y \in \tilde{V}} \underbrace{\left\|I - D^{-1}D_yF(x,y)\right\|}_{D\Phi(x,y)} \|y - z\| \\ &\leq \frac{1}{2}\|y - z\| \end{aligned} \tag{6.64}$$

Now, for $g_1, g_2 \in X$ and $x \in \tilde{U}$ we can see that

$$\begin{aligned} \|\phi(g_1)(x) - \phi(x_2)(x)\| &= \|\Phi(g_1(x)) - \Phi(g_2(x))\| \\ &\leq \frac{1}{2}\|g_1(x) - g_2(x)\| \end{aligned} \tag{6.65}$$

and by choosing the supremum over all $x \in \tilde{U}$ we can see that

$$\|\phi(g_1) - \phi(g_2)\|_\infty \leq \frac{1}{2}\|g_1 - g_2\|_\infty \tag{6.66}$$

Thus ϕ is strictly contractive on x. It is only left to show that $\phi(X) \subset X$. From the definition of ϕ we have $\forall g \in X$

$$\phi(g)(\xi) = g(\xi) = \eta \tag{6.67}$$

So $\phi(g)$ is continuous, and finally

$$\begin{aligned}\|\phi(g)(x)-\eta\| &\le \|\phi(g)(x)-\phi(\eta)(x)\| + \|\phi(\eta)(x)-\eta\| \\ &\le \frac{1}{2}\underbrace{\|g(x)-\eta\|}_{\le\frac{\epsilon}{2}} + \underbrace{\left\|D^{-1}F(x,\eta)\right\|}_{\le\frac{\epsilon}{4}} \\ &\le \frac{\epsilon}{2}\end{aligned} \tag{6.68}$$

Thus, ϕ maps X to X, and the Banach fixed-point theorem tells us

$$\exists! f \in X : \quad \phi(f) = f \iff F(x, f(x)) = 0 \ \ \forall x \in \tilde{U} \tag{6.69}$$

□

Remark 6.46 (About uniqueness). We know there is exactly one function f in X such that

$$F(x, f(x)) = 0 \ \ \forall x \in \tilde{U}$$

$f(x)$ the only solution in $\tilde{V}$, for $x \in \tilde{U}$, because if $F(x,y) = 0$ for $y \in V$, then

$$\|y - f(x)\| = \|\Phi(y) - \Phi(f(x))\| \le \frac{1}{2}\|y - f(x)\|$$

which implies $y = f(x)$

Theorem 6.47. *There is a possibly smaller neighbourhood $\tilde{U}$ around ξ on which $f \in C^1(\tilde{U}, \tilde{V})$. The derivative is given by*

$$Df(x) = -(D_yF(x, f(x)))^{-1}DxF(x, f(x))$$

Proof. Without proof. □

Corollary 6.48 (Inverse Function Theorem)**.** *Let $U \subset \mathbb{R}^n$ and $f : U \to \mathbb{R}^m$ continuously differentiable. If $Df(\xi)$ is invertible for some $\xi \in U$, then there exists a neighbourhood $\tilde{U}$ around ξ and a neighbourhood $\tilde{V}$ around $f(\xi) =: \eta$ such that f bijectively maps $\tilde{U}$ to $\tilde{V}$, and the inverse function*

$$\begin{aligned} g : \tilde{V} &\longrightarrow \tilde{U} \\ y &\longmapsto f^{-1}(y) \end{aligned}$$

is continuously differentiable. Furthermore

$$Dg(\eta) = (Df(\xi))^{-1}$$

Heuristic Proof. Use the implicit function theorem on the equation system

$$F(x,y) = f(x) - y = 0 \tag{6.70}$$

and solve that for x. □

Example 6.49 (Inverse function of the complex exponential function). Let

$$z \longmapsto \exp(z)$$

be a function $\mathbb{R}^2 \to \mathbb{R}^2$, i.e. $z = x + yi$ and

$$\exp(z) = \exp(x)\exp(yi) = \exp(x)(\cos y + i \sin y)$$

Consider

$$\begin{aligned} \phi : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x, y) &\longmapsto (\exp(x) \cdot \cos y, \exp(x) \cdot \sin y) \end{aligned}$$

This mapping is continuously differentiable (analytic even) and $D\phi(x, y)$ is invertible everywhere. Thus ϕ has a locally differentiable inverse function on $\exp(\mathbb{C})$ (the logarithm).

One can show that $\exp(\mathbb{C}) = \mathbb{C} \setminus \{0\}$. Typically, the main branch of the complex logarithm is defined as

$$\begin{aligned} \ln :& \mathbb{C} \setminus \{x \in \mathbb{R} \,|\, x \leq 0\} \\ \Longrightarrow \;& \mathbb{R} \times (-\pi, \pi) \end{aligned}$$

One can choose from many other domains, however there is no continuous logarithm on $\mathbb{C} \setminus \{0\}$.

Chapter 7

Measures and Integrals

7.1 Contents and Measures

Definition 7.1. A set M is said to be countable if there exists a surjective mapping from $\mathbb{N}$ to M, i.e.

$$\exists\,(x_n) \subset M: \quad \forall y \in M\ \exists n \in \mathbb{N}: \quad x_n = y$$

A set M is said to be countably infinite if it is countable and unbounded.

Remark 7.2. (i) Countably infinite sets are the smallest kind of infinite sets.

(ii) Subsets of countable sets are countable.

(iii) The union of two countable sets is countable. Let $(x_n) \subset M, (y_n) \subset K$ by surjective sequences, then

$$(x_1, y_1, x_2, y_2, \cdots)$$

is a surjective sequence for $M \subset K$. This argument can be used to prove $\mathbb{Z}$ is countable.

(iv) The union of countably many countable sets is countable. Let M be a countable set of countable sets, and $(A_n) \subset M$ a surjective sequence. Then $\forall n \in \mathbb{N}$ exists a surjective mapping $(x_{n_k})_{k\in\mathbb{N}} \subset A_n$

$$\begin{matrix} x_{1_1}^{\ 1} & x_{1_2}^{\ 2} & x_{1_3}^{\ 4} & \cdots \\ x_{2_1}^{\ 3} & x_{2_2}^{\ 5} & x_{2_3} & \cdots \\ x_{3_1}^{\ 6} & x_{3_2} & x_{3_3} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{matrix}$$

This sequence is surjective on

$$\bigcup_{A\in M} A$$

Especially, for countable M, K we have

$$M \times K = \bigcup_{x\in M} \{(x,y) \mid y \in K\}$$

Thus $\mathbb{N}\times\mathbb{N}$, $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$ are countable.

(v) There exist uncountable sets, like $[0,1]$, $\mathbb{R}$ and $\mathcal{P}(\mathbb{R})$.

Definition 7.3. Let Ω be a set. A family of subsets

$$(A_i)_{i\in I} \subset \mathcal{P}(\Omega) \quad (I \text{ denotes the index set})$$

is said to be pairwise disjoint is

$$A_i \cap A_j = \varnothing \quad \forall i,j \in I,\ i \neq j$$

Remark 7.4. (i) Let $\mathcal{A} \subset \mathcal{P}(\mathbb{R}^n)$ be a family of sets. A mapping

$$\mu : \mathcal{A} \to [0, \infty]$$

is said to be the content of $\mathcal{A}$, if $\forall A_1, \cdots, A_k \in \mathcal{A}$ pairwise disjoint the following holds:

$$A_1 \cup \cdots \cup A_k \in \mathcal{A} \implies \mu(A_1 \cup \cdots \cup A_k) = \sum_{l=1}^{k} \mu(A_l)$$

The content is a generalization of the concept of length ($\mathbb{R}$), area ($\mathbb{R}^2$), volume ($\mathbb{R}^3$) etc.

(ii) In the context of contents, measures and integrals we define

$$\begin{aligned} c + \infty &= \infty \quad \forall c \in \mathbb{R} \cup \{\infty\} \\ c \cdot \infty &= \infty \quad \forall c \in (0, \infty] \\ 0 \cdot \infty &= 0 \end{aligned}$$

(iii) The goal is to choose the domain of the content as big as possible. Ideal would be $\mathcal{A} = \mathcal{P}\mathbb{R}^n$. This introduces the Banach-Tarski paradox:

- Let $B_1(0) \subset \mathbb{R}^3$ be the unit sphere
- One can show: There exists a disjoint decomposition

  $$A_1 \cup \cdots \cup A_P \cup B_1 \cup \cdots \cup B_Q = B_1(0)$$

  and a set of translations and rotations

  $$D_1, \cdots, D_P, \cdots T_1, \cdots, T_Q$$

  such that

  $$\begin{aligned} D_1 A_1 \cup D_2 A_2 \cup \cdots \cup D_P A_P &= B_1(0) \\ T_1 B_1 \cup T_2 B_2 \cup \cdots \cup T_Q B_Q &= B_1(0) \end{aligned}$$

Definition 7.5. Let Ω be a set, $\mathcal{A}$ a family of subsets of Ω (so $\mathcal{A} \subset \mathcal{P}(\Omega)$). $\mathcal{A}$ is sait to be a σ-algebra, if

(i) $\varnothing \in \mathcal{A}$

(ii) $A \in \mathcal{A} \implies A^C = \Omega \setminus A \in \mathcal{A}$

(iii) For a countable subset $\{A_1, \cdots, A_n\} \subset \mathcal{A}$ follows

$$\bigcup_{i \in \mathbb{N}} A_i \subset \mathcal{A}$$

A mapping

$$\mu : \mathcal{A} \to [0, \infty]$$

is said to be a measure, if

$$\mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \mu(A_i) \quad (\sigma\text{-additivity})$$

for pairwise disjoint $(A_i)_{i\in\mathbb{N}} \subset \mathcal{A}$ and $\mu(\varnothing) = 0$. The pair $(\Omega, \mathcal{A})$ is called a measureable space, and $(\Omega, \mathcal{A}, \mu)$ is called measure space.

Example 7.6. (i) Let Ω be an arbitrary set, and let there be a disjoint decomposition

$$A_1 \cup \cdots \cup A_n = \Omega$$

Then

$$\left\{ \bigcup_{i\in I} A_i \,\middle|\, I \subset \{1, \cdots, n\} \right\}$$

is a σ-algebra.

(ii) Let Ω be arbitrary and $x \in \Omega$. Then

$$\begin{aligned} \delta_x : \mathcal{P}(\Omega) &\longrightarrow [0, \infty] \\ A &\longmapsto \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases} \end{aligned}$$

is a measure.

(iii) Let Ω be arbitrary, then

$$\begin{aligned} \# : \mathcal{P}(\Omega) &\longrightarrow [0, \infty] \\ A &\longmapsto \begin{cases} \text{Number of elements in } A, & A \text{ finite} \\ \infty, & A \text{ infinite} \end{cases} \end{aligned}$$

is the so called counting measure. It is useful for finite, countable sets.

(iv) Let Ω be countable and $(a_w)_{w\in\mathbb{R}} \subset [0, \infty]$. Then

$$\begin{aligned} \mu : \mathcal{P}(\Omega) &\longrightarrow [0, \infty] \\ A &\longmapsto \sum_{w\in A} a_w \end{aligned}$$

a measure.

(v) Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and $A \in \mathcal{A}$. Define the to A confined σ-algebra

$$\mathcal{A}|_A := \{B \cap A \,|\, B \in \mathcal{A}\}$$

Then $(A, \mathcal{A}|_A, \mu)$ is a measure space.

Remark 7.7. For countable subsets $\mathcal{A} = \{A_1, \cdots, A_n, \cdots\} \subset \sigma$-algebra we have

$$\bigcap_{i \in \mathbb{N}} A_i = \left(\bigcup_{i \in \mathbb{N}} A_i^C \right)^C \subset \mathcal{A}$$

If $A, B \in \mathcal{A} \implies A \setminus B \in \mathcal{A}$ then we can write

$$A \setminus B = A \cap B^C$$

A measure μ is monotonic, which means if $A, B \in \mathcal{A}$ and $A \subset B$, then

$$\mu(B) = \mu(B \setminus A) + \mu(A) \geq \mu(A)$$

Definition 7.8. A mapping $\mu : \mathcal{P}(\Omega) \to [0, \infty]$ is said to be an outer measure, if $\mu(\varnothing) = 0$ and

$$A \subset \bigcup_{i \in \mathbb{N}} A_i \implies \mu(A) \leq \sum_{i \in \mathbb{N}} \mu(A_i)$$

Just like measures, outer measures are monotonic. Let $\mathcal{I}$ be the family of bounded intervals, i.e.

$$\mathcal{I} = \bigcup_{\substack{x, y \in \mathbb{R} \\ x < y}} \{[x, y], [x, y), (x, y], (x, y)\}$$

We define

$$l([x, y]) := l([x, y)) := l((x, y]) := l((x, y)) = y - x$$

Theorem 7.9. *The mapping*

$$\begin{aligned} \lambda : \mathcal{P}(\mathbb{R}) &\longrightarrow [0, \infty] \\ A &\longmapsto \inf \left\{ \sum_{i=1}^{\infty} l(I_i) \,\middle|\, A \subset \bigcup_{i \in \mathbb{N}} I_i,\ I_i \in \mathcal{I}\ \forall i \in \mathbb{N} \right\} \end{aligned}$$

defines an outer measure on the real numbers. Analogously one can create outer measures on $\mathbb{R}^2, \mathbb{R}^3$.

Proof. We know

$$\lambda(\varnothing) \leq l([0, \epsilon)) = \epsilon \quad \forall \epsilon > 0 \tag{7.1}$$

which implies $\lambda(\varnothing) = 0$. We have to show that

$$A \subset \bigcup_{k \in \mathbb{N}} A_k \implies \lambda(A) \leq \sum_{k \in \mathbb{N}} \lambda(A_k) \tag{7.2}$$

If the right side is ∞ there is nothing to show. So let $\sum_{k\in\mathbb{N}} \lambda(A_k) < \infty$. Let $\epsilon > 0$, then $\forall k \in \mathbb{N} \;\; \exists (I_{k_i}) \subset \mathcal{I}$ such that

$$A_k \subset \bigcup_{i\in\mathbb{N}} I_{k_i} \text{ and } \sum_{i\in\mathbb{N}} l(I_{k_i}) \leq \left(\lambda(A_k) + \frac{\epsilon}{2^k}\right) \tag{7.3}$$

Then

$$A \subset \bigcup_{k=1}^{\infty} A_k \subset \bigcup_{i,k\in\mathbb{N}} I_{k_i} \tag{7.4}$$

and

$$\lambda(A) \leq \sum_{k,i\in\mathbb{N}} l(I_{k_i}) \leq \sum_{k\in\mathbb{N}} \left(\lambda(A_k) + \frac{\epsilon}{2^k}\right) = \sum_{k\in\mathbb{N}} \lambda(A_k) + \epsilon \tag{7.5}$$

Since this inequality holds $\forall \epsilon > 0$

$$\lambda(A) \leq \sum_{k\in\mathbb{N}} \lambda(A_k) \tag{7.6}$$

must be true. The outer measure is not additive. $\square$

Theorem 7.10. *Let μ be an outer measure on $(\Omega, \mathcal{P}(\Omega))$. Then the family of measureable sets*

$$\mathcal{A} := \left\{ A \subset \Omega \,\middle|\, \mu(E) \geq \mu(E \cap A) + \mu(E \cap A^C) \;\;\forall E \in \mathcal{P}(\Omega) \right\}$$

is a σ-algebra, and $\mu|_A$ a measure.

Theorem 7.11. *Firstly, we always have*

$$\mu(E) \leq \mu(E \cap A) + \mu(E \cap A^C) \tag{7.7}$$

which means A is measureable if and only if

$$\mu(E) = \mu(E \cap A) + \mu(E \cap A^C) \;\;\forall E \in \mathcal{P}(\Omega) \tag{7.8}$$

It's easy to see that $\varnothing$ is measurable, and that

$$A \textit{ measurable} \iff A^C \textit{ measurable} \tag{7.9}$$

We have

$$\begin{aligned} E \cap (A \cup B) &= (E \cap A) \cup (E \cap B) \\ &= (E \cap A) \cup (E \cap B \cap A^C) \end{aligned} \tag{7.10}$$

Which means that $\forall A, B$ measurable and $\forall E \in \mathcal{P}(\Omega)$:

$$\begin{aligned} \mu(E) &= \mu(E \cap A) + \mu(E \cap A^C) \\ &= \mu(E \cap A) + \mu(E \cap A^C \cap B) + \mu(E \cap A^C \cap B^C) \\ &\geq \mu(E \cap (A \cup B)) + \mu(E \cap (A \cap B)^C) \geq \mu(E) \end{aligned} \tag{7.11}$$

So $A \cup B$ is measurable and it follows for disjoint A, B

$$\mu(E \cap A) + \mu(E \cap A^C \cap B) = \mu(E \cap (A \cup B)) \tag{7.12a}$$

$$\Longrightarrow \mu(E \cap A) + \mu(E \cap B) = \mu(E \cap (A \cup B)) \tag{7.12b}$$

$$\Longrightarrow \mu \text{ is additive for measurable sets} \tag{7.12c}$$

Then by using induction we can see that finite unions of measurable sets are measurable and that for $A_1, \cdots, A_n$ measurable, pairwise disjoint sets

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu(A_i) \tag{7.13}$$

holds. Now let $(A_i)_{i\in\mathbb{N}}$ be pairwise disjoint measurable sets, and let

$$S_n := \bigcup_{i=1}^{n} A_i \qquad\qquad S := \bigcup_{i=1}^{\infty} A_i \tag{7.14}$$

Then $\forall E \in \mathcal{P}(\Omega)$

$$\mu(E \cap S_n) = \sum_{i=1}^{n} \mu(E \cap A_i) \tag{7.15}$$

To check measurability, consider

$$\begin{aligned} \mu(E) &\geq \mu(E \cap S_n) + \mu(E \cap S_n^C) \\ &\geq \sum_{i=1}^{n} \mu(E \cap A_i) + \mu(E \cap S^C) \end{aligned} \tag{7.16}$$

For $n \to \infty$:

$$\begin{aligned} \mu(E) &\geq \sum_{i=1}^{\infty} \mu(E \cap A_i) + \mu(E \cap S^C) \\ &\geq \mu\,\underbrace{(E \cap S)}_{\bigcup_{i=1}^{\infty} E\cap A_i} + \mu(E \cap S^C) \\ &\geq \mu(E) \end{aligned} \tag{7.17}$$

Thus S is measurable

$$\sum_{i=1}^{\infty} \mu(E \cap A_i) = \mu\left(E \cap \bigcup_{i=1}^{\infty} A_i\right) \tag{7.18}$$

For $E = \Omega$ the σ-additivity follows. It is left to show that for measurable (but not necessarily disjoint) A_i, that $\bigcup_{i=1}^{\infty} A_i$ is also measurable. To do that define

$$B_i = A_i \setminus \left(\bigcup_{j=1}^{i-1} A_j\right) \tag{7.19}$$

Then the B_i are disjoint and measurable. Thus

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i \tag{7.20}$$

is measurable.

Definition 7.12. Application of the previous theorem on the outer measure from $Theorem$ 7.9 gives us the σ-algebra of Lebesgue-measurable sets and the Lebesgue-measure λ.

Remark 7.13. $A \subset \mathbb{R}$ is said to be a null set if its outer measure is 0. Obviously

$$\lambda(\{0\}) = 0$$

For countable A we have

$$\lambda(A) = \lambda\left(\cup_{x\in A}\{x\}\right) \leq \sum_{x\in A} \lambda(\{x\}) = 0$$

So $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$ are null sets. Null sets are measurable, because

$$\forall E \in \mathcal{P}(\mathbb{R}): \quad \underbrace{\lambda(E \cap A)}_{0} + \lambda(E \cap A^C) = \lambda(E \cap A^C) \leq \lambda(E)$$

Theorem 7.14. *Intervals are Lebesgue measurable and*

$$\lambda([a, b]) = b - a$$

Proof. Let A be a bounded interval. Decompose $\mathbb{R}$ into the intervals

$$\mathbb{R} = I_L \cup A \cup I_R \tag{7.21}$$

For $I \in \mathcal{I}$ we have $I \cap I_L,\ I \cap A,\ I \cap I_R$ bounded (or empty) intervals. Now let $E \subset \mathcal{P}(\mathbb{R})$ and

$$E \subset \bigcup_{i\in\mathbb{N}} I_i \tag{7.22}$$

a covering. Then

$$E \cap A \subset \bigcup_{i\in\mathbb{N}} I_i \cap A \qquad\qquad E \cap A^C \subset \bigcup_{i\in\mathbb{N}} ((I_i \cap I_L) \cup (I_i \cap I_R)) \tag{7.23}$$

are coverings of countably many intervals, and we have

$$\begin{aligned}\sum_{i\in\mathbb{N}} l(I_i) &= \sum_{i\in natn} l(I_i \cap A) + \sum_{i\in\mathbb{N}} (l(I_i \cap I_L) + l(I_i \cap I_R)) \\ &\geq \lambda(E \cap A) + \lambda(E \cap A^C)\end{aligned} \tag{7.24}$$

λ is the infimum of all possible coverings

$$\lambda(E) \geq \lambda(E \cap A) + \lambda(E \cap A^C) \tag{7.25}$$

And thus A is measurable. It is left to show that

$$A = [a, b] \implies \lambda(A) = b - a \tag{7.26}$$

So let $(I_n) \subset \mathcal{I}$ such that

$$l = \sum_{n \in \mathbb{N}} (I_n) < b - a \tag{7.27}$$

First, let all I_n be open. Choose

$$A_n = A \setminus \left(\bigcup_{i=1}^{n} I_i \right) \tag{7.28}$$

Those A_n are non-empty, since A cannot be covered by finitely many intervals of length $< b - a$. Choose a sequence $x_n \in A_n \quad \forall n \in \mathbb{N}$. Since A is a compact there exists a toward $x \in A$ convergent subsequence of x_n. The point x cannot be contained in any I_n, since because the I_n are open, infinitely many x_n would be contained in I_n, which would contradict the construction of A_n.

$$\implies (I_n) \text{ do not cover } A \tag{7.29}$$

For arbitrary I_n (so not necessarily open), let (x_k) be the sequence of the (countably many) boundary points of the intervals.

$$\epsilon = \frac{b - a - l}{4} > 0 \tag{7.30}$$

And thus

$$\left\{ \mathring{I}_i \,\middle|\, i \in \mathbb{N} \right\} \cup \left\{ \left(x_k - \frac{\epsilon}{2^k}, x_k + \frac{\epsilon}{2^k} \right) \,\middle|\, \forall k \in \mathbb{N} \right\} \tag{7.31}$$

is a covering of A by countably many open intervals of total length

$$\leq l + \sum_{k=1}^{\infty} \frac{2\epsilon}{2^k} = l + \frac{b - a - l}{2} = \frac{b - a + l}{2} < b - a \tag{7.32}$$

which is impossible due to our construction above. $\square$

Theorem 7.15. *Open and closed sets are Lebesgue measurable.*

Proof. Let $O \subset \mathbb{R}$ be open. It is to show that

$$O = \bigcup_{\substack{l, r \in \mathbb{Q} \\ (l, r) \subset O}} (l, r) \implies O \text{ Lebesgue measurable} \tag{7.33}$$

Let $x \in O$, since O is open

$$\exists \epsilon > 0 : \quad (x - \epsilon, x + \epsilon) \subset O \tag{7.34}$$

Since $\mathbb{Q}$ is dense in $\mathbb{R}$

$$\exists l, r \in \mathbb{Q} : \quad x - \epsilon < l < x \text{ and } x < r < \epsilon + x \tag{7.35}$$

So $x \in (l, r) \subset O$. If C is a closed set, then $\mathbb{R} \setminus C$ is open and thus Lebesgue measurable.

$$\Longrightarrow \; C = \mathbb{R} \setminus (\mathbb{R} \setminus C) \text{ Lebesgue measurable} \tag{7.36}$$

□

Remark 7.16. The Lebesgue-σ-algebra contains many more sets. All sets that are "created by normal means" are Lebesgue measurable.

Remark 7.17. For $A \subset \mathbb{R}$ and $x \in \mathbb{R}$ we define

$$A + x := \{y + x \,|\, y \in A\}$$

A measure on $\mathbb{R}$ is said to be invariant under translation, if

$$\mu(A) = \mu(A + x) \;\; \forall A \in \mathcal{A}, \; x \in \mathbb{R}$$

Since translations of intervals result in intervals, the (outer) Lebesgue measure is invariant under translation. One can show that the Lebesgue measure is the only translational symmetric measure on $\mathbb{R}$, with

$$\lambda([0, 1]) = 1$$

Theorem 7.18. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. For a monotonically increasing sequence $(A_n) \subset \mathcal{A}$ (this means $A_n \subset A_{n+1} \;\; \forall n \in \mathbb{N}$), we have*

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mu(A_n) = \sup_{n\in\mathbb{N}} \mu(A_n)$$

For a monotonically decreasing sequence $(B_n) \subset \mathcal{A}$ we have

$$\mu\left(\bigcap_{n\in\mathbb{N}} B_n\right) = \lim_{n\to\infty} \mu(B_n) = \inf_{n\in\mathbb{N}} \mu(B_n)$$

if $\mu(B_N) < \infty$ for $N \in \mathbb{N}$

Proof. If $\mu(A_n) = \infty$ for some $n \in \mathbb{N}$ there is nothing to show. So let

$$\mu(A_n) < \infty \;\; \forall n \in \mathbb{N} \tag{7.37}$$

Set $A_0 = \varnothing$ and define

$$C_n := A_n \setminus A_{n-1} \tag{7.38}$$

These C_n are pairwise disjoint, and thus

$$\begin{aligned}\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) &= \mu\left(\bigcup_{n\in\mathbb{N}} C_n\right) = \sum_{n=1}^{\infty}\mu(C_n) = \underbrace{\sum_{n=1}^{\infty}\left(\mu(A_n)-\mu(A_{n-1})\right)}_{\text{Telescoping series}} \\ &= \lim_{n\to\infty}\mu(A_n) - \underbrace{\mu(A_0)}_{=0}\end{aligned} \tag{7.39}$$

Now let $\mu(B_N) < \infty \to \mu(B_n) < \infty \;\; \forall n \geq N$. Set

$$D_n = B_N \setminus B_n \;\; \forall n \geq N \tag{7.40}$$

(D_n) is monotonically increasing and thus

$$\bigcup_{n=N}^{\infty} D_n = \bigcup_{n=N}^{\infty} B_N \cap B_n^C = B_N \cap \underbrace{\left(\bigcap_{n=N}^{\infty} B_n\right)}_{B}^{C} = B_N \cap B^C = B_N \setminus B \tag{7.41}$$

Which in turn implies

$$\begin{aligned}\mu(B_N) - \mu(B) = \mu(B_N \setminus B) &= \lim_{n\to\infty}\underbrace{\mu(B_N \setminus B_n)}_{\mu(B_N)-\mu(B_n)} \\ &= \mu(B_N) - \lim_{n\to\infty}\mu(B_n)\end{aligned} \tag{7.42}$$

□

Remark 7.19. $\mu(B_N) < \infty$ for some $N \in \mathbb{N}$ is a necessarily requirement.

7.2 Integrals

Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. The most important example is on $\mathbb{R}$ with the Lebesgue-σ-algebra and the Lebesgue measure. We have one technical requirement, and that is that $(\Omega, \mathcal{A}, \mu)$ is a σ-finite measure space, i.e.

$$\exists\,(E_n) \subset \mathcal{A} : \quad \bigcup_{n\in\mathbb{N}} E_n = \Omega \text{ and } \mu(E_n) < \infty \;\; \forall n \in \mathbb{N}$$

On $\mathbb{R}$ this requirement is fulfilled by defining $E_n = [-n, n]$.

Remark 7.20 (Notation). Let $\Phi(x)$ be a statement depending on $x \in \Omega$. We write $[\Phi]$ for

$$\{x \in \Omega \,|\, \Phi(x)\}$$

Example: $y \in \mathbb{C}$

$$[f = y] = \{x \in \Omega \,|\, f(x) = y\} = f^{-1}(y)$$

We write "Φ holds" for "$\Phi(x)$ holds $\forall x \in \Omega$". For example "$f > g$" instead of "$f(x) > g(x) \ \ \forall x \in \Omega$".

Φ is said to hold "almost everywhere" (a.e.) if the set

$$\{x \mid \neg\Omega(x)\}$$

is a null set. For example, "$f > g$ almost everywhere" means $\mu([f \leq g]) = 0$. The sequence (f_n) converges pointwise a.e. towards f if

$$\left[\lim_{n\to\infty} f_n \neq f\right] = \left\{x \in \Omega \,\middle|\, \lim_{n\to\infty} f_n(x) \neq f(x)\right\}$$

is a null set.

Definition 7.21. Let $A \in \mathcal{A}$, then

$$\begin{aligned} \mathbb{1}_A : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto \begin{cases} 1, & x \in A \\ 0, & \text{else} \end{cases} \end{aligned}$$

is said to be the characteristic function of A. A is the support of $\mathbb{1}_A$. With this we can define the space of simple functions

$$X = \left\{ \sum_{i=1}^{n} a_i \mathbb{1}_A \,\middle|\, n \in \mathbb{N}, A_i \in \mathcal{A},\ \mu(A_i) < \infty,\ a_i \in \mathbb{C} \right\}$$

X^+ notates the non-negative, simple functions.

Remark 7.22. (i) Let $A, B \in \mathcal{A}$

$$\begin{aligned} \mathbb{1}_{A\cap B} &= \mathbb{1}_A \cdot \mathbb{1}_B \\ \mathbb{1}_{A\cup B} &= \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A\cap B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A \mathbb{1}_B \end{aligned}$$

(ii) The set X is a vector space, and the product of characteristic functions is another characteristic function, i.e.

$$f, g \in X \implies f \cdot g \in X$$

Thus X is an algebra.

(iii) If $A_1, \cdots, A_n$ is a decomposition of Ω, which means they are disjoint and

$$\bigcup_{i=1}^{n} A_i = \Omega$$

then

$$(\mathbb{1}_i) = \mathbb{1}_\Omega = \sum_{i=1}^{n} \mathbb{1}_{A_i}$$

(iv) The representation of simple functions as a linear combination is not unique

$$\mathbb{1}_{[0,2]} + \mathbb{1}[2,3] = \mathbb{1}[0,1] + \mathbb{1}[1,3]$$

(v) One can easily see that simple functions can only assume finitely many values, and their support $[f \neq 0]$ has a finite measure. The canonical representation is

$$f = \sum_{y=f(\Omega)} g \cdot \mathbb{1}_{[f=y]}$$

Definition 7.23 (Integrals of simple functions)**.** Let $f \in X$ in canonical representation

$$f = \sum_{i=1}^{\infty} a_i \mathbb{1}_{A_i}$$

Then we define

$$\int f \, \mathrm{d}\mu := \sum_{i=1}^{n} a_i \mu(A_i)$$

Remark 7.24. This sum is always finite, the only A_i with infinite measure is that where $a_i = 0$

$$a_i \cdot A_i = 0 \cdot \infty = 0$$

Let $f = \sum_{j=1}^{m} b_j \mathbb{1}_{B_j}$ be another representation of f, so $B_1, \cdots, B_m$ is a decomposition. If $A_i \cap B_j \neq \varnothing$ i.e.

$$\exists x \in A_i \cap B_j : \quad f(x) = a_i = b_j$$

Then

$$\begin{aligned}
\int f \, \mathrm{d}\mu &= \sum_{i=1}^{n} a_i \mu(A_i) = \sum_{i=1}^{n} a_i \mu \underbrace{\left(A_i \cap \bigcup_{j=1}^{m} B_j \right)}_{\bigcup_{j=1}^{m}(A_i \cap B_j)} = \sum_{i=1}^{n} a_i \sum_{j=1}^{m} \mu(A_i \cap B_j) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} b_j \mu(A_i \cap B_j) = \sum_{j=1}^{m} b_j \mu \left(\left(\bigcup_{i=1}^{n} A_i \right) \cap B_j \right) \\
&= \sum_{j=1}^{m} b_j \mu(B_j)
\end{aligned}$$

Theorem 7.25. *Let f, g be simple functions, $\alpha \in \mathbb{C}$. Then*

$$\int (f + \alpha g) \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu + \alpha \int g \, \mathrm{d}\mu$$

If f, g are real-valued and $f \leq g$ a.e., then

$$\int f \, \mathrm{d}\mu \leq \int g \, \mathrm{d}\mu$$

And especially if $f = g$ a.e.

$$\int f \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu$$

Finally, the triangle inequality holds

$$\left| \int f \,\mathrm{d}\mu \right| \leq \int |f| \,\mathrm{d}\mu$$

Proof. Let f, g be in canonical representation

$$f = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i} \tag{7.43a}$$

$$g = \sum_{j=1}^{m} b_j \mathbb{1}_{B_j} \tag{7.43b}$$

Then

$$\begin{aligned} f + \alpha g &= \sum_{i=1}^{n} a_i \mathbb{1}_{A_i} + \alpha \sum_{j=1}^{m} b_j \mathbb{1}_{B_j} \\ &= \sum_{i=1}^{n} a_i \mathbb{1}_{A_i} \underbrace{\left(\sum_{j=1}^{m} \mathbb{1}_{B_j} \right)}_{\mathbb{1}} + \alpha \sum_{j=1}^{m} b_j \mathbb{1}_{B_j} \underbrace{\left(\sum_{i=1}^{n} \mathbb{1}_{A_i} \right)}_{\mathbb{1}} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} (a_i + \alpha b_j) \mathbb{1}_{A_i \cap B_j} \end{aligned} \tag{7.44}$$

$A_i \cap B_j$ with $i \in \{1, \cdots, n\}, j \in \{1, \cdots, m\}$ is a decomposition of Ω

$$\bigcup_{\substack{i=1 \\ j=1}}^{\substack{m \\ n}} A_i \cap B_j = \bigcup_{i=1}^{n} A_i \cap \underbrace{\left(\bigcup_{j=1}^{m} B_j \right)}_{\Omega} = \Omega \tag{7.45}$$

This means that

$$\begin{aligned} \int (f + \alpha g) \,\mathrm{d}\mu &= \sum_{i=1}^{n} \sum_{j=1}^{m} (a_i + \alpha b_j) \mu(A_i \cap B_j) \\ &= \sum_{i=1}^{n} a_i \mu \left(A \cap \left(\bigcup_{j=1}^{m} B_j \right) \right) + \alpha \sum_{j=1}^{m} b_j \mu \left(\left(\bigcup_{i=1}^{n} A_i \right) \cap B_j \right) \\ &= \int f \,\mathrm{d}\mu + \alpha \int g \,\mathrm{d}\mu \end{aligned}$$

Now let $f \geq 0$ almsot everywhere, i.e. $[f < 0]$ is a null set. If $a_i < 0$, then $A_i \subset [f < 0]$, and then $\mu(A_i) = 0$ and thus the integral is a sum over non-negative values, so it is non-negative itself. If $f \leq g$ a.e., then $g - f \geq 0$ a.e.:

$$0 \leq \int (g - f) \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu - \int f \,\mathrm{d}\mu \tag{7.46}$$

Finally to show the triangle inequality

$$\left| \int f \,\mathrm{d}\mu \right| = \left| \sum_{i=1}^{n} a_i \mu(A_i) \right| \leq \sum_{i=1}^{n} |a_i| \mu(A_i) = \int |f| \,\mathrm{d}\mu \tag{7.47}$$

□

Remark 7.26. From linearity follows, that f can be in any representation

$$f = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$$

and the integral will still be

$$\int f \,\mathrm{d}\mu = \sum_{i=1}^{n} a_i \mu(A_i)$$

Remark 7.27. Notice how the integrals so far did not have any integration variables. The integrals map functions (not their values!) to numbers. If the integration variable is of concern, we can write

$$\int f(x) \,\mathrm{d}\mu(x)$$

For Lebesgue integrals we define

$$\int f(x) \,\mathrm{d}x = \int_{-\infty}^{\infty} f(x) \,\mathrm{d}x$$

Definition 7.28. $f : \Omega \to \Omega$ is said to be measurable, if there is a sequence of simple functions $(f_n) \subset X$ that converge pointwise towards f.

Remark 7.29. (i) For real-valued functions f

$$f \text{ measurable} \iff [f \leq y] \in \mathcal{A} \quad \forall y \in \mathcal{A}$$

(ii) Simple functions and characteristic functions are measurable.

(iii) Continuous functions are Lebesgue measurable.

(iv) Sums, products, quotients (if existant) of measurable sets are measurable.

(v) If (f_n) is a sequence of measurable functions, then

$$\sup_{n\in\mathbb{N}} f_n \qquad \limsup_{n\to\infty} f_n \qquad \lim_{n\to\infty} f_n$$

are measurable if they exist.

All functions from now on will be considered measurable.

Definition 7.30. Let $f : \Omega \to [0,\infty)$, then

$$\int f \,\mathrm{d}\mu := \sup\left\{ \int g \,\mathrm{d}\mu \,\middle|\, g \in X^+,\ g < f \right\}$$

Remark 7.31. (i) This integral can be ∞.

(ii) If f is a non-negative, simple function, then $\forall h$ that are non-negative, simple functions with $h \leq f$

$$\int h \,\mathrm{d}\mu \leq \int f \,\mathrm{d}\mu$$

The old integral (integral over simple functions) is identical to this one.

(iii) Let f, g be non-negative and $f \leq g$ a.e. Define $A = [f \leq g]$. Then for all simple $h < f$ we have

$$h \cdot \mathbb{1}_A \leq g$$

and

$$\int h \,\mathrm{d}\mu = \int h \cdot \mathbb{1}_A \,\mathrm{d}\mu \leq \int g \,\mathrm{d}\mu$$

Which implies

$$\int f \,\mathrm{d}\mu = \sup_h \int h \,\mathrm{d}\mu \leq \int g \,\mathrm{d}\mu$$

Especially

$$\int f \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu \ \text{ if } f = g \text{ a.e.}$$

(iv) If $[f > 0]$ is a null set, then f is the zero function a.e. and

$$\int f \,\mathrm{d}\mu = 0$$

The inverse is also true

$$\int f \,\mathrm{d}\mu = 0 \text{ and } f \geq 0 \implies f = 0 \text{ a.e.}$$

Let $A_k := [f \geq \frac{1}{k}] \in \mathcal{A}$, then

$$\frac{1}{k}\mathbb{1}_{A_k} \leq f \quad \forall k \in \mathbb{N}$$

Since $\int f \, \mathrm{d}\mu = 0$, this implies

$$\int \frac{1}{k} \mathbb{1}_{A_k} \, \mathrm{d}\mu = \frac{1}{k} \mu(A_k) = 0$$
$$\Longrightarrow \mu(A_k) = 0 \quad \forall k \in \mathbb{N}$$

The A_k are monotonically increasing, and thus due to the continuity of the measure

$$0 = \lim_{n \to \infty} \mu(A_k) = \mu \left(\bigcup_{k \in \mathbb{N}} [f \geq \frac{1}{k}] \right) = \mu([f = 0])$$

(v) The definition means $\exists \, (f_n) \subset X^+$ such that $f_n \leq f$

$$\int f_n \, \mathrm{d}\mu \xrightarrow{n \to \infty} \int f \, \mathrm{d}\mu$$

Define $g_n = \max \{f_1, \cdots, f_n\}$. These are also simple functions and $f_n \leq g_n \leq f \;\; \forall n \in \mathbb{N}$.

$$\Longrightarrow \int f_n \, \mathrm{d}\mu \leq \int g_n \, \mathrm{d}\mu \leq \int f \, \mathrm{d}\mu$$

And thus

$$\begin{array}{ccc} \int f_n \, \mathrm{d}\mu & \longrightarrow & \int f \, \mathrm{d}\mu \\ & \downarrow & \\ \int g_n \, \mathrm{d}\mu & \longrightarrow & \int f \, \mathrm{d}\mu \end{array}$$

The sequence g_n is monotonic.

(vi) Let (g_n) be convergent to $g : \Omega \to [0, \infty)$. Then

$$g \leq f \implies \int g \, \mathrm{d}\mu \leq \int f \, \mathrm{d}\mu$$

$\forall n \in \mathbb{N}$ we have $g_n \leq g$, and thus

$$\lim_{n \to \infty} \int g_n \, \mathrm{d}\mu \leq \int g \, \mathrm{d}\mu$$

$\forall f \geq 0$ there exists a monotonically increasing sequence of simple function such that

$$\int g_n \, \mathrm{d}\mu \longrightarrow \int f \, \mathrm{d}\mu$$

and thus $g = f$ a.e.

(vii)

$$\int (cf)\,\mathrm{d}\mu = c \int f\,\mathrm{d}\mu \quad c \in [0,\infty)$$
$$\int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\mu \le \int (f+g)\,\mathrm{d}\mu$$

Theorem 7.32 (Monotone Convergence Theorem)**.** *Let $f \ge 0$ and (f_n) a monotonically increasing sequence of functions converging pointwise to f a.e. Then*

$$\lim_{n\to\infty} \int f_n\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu$$

Proof. First, let $\lim_{n\to\infty} f_n = f$ everywhere. Since (f_n) is monotonic, this must also hold for $\int f_n\,\mathrm{d}\mu$, so

$$\lim_{n\to\infty} \int f_n \mathrm{d}\mu \le \int f \mathrm{d}\mu \tag{7.48}$$

First, consider the special case $(A_n) \subset \mathcal{A}$ monotonically increasing, with

$$\bigcup_{n\in\mathbb{N}} A_n = \Omega \tag{7.49}$$

Then

$$\lim_{n\to\infty} \int f \mathbb{1}_{A_n} \mathrm{d}\mu = \int f \mathrm{d}\mu \tag{7.50}$$

For $f = \mathbb{1}_B$

$$\begin{aligned} \lim_{n\to\infty} \int \underbrace{\mathbb{1}_B \mathbb{1}_{A_n}}_{\mathbb{1}_{B\cap A_n}} \mathrm{d}\mu &= \lim_{n\to\infty} \mu(B\cap A_n) \\ &= \mu(\bigcup_{n\in\mathbb{N}} B \cap A_n) \\ &= \mu(B) = \int \mathbb{1}_B \mathrm{d}\mu \end{aligned} \tag{7.51}$$

Since both sides are lienear in f (at least for simple functions), the equality holds for arbitrary simple functions. Now let $f \ge 0$ be arbitrary and $h \in X^+$, such that for $\epsilon > 0$

$$\int h \mathrm{d}\mu \ge \int f \mathrm{d}\mu - \frac{\epsilon}{2} \tag{7.52}$$

and thus $h \le f$. From this it follows that

$$\exists N \in \mathbb{N}\ \forall n \ge N : \quad \int h \mathbb{1}_{A_n} \mathrm{d}\mu \ge \int h \mathrm{d}\mu - \frac{\epsilon}{2} \tag{7.53}$$

And thus

$$\forall n \geq N: \quad \int h \mathbb{1}_{A_n} \mathrm{d}\mu \geq \int f \mathrm{d}\mu - \epsilon \tag{7.54}$$

which proves Equation (7.50) for arbitrary $f \geq 0$. Now let $c \in (0, 1)$, and set

$$A_n = [f_n \geq cf] \tag{7.55}$$

Since f_n are monotonic, the A_n are as well, and

$$\bigcup_{n \in \mathbb{N}} A_n = \Omega \tag{7.56}$$

Then

$$\int f_n \mathrm{d}\mu \geq \int f_n \mathbb{1}_{A_n} \mathrm{d}\mu \geq \int cf \mathbb{1}_{A_n} \mathrm{d}\mu = c \int f \mathbb{1}_{A_n} \mathrm{d}\mu \tag{7.57}$$

Thus

$$c \int f \mathbb{1}_{A_n} \mathrm{d}\mu \xrightarrow{n \to \infty} c \int f \mathrm{d}\mu \tag{7.58}$$

Which in turn implies

$$\lim_{n \to \infty} \int f_n \mathrm{d}\mu \geq c \int f \mathrm{d}\mu \tag{7.59}$$

For $c \to 1$ we have

$$\lim_{n \to \infty} \int f_n \mathrm{d}\mu = \int f \mathrm{d}\mu \tag{7.60}$$

And if $f_n \to f$ only a.e.

$$A = [\lim_{n \to \infty} f_n = f] \tag{7.61}$$

then $\Omega \setminus A$ is a null set.

$$\begin{aligned} \lim_{n \to \infty} \int f_n \mathrm{d}\mu &= \lim_{n \to \infty} \int f_n \mathbb{1}_A \mathrm{d}\mu \\ &= \int f \mathbb{1}_A \mathrm{d}\mu \\ &= \int f \mathrm{d}\mu \end{aligned} \tag{7.62}$$

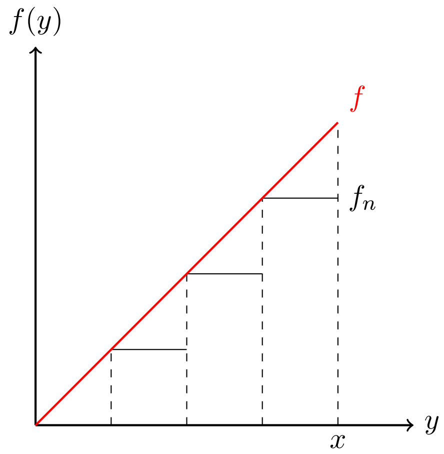
□

Example 7.33. Calculate the integral of $f(y) = y \mathbb{1}_{[0,x]}(x)$

$$f_n = \sum_{k=0}^{2^n - 1} k \frac{1}{2^n} \mathbb{1}_{[k \frac{x}{2^n}, (k+1) \frac{x}{2^n}]}$$

is a monotonically increasing sequence which converges to f on $\mathbb{R} \setminus \{x\}$.

$$\begin{aligned}\int f_n \mathrm{d}\mu = \sum_{k=0}^{2^n-1} k\frac{x}{2^n} \cdot \left(\frac{x}{2^n}\right) &= \frac{x^2}{2^{2n}} \sum_{k=0}^{2^n-1} k \\ &= \frac{x^2}{2^{2n}} \frac{2^n(2^n-1)}{2 \cdot 2^n} \\ &= \frac{x^2}{2} \frac{2n-1}{2^n} \\ &\longrightarrow \frac{x^2}{2}\end{aligned}$$



Example 7.34. Consider $f_n = n\mathbb{1}_{(0,\frac{1}{n})}$. This sequence converges pointwise to the zero function. But

$$\int f_n \mathrm{d}\mu = n \cdot \frac{1}{n} = 1 \neq 0$$

This is due to f_n not being monotonic increasing.

Example 7.35. Let $(a_n) \subset \mathbb{C}$, and define

$$f_n = a_n \mathbb{1}_{[n,n+1]}$$

This sequence converges pointwise to 0, but

$$\int f_n \mathrm{d}\lambda = a_n$$

depends on (a_n) and can converge to any value (or even diverge).

Definition 7.36. A function $f : \Omega \to \mathbb{C}$ is said to be integrable if

$$\int |f| \mathrm{d}\mu < \infty$$

A sequence of simple functions (f_n) is said to be an approximating sequence of f if

$$\int |f - f_n| \mathrm{d}\mu \xrightarrow{n\to\infty} 0$$

Corollary 7.37. *Let $f, g \geq 0$*

$$\int (f+g)\mathrm{d}\mu = \int f \mathrm{d}\mu + \int g \mathrm{d}\mu$$

Proof. Let $(f_n), (g_n) \subset X^+$ be monotone sequences with $f_n \to f$, $g_n \to g$ almost everywhere. Then $(f_n + g_n)$ is monotonically increasing as well and converge pointwise to $(f+g)$ almost everywhere.

$$\begin{aligned} & \left[\lim_{n\to\infty} f_n \neq f\right] \text{ null set, } \left[\lim_{n\to\infty} g_n \neq g\right] \text{ null set} \\ \implies & \left[\lim_{n\to\infty} f_n \neq f\right] \cup \left[\lim_{n\to\infty} g_n \neq g\right] \text{ null set} \end{aligned} \tag{7.63}$$

This implies

$$\begin{aligned} \int (f+g)\mathrm{d}\mu = \lim_{n\to\infty} \int (f_n + g_n)\mathrm{d}\mu &= \lim_{n\to\infty} \int f_n \mathrm{d}\mu + \lim_{n\to\infty} \int g_n \mathrm{d}\mu \\ &= \int f \mathrm{d}\mu + \int g \mathrm{d}\mu \end{aligned} \tag{7.64}$$

□

Remark 7.38. (i) The set of integrable functions is a vector space, because for f, g integrable and $\alpha \in \mathbb{C}$

$$\begin{aligned} \int |f + \alpha g| \mathrm{d}\mu &\leq \int |f| + |\alpha||g| \mathrm{d}\mu \\ &= \int |f| \mathrm{d}\mu + |\alpha| \int |g| \mathrm{d}\mu < \infty \end{aligned}$$

However, $f \cdot g$ is not integrable in general!

(ii) Let $f \geq 0$ be integrable, and $(f_n) \subset X^+$ such that $f_n \to f$ pointwise a.e.

$$\lim_{n\to\infty} \int f_n \mathrm{d}\mu = \int f \mathrm{d}\mu < \infty$$

$\forall n \in \mathbb{N}$:

$$\int |f - f_n| \mathrm{d}\mu = \int (f - f_n) \mathrm{d}\mu = \int f \mathrm{d}\mu - \int f_n \mathrm{d}\mu \xrightarrow{n\to\infty} 0$$

(iii) Let $f : \Omega \to \mathbb{R}$ be a function. Decompose the function into a positive and a negative part:

$$f_+ := f \cdot \mathbb{1}_{[f\geq 0]} \qquad\qquad f_- := -f \cdot \mathbb{1}_{[f\leq 0]}$$

$f_+, f_- \geq 0$, and

$$f = f_+ - f_- \qquad\qquad |f| = f_+ + f_-$$

(iv) $|\mathrm{Re}\, f| \leq |f|$, $|\mathrm{Im}\, f| \leq |f|$. If f is integrable, then $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are also integrable.

(v) Let f, g be arbitrary, and $(f_n), (g_n)$ approximating sequences for f and g. Then for $\alpha \in \mathbb{C}$:

$$\begin{aligned}\int |f + \alpha g - (f_n + \alpha g_n)| \mathrm{d}\mu &\leq \int |f - f_n| \mathrm{d}\mu + \alpha \int |g - g_n| \\ &= \int |f - f_n| \mathrm{d}\mu + |\alpha| \int |g - g_n| \mathrm{d}\mu \\ &\xrightarrow{n\to\infty} 0\end{aligned}$$

Thus $f_n + \alpha g_n$ is an approximating sequence for $f + \alpha g$

(vi) Consider

$$f = \left((\mathrm{Re}\, f)_+ - (\mathrm{Re}\, f)_-\right) + i \left((\mathrm{Im}\, f)_+ - (\mathrm{Im}\, f)_-\right)$$

If f is integrable, then all the terms are integrable as well and have approximating sequences. Thus, f has an approximating sequence.

(vii) Now let (f_n) be an approximating sequence for f. Let $\epsilon > 0$, then

$$\exists N \in \mathbb{N}\ \forall n \geq N : \quad \int |f - f_n| \mathrm{d}\mu < \frac{\epsilon}{2}$$

$\forall n, m \geq N$

$$\begin{aligned}\left|\int f_n \mathrm{d}\mu - \int f_m \mathrm{d}\mu\right| &= \left|\int (f_n + f_m) \mathrm{d}\mu\right| \\ &\leq \int |f_n - f_m| \mathrm{d}\mu \\ &\leq \int \left(|f_n - f| + |f - f_m|\right) \mathrm{d}\mu \\ &< \epsilon\end{aligned}$$

Which means $(\int f_n \mathrm{d}\mu)$ is a Cauchy sequence, so it converges to $I \in \mathbb{C}$

(viii) Let (g_n) be another approximating sequence for f

$$\begin{aligned}\left|\int f_n \mathrm{d}\mu - \int g_n \mathrm{d}\mu\right| &\leq \int |f_n - g_n| \mathrm{d}\mu \\ &\leq \int |f_n - f| \mathrm{d}\mu + \int |f - g_n| \mathrm{d}\mu \xrightarrow{n\to\infty} 0\end{aligned}$$

So the integral is invariant to the choice of approximating sequence.

Definition 7.39. Let f be integrable, and define

$$\int f \mathrm{d}\mu = \lim_{n\to\infty} \int f_n \mathrm{d}\mu$$

for some approximating sequence (f_n) of f.

Remark 7.40. If f is a simple function, then $(f_n)_{n\in\mathbb{N}}$ is an approximating sequence. The new integral definition is compatible with the integral for simple functions. and with the integral for non-negative functions.

Theorem 7.41. *Let f, g be integrable.*

(i)

$$\forall \alpha \in \mathbb{C} : \quad \int (f + \alpha g) \mathrm{d}\mu = \int f \mathrm{d}\mu + \alpha \int g \mathrm{d}\mu$$

(ii) If $f \leq g$ a.e., then

$$\int f \mathrm{d}\mu \leq \int f \mathrm{d}\mu$$

and

$$f = g \text{ a.e.} \implies \int f \mathrm{d}\mu = \int g \mathrm{d}\mu$$

(iii)

$$\left|\int f \mathrm{d}\mu\right| \leq \int |f| \mathrm{d}\mu$$

Proof. Let (f_n), (g_n) be approximating sequences for f and g. Then $(f_n + \alpha g_n)$ is an approximating sequence for $(f + \alpha g)$.

$$\int (f + \alpha g) \mathrm{d}\mu = \lim_{n\to\infty} \int (f_n + \alpha g_n) \mathrm{d}\mu = \underbrace{\lim_{n\to\infty} \int f_n \mathrm{d}\mu}_{\int f \mathrm{d}\mu} + \underbrace{\lim_{n\to\infty} \int g_n \mathrm{d}\mu}_{\int g \mathrm{d}\mu} \tag{7.65}$$

To prove the second statement, let $f \leq g$ a.e. Then $(g - f)_- = 0$ a.e.

$$\implies \int (g - f)_- = 0 \tag{7.66}$$

And thus

$$\begin{aligned}\int g\mathrm{d}\mu - \int f\mathrm{d}\mu &= \int (g-f)\mathrm{d}\mu \\ &= \int (g-f)_+\mathrm{d}\mu - \int (g-f)_-\mathrm{d}\mu \geq 0\end{aligned} \tag{7.67}$$

The final statement is proven by applying the reverse triangle inequality

$$\int ||f| - |f_n||\mathrm{d}\mu \leq \int |f - f_n|\mathrm{d}\mu \xrightarrow{n\to\infty} = 0 \tag{7.68}$$

This means if $(|f_n|)$ is an approximating sequence for $|f|$, then

$$\int |f|\mathrm{d}\mu = \lim_{n\to\infty} \int |f_n|\mathrm{d}\mu \geq \lim_{n\to\infty} \left|\int f_n\mathrm{d}\mu\right| = \left|\int f\mathrm{d}\mu\right| \tag{7.69}$$

□

Remark 7.42. For $A \subset \mathcal{A}$ we define

$$\int_A g\mathrm{d}\mu := \int g\mathbb{1}_A\mathrm{d}\mu$$

$g\mathbb{1}_A$ can be integrable, even if g isn't. The above integral doesn't depend on the behavior of g outside of A. We use $\int_A g\mathrm{d}\mu$ even if g isn't defined outside of A. Integrals are independent from the behavior on null sets, so

$$\int_{-1}^{1} \frac{1}{x}\mathrm{d}x = 0$$

is perfectly fine, even though the integrand is not defined for $x = 0$.

Example 7.43. Let $\Omega = \mathbb{N}$, $\mathcal{A} = \mathcal{P}(\Omega)$ and μ the counting measure. Let $f : \mathbb{N} \to [0,\infty)$, then

$$f_N = f\mathbb{1}_{\{1,\cdots,N\}} = \sum_{n=1}^{N} f(n)\mathbb{1}_{\{n\}}$$

is a sequence of monotonically increasing, simple functions that converge to f pointwise.

$$\int f\mathrm{d}\mu = \lim_{N\to\infty} \int f_N\mathrm{d}\mu = \lim_{N\to\infty} \sum_{n=1}^{N} f(n)\mu(\{n\}) = \sum_{n=1}^{\infty} f(n)$$

Thus we can conclude

$$f : \mathbb{N} \to \mathbb{C} \text{ integrable } \iff \int |f|\mathrm{d}\mu = \sum_{n=1}^{\infty} |f(n)| < \infty$$

and

$$\int f\mathrm{d}\mu = \sum_{n=1}^{\infty} f(n)$$

7.3 Integrals over the real numbers

Definition 7.44. Let $a, b \in \mathbb{R}$, $a < b$ and $f : [a, b] \to \mathbb{C}$ integrable. Then set

$$\int_a^b f(x)\mathrm{d}x := \int_{(a,b)} f\mathrm{d}\lambda = \int f \cdot \mathbb{1}_{(a,b)}\mathrm{d}\lambda$$

and

$$\int_b^a f(x)\mathrm{d}x = -\int_a^b f(x)\mathrm{d}x$$

Remark 7.45. Let $a, b \in \mathbb{R}$, $a < b$, then every bounded function is integrable over (a, b)

$$\int_{(a,b)} |f|\mathrm{d}\lambda \le \int_{(a,b)} \underbrace{\sup_{x\in(a,b)} |f(x)|}_{\in\mathbb{R}} \mathrm{d}\lambda = \|f\|_\infty \underbrace{\int_{(a,b)} \mathbb{1}_{(a,b)}\mathrm{d}\lambda}_{\lambda((a,b))} = \|f\|_\infty \cdot (b-a)$$

If f is continuous on $[a, b]$ then it is also bounded. Let $a < c < b$

$$\begin{aligned}\int_a^b f(x)\mathrm{d}x = \int f\dot{\mathbb{1}}_{(a,b)}\mathrm{d}\lambda &= \int f \cdot (\mathbb{1}_{(a,c)} + \mathbb{1}_{(c,b)})\mathrm{d}\lambda \\ &= \int f \cdot \mathbb{1}_{(a,c)}\mathrm{d}\lambda + \int f \cdot \mathbb{1}_{(c,b)}\mathrm{d}\lambda \\ &= \int_a^c f(x)\mathrm{d}x + \int_c^b f(x)\mathrm{d}x\end{aligned}$$

One can easily see that this formula holds for any $c \in \mathbb{R}$.

Theorem 7.46 (Mean value theorem for integrals)**.** *Let $a, b \in \mathbb{R}$, $a < b$ and $f, g : [a, b] \to \mathbb{R}$ continuous with $g \ge 0$. Then $\exists \xi \in (a, b)$ such that*

$$\int_a^b f(x)g(x)\mathrm{d}x = f(\xi)\int_a^b g(x)\mathrm{d}x$$

Especially, $\exists \eta \in (a, b)$ such that

$$\int_a^b f(x)\mathrm{d}x = f(\eta)(b-a)$$

Proof. Let f be continuous, and $[a, b]$ compact. Then define

$$c = \min_{a\le x\le b} f(x) \qquad\qquad C = \max_{a\le x\le b} f(x)$$

Thus,

$$\exists x_m, x_M \in [a, b] : \quad f(x_m) = c, \; f(x_M) = C \tag{7.70}$$

Define $\tilde{a} := \min\{x_m, x_M\}$ and $\tilde{b} := \max\{x_m, x_M\}$. Then

$$c \cdot g(x) \leq f(x)g(x) \leq Cg(x) \tag{7.71}$$

If we define

$$I = \int_a^b g(x)\mathrm{d}x \tag{7.72}$$

then we have

$$c \cdot I \leq \int_a^b \leq C \cdot I \tag{7.73}$$

Due to the mean value theorem, $\exists \xi \in (\tilde{a}, \tilde{b}) \subset (a, b)$ such that

$$f(\xi) = \frac{1}{I}\int_a^b f(x)g(x)\mathrm{d}x \tag{7.74}$$

□

Definition 7.47. Let $a, b \in \mathbb{R}$, $a < b$ and $f : [a, b] \to \mathbb{C}$. Then

$$F : [a, b] \to \mathbb{C}$$

is said to be the antiderivative of f, if it is continuous, on $[a, b]$ differentiable and $F' = f$.

Remark 7.48. Let F, G be antiderivatives of f. Then on (a, b) we have

$$(F - G)' = F' - G' = f - f = 0$$

Thus $F - G = c$ for $c \in \mathbb{C}$. Since F, G are continuous, $F - G = c$ also holds on $[a, b]$.

Theorem 7.49 (Fundamental Theorem of Calculus)**.** *Let $a, b \in \mathbb{R}$, $a < b$ and $f : [a, b] \to \mathbb{C}$ continuous. Then for arbitrary $x_0 \in [a, b]$ the function*

$$x \longmapsto \int_{x_0}^x f(y)\mathrm{d}y$$

is an antiderivative of f. Let G be an antiderivative of f, then

$$\int_a^b f(y)\mathrm{d}y = G(b) - G(a)$$

Proof. First, let f be real-vauled.

$$F(x) := \int_{x_0}^x f(y)\mathrm{d}y \tag{7.75}$$

For a fixed $x \in [a,b]$ and h such that $x+h \in [a,b]$ we have

$$\begin{aligned} F(x+h) - F(x) &= \int_{x_0}^{x+h} f(y)\mathrm{d}y - \int_{x_0}^{x} f(y)\mathrm{d}y \\ &= \int_{x}^{x+h} f(y)\mathrm{d}y = f(\xi_h) \cdot h \end{aligned} \tag{7.76}$$

with $\xi_h \in (x, x+h)$ from the mean value theorem. For $h \to 0$, the ξ_h converges to x, and thus $f(\xi_h) \to f(x)$

$$\implies \lim_{h \to 0} (F(x+h) - F(x)) = 0 \tag{7.77}$$

so F is continuous. For $x \in (a,b)$ we have $x+h \in (a,b)$ for a small enough h, and then

$$\lim_{h \to 0} \frac{F(x+h) - F(x)}{h} = \lim_{h \to 0} f(\xi_k) = f(x) = F'(x) \tag{7.78}$$

If G is another antiderivative then $G = F + c$ with $c \in \mathbb{R}$.

$$\int_a^b f(y)\mathrm{d}y = \int_a^{x_0} f(y)\mathrm{d}y + \int_{x_0}^b f(y)\mathrm{d}y = F(b) - F(a) = G(b) - G(a) \tag{7.79}$$

For complex-valued f, simply decompose f into a real and imaginary part. $\square$

Remark 7.50. The antiderivative of f is often denoted as

$$\int f(x)\mathrm{d}x \qquad \text{indefinite integral}$$

This notation is also used for

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x \qquad \text{definite integral}$$

Corollary 7.51 (Partial Integration)**.** *Let $a, b \in \mathbb{R}$, $a < b$ and $f, g : [a, c] \to \mathbb{C}$ continuously differentiable. Then*

$$\int f'(x)g(x)\mathrm{d}x = f(x)g(x) - \int f(x)g'(x)\mathrm{d}x$$

And the definite integral is

$$\int_a^b f'(x)g(x)\mathrm{d}x = f(b)g(b) - f(a)g(a) - \int_a^b f(x)g'(x)\mathrm{d}x$$

Proof. Let $H : [a,b] \to \mathbb{C}$ be the antiderivative of fg'. Then $fg - H$ is continuously differentiable, and

$$(fg - H)' = f'g + fg' - H' = f'g \tag{7.80}$$

so $fg - H$ is an antiderivative of $f'g$. From the fundamental theorem follows

$$\begin{aligned}\int_a^b f'(x)g(x)\mathrm{d}x &= (fg-H)(b) - (fg-H)(a) \\ &= f(b)g(b) - f(a)g(a) - \underbrace{(H(b)-H(a))}_{\int_a^b f(x)g'(x)\mathrm{d}x}\end{aligned} \tag{7.81}$$

$\square$

Corollary 7.52 (Substitution). *Let $a, b \in \mathbb{R}$, $a < b$ and $g : [a, b] \to \mathbb{R}$ continuously differentiable. Choose $\xi = \min g([a, b])$ and $\eta = \max g([a, b])$. Let*

$$f : [\xi, \eta] \longrightarrow \mathbb{C}$$

be continuous. Then

$$\int f(g(x))g'(x)\mathrm{d}x = \int f(y)\mathrm{d}y$$

for ($y = g(x)$), and

$$\int_a^b f(g(x))g'(x)\mathrm{d}x = \int_{g(a)}^{g(b)} f(y)\mathrm{d}y$$

Proof. Let F be an antiderivative of f, then $F \circ g$ is continuously differentiable, and due to the chain rule

$$(F \circ g)'(x) = F'(g(x))g'(x) = f(g(x))g'(x) \tag{7.82}$$

thus $F \circ g$ is an antiderivative of $(f \circ g)g'$

$$\begin{aligned}\int_a^b f(g(x))g'(x)\mathrm{d}x &= (F \circ g)(b) - (F \circ g)(a) = F(g(b)) - F(g(a)) \\ &= \int_{g(a)}^{g(b)} f(y)\mathrm{d}y\end{aligned} \tag{7.83}$$

$\square$

Example 7.53. Consider

$$\tan x = \frac{\sin x}{\cos x} = -\frac{\cos' x}{\cos x}$$

We have to determine the antiderivative of $f(y) = \frac{1}{y}$ with $y = \cos x$

$$-\int \frac{1}{y}\mathrm{d}y = -\ln y$$

After resubstituting we get

$$\int \tan x \mathrm{d}x = -\ln|\cos x|$$

The derivative of this function is identical to tan wherever it is defined. If we want to calculate definite integrals like

$$\int_a^b \tan x \mathrm{d}x$$

there cannot be any incontinuities between a and b.

Example 7.54. Consider

$$\begin{aligned} F : (0, \infty) &\longrightarrow \mathbb{R} \\ a &\longmapsto \int_0^\infty \frac{e^{-x}}{x+a} \mathrm{d}x \end{aligned}$$

Is this function continuous?

Corollary 7.55. *Let (X, d) be a metric space, $f : \Omega \times X \to \mathbb{C}$ and $\tilde{a} \in X$. Let $f(\cdot, a)$ be integrable $\forall a \in X$ and let $f(\omega, \cdot)$ be continuous in $\tilde{a}$ $\forall \omega \in \Omega$. Let U be a neighbourhood of $\tilde{a}$ and g an integrable function (independent from a) such that*

$$|f(\omega, a)| \leq g(\omega) \quad \forall \omega \in \Omega \ \forall a \in U$$

Then the function

$$\begin{aligned} F : X &\longrightarrow \mathbb{C} \\ a &\longmapsto \int f(\omega, a) \mathrm{d}\mu(\omega) \end{aligned}$$

is continuous in $\tilde{a}$.

Proof. Let $(a_n) \subset X$ be a sequence with $a_n \to \tilde{a}$. Set $f_n = f(\cdot, a_n)$. For sufficiently bit n, a_n is in the neighbourhood U, and thus

$$|f_n| = |f(\cdot, a_n)| \leq g \tag{7.84}$$

Then $\forall \omega \in \Omega$

$$\lim_{n\to\infty} f_n(\omega) = \lim_{n\to\infty} f(\omega, a_n) = f(\omega, \tilde{a}) \tag{7.85}$$

And

$$\begin{aligned} \lim_{n\to\infty} F(a_n) &= \lim_{n\to\infty} \int f_n(\omega) \mathrm{d}\mu(\omega) \\ &= \int \lim_{n\to\infty} f(\omega, a_n) \mathrm{d}\mu \\ &= \int f(\omega, \tilde{a}) \mathrm{d}\mu(\omega) \\ &= F(\tilde{a}) \end{aligned} \tag{7.86}$$

The sequence criterion for continuity tells os that F is continuous in $\tilde{a}$. □

Example 7.56. Let $\tilde{a} \in (0, \infty)$. Then

$$\forall a \in \left(\frac{\tilde{a}}{2}, \infty\right) \ \forall x \in [0, \infty) : \quad \frac{e^{-x}}{x+a} \leq \frac{e^{-x}}{\frac{\tilde{a}}{2}} = \frac{2e^{-x}}{\tilde{a}} \text{ integrable}$$

Thus, F is continuous in $\tilde{a}$. Since $\tilde{a}$ was arbitrary, F is continuous.

Corollary 7.57. *Let $X \subset \mathbb{R}^n$ be open, $f : \Omega \times X \to \mathbb{C}$ and $\tilde{a} \in X$, $f(\cdot, a)$ integrable $\forall a \in X$. Let U be a neighbourhood of $\tilde{a}$, and $f(\omega, \cdot)$ differentiable $\forall \omega \in \Omega$ in every point of U. Let g be integrable (independent from a) such that*

$$\|D_a f(\omega, a)\| \leq g(\omega)$$

Then the function

$$\begin{aligned} F : X &\longrightarrow \mathbb{C} \\ a &\longmapsto \int f(\omega, a) \mathrm{d}\mu(\omega) \end{aligned}$$

is differentiable in $\tilde{a}$ and

$$DF(\tilde{a}) = \int D_a f(\omega, \tilde{a}) \mathrm{d}\mu(\omega)$$

Proof. Without proof. $\square$

Example 7.58. The term

$$\frac{e^{-x}}{x+a}$$

is differentiable in terms of a

$$\left| \frac{\mathrm{d}}{\mathrm{d}a} \frac{e^{-x}}{x+a} \right| = \frac{e^{-x}}{(x+a)^2} \leq \frac{4}{\tilde{a}^2} e^{-x} \quad \forall a \in \left(\frac{\tilde{a}}{2}, \infty\right) \quad \forall x \in [0, \infty)$$

Thus F is differentiable and

$$F'(a) = -\int \frac{e^{-x}}{(x+\tilde{a})^2} \mathrm{d}x$$

Since $\tilde{a}$ was arbitrary, F is differentiable.

7.4 Product measures and Fubini's Theorem

Example 7.59. Let

$$f : [0,1] \times [0,1] \longrightarrow [0, \infty)$$

Question: What is the volume (or the λ^2 measure) under the graph of f, i.e.

$$\left\{ (x, y, z) \in \mathbb{R}^3 \,\middle|\, 0 \leq z \leq f(x, y) \right\}$$

The possibilities are

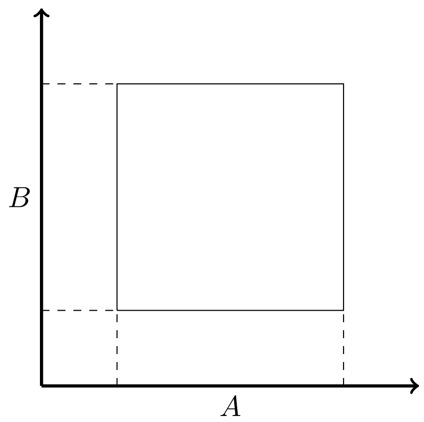
$$\int f \mathrm{d}\lambda^2$$

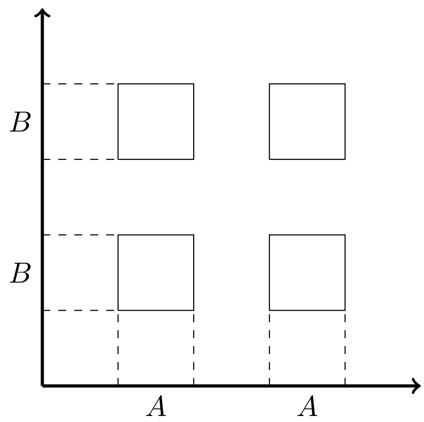
$$\int_0^1 \int_0^1 f(x,y)\mathrm{d}x\mathrm{d}y \quad \text{or} \quad \int_0^1 \int_0^1 f(x,y)\mathrm{d}y\mathrm{d}x$$

From now on we define $(\Omega, \mathcal{A}, \mu)$ and $(\Phi, \mathcal{B}, \nu)$ to be measure spaces.

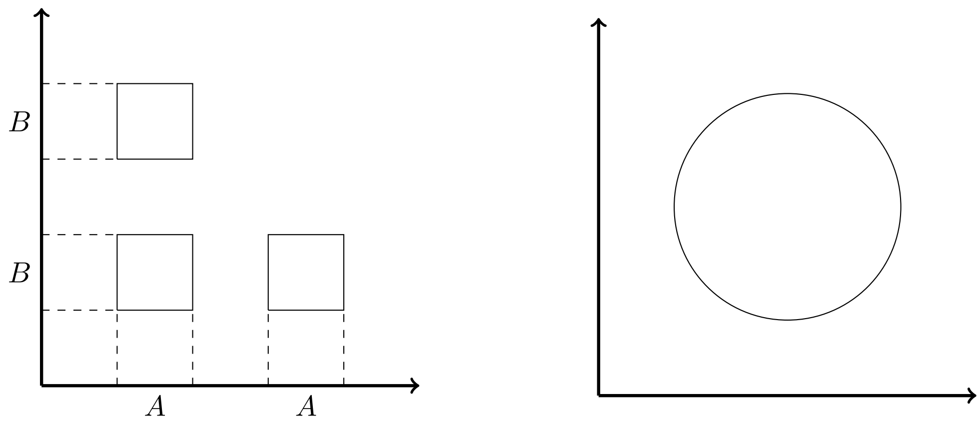
Definition 7.60. The product σ-algebra $\mathcal{A} \otimes \mathcal{B}$ is the smallest σ-algebra on $\Omega \times \Phi$ that contains all sets of type $A \times B$ for $A \in \mathcal{A}, B \in \mathcal{B}$.

Examples for $A \times B$ are





NO examples for $A \times B$ are



A measure ϑ defined on $\mathcal{A} \otimes \mathcal{B}$ is said to be a product measure of μ, ν if

$$\vartheta(A \times B) = \mu(A)\nu(B) \quad A \in \mathcal{A}, B \in \mathcal{B}$$

Remark 7.61. Product measures always exist. For σ-finite measure spaces they are unique. Notation:

$$\mu \otimes \nu \quad \text{or} \quad \mu^2 = \mu \otimes \mu$$

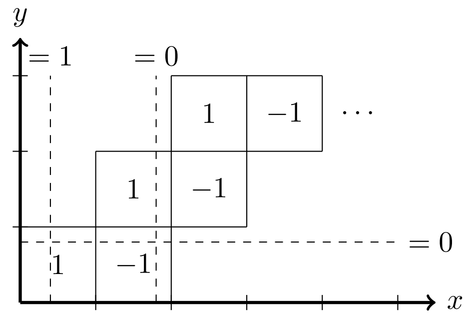
Example 7.62. $\mathbb{R}$ with Lebesgue measure λ. λ^2 is the product measure on $\mathbb{R}^2$.

$$\begin{aligned} \lambda^2([a,b] \times [c,d]) &= \lambda([a,b])\lambda([c,d]) \\ &= (b-a)(d-c) \end{aligned}$$

This means the product measure characterizes the area. Analogously this can be extended for λ^3, λ^4 etc.

Example 7.63. Consider

$$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ f &= \sum_{n=0}^{\infty} \left(\mathbb{1}_{[n,n+1)^2} - \mathbb{1}_{[n+1,n+2)\times[n,n+1)} \right) \end{aligned}$$



$$\iint f(x,y)\mathrm{d}x\mathrm{d}y = 0 \qquad\qquad \iint f(x,y)\mathrm{d}y\mathrm{d}x = 1$$

But the integral $\int f \mathrm{d}\lambda^2$ doesn't exist

$$\int |f| \mathrm{d}\lambda^2 = \sum_{n=0}^{\infty} 2 = \infty$$

Theorem 7.64 (Tonelli's Theorem)**.** *Let $f : \Omega \times \Phi \to [0,\infty)$ be measurable (in terms of $\mathcal{A} \otimes \mathcal{B}$). Then the functions*

$$\omega \longmapsto f(\omega,\phi)$$

are measurable for almost all $\phi \in \Phi$. Analogously

$$\phi \longmapsto f(\omega,\phi)$$

is measurable for almost all $\omega \in \Omega$.

$$\phi \longmapsto \int f(\omega,\phi)\mathrm{d}\mu(\omega) \ \textit{measurable}$$
$$\omega \longmapsto \int f(\omega,\phi)\mathrm{d}\mu(\phi) \ \textit{measurable}$$

and

$$\begin{aligned}\int f(\omega,\phi)\,\mathrm{d}(\mu\otimes\nu)\,(\omega,\phi) &= \iint f(\omega,\phi)\mathrm{d}\mu(\omega)\mathrm{d}\nu(\phi)\\ &= \iint f(\omega,\phi)\mathrm{d}\nu(\phi)\mathrm{d}\mu(\omega)\end{aligned}$$

Furthermore, f is integrable in terms of $\mu \otimes \nu$ is one of the above integrals is finite.

Proof. Without proof. □

Corollary 7.65 (Cavalieri's Principle)**.** *Let $A \subset \mathcal{A} \otimes \mathcal{B}$. Define*

$$A_\omega = \{\phi \in \Phi \mid (\omega,\phi) \in A\}$$

Then

$$\omega \longmapsto \nu(A_\omega)$$

is measurable and

$$(\mu\otimes\nu)(A) = \int \nu(A_\omega)\mathrm{d}\mu(A_\omega)$$

Proof. It is easy to see

$$(\omega, \phi) \in A \iff \phi \in A_\omega \tag{7.87}$$

Thus we can see

$$\mathbb{1}_A(\omega, \phi) = \mathbb{1}_{A_\omega}(\phi) \tag{7.88}$$

And then

$$\begin{aligned}(\mu \otimes \nu)(A) &= \int \mathbb{1}_A(\omega, \phi)\,\mathrm{d}(\mu \otimes \nu)\,(\omega, \phi) \\ &= \iint \underbrace{\mathbb{1}_A(\omega, \phi)}_{\mathbb{1}_{A_\omega}(\phi)}\,\mathrm{d}\nu\,(\phi)\,\mathrm{d}\mu\,(\omega) \\ &= \int \nu(A_\omega)\,\mathrm{d}\mu\,(\omega)\end{aligned} \tag{7.89}$$

□

Theorem 7.66 (Fubini's Theorem)**.** *Let $f : \Omega \times \Phi \to \mathbb{K}$ be measurable with measures μ, ν, which is integrable in terms of $\mu \otimes \nu$. Then the functions $\omega \mapsto f(\omega, \phi)$ are measurable and integrable for ν-almost every $\phi \in \Phi$, and the functions $\phi \mapsto f(\omega, \phi)$ are measurable and integrable for μ-almost every $\omega \in \Omega$. The functions*

$$\omega \longmapsto \int f(\omega, \phi)\,\mathrm{d}\nu\,(\phi) \qquad\qquad \phi \longmapsto \int f(\omega, \phi)\,\mathrm{d}\mu\,(\omega)$$

are measurable and integrable, and

$$\begin{aligned}\int f(\omega, \phi)\,\mathrm{d}(\mu \otimes \nu) &= \iint f(\omega, \phi)\,\mathrm{d}\nu\,(\phi)\,\mathrm{d}\mu\,(\omega) \\ &= \iint f(\omega, \phi)\,\mathrm{d}\mu\,(\omega)\,\mathrm{d}\nu\,(\phi)\end{aligned}$$

Proof. Without proof. □

Corollary 7.67. *Let $a_i, b_i \in \mathbb{R}$, $a_i < b_i \quad \forall i \in \{1, \cdots, n\}$.*

$$D = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$$

Let $f : D \to \mathbb{R}$ be continuous or bounded. Then

$$\int_D f\,\mathrm{d}\lambda^n = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \cdots, x_n)\,\mathrm{d}x_n \cdots \mathrm{d}x_1$$

and the order of integration is irrelevant.

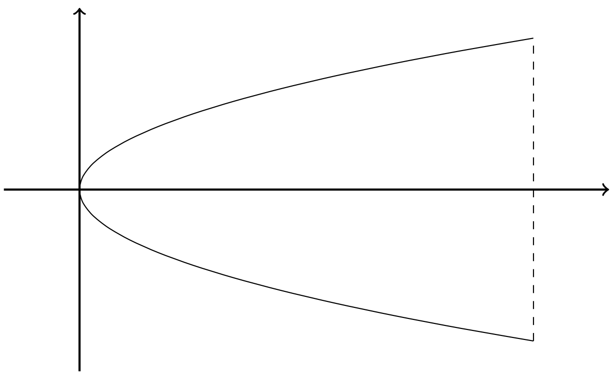
Proof. f is bounded by $k \in \mathbb{R}$ (continuous with compact domain)

$$\int_D |f| \, \mathrm{d}\lambda^n \leq \int_D k \, \mathrm{d}\lambda^n = k \cdot (b_1 - a_1)(b_2 - a_2) \cdots (b_n - a_n) < \infty \tag{7.90}$$

f is λ^n-integrable. By applying Fubini's theorem the desired statement follows. □

Example 7.68. Calculate the center of mass of

$$A = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, x \geq y^2 \wedge x \leq 1\right\}$$



The center of mass is defined by

$$\int \begin{pmatrix} x \\ y \end{pmatrix} \underbrace{\mathrm{d}\lambda^2 (x, y)}_{\mathrm{d}A}$$

In component form this is

$$\begin{aligned}
\int_A x \, \mathrm{d}\lambda^2 (x, y) &= \int x \mathbb{1}_A(x, y) \, \mathrm{d}\lambda^2 (x, y) \\
&= \int_{[0,1] \times [-1,1]} x \mathbb{1}_A(x, y) \, \mathrm{d}\lambda^2 (x, y) \\
&= \int_0^1 \int_{-1}^1 x \mathbb{1}_{[-\sqrt{x}, \sqrt{x}]}(y) \, \mathrm{d}y \, \mathrm{d}x \\
&= \int_0^1 x \cdot 2 \cdot \sqrt{x} \, \mathrm{d}x = \frac{4}{5}
\end{aligned}$$

Meaning the center of mass is at $(\frac{4}{5}, 0)$.

7.5 The Transformation Theorem

Definition 7.69. Let $U, V \subset \mathbb{R}^n$ be open. A mapping $T : U \to V$ is said to be a diffeomorphism if it is bijective and if T and T^{-1} are continuously differentiable. Analogously we define

$$\begin{gathered} C^r\text{-diffeomorphism if it is } r\text{-times differentiable} \\ C^\infty\text{-diffeomorphism if it is infinitely differentiable} \end{gathered}$$

Remark 7.70. (i) In physics, f and $f \circ T$ are often denoted with the same symbol

(ii) We can apply the chain rule to $T \circ T^{-1} = \mathrm{id}_V$

$$DT(T^{-1}(y)) \cdot DT^{-1}(y) = I_V$$

Since T^{-1} is surjective, $DT(x)$ is invertible $\forall x \in U$. According to the theorem about inverse functions, the inverse T^{-1} of a bijective mapping is continuously differentiable if $DT(x)$ is invertible

(iii) If T is a diffeomorphism, then T^{-1} is one too.

Example 7.71. (i) Polar coordinates:

$$\begin{aligned} T : (0,\infty) \times (0, 2\pi) &\longrightarrow \mathbb{R}^2 \setminus \{[0,\infty] \times \{0\}\} \\ (r, \phi) &\longmapsto (r\cos\phi, r\sin\phi) \end{aligned}$$

(ii) Another diffeomorphism would be

$$\begin{aligned} T : B_1(0) &\longrightarrow \mathbb{R}^n \\ x &\longmapsto \frac{x}{\sqrt{1 - \|x\|}} \end{aligned}$$

(iii) An example for a mapping that is no diffeomorphism would be

$$\begin{aligned} T : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto x^3 \end{aligned}$$

The Jacobian "matrix" $T'(x) = 3x^2$ is not invertible.

(iv) Another counter example would be

$$\begin{aligned} T : (0,\infty) \times \mathbb{R} &\longrightarrow \mathbb{R}^2 \setminus \{0\} \\ (r, \phi) &\longmapsto (r\cos\phi, r\sin\phi) \end{aligned}$$

This function is not injective, so it's not a diffeomorphism.

Theorem 7.72 (Transformation Theorem). *Let $U, V \subset \mathbb{R}^n$ and $T : U \to V$ a diffeomorphism. Then $f : V \to \mathbb{R}$ is integrable over V if and only if $f \circ T \cdot |\det DT|$ is integrable over U. In this case*

$$\int_V d \,\mathrm{d}\lambda^n = \int_U f \circ T \cdot |\det DT| \,\mathrm{d}\lambda^n$$

Proof. Without proof. □

Example 7.73 (Area of the unit circle). The area is defined as

$$\lambda^2(K_1(0)) = \int_{\mathbb{R}^2} \mathbb{1}_{K_1(0)} \,\mathrm{d}\lambda^2$$

We transform into polar coordinates:

$$\begin{aligned} U &= (0, \infty) \times (0, 2\pi) \\ V &= \mathbb{R}^2 \setminus \underbrace{([0, \infty] \times \{0\})}_{\lambda^2-nullset} \end{aligned}$$

We define the transformation

$$T : (r, \phi) \longmapsto (r \cos \phi, r \sin \phi)$$

Which results in

$$\begin{aligned} \det DT(r, \phi) &= r \\ \mathbb{1}_{K_1(0)} \circ T(r, \phi) &= \mathbb{1}_{(0,1]}(r) \end{aligned}$$

So we can calculate

$$\begin{aligned} \lambda^2(K_1(0)) &= \int_B \mathbb{1}_{(0,1]}(x, y) \,\mathrm{d}\lambda^2\,(x, y) \\ &= \int_U \mathbb{1}_{(0,1]}(r) \cdot r \cdot \mathrm{d}\lambda^2\,(r, \phi) \\ &= \int_0^\infty \int_0^{2\pi} \mathbb{1}_{(0,1]} r \,\mathrm{d}\phi \,\mathrm{d}r \\ &= 2\pi \int_0^\infty \mathbb{1}_{(0,1]}(r) \,\mathrm{d}r = 2\pi \int_0^1 r \,\mathrm{d}r \\ &= \pi r^2 = \pi \end{aligned}$$

Remark 7.74. (i) Consider

$$\begin{aligned} T : \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ x &\longmapsto Ax \quad A \in \mathbb{R}^{n \times n} \end{aligned}$$

If $\exists A^{-1}$, then T is a diffeomorphism with $DT = A$

$$\Longrightarrow \int f \,\mathrm{d}\lambda^2 = |\det A| \int f \circ T \,\mathrm{d}\lambda^2$$

(ii) Let A be an orthogonal matrix (so a rotation/mirroring).

$$\det A = \pm 1 \implies |\det A| = 1$$

Thus, rotations and mirrorings do not change the volume.

(iii) Let $A = \operatorname{diag}(a, a, \cdots, a)$ $\;a \in (0, \infty)$ (this is a scaling matrix). Then

$$\det A = a^n$$

which means that continuous scaling of a factor a scales the λ^n-volume by a^n.

(iv) This is a "generalization" of the substitution rule

$$\int_{\mathbb{R}} f(g(x))g'(x)\,\mathrm{d}x = \int_{\mathbb{R}} f(y)\,\mathrm{d}y$$

Example 7.75. We want to compute

$$K = \int_{\mathbb{R}} e^{-x^2}\,\mathrm{d}x$$

Consider

$$K^2 = \int_{\mathbb{R}} e^{-x^2}\,\mathrm{d}x \int_{\mathbb{R}} e^{-y^2}\,\mathrm{d}y = \int_{\mathbb{R}^2} e^{-(x^2+y^2)}\,\mathrm{d}\lambda^2(x, y)$$

By transforming $f = e^{-(x^2+y^2)}$ into polar coordinates

$$\begin{aligned}
K^2 &= \int_U f \circ T |\det DT|\,\mathrm{d}\lambda^2 \\
&= \int_V e^{-r^2} \cdot r\,\mathrm{d}\lambda^2(r, \phi) \\
&= \int_0^\infty \int_0^{2\pi} re^{-r^2}\,\mathrm{d}r\,\mathrm{d}\phi \\
&= 2\pi \int_0^\infty re^{-r^2}\,\mathrm{d}r \\
&= 2\pi \lim_{n\to\infty} \left(-\frac{1}{2}e^{-n^2} + \frac{1}{2}\right) = \pi
\end{aligned}$$

Thus $K = \sqrt{\pi}$.

Example 7.76 (Integrability of radial functions). Let $f : [0, \infty] \to \mathbb{R}$ be measureable and set

$$\begin{aligned}
F : \mathbb{R}^n &\longrightarrow \mathbb{R} \\
x &\longmapsto f(\|x\|)
\end{aligned}$$

$\|\cdot\|$ is the Euclidian norm. Under which conditions is F λ^n-integrable? Let $D := (0,\infty) \times \underbrace{(0,\pi)^{n-2} \times (0,2\pi)}_{D_\phi}$. And define

$$T : D \longrightarrow \mathbb{R}^n \setminus A$$
$$(r,\phi) \longmapsto \begin{pmatrix} r\cos\phi_1 \\ r\sin\phi_1\cos\phi_2 \\ r\sin\phi_1\sin\phi_2\cos\phi_3 \\ \vdots \\ r\sin\phi_1\cdots\sin\phi_{n-2}\cos\phi_{n-1} \\ r\sin\phi_1\cdots\sin\phi_{n-2}\sin\phi_n \end{pmatrix}^T$$

Then $\|T(r,\phi)\| = r$ and

$$|\det DT(r,\phi)| = r^{n-1}\sin^{n-2}\phi_1\sin^{n-3}\phi_2\cdots\sin\phi_{n-2} = r^{n-1}A_n(\phi)$$

Thus

$$\begin{aligned}\int_{\mathbb{R}^n} |F(x)|\,\mathrm{d}\lambda^n(x) &= \int_D \underbrace{|F\circ T(r,\phi)|}_{f(r)}|\det DT(r,\phi)|\,\mathrm{d}\lambda^n(r,\phi) \\ &= \int_{D_\phi}\int_0^\infty r^{n-1}|f(r)|A_n(\phi)\,\mathrm{d}r\,\mathrm{d}\lambda^{n-1}(\phi) \\ &= \int_0^\infty r^{n-1}|f(x)|\,\mathrm{d}r\underbrace{\int_{D_\phi}|A_n(\phi)|\,\mathrm{d}\lambda^{n-1}(\phi)}_{<\infty}\end{aligned}$$

So F is λ^n-integrable if $r^{n-1}f(x)$ is integrable over $[0,\infty)$.

7.6 Lebesgue-Stieltjes Integral

Definition 7.77. Let $F : \mathbb{R} \to \mathbb{R}$ be a monotonically increasing, continuous function. Then we set

$$\lambda_F(\varnothing) := 0 \qquad\qquad \lambda_F((a,b]) = F(b) - F(a), \quad (a,b] \in \mathcal{I}$$

Theorem 7.78. *λ_F is a measure on H.*

Proof. Without proof. □

Definition 7.79. The integral

$$\int_A f\,\mathrm{d}\lambda_F$$

is called the Lebesgue-Stieltjes integral on $\mathbb{R}$ and is denoted by

$$\int_A f(x)\,\mathrm{d}F(x) := \int_A f\,\mathrm{d}\lambda_F$$

If $A = [a, b]$, then we write

$$\int_a^b f(x)\,\mathrm{d}F(x)$$

Chapter 8

Ordinary Differential Equations

8.1 Solution Methods

Definition 8.1. An ordinary differential equation (ODE) is an equation of the form

$$F(x, y, y', \cdots, y^{(n)}) = 0$$

with $F : \mathbb{R}^{n+2} \to \mathbb{R}$. n is the order of the ODE. Let I be an open interval. A function $y : I \to \mathbb{R}$ is a solution of the ODE if $y \in C^n(\mathbb{R})$ and

$$F(x, y(x), y'(x), \cdots, y^{(n)}(x)) = 0 \quad \forall x \in I$$

Example 8.2.

$$y'' = -\frac{1}{y^2} \qquad \text{Gravitational field}$$

$$y'' = -\sin y \qquad \text{Pendulum}$$

Remark 8.3. (i) Often times F is only defined on subsets of $\mathbb{R}^{n+2}$

(ii) ODEs are not simple to solve

(iii) Even if we can't calculate explicit solutions, we can inspect the following properties

- Existence of solutions
- Uniqueness of solutions
- Dependency of solutions from initial conditions
- Sability

Example 8.4. (i) Let I be an open interval and $f : I \to \mathbb{R}$ continuous. Then the solution of

$$y' = f(x)$$

is the antiderivative of f. Let $x_0 \in I$, then

$$y(x) = \int_{x_0}^{x} f(t)\,\mathrm{d}t + c \quad c \in \mathbb{R}$$

(ii) Consider the ODE

$$y' = y$$

The functions $x \mapsto ce^x$ are solutions $\forall c \in \mathbb{R}$. Are those all the solutions that exist? Let $y : I \to \mathbb{R}$ be any solution, and consider

$$u(x) = y(x)e^{-x}$$

Then

$$\begin{aligned} u'(x) &= y'(x)e^{-x} - y(x)e^{-x} \\ &= \big(y'(x) - y(x)\big)\, e^{-x} = 0 \quad \forall x \in I \end{aligned}$$

So $u(x) = c$.

Definition 8.5 (Initial Value Problem)**.** Let $y_0, \cdots, y_{n-1} \in \mathbb{R}$ and also $F : \mathbb{R}^{n+2} \to \mathbb{R}$. The system of equations

$$F(x, y, y', \cdots, y^{(n)}) = 0 \qquad \begin{cases} y(0) = y_0 \\ y'(0) = y_1 \\ \cdots \\ y^{(n-1)}(0) = y_{n-1} \end{cases}$$

is said to be an initial value problem (IVP).

Example 8.6. Consider the problem

$$y'' = -\frac{1}{y^2} \qquad \begin{cases} y(0) = y_0 \\ y'(0) = y_1 \end{cases}$$

This describes the movement of a point mass in the gravitational field of the earth along a straight line through the center of the earth with the initial position y_0 and the initial velocity y_1.

Example 8.7. Consider the problem

$$y' = -y^2 \qquad y(0) = 1$$

Assume $y : I \to \mathbb{R}$ is a solution and $y(x) > 0 \ \ \forall x \in I$. Then

$$1 = -\frac{1}{y(t)^2}\ y'(t) \quad \forall t \in I$$

By integrating we get

$$\begin{aligned} x = -\int_0^x \frac{1}{y(t)^2} y'(t)\,\mathrm{d}t &\underset{\substack{\uparrow \\ \text{Substitution}}}{=} -\int_1^{y(x)} \frac{1}{y^2}\,\mathrm{d}y \\ &= \frac{1}{y}\bigg|_1^{y(x)} = \frac{1}{y(x)} - 1 \quad \forall x \in I \end{aligned}$$

So a solution is

$$y(x) = \frac{1}{1+x}$$

The biggest domain that makes sense is $(-1, \infty)$. Analogously one can approach equations with "separated variables", so of the form

$$y' = f(y)g(x) \qquad y(x_0) = y_0$$

Theorem 8.8 (Separation of Variables)**.** *Let I, J be open intervals, and let*

$$f : I \longrightarrow \mathbb{R} \qquad\qquad g : J \longrightarrow \mathbb{R}$$

be continuous with $0 \neq f(I)$. Let $x_0 \in J,\ y_0 \in I$. Then there exists an open interval $I_2 \subset J$ and $x_0 \in I_2$ such that the IVP

$$y' = f(y)g(x) \qquad\qquad y(x_0) = y_0$$

has exactly one solution on I_2. Set

$$F(y) = \int_{y_0}^{y} \frac{1}{f(t)} \,\mathrm{d}t$$

Then $y : I_2 \to I$ is uniquely defined by

$$F(y(x)) = \int_{x_0}^{x} g(t) \,\mathrm{d}t$$

Proof. f does not have any roots, thus w.l.o.g. $f > 0$.

$$F'(y) = \frac{1}{f(y)} > 0 \implies F \text{ strictly monotonically increasing} \tag{8.1}$$

Therefore there exists an inverse function $H : F(I) \to I$. According to the theorem about inverse functions, H is C^1 and

$$H'(z) = \frac{1}{F'(H(z))} \quad \forall z \in F(I) \tag{8.2}$$

$F(I)$ is an open interval containing the 0. Then we have

$$y(x) = H(G(x)) \quad x \in I_2 \tag{8.3}$$

where

$$G(x) = \int_{x_0}^{x} g(t) \,\mathrm{d}t \tag{8.4}$$

Now choose I_2 such that $x_0 \in I_2$ and $G(I_2) \subset F(I)$. Then

$$\begin{aligned} y'(x) &= H'(G(x)) \cdot G'(x) \\ &= \frac{1}{F'(H(G(x)))} \cdot G'(x) \\ &= \frac{1}{F'(y(x))} \cdot G'(x) \\ &= f(y(x))g(x) \end{aligned} \tag{8.5}$$

So y solves the ODE. However, if $\tilde{y} : I \to \mathbb{R}$ some solution of the IVP, then $\forall x \in I_2$

$$G(x) = \int_{x_0}^{x} g(x) \,\mathrm{d}x = \int_{x_0}^{x} \frac{\tilde{y}(x)}{f(\tilde{y}(x))} \,\mathrm{d}x = \int_{\tilde{y}(x_0)}^{\tilde{y}(x)} \frac{1}{f(y)} \,\mathrm{d}y = F(\tilde{y}(x)) \tag{8.6}$$

So $\tilde{y}(x) = H(G(x))$ □

Remark 8.9. I_2 is obviously not unique. We can find the biggest possible domain with

$$\bigcup_{\substack{x \in I_2 \\ I_2 \text{ open} \\ G(I_2) \subset F(I)}} I_2 = I_{2,\max}$$

Theorem 8.10. *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function, $a, b, c \in \mathbb{R}$ and I an open interval. Then $y : I \to \mathbb{R}$ is a solution of the ODE*

$$y' = f(ax + by + c)$$

if and only if $u(x) := ax + by + c$ is a solution of

$$u' = a + bf(u)$$

Heuristic Proof. Consider

$$u'(x) = a + by'(x)$$

□

Example 8.11 (Euler Homogeneous ODE). Let $f : \mathbb{R} \to \mathbb{R}$ be a function and I an open interval not containing the 0. Then $y : I \to \mathbb{R}$ is a solution of the ODE

$$y' = f(\frac{y}{x})$$

if and only if

$$u(x) = \frac{y(x)}{x}$$

solves the ODE

$$u' = \frac{f(u) - u}{x}$$

Example 8.12. Let $f : \mathbb{R} \to \mathbb{R}$ be continuous and $a_1, a_2, b_1, b_2, c_1, c_2 \in \mathbb{R}$ such that

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \neq 0$$

Now let $\tilde{x}, \tilde{y}$ be the solutions of the equation system

$$a_1\tilde{x} + b_1\tilde{y} + c_1 = 0$$
$$a_2\tilde{x} + b_2\tilde{y} + c_2 = 0$$

Let I be an open interval not containing the 0. Then $y : I \to \mathbb{R}$ is a solution to

$$y' = f\left(\frac{a_1x + b_1y + c_1}{a_2x + b_2y + c_2}\right)$$

if and only if

$$\begin{aligned} u : I - \tilde{x} &\longrightarrow \mathbb{R} \\ x &\longmapsto y(x + \tilde{x}) - \tilde{y} \end{aligned}$$

is a solution to

$$u' = f\left(\frac{a_1 + b_y \frac{u}{x}}{a_2 + b_2 \frac{u}{x}}\right)$$

Proof. Let $y : I \to \mathbb{R}$ be a solution to the initial equation. Then

$$\begin{aligned} u'(x) = y'(x + \tilde{x}) &= f\left(\frac{a_1(x + \tilde{x}) + b_1 y(x + \tilde{x}) + c_1}{a_2(x + \tilde{x}) + b_2 y(x + \tilde{x}) + c_2}\right) \\ &= f\left(\frac{a_1 x + b_1 u(x) + a_1 \tilde{x} + b_1 \tilde{y} + c_1}{a_2 x + b_2 u(x) + a_2 \tilde{x} + b_2 \tilde{y} + c_2}\right) \\ &= f\left(\frac{a_1 + b_1 \frac{u(x)}{x}}{a_2 + b_2 \frac{u(x)}{x}}\right) \end{aligned} \tag{8.7}$$

The other direction is left to the reader. □

Definition 8.13 (Exact ODE)**.** Let $D \subset \mathbb{R}^2$ be open, and $p, q : D \to \mathbb{R}$ continuous. The ODE

$$p(x, y) + q(x, y)y' = 0$$

is said to be exact if there exists a C^1-function $H : D \to \mathbb{R}$, such that

$$\partial_1 H = p \qquad\qquad \partial_2 H = q$$

Such a function is called a potential function.

Theorem 8.14. *Let $D \subset \mathbb{R}^2$ be open, and $p, q : D \to \mathbb{R}$ continuous. Let*

$$p(x, y) + q(x, y)y' = 0$$

be exact and H a potential function. Furthermore let I be an open interval and $y : I \to \mathbb{R}$ a C^1-function such that

$$\{(x, y(x)) \mid x \in I\} \subset D$$

Then y solves the ODE if and only if $\exists c \in \mathbb{R}$ such that

$$H(x, y(x)) = c$$

Proof.

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x} H(x, y(x)) &= \partial_1 H(x, y(x)) + \partial_2 H(x, y(x)) y'(x) \\ &= p(x, y) + q(x, y) y'(x) \end{aligned} \tag{8.8}$$

□

Theorem 8.15. *Let $D \subset \mathbb{R}^2$ be open, and $p, q : D \to \mathbb{R}$ continuously differentiable. If*

$$p(x,y) + q(x,y)y' = 0$$

is exact, then

$$\partial_2 p = \partial_1 q$$

Proof. Let H be a potential C^2-function. Then

$$\partial_2 p = \partial_2\partial_1 H = \partial_1\partial_2 H = \partial_1 q \tag{8.9}$$

□

Remark 8.16. The above condition is merely necessary! However, for "nice" D it can be considered sufficient.

Example 8.17. Consider

$$\underbrace{(2x+y^2)}_{p} + \underbrace{2xy}_{q}y' = 0 \qquad\qquad y(1) = 1$$

Then

$$\partial_2 p = 2y \qquad\qquad \partial_1 q = 2y$$

So $\partial_2 p = \partial_1 q$. If H is a potential function, then

$$\begin{aligned} \partial_1 H(x,y) &= p(x,y) = 2x + y^2 \\ \implies H(x,y) &= \int p(x,y)\,\mathrm{d}x = x^2 + y^2 x + G(y) \end{aligned}$$

and

$$\begin{aligned} &\partial_2 H(x,y) = q(x,y) = 2xy = 2xy + G'(y) \\ &\implies G(y) = c \end{aligned}$$

So the potential function is

$$H(x,y) = x^2 + y^2 x$$

We can insert the initial condition

$$H(1,1) = 2$$

So the solution has to fulfil

$$x^2 + y(x)^2 x = 2 \quad \forall x \in I$$

and thus

$$y(x) = \pm\sqrt{\frac{2}{x} - x}$$

Only the positive sign fulfils the initial conditions, so the solution is

$$y(x) = \sqrt{\frac{2}{x} - x}$$

This function is defined on $(-\infty, -\sqrt{2}] \cup (0, \sqrt{2}]$, however due to the initial conditions $(0, \sqrt{2}]$ is the only useful domain.

Remark 8.18. If

$$p(x, y) + q(x, y)y' = 0$$

is not exact one can try and find an "integrating factor", i.e. $h : D \to \mathbb{R}$ such that

$$h(x, y)p(x, y) + h(x, y)q(x, y)y' = 0$$

is exact. A necessary condition is

$$(\partial_2 h(x, y))\, p(x, y) + h(x, y)\partial_2 p(x, y) = (\partial_1 h(x, y))\, q(x, y) + h(x, y)\partial_1 q(x, y)$$

This is a partial differential equation and won't be discussed further in this chapter.

Definition 8.19 (Ordinary Differential Equation System)**.** An ordinary differential equation system (ODES) is an equation of the form

$$F(x, y, y', \cdots, y^{(n)}) = 0$$

with

$$F : \mathbb{R} \times \mathbb{R}^L \times \mathbb{R}^L \times \cdots \times \mathbb{R}^L \longrightarrow \mathbb{R}^m$$

Example 8.20. (i) Let $z = (z_1, z_2, z_3)$, then

$$z'' = -\frac{z}{\|z\|^3} = -\frac{1}{\|z\|^2}\frac{z}{\|z\|}$$

is the Kepler problem.

(ii) The equation

$$\begin{aligned} b' &= \alpha_1 b - \gamma_1 br \\ r' &= -\alpha_2 r + \gamma_2 br \end{aligned}$$

is called the "Lotka-Volterra-Equation" and it models the population of prey and predators.

Remark 8.21. The ODES

$$F(x, y, y', y'', \cdots, y^{((n)}) = 0$$

is equivalent to the ODES of first order

$$F(x, y, y_1, y_2, \cdots, y_{n-1}) = 0 \qquad \begin{cases} y_1 = y' \\ y_2 = y_1' \\ \vdots \end{cases}$$

8.2 The Picard-Lindelöf Theorem

Example 8.22. Consider the ODE

$$y' = 2\sqrt{|y|}$$

Possible solutions are

$$\begin{aligned} y(x) &= (x-c)^2 \quad c > 0 \\ y(x) &= -(x-c)^2 \quad c < 0 \\ y(x) &= 0 \end{aligned}$$

Another solution could be

$$y(x) = \begin{cases} -(x-a)^2, & x \in (-\infty, a) \\ 0, & x \in [a,b] \\ (x-b)^2, & x \in (b, \infty) \end{cases}$$

for $a, b \in \mathbb{R}$ with $a \leq b$ So the IVP $y(0) = 0$ has many solutions.

Definition 8.23. Let $D \subset \mathbb{R} \times \mathbb{R}^n$ be open, $(x_0, y_0) \in D$ and $f : D \to \mathbb{R}^n$. We say f fulfils a local Lipschitz-condition in the point (x_0, y_0) if there exists a neighbourhood U of (x_0, y_0) such that

$$\|f(x,y) - f(x,z)\| \leq L\|y - z\| \quad \forall (x,y), (x,z) \in U$$

Example 8.24. Consider

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ (x,y) &\longmapsto x^2y^2 \end{aligned}$$

Then

$$\begin{aligned} |f(x,y) - f(x,z)| = \left|x^2(y^2 - z^2)\right| &= \left|x^2(y-z)(y+z)\right| \\ &= \underbrace{\left|x^2(y+z)\right|}_{\alpha(x,y,z)} |y - z| \end{aligned}$$

The function $\alpha(x,y,z)$ is unbounded, so the global Lipschitz condition isn't satisfied. Now choose a fixed $(x_0, y_0) \in \mathbb{R} \times \mathbb{R}$, and set

$$R > \max\{|x_0|, |y_0|\}$$

Then $\forall (x,y), (x,z) \in (-R, R) \times (-R, R)$

$$\alpha(x,y,z) \leq R^2|y+z| \leq R^2\left(|y| + |z|\right) \leq 2R^3$$

So f fulfils a local Lipschitz condition in (x_0, y_0).

Definition 8.25. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, $f : \Omega \to \mathbb{R}^n$ measurable and $f_1, \cdots, f_n$ are the component functions of f. So

$$f(\omega) = (f_1(\omega), f_2(\omega), \cdots, f_n(\omega))$$

f is said to be integrable if $f_1, \cdots, f_n$ are integrable, and we define

$$\int f \mathrm{d}\mu = \left(\int f_1 \mathrm{d}\mu, \int f_2 \mathrm{d}\mu, \cdots, \int f_n \mathrm{d}\mu \right)$$

Theorem 8.26. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, define $\|\cdot\|$ to be the norm on $\mathbb{R}^n$ and let $f : \Omega \to \mathbb{R}^n$ be measurable. Then f is integrable if and only if $\|f\|$ is integrable, and*

$$\left\| \int f \mathrm{d}\mu \right\| \leq \int \|f\| \mathrm{d}\mu$$

Proof. Without proof. □

Lemma 8.27. *Let $D \subset \mathbb{R} \times \mathbb{R}^n$, $(x_0, y_0) \in D$ and $f : D \to \mathbb{R}^n$ continuous. Let I be an open interval and $y : I \to \mathbb{R}^n$ be continuously differentiable, such that $(x, y(x)) \in D \quad \forall x \in I$. Then y is a solution of the IVP*

$$y' = f(x, y) \qquad\qquad y(x_0) = y_0$$

if and only if y satisfies the integral equation

$$y(x) = y_0 + \int_{x_0}^{x} f(t, y(t))\, \mathrm{d}t$$

Proof. Let y fulfil the IVP. Then

$$y(x) - y_0 = y(x) - y(x_0) = \int_{x_0}^{x} y'(t)\, \mathrm{d}t = \int_{x_0}^{x} f(t, y(t))\, \mathrm{d}t$$

If y fulfils the integral equation, then

$$y'(x) = f(x, y(x))$$

□

Theorem 8.28 (Picard-Lindelöf Theorem)**.** *Let $D \subset \mathbb{R} \times \mathbb{R}^n$ be open, $(x_0, y_0) \in D$ and $f : D \to \mathbb{R}^n$ continuous such that f fulfils a local Lipschitz condition in y. Then $\exists \epsilon > 0$ such that the IVP*

$$y' = f(x, y) \qquad\qquad y(x_0) = y_0$$

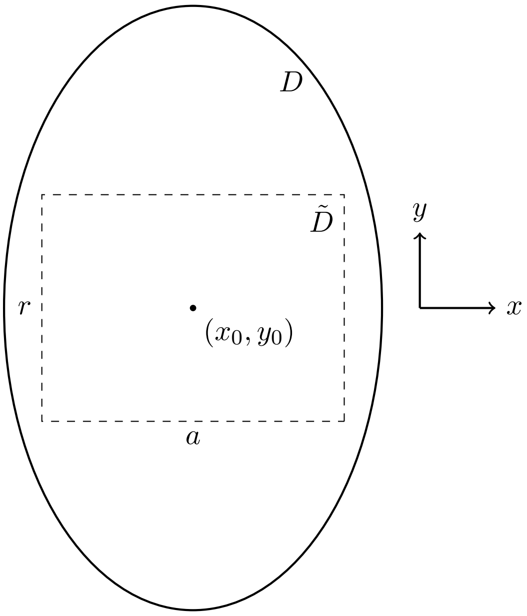
has exactly one solution on $(x_0 - \epsilon, x_0 + \epsilon)$.

Proof. Let $U \subset D$ be a neighbourhood of (x_0, y_0), such that

$$\|f(x,y) - f(x,z)\| \leq L\|y - z\| \quad \forall (x,y),(x,z) \in U \tag{8.10}$$

Choose $a, r > 0$ such that

$$\tilde{D} = [x_0 - a, x_0 + a] \times K_r(y_0) \subset U \tag{8.11}$$



$\tilde{D}$ is compact and f is continuous, i.e. f is bounded on $\tilde{D}$ by $M \in (0, \infty)$.

$$\|f(x,y)\| \leq M \quad \forall (x,y) \in \tilde{D} \tag{8.12}$$

Choose an ϵ such that $0 < \epsilon \leq a$ and such that

$$\epsilon M < r \qquad\qquad \epsilon L < 1 \tag{8.13}$$

Set $I := (x_0 - \epsilon, x_0 + \epsilon)$, and

$$X = \{y : I \to K_r(y_0) \,|\, y \text{ continuous}\} \subset C(I) \tag{8.14}$$

X is closed, and thus complete. Define $T : X \to X$ with

$$T(y)(x) := y_0 + \int_{x_0}^{x} f(t, y(t))\,\mathrm{d}t \tag{8.15}$$

We want to show $T(y) \subset X$:

$$\begin{aligned} \|T(y)(x) - y_0\| = \left\| \int_{x_0}^{x} f(t, y(t))\,\mathrm{d}t \right\| &\leq \int_{x_0}^{x} \|f(t, y(t))\|\,\mathrm{d}t \\ &\leq M \int_{x_0}^{x} \mathrm{d}t < \epsilon M < r \end{aligned} \tag{8.16}$$

Now consider

$$\begin{aligned}\|T(y)(x) - T(\tilde{y})(x)\| &= \left\| \int_{x_0}^{x} (f(t, y(t)) - f(t, \tilde{y}(t)))\,\mathrm{d}t \right\| \\ &\leq \int_{x_0}^{x} \|f(t, y(t)) - f(t, \tilde{y}(t))\|\,\mathrm{d}t \\ &\leq \int_{x_0}^{x} L \cdot \|y(t) - \tilde{y}(t)\|\,\mathrm{d}t \\ &\leq \int_{x_0}^{x} L\|y - \tilde{y}\|_\infty \leq L\|y - \tilde{y}\|_\infty \cdot \epsilon\end{aligned} \tag{8.17}$$

By taking the supremum over all $x \in I$ we get

$$\|T(y) - T(\tilde{y})\|_\infty \leq \underbrace{\epsilon L}_{<1} \|y - \tilde{y}\|_\infty \tag{8.18}$$

So $T : X \to X$ is strictly contractive. According to the Banach fixed point theorem, there exists a unique fixed point of T in x, that means $\exists y \in X$ such that

$$y_0 + \int_{x_0}^{x} f(t, y(t))\,\mathrm{d}t = T(y)(x) = y(x) \quad \forall x \in I \tag{8.19}$$

Due to Lemma 8.27, there eixsts a unique solution to the ODE. □

Remark 8.29. One can approximate a fixed point by repeatedly applying T. For example consider

$$\phi(x) = y_0$$

and define

$$\phi_0 = \phi \qquad\qquad \phi_i = T(\phi_{i-1}) = y_0 + \int_{x_0}^{x} f(t, \phi_{i-1}(t))\,\mathrm{d}t$$

This process is called Picard iteration, and the ϕ_i converge uniformly to the solution.

Example 8.30. Consider

$$y' = \sqrt{\|y\|}$$

Then

$$\lim_{y\to 0} \left(\frac{|f(x, y) - f(x, 0)|}{|y - 0|} \right) = \lim_{y\to 0} \frac{1}{\sqrt{|y|}} \longrightarrow \infty$$

Which means the local Lipschitz condition is not satisfied.

Theorem 8.31. *Let $D \subset \mathbb{R} \times \mathbb{R}^n$ be open, $f : D \to \mathbb{R}^n$ continuously differentiable. Then f satisfies a local Lipschitz condition in terms of y.*

Proof. Let $(x_0, y_0) \in D$. Choose $r > 0$ such that $K_r(x_0, y_0) \subset D$. The total derivative $D_y f$ is continuous and thus bounded on $K_r(x_0, y_0)$.

$$\exists L > 0 : \quad \|D_y f(x, y)\| \leq L \ \ \forall (x, y) \in K_r(x_0, y_0) \tag{8.20}$$

Applying the generalized mean value theorem yields

$$\begin{aligned} \|f(x, y) - f(x, z)\| &\leq \sup_{t \in [0,1]} \|D_y f(x, y + t(z - y))\| \|y - z\| \\ &\leq L\|y - z\| \end{aligned} \tag{8.21}$$

If $n = 1$ we can specify

$$|f(x, y) - f(x, z)| = |\partial_y f(x, \xi)(y - z)| \tag{8.22}$$

□

Example 8.32. Consider

$$y'' = -\frac{y}{\|y\|^3}$$

The function

$$f : \underbrace{\mathbb{R} \times \mathbb{R}^3 \setminus \{0\} \times \mathbb{R}^3}_{D} \longrightarrow \mathbb{R}^3 \times \mathbb{R}^3$$
$$(x, y, z) \longmapsto \left(z, (y_1^2 + y_2^2 + y_3^2)^{-\frac{3}{2}} \cdot y\right)$$

is continuously differentiable. So the IVP for arbitrary initial points in D has a locally unique solution.

Definition 8.33. Let $D \subset \mathbb{R} \times \mathbb{R}^n$ be open, $(x_0, y_0) \in D$. A solution $\tilde{y} : \tilde{I} \to \mathbb{R}^n$ of the IVP

$$y' = f(x, y) \qquad\qquad y(x_0) = y_0$$

is said to be a (real) continuation of the solution $y : I \to \mathbb{R}^n$ if $I \subset \tilde{I}$ and $y(x) = \tilde{y}(x) \ \forall x \in I$. A solution y is said to be a maximal solution if there are no real continuations.

Theorem 8.34. *Let $D \subset \mathbb{R} \times \mathbb{R}^n$ be open, $(x_0, y_0) \in D$ and $f : D \to \mathbb{R}^n$ continuous and satisfying a local Lipschitz condition in terms of y. Then the IVP*

$$y' = f(x, y) \qquad\qquad y(x_0) = y_0$$

has a unique solution.

Proof. First, let $y : I \to \mathbb{R}^n$ and $\tilde{y} : \tilde{I} \to \mathbb{R}^n$ be solutions of the IVP. Then $y = \tilde{y}$ on $I \cap \tilde{I} =: (a, b)$. Let

$$c = \sup \{\tilde{c} \in [x_0, b) \,|\, y = \tilde{y} \text{ on } [x_0, \tilde{c}]\} \tag{8.23}$$

According to Picard-Lindelöf, such $\tilde{c}$ exist. Then there exists a sequence $(c_n) \subset (x_0, c)$ such that $y = \tilde{y}$ on $[x_0, c_n)$ $\forall n \in \mathbb{N}$ and $c_n \to c$. If $c < b$, then

$$y(c) = \tilde{y}(c) \tag{8.24}$$

because $y(c_n) = \tilde{y}(c_n)$ $\forall n \in \mathbb{N}$. The IVP

$$u' = f(x, u) \qquad\qquad u(c) = y(c) \tag{8.25}$$

has a locally unique solution on $(c - \epsilon, c + \epsilon)$ $\epsilon > 0$ according to Picard-Lindelöf. Since the y and $\tilde{y}$ are both solutions to the IVP, they are identical on $(c - \epsilon, c + \epsilon)$. However, this contradicts the construction of c, so $c = b$.

$$\implies y = \tilde{y} \quad \text{on } [x_0, b) \tag{8.26}$$

Analogously, one can prove $y = \tilde{y}$ on $(a, x_0]$. Now let $I_{\max}$ be the union of all open intervals that are domains of the solution of the IVP. For $x \in I_{\max}$ we can choose

$$y_{\max}(x) = y(x) \tag{8.27}$$

for arbitrary solutions $y : I \to \mathbb{R}$ with $x \in I$. So

$$y_{\max} : I_{\max} \to \mathbb{R} \tag{8.28}$$

is a maximal solution that is unique. □

Example 8.35. (i) Consider

$$y' = e^{-y} \qquad\qquad y(1) = 0$$

The solution to this is

$$\begin{aligned} y : (0, \infty] &\longrightarrow \mathbb{R} \\ x &\longmapsto \ln(x) \end{aligned}$$

and is maximal.

(ii) Consider

$$y' = -i\frac{y}{x^2} \qquad\qquad y\left(\frac{1}{2\pi}\right) = 1$$

The solution to this is

$$\begin{aligned} (0, \infty) &\longrightarrow \mathbb{C} \\ x &\longmapsto e^{\frac{i}{x}} \end{aligned}$$

and is maximal.

We define (X, d) to be a metric space, $x \in X$ and $A \subset X$. Then

$$d(x, A) = \inf \{d(x, y) \mid y \in A\}$$

Theorem 8.36. *Let $D \subset \mathbb{R} \times \mathbb{R}^n$ be open, $(x_0, y_0) \subset D$ and $f : D \to \mathbb{R}^n$ continuous and satisfying the local Lipschitz condition in terms of y. Let $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$ such that*

$$-\infty \leq a < x_0 < b \infty\infty$$

and let

$$y : (a, b) \to \mathbb{R}$$

a solution of the IVP

$$y' = f(x, y) \qquad y(x_0) = y_0$$

Then y is the maximal solution of the IVP if and only if one of these conditions

$$\begin{aligned} &(i) && b = \infty \\ &(ii) && \lim_{x \to b} \|y(x)\| = \infty \\ &(iii) && \lim_{x \to b} d((x, y(x)), \partial D) = 0 \end{aligned}$$

and one of these

$$\begin{aligned} &(i) && a = -\infty \\ &(ii) && \lim_{x \to a} \|y(x)\| = \infty \\ &(iii) && \lim_{x \to a} d((x, y(x)), \partial D) = 0 \end{aligned}$$

is fulfilled.

8.3 Linear Differential Equation Systems

Definition 8.37. Let I be an open interval, and $A : I \to \mathbb{R}^{n \times n}$, $b : I \to \mathbb{R}^n$. Then the ODES

$$y' = A(x)y + b(x)$$

is said to be a linear differential equation system. If b is the zero function, then the system is homogeneous (otherwise it's inhomogeneous). If $A(x) = \text{const.}$, then the system is said to have constant coefficients.

Remark 8.38. (i) By using substitution we can transform the equation

$$y^{(n)} = a_{n-1}(x)y^{(n-1)} + a_{n-2}(x)y^{(n-2)} + \cdots + a_0 y + b(x)$$

into the system

$$\begin{aligned} y_{n-1}' &= a_{n-1}(x)y_{n-2} + a_{n-2}(x)y_{n-3} + \dots + a_0 y + b(x) \\ y_1 &= y' \\ y_2 &= y_1' \\ &\vdots \\ y_{n-1} &= y_{n-2}' \end{aligned}$$

(ii) Let y, z be solutions of $y' = A(x)y + b(x)$, then $y - z$ is the solution of the related homogeneous equation $y' = A(x)y$. This follows from

$$\begin{aligned} (y-z)'(x) &= A(x)y(x) + b(x) - (A(x)z(x) + b(x)) \\ &= A(x)(y-z)(x) \end{aligned}$$

Lemma 8.39 (Grönwall's Lemma)**.** *Let I be an open interval, $x_0 \in I$, $y : I \to [0, \infty)$ continuous, $a, b \geq 0$ and*

$$y(x) \leq a + b\left|\int_{x_0}^{x} y(t)\,\mathrm{d}t\right|$$

Then

$$y(x) \leq ae^{b|x-x_0|}$$

Proof. Here we only prove $x > x_0$, but the proof for $x \leq x_0$ works analogously. Let $\epsilon > 0$ be arbitrary and choose

$$z(x) := a + \epsilon + b\int_{x_0}^{x} y(t)\,\mathrm{d}t \tag{8.29}$$

Then

$$z'(x) = by(x) \leq bz(x) \quad \forall x \in I \tag{8.30}$$

And since

$$z(t) \geq a + \epsilon > 0 \tag{8.31}$$

we get

$$\int_{x_0}^{x} \frac{z'(t)}{z(t)}\,\mathrm{d}t \leq b(x - x_0) \tag{8.32}$$

$$\int_{x_0}^{x} \frac{z'(t)}{z(t)}\,\mathrm{d}t = \ln(x) - \ln(z) \tag{8.33}$$

Due to the monotony of the exponential function

$$z(x) \leq z(x_0)e^{b(x-x_0)} = (a+\epsilon)e^{b(x-x_0)} \tag{8.34}$$

So

$$y(x) \leq z(x) \leq (a+\epsilon)e^{b(x-x_0)} \leq ae^{b(x-x_0)} \quad \forall x \in I \tag{8.35}$$

□

From now on I will always be an open interval, and

$$\begin{aligned} A &: I \to \mathbb{R}^{n\times n} \\ b &: I \to \mathbb{R}^n \end{aligned}$$

are continuous, $x_0 \in I$ and $y_0 \in \mathbb{R}$.

Corollary 8.40. *The IVP*

$$y' = A(x)y + b(x) \qquad\qquad y(x_0) = y_0$$

has a unique maximal solution that is defined on all of I.

Proof.

$$\begin{aligned} f : I \times \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ (x, y) &\longmapsto A(x)y + b(x) \end{aligned} \tag{8.36}$$

We need to show that f fulfils a local Lipschitz condition in y. Let $(x_1, y_1) \in I \times \mathbb{R}^n$. Choose a compact I_1 such that $x_1 \in I_1 \subset I$. Then $A(x)$ is bounded on I_1, i.e.

$$\exists L > 0 : \quad \|A(x)\| \leq L \quad \forall x \in I_1 \tag{8.37}$$

And then $\forall (x, y), (x, z) \in I_1 \times \mathbb{R}^n$

$$\|f(x, y) - f(x, z)\| = \|A(x)(y - z)\| \leq \|A(x)\|\|y - z\| \leq L\|y - z\| \tag{8.38}$$

So f fulfils a local Lipschitz condition, and thus there exists a unique maximal solution. Let $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$ such that $y : (a, b) \to \mathbb{R}^n$ is the maximal solution. Assume $b \in I$ (so y isn't defined on all of I). Then there exists $M, K > 0$ such that $\|A(x)\| \leq M$ and $\|b(x)\| \leq K$ and $[x_0, b]$ and

$$\begin{aligned} \|y(x)\| = \left\| y_0 + \int_{x_0}^{x} y'(t)\,\mathrm{d}t \right\| &= \left\| y_0 + \int_{x_0}^{x} A(t)y(t) + b(t)\,\mathrm{d}t \right\| \\ &\leq \|y_0\| + \int_{x_0}^{x} \|A(t)\|\|y(t)\|\,\mathrm{d}t + \int_{x_0}^{x} \|b(t)\|\,\mathrm{d}t \\ &\leq \|y_0\| + K(b - x_0) + M \int_{x_0}^{x} \|y(t)\|\,\mathrm{d}t \end{aligned} \tag{8.39}$$

Applying Grönwall's Lemma onto $\|y(t)\|$ yields

$$\|y(x)\| \leq (\|y_0\| + K(b - x_0))\, e^{M|x - x_0|} \leq (\|y_0\| + K(b - x_0))\, e^{M(b - x_0)} \tag{8.40}$$

and thus y is bounded on $[x_0, b)$. So none of the conditions from Theorem 8.36 are satisfied, and therefore $b \notin I$. This mean that y is defined up to the right boundary of I. $\square$

Remark 8.41. One can show that for linear systems, the Picard iteration leads to a solution that converges on all of I. This would lead to an alternative proof.

Corollary 8.42. *Let $y, z : I \to \mathbb{R}^n$ be solutions of the ODES*

$$y' = A(x)y + b(x)$$

Then the following are equivalent

(i) $y(x) = z(x) \quad \forall x \in I$

(ii) $y(x_0) = z(x_0)$

(iii) $y(x) = z(x) \quad$ *for some* $x \in I$

Proof. $(i) \implies (ii)$, $(ii) \implies (iii)$ is trivial. To prove $(iii) \implies (i)$, let $x_1 \in I$ such that $y_1 = y(x_1) = z(x_1)$. Then y, z are solutions to the IVP

$$y' = A(x)y + b(c) \qquad\qquad y(x_1) = y_1 \tag{8.41}$$

Since this problem has unique solutions

$$y = z \tag{8.42}$$

must hold. □

Theorem 8.43. *The solution set of the homogeneous ODES*

$$y' = A(x)y$$

so

$$V := \left\{ y : I \to \mathbb{R}^n \,\middle|\, y'(x) = A(x)y(x) \quad \forall x \in I \right\}$$

is an n-dimensional linear subspace of $C^1(I, \mathbb{R}^n)$.

Proof. Proving that V is a vector space is trivial. So let $e_1, \cdots, e_n$ be a basis of $\mathbb{R}^n$ and let y_i be the unique solutions of the initial value problem

$$y' = A(x)y \qquad\qquad y(x_0) = e_i \quad i \in \{1, \cdots, n\}$$

Then $y_1, \cdots, y_n$ is a basis of V. To prove their linear independence, let $\alpha_1, \cdots, \alpha_n \in \mathbb{R}$ such that

$$\alpha_1 y_1 + \cdots + \alpha_n y_n = 0 \tag{8.43}$$

then

$$\alpha_1 y_1(x_0) + \cdots + \alpha_n y_n(x_0) = \alpha_1 e_1 + \cdots + \alpha_n e_n = 0 \tag{8.44}$$

Since the $e_1, \cdots, e_n$ are linear independent

$$\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0 \tag{8.45}$$

To prove that the $y_1, \cdots, y_n$ span V, set $z \in V$ and choose $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ such that

$$\alpha_1 e_1 + \alpha_2 e_2 + \cdots + \alpha_n e_n = z(x_0) \tag{8.46}$$

Then the z and $\alpha_1 y_1 + \cdots + \alpha_n y_n$ are maximal solutions of the ODES that are equal in x_0. Thus

$$z = \alpha_1 y_1 + \cdots + \alpha_n y_n \tag{8.47}$$

□

Definition 8.44. A basis $y_1, \cdots, y_n$ of V is said to be a fundamental system of the ODES

$$y' = A(x)y$$

Analogously, n linearly independent solutions of the equation

$$y^{(n)} = a_{n-1}(x)y^{(n-1)} + a_{n-2}(x)y^{(n-2)} + \cdots + a_0 y$$

are said to be a fundamental system.

Example 8.45. Consider the inhomogeneous equation

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

First, find the solutions to the homogeneous equation

$$\frac{y'}{y} = \sin(x)$$

This can be done via integration

$$\int \frac{y'(t)}{y(t)}\,\mathrm{d}t = -\cos(x) + c$$
$$\ln y + c = -\cos(x) + c$$

Then the solution is

$$y = Ke^{-\cos(x)}$$

The fundamental system in this case is $e^{-\cos x}$. We can use a technique called "variation of the constant" to find a solution of the inhomogeneous equation. Define

$$y(x) = C(x)e^{-\cos(x)}$$

Deriving this gives

$$y'(x) = C'(x)e^{-\cos(x)} - C(x)\sin(x)e^{-\cos(x)}$$

Resubstituting this into the initial equation yields

$$\begin{aligned} C'(x)e^{-\cos(x)} + \cancel{C(x)\sin(x)e^{-\cos(x)}} &= \cancel{C(x)\sin(x)e^{-\cos(x)}} + \sin(x)\cos(x) \\ C'(x)e^{-\cos(x)} &= \sin(x)\cos(x) \\ C'(x) &= \sin(x)\cos(x)e^{\cos(x)} \\ C(x) &= (1-\cos(x))e^{\cos(x)} \end{aligned}$$

So the general solution to the ODE is

$$y(x) = 1 - \cos(x) + Ke^{-\cos(x)}$$

Theorem 8.46. *Let $y_1, \cdots, y_n$ be a fundamental system for $y' = A(x)y$. Define an $n \times n$-matrix*

$$W(x) := (y_1(x), y_2(x), \dots, y_n(x))$$

Then $W(x)$ is invertible $\forall x \in I$ and

$$\begin{aligned} z : I &\longrightarrow \mathbb{R}^n \\ x &\longmapsto W(x) \int_{x_0}^{x} W(t)^{-1} b(t)\, \mathrm{d}t \end{aligned}$$

is a solution to the inhomogeneous system

$$y' = A(x)y + b(x)$$

Proof. According to the prerequisites the $y_1, \cdots, y_n$ are linearly independent, so the $y_1(x), \dots, y_n(x)$ are also linearly independent in $\mathbb{R}^n$. Thus

$$\det W(x) \neq 0 \implies W(x) \text{ invertible} \tag{8.48}$$

Deriving this yields

$$W'(x) = A(x)W(x) \tag{8.49}$$

which means the i-th column of this equation is $y_i'(x) = A(x)y_i(x)$. Deriving z gives us

$$\begin{aligned} z'(x) &= W'(x) \int_{x_0}^{x} W(t)^{-1} b(t)\, \mathrm{d}t + W(x)W(x)^{-1}b(x) \\ &= A(x)z(x) + b(x) \end{aligned} \tag{8.50}$$

To apply the fundamental theorem, $W(t)b(t)$ should be continuous. The mapping $A \mapsto A^{-1}$ is continuous on $Gl(n)$ (space of invertible matrices). □

Example 8.47. Consider the system

$$u' = v + \sin(x) \qquad\qquad v' = -u + \cos(x)$$

The homogeneous system in this case is

$$\begin{pmatrix} u \\ v \end{pmatrix}' = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

The fundamental system is

$$y_1(x) = \begin{pmatrix} \sin \\ \cos \end{pmatrix} (x) \qquad\qquad y_2 = \begin{pmatrix} \cos \\ -\sin \end{pmatrix} (x)$$

Then define

$$\begin{aligned} z(x) &= C_1(x) y_1(x) + C_2(x) y_2(x) \\ &= \underbrace{\begin{pmatrix} \sin(x) & \cos(x) \\ \cos(x) & -\sin(x) \end{pmatrix}}_{W(x)} \begin{pmatrix} C_1(x) \\ C_2(x) \end{pmatrix} \end{aligned}$$

Deriving this yields

$$\begin{aligned} z'(x) &= C_1'(x) y_1(x) + \cancel{C_1(x) y_1'(x)} + C_2'(x) y_2(x) + \cancel{C_2(x) y_2'(x)} \\ &= \cancel{C_1(x) A y_1(x)} + \cancel{C_2(x) A y_2(x)} + b(x) \\ &= b(x) \end{aligned}$$

This can be explicitly solved

$$\begin{aligned} C_1'(x) \sin(x) + C_2'(x) \cos(x) &= \sin(x) \\ C_1'(x) \cos(x) - C_2'(x) \sin(x) &= \cos(x) \end{aligned}$$

Leading to

$$\begin{aligned} C_1'(x) &= C_1'(x)(\sin^2(x) + \cos^2(x)) = \sin^2(x) + \cos^2(x) = 1 \\ C_2'(x) &= C_2'(x)(\cos^2(x) - \sin^2(x)) = 0 \end{aligned}$$

Thus

$$\begin{aligned} C_1(x) &= x \\ C_2(x) &= 0 \end{aligned}$$

So the general solution of the homogeneous equation is

$$y_h = \begin{pmatrix} x \sin(x) \\ x \cos(x) \end{pmatrix}$$

Our next goal is to find a solution of $y' = Ay$ with $A \in \mathbb{R}^{n \times n}$ constant. In one dimension the solution would be

$$y = Ce^{Ax}$$

Does this also hold for $n > 1$?

Remark 8.48. Let $A \in \mathbb{R}^{n\times n}$

$$e^{Ax} = \sum_{k=0}^{\infty} \frac{1}{k!}(Ax)^k = \sum_{k=0}^{\infty} \frac{1}{k!} A^k x^k$$

We have

$$\sum_{k=0}^{\infty} \frac{1}{k!} \left\| A^k x^k \right\| \le \sum_{k=0}^{\infty} \frac{|x|^k}{k!} \|A\|^k = e^{\|x\|\|A\|} < \infty$$

Thus, e^{Ax} is defined $\forall A \in \mathbb{R}^{n\times n}$, $\forall x \in \mathbb{R}$. Deriving this yields

$$\frac{\mathrm{d}}{\mathrm{d}x} e^{Ax} = \sum_{k=1}^{\infty} \frac{1}{k!} A^k x^{k-1} = A \sum_{k=1}^{\infty} \frac{1}{(k-1)!} A^{k-1} x^{k-1} = Ae^{Ax}$$

Theorem 8.49. *Let $A \in \mathbb{R}^{n\times n}$. The IVP*

$$y' = Ay \qquad\qquad y(x_0) = y_0$$

is solved exactly by

$$y(x) = e^{A(x-x_0)} y_0$$

Proof. Without proof. □

Remark 8.50. (i) The problem of solving IVPs can be reduced to a problem of calculating a matrix exponential.

(ii) The following does NOT generall hold

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} e^{A(x)} &= A'(x) e^{A(x)} \\ e^{A+B} &= e^A e^B \end{aligned}$$

(iii) Let v be an eigenvector of A to the eigenvalue λ. Then

$$\begin{aligned} e^{Ax} v = \left(\sum_{k=0}^{\infty} \frac{1}{k!} A^k x^k \right) v &= \sum_{k=0}^{\infty} \frac{x^k}{k!} A^k v \\ &= \left(\sum_{k=0}^{\infty} \frac{x^k}{k!} \lambda^k \right) v = e^{\lambda x} v \end{aligned}$$

Example 8.51. Consider the IVP

$$\begin{pmatrix} y \\ z \end{pmatrix}' = \underbrace{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}}_{A} \begin{pmatrix} y \\ z \end{pmatrix} \qquad\qquad y_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

This A is diagonalizable and has the eigenvalues

$$\lambda_1 = -1 \qquad \lambda_2 = 1$$

and the eigenvectors

$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

So we can solve this ODES by calculating

$$\begin{aligned} e^{Ax} y_0 = e^{Ax} \cdot \frac{1}{2}(v_1 + v_2) &= \frac{1}{2}\left(e^{\lambda_1 x} v_1 + e^{\lambda_2 x} v_2\right) \\ &= \frac{1}{2}\left(e^{x} v_1 + e^{-x} v_2\right) \end{aligned}$$

And thus

$$y(x) = \frac{1}{2}\left(e^x + e^{-x}\right) \qquad z(x) = \frac{1}{2}\left(e^x - e^{-x}\right)$$

Remark 8.52. Often the process above is formulated as follows: Start by defining

$$y(x) = c \cdot e^{\lambda x} v \quad c, \lambda \in \mathbb{K} \text{ and } v \in \mathbb{R}$$

Insert this into the ODE

$$c\lambda e^{\lambda x} = c e^{\lambda x} A v$$

So λ is an eigenvalue of A to the eigenvector v.

Theorem 8.53. *Let $A \in \mathbb{R}^{n\times n}$ be diagonalizable, and $v_1, \cdots, v_n$ is a basis of eigenvectors to the eigenvalues $\lambda_1, \cdots, \lambda_n$. Then the functions*

$$y_i(x) = e^{\lambda_i x} v_i \quad i \in \{1, \cdots, n\}$$

are a fundamental system to the ODES

$$y' = Ay$$

Proof. We have

$$e^{Ax} v_i = e^{\lambda_i x} v_i \tag{8.51}$$

In $x = 0$ the

$$y_1(0) = v_1,\ y_2(0) = v_2,\ \cdots,\ y_n(0) = v_n \tag{8.52}$$

are linearly independent, so the $y_1, \cdots, y_n$ are also linearly independent. ∎

Remark 8.54. (i) There is a special case, where $A \in \mathbb{R}^{n\times n}$ is not diagonalizable in the real number space, but in the complex number space. Let $\lambda = \lambda_r + \lambda_i$ be the eigenvalue to the eigenvector $v = v_r + v_i$. Then

$$e^{\lambda_r x}(v_r \sin(\lambda_i x) + v_i \cos(\lambda_i x))$$
$$e^{\lambda_r x}(v_r \cos(\lambda_i x) + v_i \sin(\lambda_i x))$$

be linearly independent, real-valued solutions. To solve the IVP

$$y(x) = Ce^{\lambda x}v \qquad\qquad y(0) = y_0$$

we want to transform it into an eigenvalue problem and find a solution to that. Doing that gives us

$$y(x) = C_1 e^{\lambda_1 x} v_1 + \cdots + C_n e^{\lambda_n x} v_n$$

By inserting the initial condition we can find

$$C_1 v_1 + C_2 v_2 + \cdots + C_n v_n = y_0$$

Finding the $C_1, \cdots, C_n$ shows us that the solution is automatically real.

(ii) If A is not diagonalizable one can try and bring A into Jordan normal form.

Example 8.55. Consider the IVP

$$\begin{pmatrix} y \\ z \end{pmatrix}' = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} \qquad\qquad y_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

The eigenvalues and eigenvectors are

$$\lambda_1 = i \qquad\qquad \lambda_2 = -i$$
$$v_1 = \begin{pmatrix} 1+i \\ -1+i \end{pmatrix} \qquad\qquad v_2 = \begin{pmatrix} 1-i \\ -1-i \end{pmatrix}$$

Thus we have the general solution

$$C_1 e^{ix} v_1 + C_2 e^{-ix} v_2$$

which expands to

$$(i+1)C_1\cancel{e^{i0}} + (1-i)C_2\cancel{e^{-i0}} = 1$$
$$(i-1)C_1\cancel{e^{i0}} + (-1-i)C_2\cancel{e^{-i0}} = 0$$

and solves to

$$C_1 = \frac{1}{4}(1-i) \qquad\qquad C_2 = \frac{1}{4}(1+i)$$

So the solution to the IVP is

$$y(x) = \cos(x) \qquad\qquad z(x) = -\sin(x)$$

Theorem 8.56. *Let* $a_1, \cdots, a_{n-1} \in \mathbb{C}$. *Let* $\lambda_1, \cdots, \lambda_k$ *be the roots of the polynomial*

$$a_0 + a_1\lambda + \cdots + a_{n-1}\lambda^{n-1} + \lambda^n$$

and $\nu_1, \cdots, \nu_k$ *their multiples. Then the functions*

$$x \longmapsto x^l e^{\lambda_i x} \quad i \in \{1, \cdots, k\}, l \in \{0, \cdots, \nu_{i_1}\}$$

form a fundamental system for

$$a_0 y + a_1 y' + \cdots + a_{n-1} y^{(n-1)} + y^{(n)}$$

Chapter 9

Integration over Submanifolds

9.1 Line Integrals

Definition 9.1. Let I be an interval and $n \in \mathbb{N}$. A parametrized curve (or path) in $\mathbb{R}^n$ is a continuous mapping

$$\gamma : I \longrightarrow \mathbb{R}^n$$

A parametrized curve is said to be regular if it is C^1 and $\gamma'(t) \neq 0 \quad \forall t \in I$. It is said to be piecewise regular if there is a disjoint decomposition

$$I = I_1 \cup I_2 \cup \cdots \cup I_n$$

into partial intervals such that γ is regular on each partial interval.

A curve is a subset of $\mathbb{R}^n$ that is the image of a parametrized curve. If $\mathcal{C}$ is a curve, then

$$\gamma : I \longrightarrow \mathbb{R}^n$$

is said to be the parametrization of $\mathcal{C}$, if $\gamma(I) = \mathcal{C}$ and if γ is injective on $\mathring{I}$. The curves in this chapter will always be regular.

Example 9.2. (i) $\alpha, \kappa > 0$:

$$\begin{aligned} \gamma : \mathbb{R} &\longrightarrow \mathbb{R}^3 \\ t &\longmapsto (\cos(\alpha t), \sin(\alpha t), \kappa t) \end{aligned}$$

This is the parametrization of a screw curve.

(ii) The unit circle

$$\left\{(x, y) \in \mathbb{R}^2 \,\middle|\, x^2 + y^2 = 1\right\}$$

is a curve with the parametrization

$$\begin{aligned} \gamma : [0, 2\pi] &\longrightarrow \mathbb{R}^2 \\ t &\longmapsto (\cos t, \sin t) \end{aligned}$$

(iii) A square

$$\left\{(x, y) \in \mathbb{R}^2 \,\middle|\, \max\{|x_1|, |x_2|\} = 1\right\}$$

is a piecewise regular curve.

Remark 9.3. Let $\gamma : I \to \mathbb{R}^n$ be regular, $f : \gamma(I) \to \mathbb{R}$ be continuous and $a, b \in \mathring{I}$. A decomposition Z is given by the grid points

$$a = t_0 < t_1 < \cdots < t_n = b$$

The fineness of Z is given by

$$m(Z) := \max_{t \in \{0,1,\cdots,n-1\}} (t_{i+1} - t_i)$$

We can represent I in terms of Z via

$$I(Z) := \sum_{i=0}^{n-1} f(\gamma(t_i))\|\gamma(t_{i+1}) - \gamma(t_i)\|$$

Or in integral representation

$$I(Z) = \int_a^b \underbrace{\sum_{i=0}^{n-1} f(\gamma(t_i))\frac{\|\gamma(t_{i+1}) - \gamma(t_i)\|}{\|t_{i+1} - t_i\|}\mathbb{1}_{[t_i,t_{i+1})}(t)}_{g_Z(t)}\,\mathrm{d}t$$

So let (Z_j) be a sequence of decompositions with

$$m(Z_j) \xrightarrow{j\to\infty} 0$$

Let $t \in [a, b]$ not be a grid point of any Z_j. Then there exists a unique grid poiont t_{j,i_j} such that $t \in [t_{j,i_j}, t_{j,i_{j+1}}]$. Then

$$\lim_{j\to\infty} t_{j,i_j} = \lim_{j\to\infty} t_{j,i_{j+1}} = t$$

And thus

$$\lim_{j\to\infty} g_{Z_j}(t) = f(\gamma(t))\big\|\gamma'(t)\big\|$$

$\forall t$ that are not grid points of Z_j, this means tahat

$$g_{Z_j} \xrightarrow{j\to\infty} f\big\|\gamma'\big\|$$

almost everywhere. The dominated convergence theorem then tells us

$$I(Z_j) = \int_a^b g_{Z_j}(t)\,\mathrm{d}t \xrightarrow{j\to\infty} \int_a^b f(\gamma(t))\big\|\gamma'(t)\big\|\,\mathrm{d}t$$

Special case: For $f \equiv 1$ one gets the arc length.

Definition 9.4 (Line Integrals, Arc Length)**.** Let I be an interval and $\gamma : I \to \mathbb{R}^n$ a parametrized curve. Define the functions

$$f : \gamma(I) \longrightarrow \mathbb{R} \qquad\qquad E : \gamma(I) \longrightarrow \mathbb{R}^n$$

Then

$$\int_\gamma f\,\mathrm{d}s := \int_I f(\gamma(t))\big\|\gamma'(t)\big\|\,\mathrm{d}t$$

is said to be a scalar line integral (line integral of first kind), and

$$\int_\gamma \langle E|\mathrm{d}s\rangle := \int_I \big\langle E(\gamma(t))\big|\gamma'(t)\big\rangle\,\mathrm{d}t$$

is said to be a vector line integral (line integral of second kind). The function f or the vector field E are integrable along γ if the according integral exists. The integral

$$\int_\gamma \mathrm{d}s$$

is the arc length of γ, and γ is said to be rectifiable if this integral is finite.

If the curve γ is closed, i.e. if $I = [a, b] \quad a, b \in \mathbb{R}$ and

$$\gamma(a) = \gamma(b)$$

Then the above integrals are often notated as

$$\oint_\gamma \mathrm{d}s \qquad\qquad \oint_\gamma \langle E|\mathrm{d}s\rangle$$

to emphasize that the curve is closed. This changes nothing about the formulas, it is merely visual. I will try to adhere to this style.

Example 9.5 (Circumference of the unit circle)*.* Define

$$\begin{aligned}\gamma : [0, 2\pi] &\longrightarrow \mathbb{R}^2\\ t &\longrightarrow (\cos(t), \sin(t))\end{aligned}$$

and derive this function

$$\gamma'(t) = (-\sin(t), \cos(t)) \implies \left\|\gamma'(t)\right\| = 1$$

Then the circumference is

$$\oint_\gamma \mathrm{d}s = \int_0^{2\pi} \mathrm{d}t = 2\pi$$

Remark 9.6*.*

(i) If γ is only piecewise regular then the integrands might not be defined for all t.

(ii) Line integrals don't depend on the chosen parametrization. This means if $\mathcal{C}$ is a curve and

$$\gamma : I \to \mathcal{C} \qquad\qquad \rho : J \to \mathcal{C}$$

are parametrizations, then

$$\int_\gamma f\,\mathrm{d}s = \int_\rho f\,\mathrm{d}s$$

We also write

$$\int_{\mathcal{C}} f\,\mathrm{d}s$$

The same holds for vector integrals.

(iii) Both kinds of integrals depend on the scalar product.

(iv) Both kinds of integrals are special cases of integrals over so called One-forms

Theorem 9.7. *Let $\gamma : I \to \mathbb{R}^n$ be a parametrized curve, and $\vartheta : J \to I$ a diffeomorphism (so $\vartheta \in C^1$ and $\vartheta'(t) \neq 0 \ \ \forall t \in J$). Let $f : \gamma(I) \to \mathbb{R}$, then*

$$\int_\gamma f \, \mathrm{d}s = \int_{\gamma \circ \vartheta} f \, \mathrm{d}s$$

Proof. We can assume I, J to be open, since the endpoints of the integrals are a null set and thus don't matter. W.l.o.g. let γ be regular. Then

$$\begin{aligned} \int_{\gamma \circ \vartheta} f \, \mathrm{d}s &= \int_J f(\gamma \circ \vartheta)(t) \big\| (\gamma \circ \vartheta)'(t) \big\| \, \mathrm{d}t \\ &= \int_J f(\gamma(\vartheta(t))) \big\| \gamma'(\vartheta(t)) \vartheta'(t) \big\| \, \mathrm{d}t \\ &= \int_J f(\gamma(\vartheta(t))) \big\| \gamma'(\vartheta(t)) \big\| \big| \vartheta'(t) \big| \, \mathrm{d}t \qquad (9.1) \\ &= \int_I f(\gamma(\tau)) \big\| \gamma'(\tau) \big\| \, \mathrm{d}\tau \\ &= \int_\gamma f \, \mathrm{d}s \end{aligned}$$

□

Remark 9.8. (i) One can show that for a curve $\mathcal{C}$ and parametrizations

$$\gamma : I \to \mathcal{C} \qquad\qquad \rho : J \to \mathcal{C}$$

there exists a diffeomorphism $\vartheta : J \to I$ such that

$$\rho = \gamma \circ \vartheta$$

So the line integral of first degree doesn't depend on the parametrization.

(ii) A line integral of second degree doesn't depend on the parametrization if the parametrizations run along the curve in the same direction. So if $\vartheta' > 0$, ϑ is said to conserve orientation. If $\vartheta' < 0$ then the integral switches sign.

Example 9.9. Let $\gamma : I \to \mathbb{R}^3$ be the trajectory of a point mass, and $F : \mathbb{R}^3 \to \mathbb{R}^3$ a time-independent forcefield. The work done is then given by

$$W := \int_\gamma \langle F | \mathrm{d}s \rangle$$

The fact that the parametrization can be chosen arbitrarily means that the work done in a forcefield is independent from the velocity of the point mass.

Remark 9.10. (i) Line integrals are linear in f or E, meaning for

$$f, g : \gamma(I) \to \mathbb{R}, \quad \lambda \in \mathbb{R}$$

we have

$$\int_\gamma (g + \lambda g) \, \mathrm{d}s = \int_\gamma f \, \mathrm{d}s + \lambda \int_\gamma g \, \mathrm{d}s$$

(ii) Parametrized curves over compact intervals can be reparametrized so that $I = [0, 1]$.

(iii) Let

$$\gamma : [0, 1] \to \mathbb{R}^n \qquad\qquad \rho : [0, 1] \to \mathbb{R}^n$$

be curves with $\gamma(1) = \rho(0)$. Define

$$\begin{aligned} \gamma^{-1} : [0,1] &\longrightarrow \mathbb{R}^n \\ t &\longrightarrow \gamma(1-t) \end{aligned} \qquad\qquad \begin{aligned} \gamma\rho : [0,1] &\longrightarrow \mathbb{R}^n \\ t &\longrightarrow \begin{cases} \gamma(2t), & t \leq 0.5 \\ \rho(2t+1), & t > 0.5 \end{cases} \end{aligned}$$

Then we have

$$\begin{aligned} \int_{\gamma^{-1}} f \, \mathrm{d}s &= \int_\gamma f \, \mathrm{d}s \\ \int_{\gamma\rho} f \, \mathrm{d}s &= \int_\gamma f \, \mathrm{d}s + \int_\rho f \, \mathrm{d}s \\ \int_{\gamma^{-1}} \langle E | \mathrm{d}s \rangle &= - \int_\gamma \langle E | \mathrm{d}s \rangle \\ \int_{\gamma\rho} \langle E | \mathrm{d}s \rangle &= \int_\gamma \langle E | \mathrm{d}s \rangle + \int_\rho \langle E | \mathrm{d}s \rangle \end{aligned}$$

Definition 9.11. Let $U \subset \mathbb{R}^n$ be open and $f : U \to \mathbb{R}$ a C^1-function. Define

$$\vec{\nabla} f = (\partial_1 f, \partial_2 f, \cdots, \partial_m f)$$

The vector field $E : U \to \mathbb{R}^n$ is said to be conservative if there is a function $g : U \to \mathbb{R}$ such that

$$E = \vec{\nabla} g$$

g is the potential of E.

Remark 9.12. (i) In physics the sign is typically switched, so

$$E = -\vec{\nabla} g$$

(ii) The IDE

$$p(x,y) + q(x,y)y' = 0$$

is exact if and only if the vector field (p,q) is conservative.

(iii) If E is conservative and C^1, then

$$\partial_i E_j = \partial_j E_i$$

This condition is not sufficient in general.

(iv) If g is a potential for E, then the functions

$$g + c \quad c \in \mathbb{R}$$

are also potentials.

(v) If E is conservative, g a potential and $\gamma : [a,b] \to \mathbb{R}^n$ a curve, then

$$\begin{aligned}\int_\gamma \langle E|\mathrm{d}s\rangle &= \int_a^b \big\langle E(\gamma(t))\big|\gamma'(t)\big\rangle \,\mathrm{d}t \\ &= \int_a^b \big(\partial_1 g(\gamma(t))\gamma_1'(t) + \cdots + \partial_n g(\gamma(t))\gamma_n'(t)\big)\,\mathrm{d}t \\ &= \int_a^b (g\circ\gamma)'(t)\,\mathrm{d}t = g(\gamma(b)) - g(\gamma(a))\end{aligned}$$

The vector line integral over conservative fields is independent from the chosen path (it only depends on the start and end points).

(vi) Let U be open, path-connected and $E : U \to \mathbb{R}^n$ a conservative vector field. Choose a fixed $x_0 \in U$, and for $x \in U$ choose a parametrized curve γ_x from x_0 to x. Then

$$x \longmapsto \int_{\gamma_x} \langle E|\mathrm{d}s\rangle$$

is a potential, because if g is an arbitrary potential we have

$$\int_{\gamma_x} \langle E|\mathrm{d}s\rangle = g(x) - g(x_0) \quad \forall x \in U$$

Example 9.13. (i) Let

$$\begin{aligned}E : \mathbb{R}^3 \setminus \{0\} &\longrightarrow \mathbb{R}^3 \\ x &\longmapsto -\frac{x}{\|x\|^3}\end{aligned}$$

This field is conservative, with the potential

$$\begin{aligned}\phi : \mathbb{R}^3 \setminus \{0\} &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{1}{\|x\|}\end{aligned}$$

(ii) Let

$$\begin{aligned} E : \mathbb{R}^2 \setminus \{0\} &\longrightarrow \mathbb{R}^2 \\ (x, y) &\longmapsto \left(-\frac{y}{x^2 + y^2}, \frac{x}{x^2 + y^2} \right) \end{aligned}$$

Then

$$\partial_1 E_2 = \frac{1}{x^2 + y^2} - \frac{2x^2}{(x^2 + y^2)} = \frac{y^2 - x^2}{(x^2 + y^2)^2} = \partial_2 E_1$$

We can calculate the line integral of E along the unit circle

$$\begin{aligned} \gamma : [0, 2\pi] &\longrightarrow \mathbb{R}^2 \\ t &\longmapsto (\cos t, \sin t) \end{aligned}$$

Then

$$E(\gamma(t)) = (-\sin t, \cos t) = \gamma'(t)$$

The integral is then

$$\int_\gamma \langle E | \mathrm{d}s \rangle = \int_0^{2\pi} \|(-\sin t, \cos t)\|^2 \, \mathrm{d}t = 2\pi \neq 0$$

(iii) In the chapter about differential equations we looked at an exact equation in Example 8.17:

$$(2x + y^2) + (2xy)y' = 0$$

We can now use curve integrals to calculate the potential function more easily. For that let $x_0 = (0, 0)$. Then for (ξ, η) we can define a curve connecting x_0 and (ξ, η) for $t \in [0, 1]$:

$$t \longmapsto (\xi t, \eta t)$$

Consider the vector field

$$E(x, y) = (2x + y^2, 2xy)$$

Then

$$\begin{aligned} (\xi, \eta) \longmapsto \int_\gamma \langle E | \mathrm{d}s \rangle &= \int_0^1 \langle E(\xi t, \eta t) | (\xi, \eta) \rangle \, \mathrm{d}t \\ &= \int_0^1 (2\xi^2 t + \eta^2 \xi t^2 + 2\xi \eta^2 t^2) \, \mathrm{d}t \\ &= \xi^2 + \eta^2 \xi \end{aligned}$$

Theorem 9.14. *Let $U \subset \mathbb{R}^n$ be an open subset. A continuous vector field $E : U \to \mathbb{R}^n$ is conservative if and only if for every closed curve $\gamma : [0, 1] \to U$ the following holds*

$$\oint_\gamma \langle E | \mathrm{d}s \rangle = 0$$

Proof. Line integrals over E are path independent. Let $\gamma, \rho : [0,1] \to U$ be paths with

$$\gamma(0) = \rho(0) \qquad \gamma(1) = \rho(1) \tag{9.2}$$

Then $\gamma\rho^{-1}$ is closed, so

$$0 = \int_{\gamma\rho^{-1}} \langle E|\mathrm{d}s\rangle = \int_{\gamma} \langle E|\mathrm{d}s\rangle - \int_{\rho} \langle E|\mathrm{d}s\rangle \tag{9.3}$$

Assume that U is path continuous. Choose a fixed $x_0 \in U$ and let $g : U \to \mathbb{R}$. Then

$$g(x) = \int_{x_0}^{x} \langle E|\mathrm{d}s\rangle \tag{9.4}$$

Performing a directional derivation in direction $h \in \mathbb{R}^n$ yields

$$\begin{aligned} g(x+ah) - g(x) &= \int_{x_0}^{x+ah} \langle E|\mathrm{d}s\rangle - \int_{x_0}^{x} \langle E|\mathrm{d}s\rangle \\ &= \int_{x}^{x+ah} \langle E|\mathrm{d}s\rangle \\ &= \int_0^a \langle E(x+th)|h\rangle \,\mathrm{d}t \end{aligned} \tag{9.5}$$

Here we have chosen a linear path of integration between x_0 and x, and between x and $x + ah$. In other words, we're integrating along

$$t \longmapsto x + th \tag{9.6}$$

Using the intermediate value theorem, we can find that $\exists \xi_a \in (0, a)$ such that

$$\int_0^a \langle E(x+th)|h\rangle \,\mathrm{d}t = \langle E(x+\xi_a h)|h\rangle \cdot a \tag{9.7}$$

Then we have

$$\partial_h g(x) = \lim_{a\to 0} \frac{g(x+ah) - g(x)}{a} = \lim_{a\to 0} \langle E(x+\xi_a h)|h\rangle = \langle E(x)|h\rangle \tag{9.8}$$

So if h is a standard basis e_i, then

$$\partial_i g(x) = E_i(x) \tag{9.9}$$

Thus the partial derivative of g is continuous, and therefore g is continuously differentiable, and thus a potential. □

9.2 Surface Integrals

In this section we will exclusively look at surfaces in $\mathbb{R}^3$.

Definition 9.15. Let $V \subset \mathbb{R}^2$ be open. A mapping $\phi : V \to \mathbb{R}^3$ is said to be a parametrized surface if it is C^1 and if $\partial_1\phi(t),\ \partial_2\phi(t)$ are linearly independent $\forall t \in V$. A subset $S \subset \mathbb{R}3$ is said to be a regular surface, if there exist:

- open subsets $U_1, \cdots, U_n \subset \mathbb{R}^3$
- open subsets $V_1, \cdots, v_n \subset \mathbb{R}^2$
- mappings $\phi_i : V_i \longrightarrow U_i \cap S$

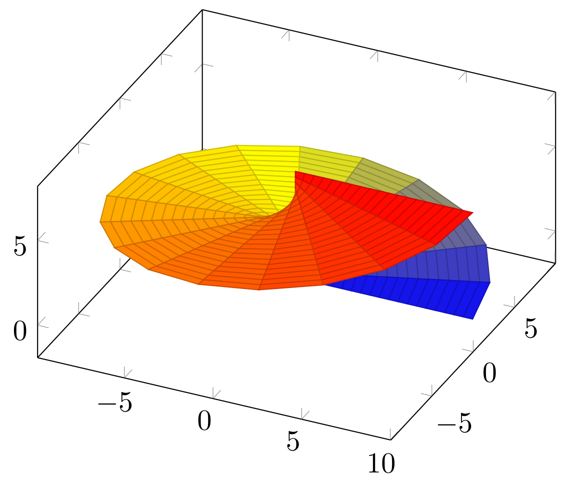
such that the ϕ_i are parametrized surfaces, bijective and have a continuous ϕ^{-1}. These S are also said to be embedded, two-dimensional manifolds, and the ϕ_i are then called maps. The collection of all maps ϕ_i are called atlas.

$S \subset \mathbb{R}^3$ is said to be a piecewise regular surface if there exist parametrized surfaces $\phi_1, \cdots, \phi_n$, parametrized paths $\gamma_1, \cdots, \gamma_k$ and points $P_1, \cdots, P_l$ such that

$$S = \phi_1(V_1) \cup \cdots \cup \phi_n(V_n) \cup \gamma(I_1) \cup \cdots \cup \gamma(I_k) \cup \{P_1, \cdots, P_l\}$$

Example 9.16. (i) Consider

$$\begin{aligned} \phi : (0, \infty) \times \mathbb{R} &\longrightarrow \mathbb{R}^3 \\ (s, t) &\longmapsto (s\cos t, s\sin t, t) \end{aligned}$$

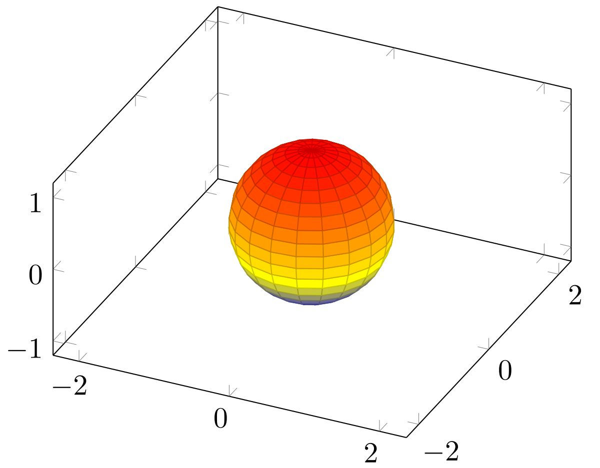


(ii) The set

$$S^2 := \left\{ (x, y, z) \in \mathbb{R}^3 \,\middle|\, x^2 + y^2 + z^2 = 1 \right\}$$

is a regular surface. A map describing this surface would be

$$\begin{aligned} \phi : (0, 2\pi) \times (0, \pi) &\longrightarrow \mathbb{R}^3 \\ (s, t) &\longmapsto (\cos(s)\sin(t), \sin(s), \sin(t), \cos(t)) \end{aligned}$$



(iii) The unit cube

$$\left\{(x,y,z)\in\mathbb{R}^3 \,\middle|\, \max\{|x|,|y|,|z|\}\right\}$$

is a piecewise regular surface.

Remark 9.17. Our definition of regular curves is not equal to the definition of one-dimensional embedded manifolds, because regular curves are not allowed to intersect themselves.

Definition 9.18 (Cross Product)**.** Define the vectors $v=(v_1,v_2,v_3)$ and $w=(w_1,w_2,w_3)\in\mathbb{R}^3$. Then

$$v\times w=\begin{pmatrix} v_2w_3-v_3w_2\\ v_3w_1-v_1w_3\\ v_1w_2-v_2w_1\end{pmatrix}^T$$

is the cross product of v and w.

Remark 9.19. (i) The cross product is linear in v and w, with

$$v\times w=w\times v$$

(ii) $v\times w$ is orthogonal to v and w.

(iii) The cross product is not associative, but it fulfils the Jacobi-identity:

$$u\times(v\times w)+v\times(w\times u)+w\times(u\times v)=0$$

(iv) v and w are linearly dependent if and only if $v\times w=0$

(v) The definition depends on the coordinates of v and w. So the choice of a basis matters. In reality the cross product depends on the scalar product.

(vi) Consider the space of anti-symmetric matrices

$$V = \left\{ A \in \mathbb{R}^{n \times n} \,\middle|\, A^T = -A \right\} \qquad \dim V = \frac{1}{2}n(n-1)$$

The cross product is an outer product on V. V can be interpreted as an anti-symmetric bilinear form, or as the space of infinitesimal rotations (Lie-algebra to the Lie-group of rotations). This is not relevant.

Definition 9.20. Let $V \subset \mathbb{R}^2$ be open and $\phi : V \to \mathbb{R}^3$ a parametrized surface. Then

$$\sigma_\phi(t) = \partial_1 \phi(t) \times \partial_2 \phi(t)$$

is said to be a vector surface element of ϕ, and $\|\sigma_\phi(t)\|$ is the scalar surface element at the point $\phi(t)$.

Remark 9.21. The surface element can be defined for arbitrary C^1-mappings. ϕ is a parametrized surface if and only if $\sigma_\phi(t) \neq 0$ or $\|\sigma_\phi(t)\| \neq 0 \;\; \forall t \in V$.

Example 9.22. (i) Consider the unit sphere

$$\begin{aligned} \phi : (0, 2\pi) \times (0, \pi) &\longrightarrow \mathbb{R}^3 \\ (s, t) &\longmapsto (\cos(s)\sin(t), \sin(s), \sin(t), \cos(t)) \end{aligned}$$

The derivatives of ϕ are

$$\begin{aligned} \partial_1 \phi(s, t) &= (-\sin(t)\sin(s), \sin(t)\cos(s), \cos(t)) \\ \partial_2 \phi(s, t) &= (\cos(t)\cos(s), \cos(t)\sin(s), -\sin(t)) \end{aligned}$$

Then the surface elements are

$$\begin{aligned} \sigma_\phi(s, t) &= (-\sin^2(t)\cos(s), -\sin^2(t)\sin(s), -\sin(t)\cos(t)) \\ \|\sigma_\phi(s, t)\| &= \sin(t) \end{aligned}$$

(ii) Let $U \subset \mathbb{R}^2$ be open, and $f : U \to \mathbb{R}$ a continuously differentiable function. Then

$$\phi(s, t) = (s, t, f(s, t))$$

is a parametrization of the graph of f. The derivatives are

$$\partial_1 \phi(s, t) = (1, 0, \partial_1 f(s, t)) \qquad \partial_2 \phi(s, t) = (0, 1, \partial_2 f(s, t))$$

And the surface elements are

$$\begin{aligned} \sigma_\phi(s, t) &= (-\partial_1 f(s, t), -\partial_2 f(s, t), 1) \\ \|\sigma_\phi(s, t)\| &= \sqrt{(\partial_1 f(s, t))^2 + (\partial_2 f(s, t))^2 + 1} \end{aligned}$$

Definition 9.23. Let $V \subset \mathbb{R}^2$ be open, $\phi : V \to \mathbb{R}^3$ a parametrized surface and $f : \phi(V) \to \mathbb{R}$. Then

$$\iint_\phi f \,\mathrm{d}\sigma := \iint_V f(\phi(t)) \|\sigma_\phi(t)\| \,\mathrm{d}\lambda^2(t)$$

is said to be the scalar surface integral of f over ϕ. The integral

$$\iint_\phi \mathrm{d}\sigma$$

is said to be the surface of $\phi(V)$.

Lemma 9.24. *Let $V, \tilde{V} \subset \mathbb{R}^2$ be open, $\phi : V \to \mathbb{R}^3$ a parametrized surface and $T : \tilde{V} \to V$ a diffeomorphismus. Set $\psi = \phi \circ T$, then the surface element is*

$$\sigma_\psi(t) = \det(DT(t)) \cdot \sigma_\phi(T(t))$$

Proof. Calculate

$$D\psi(t) = D\phi(T(t))DT(t) \tag{9.10}$$

Or if we consider each column of the derivative separately

$$\partial_1\psi = \partial_1 T_1 \cdot \partial_2\phi + \partial_1 T_2 \cdot \partial_2\phi \tag{9.11a}$$

$$\partial_2\psi = \partial_2 T_1 \cdot \partial_1\phi + \partial_2 T_2 \cdot \partial_1\phi \tag{9.11b}$$

Then

$$\begin{aligned}\sigma_\psi = \partial_1\psi \times \partial_2\psi &= (\partial_1 T_1)(\partial_2 T_2)\partial_1\phi \times \partial_2\phi + (\partial_1 T_2)(\partial_2 T_1)\partial_2\phi \times \partial_1\phi \\ &= (\det DT)\sigma_\phi\end{aligned} \tag{9.12}$$

□

Remark 9.25. Let there be the same notation as above, and $f : \phi(V) \to \mathbb{R}$

$$\begin{aligned}\iint_\psi f \,\mathrm{d}\sigma &= \iint_{\tilde{V}} f(\psi(t)) \|\sigma_\theta(t)\| \,\mathrm{d}\lambda^2(t) \\ &= \iint_{\tilde{V}} f(\phi \circ T(t)) \|\sigma_\psi(T(t))\| \det(DT(t)) \,\mathrm{d}\lambda^2(t) \\ &= \iint_V f(\phi(s)) \|\sigma_\phi(s)\| \,\mathrm{d}\lambda^2(s) = \iint_\phi f \,\mathrm{d}\sigma\end{aligned}$$

In general we have to decompose a (piecewise) regular surface into disjoint regular pieces and parametrize them. The surface integral – so the sum of integrals over the pieces – is independent of the chosen decomposition and parametrization. Structures of lower dimensions (curves, points) don't contribute to surface integrals.

Example 9.26. (i) We want to calculate the surface of the unit sphere. Using the parametrization we established earlier, we can get

$$\begin{aligned}\iint_\phi \mathrm{d}\sigma = \iint_{(0,2\pi)\times(0,\pi)} \sin(t)\,\mathrm{d}\lambda^2(s,t) &= \int_0^\pi \int_0^{2\pi} \sin(t)\,\mathrm{d}s\,\mathrm{d}t \\ &= \int_0^\pi 2\pi\sin(t)\,\mathrm{d}t \\ &= 4\pi\end{aligned}$$

(ii) Let $U \subset \mathbb{R}^2$ be open and

$$\begin{aligned}\phi : U &\longrightarrow \mathbb{R}^3 \\ (s,t) &\longmapsto (s,t,0)\end{aligned}$$

Then $\|\sigma_\phi\| = 1$, and let $f : \mathbb{R}^2 \times 0 \to \mathbb{R}$:

$$\iint_\phi f\,\mathrm{d}\sigma = \iint_U f(s,t,0)\,\mathrm{d}\lambda^2(s,t)$$

Definition 9.27. Let $V \subset \mathbb{R}^2$ be open, $\phi : V \to \mathbb{R}^3$ a parametrized surface and let $E : \phi(V) \to \mathbb{R}^3$. Then

$$\iint_\phi \langle E|\mathrm{d}\sigma\rangle := \iint_V \langle E(\phi(t))|\sigma_\phi(t)\rangle\,\mathrm{d}\lambda^2(t)$$

is said to be the vector surface integral of E over ϕ.

Remark 9.28. This integral is independent from the parametrization if the determinant $\det DT$ is positive. Then T is said to conserve orientation. Otherwise the integral is switching signs.

For general (piecewise) regular surfaces one has to watch out that the parametrizations are consistent. There are surfaces (regular surfaces even) where that isn't possible (so called non-orientable surfaces). For these kinds of surfaces the vector surface integral isn't properly defined.

If a surface splilts $\mathbb{R}^3$ into an "outside" and an "inside", then we typically choose the parametrization where the surface elements point outwards.

Example 9.29. We want to integrate

$$E(x,y,z) := \left(0, 0, \frac{1}{1+z^2}(x\sin y + y\cos x)\right)$$

over the surface of the unit cube. E points in z-direction, so the integrals over the sides disappear. So we can parametrize the "lid"

$$(s,t) \longmapsto (s,t,1) \quad s,t \in [-1,1]$$

and calculate the integral

$$\iint \langle E|\mathrm{d}\sigma\rangle = \iint_{(-1,1)^2} \frac{1}{2}(s\sin t + t\cos s)\,\mathrm{d}\lambda^2(s,t)$$

Doing this for the base yields the same result, just with a different sign. So the surface integral over the cube is 0.

9.3 Integral Theorems

Definition 9.30. Let $U \subset \mathbb{R}^3$. We define the following mappings

$$\begin{aligned} \text{Gradient} \quad & \vec{\nabla} : C^1(U) \longrightarrow C^1(U,\mathbb{R}^3) \\ \text{Divergence} \quad & \vec{\nabla}\cdot : C^1(U,\mathbb{R}^3) \longrightarrow C(U) \\ \text{Curl} \quad & \vec{\nabla}\times : C^1(U,\mathbb{R}^3) \longrightarrow C^1(U,\mathbb{R}^3) \\ \text{Laplacian} \quad & \nabla^2 : C^2(U) \longrightarrow C(U) \end{aligned}$$

And define the operations for $f \in C^1(U)$, $g \in C^2(U)$, $E \in C^1(U,\mathbb{R}^3)$

$$\begin{aligned} \vec{\nabla} f &:= (\partial_1 f, \partial_2 f, \partial_3 f) \\ \vec{\nabla}\cdot E &:= \partial_1 E_1 + \partial_2 E_2 + \partial_3 E_3 \\ \vec{\nabla}\times E &:= (\partial_2 E_3 - \partial_3 E_2, \partial_3 E_1 - \partial_1 E_3, \partial_1 E_2 - \partial_2 E_1) \\ \nabla^2 g &:= \partial_1^2 g + \partial_2^2 g + \partial_3^2 g \end{aligned}$$

∇ is called the Nabla operator and it's defined as

$$\nabla = (\partial_1, \partial_2, \partial_3)$$

and subsequently the Laplacian can be defined as

$$\nabla^2 = \vec{\nabla}\cdot\vec{\nabla}$$

Remark 9.31. (i) All of these operations are linear. Typically they operate on everything to their right up until the next $+$ or $-$.

(ii) $\vec{\nabla}$, $\vec{\nabla}\cdot$, ∇^2 can all be extended to $\mathbb{R}^n$, however because the cross product isn't sensibly defined outside of $\mathbb{R}^3$, $\vec{\nabla}\times$ can't be extended to $\mathbb{R}^n$.

(iii) There are some identities:

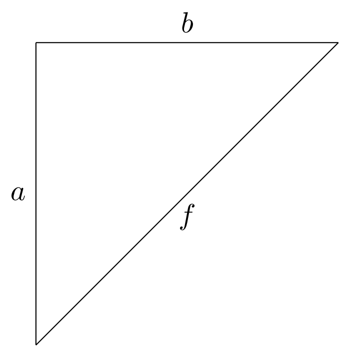
$$\begin{aligned}
\vec{\nabla} \cdot \left(\vec{\nabla} \times E\right) &= 0 = \vec{\nabla} \times \left(\vec{\nabla} E\right) \\
\vec{\nabla} \cdot \left(\vec{\nabla} f\right) &= \nabla^2 f \\
\vec{\nabla}(fg) &= (\vec{\nabla} f)g + f(\vec{\nabla} g) \\
\vec{\nabla} \cdot (fE) &= \left\langle \vec{\nabla} f \middle| E \right\rangle + f(\vec{\nabla} \cdot E) \\
\vec{\nabla} \times (fE) &= (\vec{\nabla} f) \times E + f(\vec{\nabla} \times E) \\
\nabla^2(fE) &= (\nabla^2 f)g + 2\left\langle \vec{\nabla} f \middle| \vec{\nabla} g \right\rangle + f(\nabla^2 g)
\end{aligned}$$

Remark 9.32. A parametrized curve $\gamma : [0,1] \to \mathbb{R}^n$ is said to be simple closed if it doesn't intersect itself (γ is injective on $[0,1)$). In $\mathbb{R}^2$ these kinds of curves split the space into a bounded part U and an unbounded part. We assums γ to be positive oriented (meaning U is "left" of the curve).

Theorem 9.33 (Green's Theorem)**.** *Let $\gamma : [0,1] \to \mathbb{R}^2$ be a simple closed curve, more specifically the boundary of U. Let $E : \mathbb{R}^2 \to \mathbb{R}^2$ be a C^1-vector-field. Then*

$$\iint_U (\partial_1 E_2 - \partial_2 E_1)\, \mathrm{d}\lambda^2 = \oint_\gamma \langle E | \mathrm{d}s \rangle$$

Heuristic Proof. Consider the following special case



So $f : [0,b] \to \mathbb{R}^2$ strictly monotonically increasing, C^1, $f(0) = a < 0$ and $f(b) = 0$. Then define the curves

$$C_1 = [0,b] \times \{0\} \qquad\qquad C_2 = \{0\} \times [0,a] \tag{9.13}$$

And C_3 the graph of f parametrized by

$$t \longmapsto (t, f(t)) \tag{9.14}$$

Since f is monotonically increasing, there exists an inverse function g that us continuously differentiable. Then

$$\begin{aligned}
&\iint_U (\partial_1 E_2 - \partial_2 E_1)\,\mathrm{d}\lambda^2 \\
&= \int_a^0 \int_0^{g(y)} \partial_1 E_2(x,y)\,\mathrm{d}x\,\mathrm{d}y - \int_0^b \int_{f(x)}^0 \partial_2 E_1(x,y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_0^a (E_2(g(y),y)) - E_2(0,y))\,\mathrm{d}y - \int_0^b (E_1(t,f(t)) - E(x,f(x)))\,\mathrm{d}x \\
&= \underbrace{-\int_0^b E_1(t,0)\,\mathrm{d}t}_{\int_{C_1}\langle E|\mathrm{d}s\rangle} + \underbrace{\int_0^a E_2(0,t)\,\mathrm{d}t}_{\int_{C_2}\langle E|\mathrm{d}s\rangle} + \underbrace{\int_0^b (E_1(t,f(t)) + E_2(t,f(t)))f'(t)\,\mathrm{d}t}_{\int_{C_3}\langle E|\mathrm{d}s\rangle} \\
&= \oint_{C_1C_2C_3} \langle E|\mathrm{d}s\rangle
\end{aligned} \tag{9.15}$$

□

Corollary 9.34 (Divergence Theorem in 2D)**.** *Let $E \in C^1(\mathbb{R}^2, \mathbb{R}^2)$ and define $\gamma : [0,1] \to \mathbb{R}^2$ simple closed to be the boundary of U. We set*

$$\sigma(t) = (\gamma'(t), -\gamma'(t))$$

Then

$$\iint_U \vec{\nabla} \cdot E\,\mathrm{d}\lambda^2 = \oint_\gamma \langle E|\mathrm{d}s\rangle = \int_0^1 \langle E(\gamma(t))|\sigma(t)\rangle\,\mathrm{d}t$$

Proof. Set $\tilde{E} = (-E_2, E_1)$ and apply Green's theorem:

$$\begin{aligned}
\iint_U \vec{\nabla} \cdot E\,\mathrm{d}\lambda^2 = \iint_U (\partial_1 E_2 - \partial_2 E_1)\,\mathrm{d}\lambda^2 &= \oint_\gamma \langle E|\mathrm{d}s\rangle \\
&= \int_0^1 \langle E(\gamma(t))|\sigma(t)\rangle\,\mathrm{d}t
\end{aligned} \tag{9.16}$$

□

Corollary 9.35 (Stokes' Theorem in the x-y-plane)**.** *Let $\tilde{E} : \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field, $\gamma : [0,1] \to \mathbb{R}^2$ the simple closed boundary of U. Set $\tilde{\gamma(t)} = (\gamma(t), 0)$ and $\tilde{U} = U \times \{0\}$*

$$\iint_{\tilde{U}} \left\langle \vec{\nabla} \times E \middle| \mathrm{d}\sigma \right\rangle = \oint_{\tilde{\gamma}} \left\langle \tilde{E} \middle| \mathrm{d}s \right\rangle$$

Proof. Choose

$$(x,y) \longmapsto (x,y,0)$$

as a parametrization of $\tilde{U}$ with $\sigma = (0,0,1)$. Set

$$E(x,y) = (\tilde{E}_1(x,y,0), \tilde{E}_2(x,y,0))$$

Then

$$\begin{aligned}\iint_{\tilde{U}} \left\langle \vec{\nabla} \times \tilde{E} \middle| \mathrm{d}\sigma \right\rangle &= \iint_{U} \left\langle \vec{\nabla} \times \tilde{E}(x,y,0) \middle| (0,0,2) \right\rangle \mathrm{d}\lambda^2(x,y) \\ &= \iint_{U} \partial_1 E_2(x,y) - \partial_2 E_1(x,y)\, \mathrm{d}\lambda(x,y) \\ &= \oint_{\gamma} \langle E | \mathrm{d}s \rangle = \oint_{\tilde{\gamma}} \left\langle \tilde{E} \middle| \mathrm{d}s \right\rangle\end{aligned} \tag{9.17}$$

□

Remark 9.36. A set $U \subset \mathbb{R}^n$ is said to be simply connected, if for every closed curve $\gamma : [0,1] \to U$ there exists a continuous mapping $\vartheta : [0,1]^2 \to U$, such that

$$\vartheta(1,t) = \gamma(t) \quad \vartheta(0,t) = \gamma(0) \qquad\qquad \forall t \in [0,1]$$

ϑ is said to be a homotopy.

Theorem 9.37 (Stokes' Theorem)**.** *Let $U \subset \mathbb{R}^3$ be a simply connected, orientable surface whose boundary is a closed curve γ. For U let there be an orientation (so a continuous normal vector field), and orientate γ such that U is to the left of γ relative to the normal direction. Let $E \in C^1(\mathbb{R}^3, \mathbb{R}^3)$ be a vector field, then*

$$\iint_{U} \left\langle \vec{\nabla} \times E \middle| \mathrm{d}\sigma \right\rangle = \oint_{\gamma} \langle E | \mathrm{d}s \rangle$$

Proof. Without proof. □

Example 9.38. The condition that U is simply connected is necessary:

$$(x,y,z) \longmapsto \left(\frac{-y}{x^2+y^2}, \frac{x}{x^2+y^2}, 0 \right)$$

is free of curl. Curve integrals "around the z-axis" can be non-zero.

Remark 9.39. Let $U \subset \mathbb{R}^3$ be simply connected, $E \in C^1(U, \mathbb{R}^3)$. Then

$$E \text{ conservative} \iff \vec{\nabla} \times E = 0$$

If $\vec{\nabla} \times E = 0$, U is simply connected and γ is a closed curve in U, then there exists a surface in U that is bounded by γ. Using Stokes' theorem one can then see that

$$\oint_{\gamma} \langle E | \mathrm{d}s \rangle = 0 \ \ \forall \gamma \text{ closed}$$

And thus E is conservative. A surface A is said to be closed, if it splits $\mathbb{R}^3$ into a bounded and an unbounded part. The bounded part shall be named U and is oriented such that the normals point outwards.

Theorem 9.40 (Divergence Theorem)**.** *Let M be a closed surface and $E \in C^1(\mathbb{R}^3, \mathbb{R}^3)$. Then*

$$\iiint_U \vec{\nabla} \cdot E \, \mathrm{d}\lambda^3 = \oiint_M \langle E | \mathrm{d}\sigma \rangle$$

Proof. Without proof. □

Corollary 9.41 (Green's Identities)**.** *Let M be a closed surface, let $f, g \in C^2(U, \mathbb{R})$, and n an orientation (continuous normal vector field). Then*

$$\iiint_U f\nabla^2 g + \left\langle \vec{\nabla} f \middle| \vec{\nabla} g \right\rangle \mathrm{d}\lambda^3 = \oiint_M \left\langle f\vec{\nabla} g \middle| \mathrm{d}\sigma \right\rangle$$

$$\begin{aligned} \iiint_U f\vec{\nabla} g - g\vec{\nabla} f \, \mathrm{d}\lambda^3 &= \oiint_M \left\langle f\vec{\nabla} g - g\vec{\nabla} f \middle| \mathrm{d}\sigma \right\rangle \\ &= \oiint_M (f\partial_n g - g\partial_n f) \, \mathrm{d}\sigma \end{aligned}$$

Proof. Apply the divergence theorem to $f\vec{\nabla} g$:

$$\vec{\nabla} \cdot \left(f\vec{\nabla} g \right) = \left\langle \vec{\nabla} f \middle| \vec{\nabla} g \right\rangle + f\nabla^2 g \tag{9.18}$$

Swapping and subtracting f and g yields the second equation. Let $\phi : V \to M$ be a parametrization. Then

$$\begin{aligned} \oiint_\phi \left\langle f\vec{\nabla} g \middle| \mathrm{d}\sigma \right\rangle &= \iint_V \left\langle f(\phi(t))\vec{\nabla} g(\phi(t)) \middle| \sigma_\phi(t) \right\rangle \mathrm{d}\lambda^2(t) \\ &= \iint_V f(\phi(t)) \underbrace{\left\langle \vec{\nabla} g(\phi(t)) \middle| n(\phi(t)) \right\rangle}_{\partial_n g(\phi(t))} \|\sigma_\phi(t)\| \qquad (9.19) \\ &= \oiint_M f\partial_n g \, \mathrm{d}\sigma \end{aligned}$$

□

Example 9.42. Let $U \subset \mathbb{R}^3$ be bounded with a given volume V, and a "nice" boundary M with area A. Set

$$R = \sup \{ \|(x, y, z)\| \mid (x, y, z) \in M \}$$

Let $E(x, y, z) = (x, y, z)$ and $\phi : W \to M$ a parametrization. Then

$$\begin{aligned} 3V = \iiint_U E \, \mathrm{d}\lambda^3 &= \oiint_M \langle E | \mathrm{d}\sigma \rangle \\ &= \iint_W \langle E(\phi(t)) | \sigma_\phi(t) \rangle \, \mathrm{d}\lambda^2(t) \\ &\le \iint_W |\langle \cdots \rangle| \, \mathrm{d}\lambda^2(t) \\ &\le \iint_W \underbrace{\|E(\phi(t))\|}_{\le R} \|\sigma_\phi(t)\| \, \mathrm{d}\lambda^2(t) \le R \cdot A \end{aligned}$$

For the sphere with radius R we have equality.

Chapter 10

Complex Analysis

10.1 Complex Differentiability

Definition 10.1. Let $f : U \to \mathbb{C}$, with $U \subset \mathbb{C}$ open. f is said to be complex differentiable in $z_0 \in U$ if

$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0} =: f'(z_0)$$

exists. If f is complex differentiable on all of U, f is said to be holomorphic. A funciton that is holomorphic on all of $\mathbb{C}$ is entire.

An equivalent formulation wiould be

$$\forall \epsilon > 0 \; \exists \delta > 0 : \quad |z - z_0| < \delta \implies |f(z) - f(z_0) - a(z - z_0)| < \epsilon$$

In this case $a = f'(z_0)$.

Theorem 10.2. *(i) f complex differentiable in $z_0 \in \mathbb{C} \implies f$ continuous in z_0*

(ii) f, g complex differentiable in z_0, then $f + g$ and $f \cdot g$ are complex differentiable in z_0, and

$$\begin{aligned} (f+g)'(z_0) &= f'(z_0) + g'(z_0) \\ (fg)'(z_0) &= f'(z_0)g(z_0) + f(z_0)g'(z_0) \end{aligned}$$

If $g(z_0) \neq 0$, then $\frac{f}{g}$ is complex differentiable and

$$\left(\frac{f}{g}\right)'(z_0) = \frac{g(z_0)f'(z_0) - g'(z_0)f(z_0)}{g(z_0)^2}$$

(iii) Let $f : U \to \mathbb{C}$, $U \subset \mathbb{C}$ open and $C \subset \mathbb{C}$ open with $f(U) \subset V$, and let $g : V \to \mathbb{C}$. Then $g \circ f : U \to \mathbb{C}$. If f is complex differentiable in z_0, and g is complex differentiable in $f(z_0)$, then $g \circ f$ is complex differentiable in z_0 with

$$(g \circ f)'(z_0) = g'(f(z_0))f'(z_0)$$

(iv) If f is complex differentiable in z_0, $f'(z_0) \neq 0$ and if $\exists \delta > 0$ such that $f : B_\delta(z_0) \to U \subset \mathbb{C}$ is bijective, then the inverse function g is complex differentiable in $f(z_0)$, with

$$g'(f(z_0)) = \frac{1}{f'(z_0)}$$

Proof. Left as an exercise for the reader. $\square$

Remark 10.3 (Complex vs. Real Differentiability). Consider $f : U \to \mathbb{C}$, $U \subset \mathbb{C}$ open. Let

$$x = \operatorname{Re} z \qquad\qquad y = \operatorname{Im} z$$

and define

$$\tilde{U} = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, x + iy \in U\right\}$$

and

$$\begin{aligned}\tilde{f} : \tilde{U} &\longrightarrow \mathbb{R}^2 \\ (x, y) &\longmapsto (\mathrm{Re}(f(x + iy)), \mathrm{Im}(f(x + iy))) =: (u(x, y), v(x, y))\end{aligned}$$

Then f is complex differentiable in $z = x + iy$.

(i) We have

$$\begin{aligned}f'(z) &= \lim_{h\to 0} \frac{f(z + h) - f(z)}{h} \\ &= \lim_{h\to 0} \frac{u(x + h, y) + iv(x, y + h) - u(x, y) - iv(x, y)}{h} \\ &= \lim_{h\to 0} \frac{u(x + h, y) - u(x, y)}{h} + i \lim_{h\to 0} \frac{v(x + h, y) - v(x, y)}{h} \\ &= \frac{\partial}{\partial x} u(x, y) + i \frac{\partial}{\partial x} v(x, y)\end{aligned}$$

(ii) And also

$$\begin{aligned}f'(z) &= \lim_{h\to 0} \frac{f(z + ih) - f(z)}{ih} \\ &= -i \lim_{h\to 0} \frac{u(x, y + h) + iv(x, y + h) - u(x, y) - iv(x, y)}{h} \\ &= -i \lim_{h\to 0} \frac{u(x, y + h) - u(x, y)}{h} + \lim_{h\to 0} \frac{v(x, y + h) - v(x, y)}{h} \\ &= -\frac{\partial}{\partial y} u(x, y) + \frac{\partial}{\partial y} (x, y)\end{aligned}$$

This results in the Cauchy-Riemann equations:

$$\begin{aligned}\frac{\partial}{\partial x} u(x, y) &= \frac{\partial}{\partial y} v(x, y) \\ \frac{\partial}{\partial y} u(x, y) &= -\frac{\partial}{\partial x} v(x, y)\end{aligned}$$

if f is complex differentiable in $z = x + iy$.

From the Cauchy-Riemann equations and the real differentiability of the function $\tilde{f} : \tilde{U} \to \mathbb{R}^2$ follows

$$\begin{aligned}D\tilde{f}(x, y) = \begin{pmatrix} \partial_x u(x, y) & \partial_y u(x, y) \\ \partial_x v(x, y) & \partial_y v(x, y) \end{pmatrix} &= \begin{pmatrix} \partial_x u(x, y) & -\partial_x v(x, y) \\ \partial_x v(x, y) & \partial_x u(x, y) \end{pmatrix} \\ &=: \begin{pmatrix} a & -b \\ b & a \end{pmatrix}\end{aligned}$$

and thus for $h = (h_1, h_2) \in \mathbb{R}^2$

$$\begin{aligned}\tilde{f}(x + h_1, y + h_2) - \tilde{f}(x, y) &= D\tilde{f}(x, y)h + \mathcal{O}(|h|) \\ &= \begin{pmatrix} ah_1 - bh_2 \\ bh_1 + ah_2 \end{pmatrix} + \mathcal{O}(|h|)\end{aligned}$$

A side calculation:

$$\begin{aligned}(a + ib)(h_1 + ih_2) &= ah_1 - bh_2 + i(bh_1 + ah_2) \\ \implies \begin{pmatrix} ah_1 - bh_2 \\ bh_1 + ah_2 \end{pmatrix} + \mathcal{O}(|h|) &= \begin{pmatrix} \mathrm{Re}(a + ib)(h1 + ih_2) \\ \mathrm{Im}(a + ib)(h_1 + ih_2) \end{pmatrix} + \mathcal{O}(|h|)\end{aligned}$$

So for $h = h_1 + ih_2$ we get

$$f(z + h) - f(z) = (a + ib)h + \mathcal{O}(|h|)$$

So f is complex differentiable in z with $f'(z) = a+ib$. In short, we have shown the following theorem.

Theorem 10.4. *Let $f : U \to \mathbb{C}$ with $U \subset \mathbb{C}$ open. f is complex differentiable in $z \in U$ if and only if $\tilde{f} : \tilde{U} \to \mathbb{R}^2$ is real differentiable in $(x, y) \in \tilde{U}$, and if the Cauchy-Riemann equations are satisfied.*

Proof. Proof is in the previous remark. □

Example 10.5. (i) Power series like

$$f(z) = \sum_{n=0}^{\infty} a_n z^n, \quad (a_n) \subset \mathbb{C}$$

with convergence radius $\rho \in [0, \infty]$ are holomorphic on $B_\rho(0)$. The following holds

$$f'(z) = \sum_{n=0}^{\infty} n a_n z^{n-1}$$

Especially, the funciton

$$f(z) = e^{\alpha z}, \quad \alpha \in \mathbb{C}$$

is holomorphic on all of $\mathbb{C}$ with

$$f'(z) = \alpha e^{\alpha z}$$

(ii) The function

$$f(z) = \frac{1}{z^n}$$

is holomorphic $\mathbb{C} \setminus \{0\}$ with

$$f'(z) = -n\frac{1}{z^{n+1}}$$

(iii) Functions that are not complex differentiable include

$$f(z) = \overline{z} \qquad\qquad f(z) = z\overline{z}$$

$$(\partial_x u = 1 \neq \partial_y v = -1) \qquad\qquad (\partial_x u = 2x^2 \neq \partial_y v = 0)$$

10.2 Contour Integrals

Definition 10.6 (Contour integrals)**.** Let $U \subset \mathbb{C}$ be open, $\gamma = C([a,b],U)$ a curve in U and $f : U \to \mathbb{C}$ continuous. Then

$$\int_\gamma f(z)\,\mathrm{d}z := \int_a^b f(\gamma(t))\gamma'(t)\,\mathrm{d}t$$

Example 10.7. Consider the path

$$\gamma(t) = re^{it}, \qquad t \in [0, 2\pi], r > 0$$

we want to take the contout integral along the path γ of the function z^n

$$\begin{aligned}\int_\gamma z^n\,\mathrm{d}z &= \int_0^{2\pi} (re^{it})^n ire^{it}\,\mathrm{d}t \\ &= ir^{n+1}\int_0^{2\pi} e^{it(n+1)}\,\mathrm{d}t = ir^{n+1}\begin{cases}2\pi, & n=-1\\ 0, & n\neq -1\end{cases}\end{aligned}$$

Lemma 10.8 (Estimation Lemma)**.** *For every curve $\gamma \in C([0,1],U)$ and every continuous function $f : U \to \mathbb{C}$ we have*

$$\left|\int_\gamma f(z)\,\mathrm{d}z\right| \le \sup_{z\in\gamma}|f(z)|\int_0^1 \left|\gamma'(t)\right|\mathrm{d}t$$

Proof.

$$\begin{aligned}\left|\int_\gamma f(z)\,\mathrm{d}z\right| = \left|\int_0^1 f(\gamma(t))\gamma'(t)\,\mathrm{d}t\right| &\le \int_0^1 |f(\gamma(t))|\left|\gamma'(t)\right|\mathrm{d}t \\ &\le \sup_{t\in[0,1]}|f(\gamma(t))|\int_0^1\left|\gamma'(t)\right|\mathrm{d}t\end{aligned} \tag{10.1}$$

□

Corollary 10.9. *Let $\gamma \in C([0,1],U)$ be a simple closed curve, $U \subset \mathbb{C}$, and let $f : U \to \mathbb{C}$ a holomorphic function with*

$$u = \operatorname{Re} f \qquad\qquad v = \operatorname{Im} f$$

Then

$$\oint_\gamma f(z)\,\mathrm{d}z = 0$$

Proof. Let $A \subset U$ be the surface bounded by γ. Then

$$\oint_\gamma f(z)\,\mathrm{d}z = \int_0^1 f(\gamma(t))\gamma'(t)\,\mathrm{d}t \tag{10.2}$$

We can split γ into a real and an imaginary part, like this

$$\gamma(t) = \gamma_1(t) + i\gamma_2(t), \quad \gamma_1, \gamma_2 : [0,1] \to \mathbb{R} \tag{10.3}$$

Then we can calculate

$$\begin{aligned}
\oint_\gamma f(z)\,\mathrm{d}z &= \int_0^1 \big(u(\gamma_1(t),\gamma_2(t)) + iv(\gamma_1(t),\gamma_2(t))(\gamma_1'(t) + i\gamma_2'(t))\big)\,\mathrm{d}t \\
&= \int_0^1 u(\gamma_1(t),\gamma_2(t))\gamma_1'(t) - v(\gamma_1(t),\gamma_2(t))\gamma_2'(t)\,\mathrm{d}t \\
&\quad + i\int_0^1 u(\gamma_1(t),\gamma_2(t))\gamma_2'(t) + v(\gamma_1(t),\gamma_2(t))\gamma_1'(t)\,\mathrm{d}t \\
&= \int_0^1 \begin{pmatrix} u(\gamma(t)) \\ -v(\gamma(t)) \end{pmatrix} \begin{pmatrix} \gamma_1'(t) \\ \gamma_2'(t) \end{pmatrix} \mathrm{d}t + i\int_0^1 \begin{pmatrix} v(\gamma(t)) \\ u(\gamma(t)) \end{pmatrix} \begin{pmatrix} \gamma_1'(t) \\ \gamma_2'(t) \end{pmatrix} \mathrm{d}t \\
&= \oint_{\partial A} \begin{pmatrix} u \\ -v \end{pmatrix} \mathrm{d}s + i\oint_{\partial A} \begin{pmatrix} v \\ u \end{pmatrix} \\
&= \int_A (-\partial_x v - \partial_y u)\,\mathrm{d}\lambda^2 + i\int_A (\partial_x u - \partial_y v)\,\mathrm{d}\lambda^2
\end{aligned} \tag{10.4}$$

Because f is holomorphic we can apply the Cauchy-Riemann equation

$$\oint_\gamma f(z)\,\mathrm{d}z = 0 \tag{10.5}$$

$\square$

Definition 10.10. (i) A closed curve $\gamma : [a,b] \to U$ with $U \subset \mathbb{C}$ is said to be null-homotopic, if it can be continuously deformed into a point within the set U.

(ii) Two curves $\gamma_1, \gamma_2 : [0,1] \to U$ with identical boundary points

$$\gamma_1(0) = \gamma_2(0) \;\wedge\; \gamma_1(1) = \gamma_2(1)$$

is said to be homotopic in U if the concatenation

$$\begin{aligned}
\gamma : [0,2] &\longrightarrow U \\
\gamma(t) &= \begin{cases} \gamma_1(t), & t \in [0,1] \\ \gamma_2(2-t) & t \in [1,2] \end{cases}
\end{aligned}$$

is null-homotopic.

(iii) Two closed surves γ_0, γ_1 are said to be free-homotopic in U if they can be continuously transformed into each other.

Definition 10.11. A non-empty set $U \subset \mathbb{C}$ is said to be

(i) *connected* if any two points in U can be connected by a curve in U.

(ii) *simply connected* if U is connected and every closed surve in U is null-homotopic.

(iii) a *domain* if it is open and connected.

Theorem 10.12 (Cauchy's Integral Theorem)**.** *Let $f : U \to \mathbb{C}$ be holomorphic and γ a closed, null-homotopic curve in $U \subset \mathbb{C}$ open. Then*

$$\oint_\gamma f(z)\,\mathrm{d}z = 0$$

Proof. Without proof. □

Corollary 10.13. *(i) Let γ_1, γ_2 be holomorphic curves with the same endpoints on the open set $U \subset \mathbb{C}$. Then*

$$\int_{\gamma_1} f(z)\,\mathrm{d}z = \int_{\gamma_2} f(z)\,\mathrm{d}z$$

for all holomorphic $f : U \to \mathbb{C}$.

(ii) For $f : U \to \mathbb{C}$ holomorphic, with $U \subset \mathbb{C}$ open and simple connected. Then $\forall z_0 \in U$

$$F(z) := \int_{z_0}^{z} f(\zeta)\,\mathrm{d}\zeta = \int_{\gamma=\gamma_0} f(\zeta)\,\mathrm{d}\zeta$$

is a holomorphic anti-derivative of f, i.e.

$$F'(z) = f(z) \quad \forall z \in U$$

Proof. First we prove (i). The concatenation $\gamma := \gamma_1\gamma_2$ is a null-homotopic curve, so together with the holomorphy of f we can apply the Cauchy integral theorem

$$\begin{aligned} 0 = \oint_\gamma f(z)\,\mathrm{d}z &= \int_0^2 f(\gamma(t))\dot{\gamma}(t)\,\mathrm{d}t \\ &= \int_0^1 f(\gamma_1(t))\dot{\gamma_1}(t)\,\mathrm{d}t - \int_1^2 f(\gamma_2(2-t))\dot{\gamma_2}(2-t)\,\mathrm{d}t \end{aligned} \tag{10.6}$$

Substitute $s = 2 - t$ with $\mathrm{d}s = -\,\mathrm{d}t$:

$$\begin{aligned} &= \int_0^1 f(\gamma_1(t))\dot{\gamma}(t)\,\mathrm{d}t - \gamma_0^1 f(\gamma_2(s))\dot{\gamma_2}(s)\,\mathrm{d}s \\ &= \int_{\gamma_1} f(z)\,\mathrm{d}z - \int_{\gamma_2} f(z)\,\mathrm{d}z \end{aligned} \tag{10.7}$$

Now we prove (ii). According to (i), we have

$$F(z+h) = F(z) + \int_{\gamma_{z+h,z}} f(z)\,\mathrm{d}z \tag{10.8}$$

We choose $\gamma_{z+h,z}$ to be a straight line, i.e.

$$\gamma_{z+h,z}(t) = t(z+h) + (1-t)z, \quad t \in [0,1] \tag{10.9}$$

Then

$$\int_{\gamma_{z+h,z}} 1\,\mathrm{d}\zeta = \int_0^1 \dot{\gamma}(t)\,\mathrm{d}t = h \tag{10.10}$$

Thus follows

$$F(z+h) - F(z) = \int_{\gamma_{z+h,z}} f(\zeta)\,\mathrm{d}\zeta \iff \frac{F(z+h)-F(z)}{h} = \frac{1}{h} \in f(\zeta)\,\mathrm{d}\zeta \tag{10.11}$$

and therefore

$$\begin{aligned}
\left|\frac{F(z+h)-F(z)}{h} - f(z)\right| &= \left|\frac{1}{h}\int_{\gamma_{z+h,z}} f(\zeta)\,\mathrm{d}\zeta - f(z)\right| \\
&= \left|\frac{1}{h}\int_{\gamma_{z+h,z}} f(\zeta) - f(z)\,\mathrm{d}\zeta\right| \\
&= \frac{1}{|h|}\int_0^1 |f(\gamma_{z+h,z}(t)) - f(z)|\,|\dot{\gamma_{z+h,z}}(t)|\,\mathrm{d}t \\
&\leq \frac{1}{h}\sup_{t\in[0,1]} |f(\gamma_{z+h,z}(t)) - f(z)| \cdot \underbrace{\int |\dot{\gamma_{z+h,z}}(t)|\,\mathrm{d}t}_{|h|} \\
&= \sup_{t\in[0,1]} |f(\gamma_{z+h,z}(t)) - f(z)| \\
&\xrightarrow{k\to 0} 0
\end{aligned} \tag{10.12}$$

□

Example 10.14 (The complex logarithm). Consider $t \mapsto e^{it}$, $t \in \mathbb{R}$. This is a 2π-periodic function, that means

$$e^{it} = e^{i(t+2\pi n)}, \quad n \in \mathbb{Z}$$

The function

$$\begin{aligned}
f : \mathbb{C}\setminus\{0\} &\longrightarrow \mathbb{C} \\
z &\longmapsto \frac{1}{z}
\end{aligned}$$

is holomorphic, and does not have an anti-derivative on $\mathbb{C} \setminus \{0\}$. If it did, then

$$\int_\gamma f(z)\,\mathrm{d}z = F(\gamma(2\pi)) - F(\gamma(0)) = 0$$

would have to hold, but we know that

$$\int_\gamma \frac{\mathrm{d}z}{z} = 2\pi i$$

This is a contradiction. However f does have an anti-derivative on $\mathbb{C}_-$ (the complex numbers without the negative real axis) , since $\mathbb{C}_-$ is simple connected and f is holomorphic. Thus we can define

$$\begin{aligned} \operatorname{Log} : \mathbb{C}_- &\longrightarrow \mathbb{C} \\ z &\longmapsto \int_{\gamma:[0,1]\to z} \frac{\mathrm{d}\zeta}{\zeta} \end{aligned}$$

It can also be defined as

$$\operatorname{Log} z = \begin{cases} 0, & 1 \\ \log|z| + i\arg(z), & \text{else} \end{cases}$$

The function arg is defined as

$$\begin{aligned} \arg : \mathbb{C}_- &\longrightarrow (-\pi, \pi) \\ z &\longmapsto \phi \text{ for } z = |z|e^{i\phi} \end{aligned}$$

Log is said to be the main branch of the complex logarithm, and

$$\operatorname{Log} z = \log|z| + i(\arg(z) + 2\pi n), \quad n \in \mathbb{Z}$$

the secondary branches.

Example 10.15 (Fresnel Integrals). Consider the integrals

$$\int_0^\infty \cos\left(t^2\right)\mathrm{d}t \qquad\qquad \int_0^\infty \sin\left(t^2\right)\mathrm{d}t$$

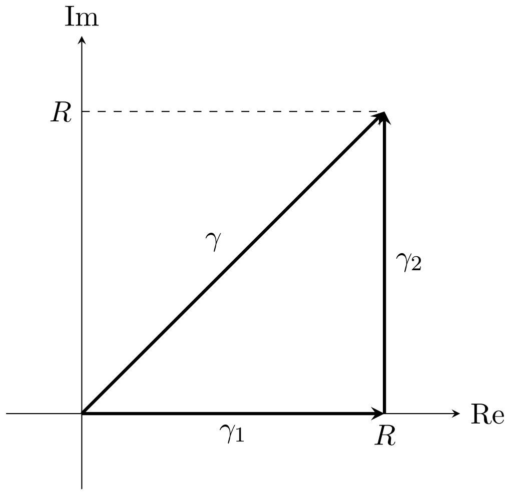
The way these integrals are supposed to be interpreted is as

$$\int_0^\infty f(t)\,\mathrm{d}t = \lim_{N\to\infty}\int_0^N f(t)\,\mathrm{d}t$$

We realize that

$$\cos\left(t^2\right) = \operatorname{Re} e^{-it^2} \qquad\qquad \sin\left(t^2\right) = -\operatorname{Im} e^{-it^2}$$

Now, consider these paths



So it becomes apparent that

$$\int_0^R \cos\left(t^2\right) \mathrm{d}t = \operatorname{Re} \int_0^R e^{-it^2} \mathrm{d}t = \operatorname{Re} \int_\gamma e^{-z^2} \mathrm{d}z$$

We can define a new (closed) path

$$\Gamma = \gamma_1 \gamma_2 (-\gamma)$$

and with Cauchy's theorem we can realize that

$$0 = \oint_\Gamma e^{-z^2} \mathrm{d}z = \int_{\gamma_1} e^{-z^2} \mathrm{d}z + \int_{\gamma_2} e^{-z^2} \mathrm{d}z - \int_\gamma e^{-z^2} \mathrm{d}z$$

The next step is to evaluate each of the integrals in the last term, starting with the integral over γ.

$$\begin{aligned} \int_\gamma e^{-z^2} \mathrm{d}z = & \int_0^R e^{-((1+i)t)^2} (1+i) \, \mathrm{d}t \\ =& (1+i) \int_0^R e^{-2it^2} \mathrm{d}t \\ =& \frac{1+i}{\sqrt{2}} \int_0^{\sqrt{2}R} e^{-is^2} \mathrm{d}s \end{aligned}$$

The integrall over γ_1 evaluates to

$$\int_{\gamma_1} e^{-z^2} \mathrm{d}z = \int_0^R e^{-t^2} \mathrm{d}t \xrightarrow{R\to\infty} \int_0^\infty e^{-t^2} \mathrm{d}t = \frac{1}{2} \int_{-\infty}^\infty e^{-t^2} \mathrm{d}t = \frac{\sqrt{\pi}}{2}$$

And the one over γ_2 to

$$\int_{\gamma_2} e^{-z^2}\,\mathrm{d}z = \int_0^R e^{-(r+it)^2} i\,\mathrm{d}t = i\int_0^R e^{-R^2+t^2} e^{-2irt}\,\mathrm{d}t$$

To evaluate this we need to consider the absolute value of this integral

$$\begin{aligned}\left|\int_{\gamma_2} e^{-z^2}\,\mathrm{d}z\right| \le & e^{-R^2}\int_0^R e^{t^2} \underbrace{\left|e^{-2iRt}\right|}_{=1}\mathrm{d}t \\ =& e^{-R^2}\int_0^R e^{t^2}\,\mathrm{d}t \le e^{-R^2}\int_0^R e^{tR}\,\mathrm{d}t \\ =& e^{-R^2}\left[\frac{1}{R}e^{tR}\right]_0^R = \frac{e^{-R^2}}{R}\left(e^{R^2}-1\right)\end{aligned}$$

so

$$\left|\int_{\gamma_2} e^{-z^2}\,\mathrm{d}z\right| \le \frac{1}{R}\left(1-e^{-R^2}\right) \xrightarrow{R\to\infty} 0$$

Thus we can calculate

$$\int_{\gamma} e^{-z^2}\,\mathrm{d}z = \frac{1+i}{\sqrt{2}}\int_0^{\sqrt{2}R} e^{-it^2}\,\mathrm{d}t = \int_{\gamma_1} e^{-z^2}\,\mathrm{d}z + \int_{\gamma_2} e^{-z^2}\,\mathrm{d}z$$

And finally

$$\begin{aligned}\lim_{R\to\infty}\int_0^\infty e^{-it^2}\,\mathrm{d}t = & \lim_{R\to\infty}\int_0^{\sqrt{2}R} e^{-t^2}\,\mathrm{d}t \\ =& \frac{\sqrt{2}}{1+i}\left(\lim_{R\to\infty}\left(\int_{\gamma_1} e^{-z^2}\,\mathrm{d}z + \int_{\gamma_2} e^{-z^2}\,\mathrm{d}z\right)\right) \\ =& \frac{\sqrt{2}}{1+i}\left(\frac{\pi}{2}+0\right) \\ =& \sqrt{\frac{\pi}{2}}\frac{1-i}{2} = \sqrt{\frac{\pi}{8}}(1-i)\end{aligned}$$

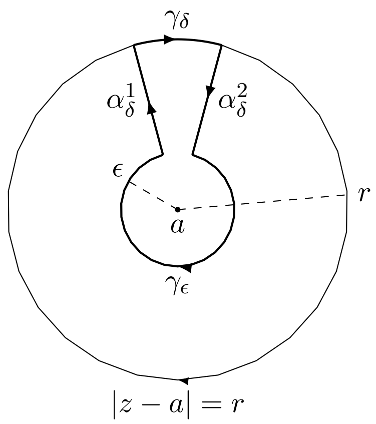
So we can calculate the Fresnel integrals

$$\int_0^\infty \cos\left(t^2\right)\mathrm{d}t = \sqrt{\frac{\pi}{8}}$$
$$\int_0^\infty \sin\left(t^2\right)\mathrm{d}t = \sqrt{\frac{\pi}{8}}$$

Theorem 10.16 (Cauchy's Theorem for circular disks)**.** *Let $f : U \to \mathbb{C}$ be holomorphic, $U \subset \mathbb{C}$ open and $K_r(a) \subset U$. Then*

$$f(a) = \frac{1}{2\pi i}\int_{|z-a|=r} \frac{f(z)}{z-a}\,\mathrm{d}z$$

Proof. Consider the following path



According to the first corollary of Cauchy's theorem we have

$$\begin{aligned}\int_{|z-a|=r} \frac{f(z)}{z-a}\,\mathrm{d}z &= \lim_{\delta\to 0} \int_{\gamma_{\epsilon,\delta}} \frac{f(z)}{z-a}\,\mathrm{d}z + \underbrace{\int_{\alpha_\delta^1} \frac{f(z)}{z-a}\,\mathrm{d}z + \int_{\alpha_\delta^2} \frac{f(z)}{z-a}\,\mathrm{d}z}_{\xrightarrow{\delta\to 0} 0} \\ &= \int_{\gamma_\epsilon} \frac{f(z)}{z-a}\,\mathrm{d}z\end{aligned} \tag{10.13}$$

Thus we conclude

$$\begin{aligned}\int_{|z-a|=r} \frac{f(z)}{z-a}\,\mathrm{d}z = \int_{\gamma_\epsilon} \frac{f(z)}{z-a}\,\mathrm{d}z &= \int_{\gamma_\epsilon} \frac{f(z)-f(a)}{z-a}\,\mathrm{d}z + \int_{\gamma_\epsilon} \frac{f(a)}{z-a}\,\mathrm{d}z \\ &= \int_{\gamma_\epsilon} \frac{f(z)-f(a)}{z-a}\,\mathrm{d}z + f(a)\int_{\gamma_\epsilon} \frac{\mathrm{d}z}{z-a}\end{aligned} \tag{10.14}$$

We also know that

$$\int_{\gamma_\epsilon} \frac{\mathrm{d}z}{z-a} = 2\pi i \tag{10.15}$$

Since f is holomorphic we can realize

$$\sup_{K_r(a)} \left| \frac{f(z)-f(a)}{z-a} \right| = M_r < \infty \tag{10.16}$$

Which results in

$$\left| \int_{\gamma_\epsilon} \frac{f(z)-f(a)}{z-a}\,\mathrm{d}z \right| \leq M_r \underbrace{\int_0^{2\pi} |\dot{\gamma_\epsilon}(t)|\,\mathrm{d}t}_{2\pi\epsilon} \xrightarrow{\epsilon\to 0} 0 \tag{10.17}$$

Thus follows

$$\int_{|z-a|=r} \frac{f(z)}{z-a}\,\mathrm{d}z = \int_{\gamma_\epsilon} \frac{f(z)-f(a)}{z-a}\,\mathrm{d}z + 2\pi i f(a) \xrightarrow{\epsilon\to 0} 2\pi i f(a) \tag{10.18}$$

Or short

$$\int_{|z-a|=r} \frac{f(z)}{z-a}\,\mathrm{d}z = 2\pi i f(a) \tag{10.19}$$

□

Corollary 10.17. *Let $f : U \to \mathbb{C}$ be a holomorphic function and $U \subset \mathbb{C}$ an open set such that $K_r(a) \subset U$. Then*

$$f(a) = \frac{1}{2\pi}\int_0^{2\pi} f(a + re^{it})\,\mathrm{d}t$$

Proof. Left as an exercise for the reader. □

Definition 10.18 (Analytic functions)**.** Let $f : U \to \mathbb{C}$ be a function and $U \subset \mathbb{C}$ a domain. f is said to be analytic in $z_0 \in U$ if and only if there exists a power series

$$\sum_{n=0}^{\infty} a_n \zeta^n$$

with convergence radius

$$\rho = \left(\limsup |a_n|^{\frac{1}{n}}\right)^{-1} > 0$$

and $\delta \in (0, \rho)$ such that $B_\delta(z_0) \subset U$ and

$$f(z) = \int_{k=0}^{\infty} a_n (z - z_0)^n, \quad \forall z \in B_\delta(z_0)$$

f is said to be analytic on U if f is analytic $\forall z_0 \in U$.

Theorem 10.19 (Power series expansion)**.** *If f is holomorphic on a circular disk $B_r(z_0)$ for some $r > 0$, then f is analytic in z_0. f can be represented with the on $B_\rho(z_0)$ convergent power series*

$$f(z) = \sum_{n=0}^{\infty} c_n (z - z_0)^n, \quad z \in B_\rho(z_0)$$

with

$$c_n = \frac{1}{2\pi i}\int_{|z-z_0|=r} \frac{f(z)}{(z-z_0)^{n+1}}\,\mathrm{d}z\,, \quad \forall \rho \in (0, r)$$

Proof. Without proof. □

Remark 10.20. If f is holomorphic then f can be infinitely often differentiated on $\mathbb{C}$ with

$$f^{(n)}(z) = n! c_n = \frac{n!}{2\pi i}\int_{|z-z_0|=\rho} \frac{f(z)}{(z-z_0)^{n+1}}\,\mathrm{d}z$$

By employing the estimation lemma we can then find that

$$\begin{aligned}|c_n| \leq \frac{1}{2\pi}\left|\int_{|z-z_0|=\rho} \frac{f(z)}{(z-z_0)^{n+1}}\,\mathrm{d}z\right| \leq & \frac{1}{2\pi} \sup_{|z-z_0|=\rho} \left|\frac{f(z)}{|z-z_0|^{n+1}}\right| \cdot 2\pi\rho \\ =& \frac{1}{\rho^n} \sup_{z\in B_r(z_0)} |f(z)| \\ =& \frac{M_r}{\rho^n}, \quad M_r < \infty\end{aligned}$$

This is Cauchy's estimate.

Theorem 10.21 (Liouville's Theorem)**.** *Every bounded entire function is constant.*

Proof. According to the power series expansion theorem, f can be represented by a power series on all of $\mathbb{C}$:

$$f(z) = \sum_{n=0}^{\infty} c_n z^n \tag{10.20}$$

and the coefficients satisfy the Cauchy estimate

$$|c_n| \leq \frac{1}{\rho^n} \sup_{|z|=\rho} |f(z)| \leq \frac{1}{\rho^n} \underbrace{\sup_{z\in\mathbb{C}} |f(z)|}_{<\infty} \tag{10.21}$$

This inequality tends to 0 if ρ tends to ∞ for all $n \geq 1$, thus we can find

$$c_n = 0, \quad \forall n \geq 1 \tag{10.22}$$

Thus

$$f(z) = c_0 = \text{const}\,. \tag{10.23}$$

□

Theorem 10.22 (Fundamental Theorem of Algebra)**.** *Every polynomial of degree* $n \geq 1$

$$f(z) = \sum_{k=0}^{n} c_k z^k, \quad c_n \neq 0$$

has a root, i.e.

$$\exists z_0 \in \mathbb{C} : \quad f(z_0) = 0$$

Proof. Assume there exists no root. Then the function

$$z \longmapsto \frac{1}{f(z)} \tag{10.24}$$

would be holomorphic on all of $\mathbb{C}$, since $z \mapsto \frac{1}{z}$ is holomorphic on $\mathbb{C} \setminus \{0\}$. Furthermore we find that

$$\exists R \geq 0: \quad |z| \geq R \implies |f(z)| \geq |f(0)| > 0 \tag{10.25}$$

which implies

$$\sup_{z\in\mathbb{C}} \frac{1}{|f(z)|} = \sup_{|z|<R} \frac{1}{|f(z)|} = \max_{|z|\leq R} \frac{1}{|f(z)|} < \infty \tag{10.26}$$

since f doesn't have a root. According to Liouville's theorem $\frac{1}{f}$ has to be constant, and thus f must be constant. This implies that $c_n = 0$, which contradicts the assumption. So f has to have a root. □

Corollary 10.23 (Polynomial Decomposition)**.** *Let*

$$f(z) = \sum_{k=0}^{n} c_k z^k, \quad n \in \mathbb{N}, c_k \in \mathbb{C}, c_n = 1$$

Then $\exists z_j \in \mathbb{C}, \quad j = 1, \cdots, n$ *such that*

$$f(z) = \prod_{j=1}^{n} (z - z_j)$$

10.3 Identity Theorem & Analytic Continuation

Definition 10.24. Let $f : U \to \mathbb{C}$ be a function on $U \subset \mathbb{C}$ and $n \in \mathbb{N}$. f has a root with multiplicity n at z_0, if

$$\begin{aligned} f^{(k)}(z_0) &= 0, \quad \forall k = 0, \cdots, n-1 \\ f^{(n)}(z_0) &= 0 \end{aligned}$$

If f is holomorphic it can be written as

$$f(z) = \sum_{k=n}^{\infty} c_n (z - z_0)^k$$

Theorem 10.25 (Identity Theorem)**.** *Let* $U \subset \mathbb{C}$ *be a domain and* $f : U \to \mathbb{C}$ *analytic. If*

$$\{z \in \mathbb{C} \,|\, f(z) = 0\}$$

has an accumulation point, i.e.

$$f(z_n) = 0, \quad (z_n)_{n\in\mathbb{N}} \subset U, \ (z_n) \xrightarrow{n\to\infty} z_\infty \in U$$

then $f = 0$ *on* U*.*

Proof. Since f is analytic in $z_0 \in U$, $\exists \delta > 0$ such that

$$f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k, \quad \forall z \in B_\delta(z_0) \tag{10.27}$$

Because $z_0 \in U$ is a root of f we can find that $a_0 = 0$. If $a_k \neq 0$ for some $k \geq 1$ then we can consider

$$m = \min \{k \geq 1 \,|\, a_k \neq 0\} \tag{10.28}$$

Define

$$g(z) = \sum_{n=0}^{\infty} a_{n+m} (z - z_0)^n \tag{10.29}$$

Then $g(z_0) \neq 0$ and

$$f(z) = (z - z_0)^m g(z) \tag{10.30}$$

This function g is analytic in $B_\delta(z_0)$, and thus continuous. This means $\exists \delta' < \delta$ such that g doesn't vanish on $B_{\delta'}(z_0)$. We can conclude that f doesn't vanish on $B_{\delta'}(z_0) \setminus \{z_0\}$ either. If $a_k = 0 \;\; \forall k \in \mathbb{N}$, then $f = 0$ on $B_{\delta(z_0)}$.

Now define the set

$$A = \left\{ z \in U \,\middle|\, f^{(k)}(z) = 0, \quad \forall k \in \mathbb{N}_0 \right\} \tag{10.31}$$

Since $f^{(n)}$ is continuous for all $n \in \mathbb{N}_0$, we find

$$\begin{aligned} A &= \bigcap_{n \in \mathbb{N}_0} \left\{ z \in U \,\middle|\, f^{(n)}(z) = 0 \right\} \\ &= \underbrace{\bigcap_{n \in \mathbb{N}_0} \underbrace{\underbrace{\left(f^{(n)}\right)^{-1}}_{\text{continuous}} (\underbrace{\{0\}}_{\text{closed}})}_{\text{closed}}}_{\text{closed}} \end{aligned} \tag{10.32}$$

But A is also open. To prove this we consider a point $z_1 \in A$. Then the Taylor series of f in z_1 is identical to the zero-function. But then $f = 0$ on a neighbourhood V of z_1. However, since $f^{(n)}(z) = 0 \;\; \forall n \in \mathbb{N}_0$ and $z \in V$, we can use our previous results to conclude that $V \subset A$, making A a closed set.

U can now be represented in terms of A:

$$U = A \cup (U \setminus A) \tag{10.33}$$

This is the disjoint union of two open sets. Since U is a domain (and thus connected) this can only be the case if

$$A = \{U, \varnothing\} \tag{10.34}$$

Since $z_0 = 0$ we can conclude $A = U$. $\square$

Definition 10.26. If $V \subset U \subset \mathbb{C}$, and there exist two holomorphic functions

$$\begin{aligned} f &: V \longrightarrow \mathbb{C} \\ \tilde{f} &: U \longrightarrow \mathbb{C} \end{aligned}$$

with the property

$$f(z) = \tilde{f}(z), \quad \forall z \in V$$

then $\tilde{f}$ is said to be the analytic continuation of f on U.

Remark 10.27. If the set V has an accumulation point and if U is a domain, then the analytic continuation $\tilde{f}$ of f on U is unique (This follows from the identity theorem).

Example 10.28.

(i) $f(z) = \sum_{n=0}^{\infty} z^n$ is holomorphic on $\{z \in \mathbb{C} \mid |z| < 1\}$. The function

$$\tilde{f}(z) = \frac{1}{1-z}$$

is an analytic continuation of f on $\mathbb{C} \setminus \{1\}$.

(ii) We can also find the analytic continuation along a chain of circular disks: for $j \in \mathbb{N}$ define the power series

$$f_j(z) := \sum_{n=0}^{\infty} a_n(j)(z - z_j)^n$$

around $z_j \in \mathbb{C}$ with convergence radius $\rho_j \in (0, \infty]$. If the disks overlap and the functions are compatible, i.e.

$$f_j(z) = f_k(z), \quad \forall z \in B_{\rho_j}(z_j) \cap B_{\rho_k}(z_k)$$

then there is a unique holomorphic continuation on

$$\bigcup_{j \in \mathbb{N}} B_{\rho_j}(z_j)$$

Definition 10.29 (Analytic continuation along curves)**.** Let $\gamma : [t_0, t_1] \to \mathbb{C}$ be a continuous curve and

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$$

a converging power series around $z_0 = \gamma(t_0)$. Then the family of functions

$$f_t(z) := \sum_{n=0}^{\infty} a_n(t)(z - \gamma(t))^n, \quad t \in [t_0, t_1]$$

is an analytic continuation of f along γ if

- $f_{t_0} = f$
- $\forall t \in [t_0, t_1]$ exists a $\epsilon > 0$ such that for all $|\tau| < \epsilon$ the functions f_t and $f_{\min\{t,\tau,t_1\}}$ are compatible.

Example 10.30 (Complex Logarithm). The family

$$L_t(z) := it + \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \left(e^{it}z - 1\right)^n, \quad t \in [0, \infty)$$

is an analytic continuation of the main branch of the complex logarithm $L_0(z) = \mathrm{Log}(z)$ along the unit circle. This yields the secondary branches of the complex logarithm:

$$L_{2\pi n}(z) = 2\pi i n + \mathrm{Log}(z)$$

10.4 Laurent Series

Definition 10.31 (Classification of isolated singularities)**.** Let $f : U \to \mathbb{C}$ and $U \subset \mathbb{C}$ open. Then $z_0 \in \mathbb{C} \setminus \{U\}$ is said to be an isolated singularity if there exists an $\epsilon > 0$ such that $B_\epsilon(z_0) \setminus \{z_0\} \subset U$.

An isolated singularity z_0 is said to be

(i) *removable* if f can be analytically continued on $U \cup \{z_0\}$

(ii) a *pole* if $\exists m \geq 1$ such that
$$(z - z_0)^m f(z)$$
has a removable singularity in z_0. The smallest such m is the order of the pole.

(iii) *essential* if it is neither removable nor a pole of finite degree.

Example 10.32. (i) The function $f(z) = \frac{\sin z}{z}$ is holomorphic on $\mathbb{C} \setminus \{0\}$, and has a removable singularity in $z_0 = 0$. An analytic continuation of f on all of $\mathbb{C}$ is given by
$$z \longmapsto \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n+1)!}$$

(ii) Let $g : U \to \mathbb{C}$ be holomorphic with $g(z_0) \neq 0$ for $z_0 \in U$. The function
$$f(z) = \frac{g(z)}{(z - z_0)^m}$$
has a pole of m-th degree in z_0.

(iii) Consider the function

$$\begin{aligned} f : \mathbb{C} \setminus \{0\} &\longrightarrow \mathbb{C} \\ z &\longmapsto e^{\frac{1}{z}} \end{aligned}$$

f has an essential singularity in $z_0 = 0$. The power series representation of f is

$$f(z) = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{1}{z^n}$$

This doesn't remove the sungularity in z_0, and the pole is of infinite order

$$z^k |f(z)| \xrightarrow{z \to 0} \infty, \quad \forall k \in \mathbb{N}$$

Theorem 10.33 (Riemann's Theorem)**.** *An isolated singularity $z_0 \in U$ of a holomorphic function $f : U \setminus \{z_0\} \to \mathbb{C}$ is removable if and only if f is bounded in a punctured neighbourhood of z_0, i.e.*

$$\exists \epsilon > 0, c \geq 0 : \quad |f(z)| \leq c \quad \forall z \in \{\zeta \in \mathbb{C} \,|\, 0 < |\zeta - z_0| < \epsilon\}$$

Proof. If f can be analytically continued on $U \cup \{z_0\}$, then this continuation is continuous in z_0 and thus bounded in a neighbourhood of z_0. Inversely, if there exists some $c \geq 0$ and $\epsilon > 0$ such that

$$|f(z)| \leq c \quad \forall z \in \{\zeta \in \mathbb{C} \,|\, 0 < |\zeta - z_0| < \epsilon\} \tag{10.35}$$

Define the function

$$\begin{aligned} g : U &\longrightarrow \mathbb{C} \\ z &\longmapsto \begin{cases} (z - z_0)^2 f(z), & z \neq z_0 \\ 0, & z = z_0 \end{cases} \end{aligned} \tag{10.36}$$

Then

$$\lim_{z \to z_0} \frac{|g(z) - g(z_0)|}{|z - z_0|} = \lim_{z \to z_0} \frac{|z - z_0|^2 |f(z)|}{|z - z_0|} = \lim_{z \to z_0} (|z - z_0||f(z)|) = 0 \tag{10.37}$$

Thus g is holomorphic on U with $g(z_0) = g'(z_0) = 0$, meaning that

$$g(z) = \sum_{n=2}^{\infty} c_n (z - z_0)^n \tag{10.38}$$

with $c_n \in \mathbb{C}$. So the function

$$\tilde{f} : U \longrightarrow \mathbb{C} \tag{10.39}$$

$$z \longmapsto \sum_{n=2}^{\infty} c_n (z - z_0)^{n-2} = \sum_{n=0}^{\infty} c_{n+2} (z - z_n)^n \tag{10.40}$$

is a holomorphic continuation of f on $U \cup \{z_0\}$. □

Definition 10.34 (Laurent Series)**.** If we define the coefficients $c_n \in \mathbb{C}$ for $n \in \mathbb{Z}$, and $z, z_0 \in \mathbb{C}$, then the series

$$\sum_{n\in\mathbb{Z}} c_n(z-z_0)^n := \underbrace{\sum_{n=1}^{\infty} c_{-n}(z-z_0)^{-n}}_{\text{Analytic part}} + \underbrace{\sum_{n=0}^{\infty} c_n(z-z_0)^n}_{\text{Principal part}}$$

is said to be a Laurent series. It converges absolutely if the parts do so.

If $\frac{1}{r} \in [0,\infty]$ is the convergence radius of the principal part and $R \in [0,\infty]$ the convergence radius of the analytic branch, then the Laurent series converges on the annulus

$$K_{r,R}(z_0) := \{z \in \mathbb{C} \,|\, r < |z-z_0| < R\}$$

and is holomorphic.

Lemma 10.35. *If the series $f(z) := \sum_{n\in\mathbb{Z}} c_n(z-z_0)^n$ converges on $K_{r,R}(z_0)$, then for $\rho \in (r,R)$*

$$c_n = \frac{1}{2\pi i}\oint_{|z-z_0|=\rho} \frac{f(z)}{(z-z_0)^{n+1}}\,\mathrm{d}z\,, \quad n \in \mathbb{Z}$$

Proof. Due to the uniform convergence of the series on $K_{r,R}(z_0)$, we have

$$\begin{aligned}\oint_{|z-z_0|=\rho} \frac{f(z)}{(z-z_0)^{n+1}}\,\mathrm{d}z &= \sum_{k\in\mathbb{Z}} c_k \oint_{|z-z_0|=\rho} (z-z_0)^{k-n-1}\,\mathrm{d}z \\ &= \sum_{k\in\mathbb{Z}} c_k \cdot 2\pi i \delta_{k-n-1,-1} = 2\pi i \cdot c_n\end{aligned} \tag{10.41}$$

with $\delta_{i,j}$ the Kronecker delta, defined as

$$\delta_{i,j} := \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{10.42}$$

□

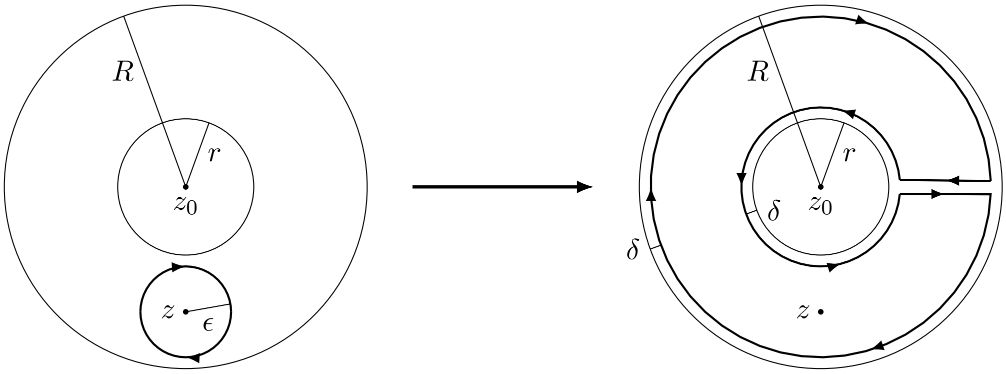
Theorem 10.36. *Let $f : K_{r,R}(z_0) \to \mathbb{C}$ be holomorphic, then*

$$f(z) = \sum_{n\in\mathbb{Z}} c_n(z-z_0)^n$$

with

$$c_n = \frac{1}{2\pi i}\oint_{|z-z_0|=\rho} \frac{f(z)}{(z-z_0)^{n+1}}\,\mathrm{d}z\,, \quad n \in \mathbb{Z},\ \rho \in (r,R)$$

Proof. W.l.o.g. we set $z_0 = 0$. Similar to the proof of Cauchy's theorem, we can prove Cauchy's theorem for annuli. To do that we define the following integration path



The two parallel path segments in the right figure are actually overlapping. They have been drawn next to each other for visual clarity. Now we can write

$$\begin{aligned} f(z) =& \frac{1}{2\pi i} \oint_{|\zeta - z| = \epsilon} \frac{f(\zeta)}{\zeta - z} \, \mathrm{d}\zeta \\ =& \frac{1}{2\pi i} \oint_{|\zeta| = R-\delta} \frac{f(\zeta)}{\zeta - z} \, \mathrm{d}\zeta - \frac{1}{2\pi i} \oint_{|\zeta| = r+\delta} \frac{f(\zeta)}{\zeta - z} \, \mathrm{d}\zeta \\ =& \frac{1}{2\pi i} \oint_{|\zeta| = R-\delta} \frac{f(\zeta)}{\zeta} \frac{1}{1 - \frac{z}{\zeta}} \, \mathrm{d}\zeta + \frac{1}{2\pi i} \frac{1}{z} \oint_{|\zeta| = r+\delta} f(\zeta) \frac{1}{1 - \frac{\zeta}{z}} \, \mathrm{d}\zeta \end{aligned} \tag{10.43}$$

We can now make use of the geometric series:

$$\frac{1}{1 - \frac{z}{\zeta}} = \sum_{n=0}^{\infty} \left(\frac{z}{\zeta} \right)^n, \quad |z| < |\zeta| \tag{10.44a}$$

$$\frac{1}{1 - \frac{\zeta}{z}} = \sum_{n=0}^{\infty} \left(\frac{\zeta}{z} \right)^n, \quad |\zeta| < |z| \tag{10.44b}$$

Thus we get

$$\begin{aligned} f(z) =& \frac{1}{2\pi i} \oint_{|\zeta| = R-\delta} \frac{f(\zeta)}{\zeta} \sum_{n=0}^{\infty} \frac{z^n}{\zeta^n} \, \mathrm{d}\zeta + \frac{1}{2\pi i} \frac{1}{z} \oint_{|\zeta| = r+\delta} f(\zeta) \sum_{n=0}^{\infty} \frac{\zeta^n}{z^n} \, \mathrm{d}\zeta \\ =& \sum_{n=0}^{\infty} z^n \left(\frac{1}{2\pi i} \oint_{|\zeta| = R-\delta} \frac{f(\zeta)}{\zeta^{n+1}} \, \mathrm{d}\zeta \right) + \sum_{n=0}^{\infty} \frac{1}{z^{n+1}} \left(\frac{1}{2\pi i} \oint_{|\zeta| = r+\delta} f(\zeta) \zeta^n \, \mathrm{d}\zeta \right) \\ =& \sum_{n=0}^{\infty} c_n z^n + \sum_{n=1}^{\infty} z^{-n} \underbrace{\left(\frac{1}{2\pi i} \oint_{|\zeta| = r+\delta} \frac{f(\zeta)}{\zeta^{-n+1}} \, \mathrm{d}\zeta \right)}_{= c_{-n}} \end{aligned} \tag{10.45}$$

□

Example 10.37. Consider

$$f(z) = \frac{1}{z(z-1)} = \frac{1}{z-1} - \frac{1}{z}$$

Using the geometric series we can then find for $K_{0,1}(0)$

$$f(z) = -\frac{1}{z} - \sum_{n=0}^{\infty} z^n$$

the Laurent series of f around $z_0 = 0$. For $K_{0,1}(1)$ we get

$$\begin{aligned} f(z) &= \frac{1}{z-1} - \frac{1}{z-1+1} = \frac{1}{z-1} - \frac{1}{1-(1-z)} \\ &= \underbrace{\frac{1}{z-1}}_{\text{Principal part}} \underbrace{-\sum_{n=0}^{\infty}(1-z)^n}_{\text{Analytic part}} \end{aligned}$$

Example 10.38.

$$f(z) = e^{\frac{1}{z}} = \sum_{n=0}^{\infty} \frac{1}{n!}\left(\frac{1}{z}\right)^n = 1 + \underbrace{\sum_{n=1}^{\infty} \frac{1}{n!}\frac{1}{z^n}}_{\text{Principal part}}$$

converges on $K_{0,\infty}(0)$.

Theorem 10.39. *If $f : U \setminus \{z_0\} \to \mathbb{C}$ has an essential singularity in $z_0 \in U$, then for every $\epsilon > 0$ the image $f(B_\epsilon(z_0) \setminus \{z_0\})$ is dense in $\mathbb{C}$, i.e.*

$$\forall \alpha \in \mathbb{C}\ \exists\, (z_n) \subset U \setminus \{z_0\} : \quad z_n \longrightarrow z_0 \implies f(z_n) \longrightarrow \alpha$$

Proof. Left as an exercise for the reader. □

Remark 10.40. We have essentially noticed three things:

(i)

$$\begin{aligned} & f \text{ has a removable singularity in } z_0 \\ \iff & f \text{ is bounded in a neighbourhood of } z_0 \\ \iff & \lim_{|z-z_0|\to 0} f(z) \text{ exists and is bounded} \end{aligned}$$

(ii)

$$\begin{aligned} & f \text{ has a pole of order } m \geq 1 \text{ in } z_0 \\ \iff & \lim_{|z-z_0|\to 0} |f(z)| = \infty \text{ and } \lim_{|z-z_0|\to 0} (z-z_0)^m f(z) < \infty \end{aligned}$$

(iii)

$$\begin{aligned} &f \text{ has an essential singularity in } z_0 \\ \Longleftrightarrow\ &\text{the set of accumulation points of } f(z) \text{ for } z \longrightarrow z_0 \text{ is all of } \mathbb{C} \end{aligned}$$

Definition 10.41. Let $U \subset \mathbb{C}$ be a domain. For holomorphic $g, h : U \to \mathbb{C}$ with $h \neq 0$ the function

$$\begin{aligned} f : U \setminus \{z \in U \,|\, h(z) = 0\} &\longrightarrow \mathbb{C} \\ z &\longmapsto \frac{g(z)}{h(z)} \end{aligned}$$

is said to be a meromorphic function. Meromorphic functions are holomorphic on $U \setminus \{h(z) = 0\}$. If $z_0 \in U$ a root of order $m \in \mathbb{N}$ of h and a root of order $k \in \mathbb{N}_0$ of g, then the isolated singularity in z_0 of f is

- removable for $k \geq m$
- a pole of order $m - k$ for $k < m$

10.5 Residual Calculus

Definition 10.42 (Residue)**.** Let $r > 0$, $z_0 \in U$ and $f : K_{0,r}(z_0) \to \mathbb{C}$ holomorphic. Then for $\rho \in (0, r)$ the number

$$\operatorname{Res}_{z_0} f = \frac{1}{2\pi i} \oint_{\partial B_\rho(z_0)} f(z)\,\mathrm{d}z$$

is said to be the residue of f at z_0.

Lemma 10.43. *If f is a function as defined in Definition 10.42 with the Laurent series expansion $f(z) = \sum_{n \in \mathbb{Z}} c_n (z - z_0)^n$, then*

$$\operatorname{Res}_{z_0} f = c_{-1}$$

Proof.

$$\operatorname{Res}_{z_0} f = \frac{1}{2\pi i} \oint_{\partial B_r(z_0)} f(z)\mathrm{d}z = \sum_{n \in \mathbb{Z}} c_n \underbrace{\frac{1}{2\pi i} \oint_{\partial B_r(z_0)} (z - z_0)^n \mathrm{d}z}_{=\delta_{n,-1}} \tag{10.46}$$

□

Example 10.44. If f has a pole of order h at z_0, then the Laurent series of f around z_0 is

$$f(z) = a_{-k} \frac{1}{(z - z_0)^h} + a_{-(h-1)} \frac{1}{(z - z_0)^{h-1}} + \cdots + a_{-1} \frac{1}{(z - z_0)} + \sum_{n=0}^{\infty} a_n (z - z_0)^n$$

and thus

$$(z-z_0)^k f(z) = a_{-k} + a_{-(k-1)}(z-z_0) + \cdots a_{-1}(z-z_0)^{h-1} + \sum_{n=0}^{\infty} a_n (z-z_0)^{n+h}$$

From this follows

$$\operatorname{Res}_{z_0} f = \lim_{z \to z_0} \frac{1}{(h-1)!} \frac{\mathrm{d}^{h-1}}{\mathrm{d}z^{h-1}} (z-z_0)^h f(z)$$

Example 10.45. If $f(z)$ is meromorphic with a root of order 1 of h at z_0 and $g(z_0) \neq 0$, then

$$\begin{aligned} \operatorname{Res}_{z_0} f &= \lim_{z \to z_0} (z-z_0) f(z) = \lim_{z \to z_0} \frac{(z-z_0)}{h(z_0)} g(z_0) \\ &= \lim_{z \to z_0} \frac{z-z_0}{h(z)-h(z_0)} g(z_0) = \frac{g(z_0)}{h'(z_0)} \end{aligned}$$

Example 10.46. Consider $f(z) = \frac{1}{\sin z}$. f is meromorphic with poles of order 1 in $z_n = n\pi,\ n \in \mathbb{Z}$. With the previous example we can thus get

$$\operatorname{Res}_{z_n} f = \frac{1}{\cos(z_n)} = (-1)^n$$

Theorem 10.47 (Residue Theorem)**.** *Let $U \subset \mathbb{C}$ be open and $S := \{z_1, \cdots, z_n\} \subset U$ a set of pairwise disjoint points. Let $\gamma : [0,1] \to U \setminus S$ be a closed, piecewise differentiable curve without intersections that is null-homotopic in U and surrounds S in a positive orientation. Then for any holomorphic function $f : U \setminus S \longrightarrow \mathbb{C}$ the following holds*

$$\oint_\gamma f(z)\mathrm{d}z = 2\pi i \sum_{j=1}^{n} \operatorname{Res}_{z_j} f$$

Heuristic Proof. γ is null-homotopic, so we can choose arbitrary curves around the residues. Consider a path $\tilde{\gamma}$ that surrounds every residue with a disk of radius ρ_i, and connects them to each other using straight segments that cancel each other in the limit. Then

$$\oint_\gamma f(z)\mathrm{d}z = \oint_{\tilde{\gamma}} f(z)\mathrm{d}z = \sum_{j=1}^{n} \oint_{\partial B_{\rho_j}(z_0)} f(z)\mathrm{d}z = 2\pi i \sum_{j=1}^{n} \operatorname{Res}_{z_j} f$$

□

Example 10.48. The residue theorem can be used to calculate integrals such as

$$\int_{-\infty}^{\infty} \frac{1}{1+x^4}\mathrm{d}x$$

This integral is to be interpreted as

$$\int_{-\infty}^{\infty} \frac{1}{1+x^4}\mathrm{d}x = \lim_{R \to \infty} \int_{-R}^{R} \frac{\mathrm{d}x}{1+x^4} = \lim_{R \to \infty} I_R$$

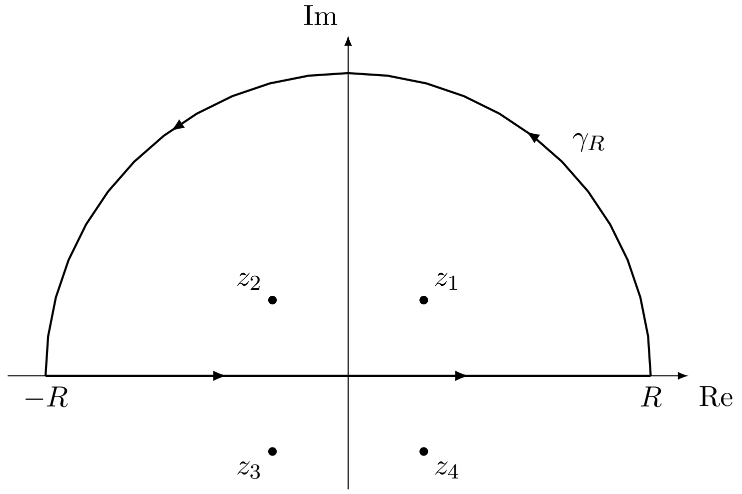
The poles of the integrand in $\mathbb{C}$ are the roots of the numerator:

$$1 + z^4 = 0 \iff z^4 = -1 = e^{-i\pi}$$

Which gives us the position of the isolated singularities

$$z_1 = e^{i\frac{\pi}{4}} \qquad z_2 = e^{i\frac{3\pi}{4}} \qquad z_3 = e^{i\frac{5\pi}{4}} \qquad z_4 = e^{i\frac{7\pi}{4}}$$

We want to integrate along the following path



Using the residue theorem we get

$$I_R + \int_{\gamma_R} \frac{\mathrm{d}z}{1+z^4} = 2\pi i \sum_{j=1}^{2} \mathrm{Res}_{z_j} f$$

So our next task is to calculate the residues of the poles z_1 and z_2.

$$\begin{aligned}
\mathrm{Res}_{z_1}\left(\frac{1}{1+z^4}\right) &= \lim_{z\to z_1}(z-z_1)\frac{1}{1+z^4} = \lim_{z\to z_1}\left(\frac{1}{z-z_2}\frac{1}{z-z_3}\frac{1}{z-z_4}\right) \\
&= \frac{1}{z_1-z_2}\frac{1}{z_1-z_3}\frac{1}{z_1-z_4} = \frac{1}{z_1^3}\left(\frac{1}{1-\frac{z_2}{z_1}}\frac{1}{1-\frac{z_3}{z_1}}\frac{1}{1-\frac{z_4}{z_1}}\right) \\
&= e^{-i\frac{3\pi}{4}}\underbrace{\frac{1}{1-e^{i\frac{\pi}{2}}}}_{\frac{1}{1-i}}\underbrace{\frac{1}{1-e^{i\pi}}}_{\frac{1}{2}}\underbrace{\frac{1}{1-e^{i\frac{3\pi}{2}}}}_{\frac{1}{1+i}} \\
&= e^{-i\frac{3\pi}{4}}\frac{1}{2}\left(\frac{1}{1+i}\frac{1}{1-i}\right) \\
&= \frac{1}{4}e^{i\frac{3\pi}{4}}
\end{aligned}$$

Analogously for z_2:

$$\operatorname{Res}_{z_2} f = \frac{1}{4} e^{-i\frac{\pi}{4}}$$

Thus we can evaluate the sum

$$\begin{aligned}
2\pi i \sum_{j=1}^{2} \operatorname{Res}_{z_j} f &= 2\pi i \left(\frac{e^{-i\frac{3\pi}{4}}}{4} + \frac{e^{-i\frac{\pi}{4}}}{4} \right) \\
&= \frac{\pi i}{2} \underbrace{e^{-i\frac{\pi}{4}}}_{\frac{1-i}{\sqrt{2}}} \underbrace{\left(1 + e^{-i\frac{\pi}{2}}\right)}_{1-i} \\
&= \frac{\pi i}{2} \frac{(1-i)}{\sqrt{2}} (1-i) = \frac{\pi}{\sqrt{2}}
\end{aligned}$$

We also need to calculate the integral along the curve γ_R. We use the parametrization

$$\gamma_R(t) = Re^{it}, \quad t \in [0, \pi]$$

$$\begin{aligned}
\left| \int_{\gamma_R} \frac{1}{1+z^4} \mathrm{d}z \right| = \left| \int_0^\pi \frac{1}{1 + R^4 e^{4it}} Rie^{it} \mathrm{d}t \right| &\leq \int_0^\pi \frac{R}{|1 + R^4 e^{4it}| \mathrm{d}t} \\
&= \int_0^\pi \frac{1}{R^3} \underbrace{\frac{1}{\left| e^{4it} + \frac{1}{R} \right|}}_{\leq 1} \\
&\leq \frac{1}{R^3} \pi \xrightarrow{R \to \infty} 0
\end{aligned}$$

So in total, we get

$$\int_{\mathbb{R}} \frac{\mathrm{d}x}{1+x^4} = \lim_{R \to \infty} I_R = 2\pi i \sum_{j=1}^{2} \operatorname{Res}_{z_j} f = \frac{\pi}{\sqrt{2}}$$

Example 10.49 (Fourier integrals of rational functions). We want to inspect Fourier integrals of the form

$$\int_{-\infty}^{\infty} \frac{P(x)}{Q(x)} e^{ix} \mathrm{d}x$$

where P and Q are polynomials such that the degree of Q is greater than the degree of P. The roots $z_1, \cdots, z_n \in \mathbb{C}$ of Q can not lie on the real axis. For the integration we'll use the same path as in the previous example. Using the residue theorem we get

$$\int_{-\infty}^{\infty} \frac{P(z)}{Q(z)} e^{iz} \mathrm{d}z + \int_{\gamma_R} \frac{P(z)}{Q(z)} e^{iz} \mathrm{d}z = 2\pi i \sum_{\operatorname{Im}(z_j) > 0} \operatorname{Res}_{z_j} \left(\frac{P(z)}{Q(z)} e^{iz} \right)$$

The integral along γ_R doesn't contribute anything to the total calculation:

$$\begin{aligned}\left|\int_{\gamma_R} \frac{P(z)}{Q(z)} e^{iz} \mathrm{d}z\right| &= \left|\int_0^\pi \frac{P(Re^{it})}{Q(Re^{it})} e^{iRe^{it}} \mathrm{d}t\right| \\ &\leq \int_0^\pi \frac{\left|P(Re^{it})\right|}{\left|Q(Re^{it})\right|} \left|e^{iRe^{it}}\right| \mathrm{d}t\end{aligned}$$

I want to insert a quick calculation:

$$\left|e^{iRe^{it}}\right| = \left|e^{iR\cos t} e^{-R\sin t}\right| = e^{-R\sin t}, \quad t \in (0,\pi)$$

Using this we can continue our calculations

$$\begin{aligned}&\leq \int_0^\pi \underbrace{\frac{\left|P(Re^{it})\right|}{\left|Q(Re^{it})\right|}}_{\leq M,\ M>0} e^{-R\sin t} \mathrm{d}t \\ &\leq M \int_0^\pi \underbrace{e^{-R\sin t}}_{\leq 1} \mathrm{d}t \xrightarrow{R\to\infty} 0\end{aligned}$$

Thus we showed that

$$\int_{-\infty}^{\infty} \underbrace{\frac{P(x)}{Q(x)}}_{f(x)} e^{ix} \mathrm{d}x = 2\pi i \sum_{\mathrm{Im}(z_j)>0} \mathrm{Res}_{z_j}\left(\frac{P(z)}{Q(z)} e^{iz}\right)$$

10.6 Application: Potential Theory

Definition 10.50 (Harmonic Function)**.** A function $\phi : U \to \mathbb{R}$ with $U \subset \mathbb{R}^d$ open, $d \in \mathbb{N}$ is said to be harmonic if

$$\nabla^2 \phi(x) = \sum_{j=1}^{d} \partial_j^2 \phi(x) = 0, \quad x \in \mathbb{R}^d$$

Theorem 10.51. *If $f : U \to \mathbb{C}$, $U \subset \mathbb{C}$ is holomorphic, then* $\mathrm{Re}(f)$ *and* $\mathrm{Im}(f)$ *are harmonic on U.*

Proof. If f is holomorphic on U, then f is also analytic, which means it is infinitely differentiable on U. Using the Cauchy-Riemann equations the desired statement can be shown. □

Example 10.52 (Potential problems in $\mathbb{R}^2$). Let $U \subset \mathbb{R}^2$ be a domain with smooth boundary. A typical problem from electrostatics is

$$\begin{aligned}\nabla^2 \phi(x,y) &= 0 & (x,y) &\in U \\ \phi(x,y) &= \phi_0(x,y) & (x,y) &\in \partial U\end{aligned}$$

where $\phi : \overline{U} \to \mathbb{R}$ is the desired function with given boundary values $\phi_0 \in C(\partial U, \mathbb{R})$. Such a boundary value problem is known as the Dirichlet problem.

An important example is the Dirichlet problem for the upper half plane

$$\left\{(x,y) \in \mathbb{R}^2 \,\middle|\, y > 0\right\}$$

with $\phi_0 \in C(\mathbb{R})$. We want to assume that ϕ_0 decreases like

$$|\phi_0(x)| \leq \frac{c}{1+|x|}, \quad x \in \mathbb{R}, c > 0$$

near infinity.

Theorem 10.53 (Poisson integral formula for the upper half plane)**.** *The function*

$$\phi(x,y) = \frac{1}{\pi} \int_{-\infty}^{\infty} \phi_0(t) \frac{y}{(x-t)^2 + y^2} \,\mathrm{d}t$$

solves the Dirichlet problem for the upper half plane

$$\begin{aligned} \nabla^2 \phi(x,y) &= 0 & \quad (x,y) &\in \left\{(x,y) \in \mathbb{R}^2 \,\middle|\, y > 0\right\} \\ \phi(x,0) &= \phi_0(x) & x &\in \mathbb{R} \end{aligned}$$

Proof. The function

$$f(z) := \frac{1}{\pi i} \int_{-\infty}^{\infty} \phi_0(t) \frac{1}{t-z} \,\mathrm{d}t \tag{10.47}$$

is holomorphic on $\mathbb{C}_+$ with

$$\mathrm{Re}(f(x+iy)) = \frac{1}{\pi} \int_{-\infty}^{\infty} \phi_0(t) \,\mathrm{Im}\left(\frac{1}{t-z}\right) \mathrm{d}t \tag{10.48}$$

We find that

$$\mathrm{Im}\left(\frac{1}{t-x-iy}\right) = \mathrm{Im}\left(\frac{1}{(t-x)-iy}\right) = \mathrm{Im}\left(\frac{(t-x)+iy}{(t-x)^2+y^2}\right) = \frac{y}{(t-x)^2+y^2} \tag{10.49}$$

and thus

$$\mathrm{Re}(f(x+iy)) = \phi(x,y) \tag{10.50}$$

for $y > 0$. However according to Theorem 10.51 this means that ϕ is harmonic for $y > 0$:

$$\nabla^2 \phi(x,y) = 0, \quad \forall (x,y) \in \left\{(x,y) \in \mathbb{R}^2 \,\middle|\, y > 0\right\} \tag{10.51}$$

It remains to be shown that the boundary values are being accounted for. For that we rewrite the Poisson integral formula as

$$\phi(x,y) = (\phi_0 * P_y)(x) := \int_{\mathbb{R}} \phi(t) P_y(x-t) \,\mathrm{d}t \tag{10.52}$$

with the Poisson kernel $P_y(x) = \frac{1}{\pi} \frac{y}{x^2+y^2}$. We will finish this proof later. $\square$

Definition 10.54 (Convolution)**.** We define

$$L^1(\mathbb{R}^d) := \left\{ f : \mathbb{R}^d \to \mathbb{R} \text{ measurable} \,\middle|\, \int_{\mathbb{R}^d} |f(x)| \,\mathrm{d}x < \infty \right\}$$

the Lebesgue space of absolutely integrable functions. It is a complete, normed space with

$$\|f\|_{L^1} = \int_{\mathbb{R}^d} |f| \,\mathrm{d}x$$

It induces a metric space with the metric

$$d(f,g) := \|f - g\|_{L^1}$$

Let $(f_n) \subset L^1(\mathbb{R}^d)$. This sequence converges to $f \in L^1(\mathbb{R}^d)$ if

$$\|f_n - f\|_{L^1} \xrightarrow{n\to\infty} 0$$

Since L^1 is complete, every Cauchy sequence converges. For $f, g \in L^1(\mathbb{R}^d)$

$$(f * g)(x) := \int_{\mathbb{R}^d} f(y) g(x-y) \,\mathrm{d}y$$

is said to be the convolution of f and g.

Theorem 10.55. *The convolution is well defined as a mapping*

$$* : L^1(\mathbb{R}^d) \times L^1(\mathbb{R}^d) \longrightarrow L^1(\mathbb{R}^d)$$

with $\|f * g\|_{L^1} \le \|f\|_{L^1} \|g\|_{L^1}$. *The space* $(L^1(\mathbb{R}^d), *)$ *is a commutative and associative algebra, i.e.* $\forall f, g, h \in L^1(\mathbb{R}^d)$*:*

(i) $f * g = g * f$

(ii) $(f * g) * h = f * (g * h)$

(iii) $f * (g + h) = f * g + f * h$

(iv) $\forall \lambda \in \mathbb{C} : \quad \lambda(f * g) = (\lambda f) * g = f * (\lambda g)$

Proof. We will only be proving that the mapping is well defined and that the inequality holds. First, let

$$f, g \in L^1 \cap L^\infty := \left\{ f : \mathbb{R}^d \to \mathbb{R} \text{ measurable} \,\middle|\, \operatorname*{ess\,sup}_{x \in \mathbb{R}^d} |f(x)| < \infty \right\} \tag{10.53}$$

Then the convolution $f * g$ is well defined (pointwise almost everwhere), because

$$|f(y) g(x-y)| \underset{\substack{\uparrow \\ g \in L^\infty}}{\le} C|f(y)| \implies f \in L^1 \implies \text{integrable} \tag{10.54}$$

We then find that

$$\begin{aligned}
\|f * g\|_{L^1} &= \int_{\mathbb{R}^d} |(f * g)(x)| \,\mathrm{d}x = \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} f(y) g(x-y) \,\mathrm{d}y \right| \mathrm{d}x \\
&\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(y)||g(x-y)| \,\mathrm{d}y \,\mathrm{d}x \qquad (10.55) \\
&= \int_{\mathbb{R}^d} |f(y)| \int_{\mathbb{R}^d} |g(x-y)| \,\mathrm{d}x \,\mathrm{d}y
\end{aligned}$$

By substituting $z = x - y$ we get

$$= \int_{\mathbb{R}^d} |f(y)| \int_{\mathbb{R}^d} |g(z)| \,\mathrm{d}z \,\mathrm{d}y = \|f\|_{L^1} \|g\|_{L^1} \qquad (10.56)$$

For more general $f, g \in L^1$ we can approximate f via a function sequence

$$f_n := \min\{f, n\} \in L^1 \cap L^\infty \qquad (10.57)$$

Then $f_n \xrightarrow{n\to\infty} f$ in L^1, since

$$|f_n(x)| \leq |f(x)| \quad \forall x \in \mathbb{R}^d \qquad (10.58)$$

By using the previous results we can conclude

$$\|f_n * g - f_m * g\|_{L^1} = \|(f_n - f_m) * g\|_{L^1} \leq \underbrace{\|f_n - f_m\|_{L^1}}_{\xrightarrow{n,m\to\infty} 0} \|g\|_{L^1} \qquad (10.59)$$

So $(f_n * g)_{n\in\mathbb{N}}$ is a Cauchy sequence in L^1 and thus

$$f * g := \lim_{n\to\infty} f_n * g \qquad (10.60)$$

□

Remark 10.56. One can show that $(L^1(\mathbb{R}^d), *)$ does not have a neutral element, i.e.

$$\nexists \delta \in L^1(\mathbb{R}^d) : \quad f * \delta = f \quad \forall f \in L^1(\mathbb{R}^d)$$

Definition 10.57 (Good kernels, Approximative identity)**.** A sequence of convolution kernels $(K_n) \subset L^1(\mathbb{R}^d)$ is said to be a class of good kernels if

$$\begin{aligned}
\forall n \in \mathbb{N} : &\quad \int_{\mathbb{R}^d} K_n(x) \,\mathrm{d}x = 1 \\
\exists M > 0 \ \forall n \in \mathbb{N} : &\quad \int_{\mathbb{R}^d} |K_n(x)| \,\mathrm{d}x \leq M \\
\forall \delta > 0 : &\quad \lim_{n\to\infty} \int_{|x|>\delta} |K_n(x)| \,\mathrm{d}x = 0
\end{aligned}$$

A sequence of good kernels with $K_n \geq 0$ for all $n \in \mathbb{N}$ is called Dirac sequence.

Theorem 10.58 (Smoothing by convolution with good kernels)**.** *Let $(K_n) \subset L^1(\mathbb{R}^d)$ be a class of good kernels. Then:*

(i) If $f \in L^1(\mathbb{R}^d)$ then
$$\|f * K_n - f\|_{L^1(\mathbb{R}^d)} = 0$$

(ii) If $K_n \subset C^m(\mathbb{R}^d)$ $\forall n \in \mathbb{N}$ and if the partial derivatives $\partial^\alpha K_n$, $|\alpha| \le m$ are bounded, then
$$f * K_n \in C^m(\mathbb{R}^d) \text{ and } \partial^\alpha (f * K_n) = f * \partial^\alpha K_n$$

(iii) If $f \in C(\mathbb{R}^d)$ is bounded, then
$$\lim_{n\to\infty} (f * K_n)(x) = f(x), \quad \forall x \in \mathbb{R}^d$$

Example 10.59. Let $(\epsilon_n) \subset (0, \infty)$ be a null sequence. Then

$$\text{Poisson kernels} \qquad P_k(x) := \frac{1}{\pi}\frac{\epsilon_k}{x^2 + \epsilon_k^2}$$

$$\text{Gauß kernels} \qquad \delta_k(x) := (2\pi\epsilon_k^2)^{-\frac{d}{2}} e^{-\frac{|x|^2}{2\epsilon_k^2}}, \quad x \in \mathbb{R}^d$$

are classes of good kernels. Now let $0 \le \phi \in L^1(\mathbb{R}^d)$ with $\|\phi\|_{L^1} = 1$. Then
$$\phi_k(x) = \frac{1}{\epsilon_k^d}\phi\left(\frac{x}{\epsilon}\right)$$
is a class of good kernels. We can show that
$$P_{\epsilon_k}(x) = \frac{1}{\pi}\frac{1}{\epsilon_k^2}\frac{\epsilon_k}{\left(\frac{x}{\epsilon_k}\right)^2 + 1} = \frac{1}{\epsilon_k}\frac{1}{\pi}\frac{1}{\left(\frac{x}{\epsilon_k}\right)^2 + 1} = \frac{1}{\epsilon_k} P\left(\frac{x}{\epsilon_k}\right)$$

One has to show that
$$P(x) = \frac{1}{\pi}\frac{1}{1 + x^2} \in L^1$$
with
$$\int_{\mathbb{R}} P(x)\,\mathrm{d}x = 1$$
To do that we can calculate
$$\int_{\mathbb{R}} \underbrace{P(x)}_{\ge 0}\,\mathrm{d}x = \frac{1}{\pi}\int_{\mathbb{R}} \frac{1}{1 + x^2}\,\mathrm{d}x = \frac{1}{\pi}(\arctan(\infty) - \arctan(-\infty)) = \frac{1}{\pi}\left(\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right) = 1$$

Proof of Theorem 10.58 (iii). Consider $f * K_n(x) - f(x)$. We can calculate

$$\begin{aligned} f * K_n(x) - f(x) &= \int_{\mathbb{R}} f(x-y)K_n(y)\,\mathrm{d}y - f(x)\underbrace{\int_{\mathbb{R}^d} K_n(y)\,\mathrm{d}y}_{=1} \\ &= \int_{\mathbb{R}^d} \left(f(x-y) - f(x)\right) K_n(y)\,\mathrm{d}y \end{aligned} \tag{10.61}$$

Since f is continuous in x,

$$\forall \epsilon > 0\ \exists \delta > 0 : \quad |f(x-y) - f(x)| < \epsilon \quad \forall |y| < \delta \tag{10.62}$$

From the definition of good kernels follows that

$$\exists M > 0 : \quad \int_{\mathbb{R}^d} |K_n(y)|\,\mathrm{d}y \leq M \tag{10.63}$$

which lets us conclude

$$\int_{|y|<\delta} \underbrace{|f(x-y) - f(x)|}_{<\epsilon} |K_n(y)|\,\mathrm{d}y < \epsilon \int_{|y|<\delta} |K_n(x)|\,\mathrm{d}y \leq \epsilon M \tag{10.64}$$

By utilising the boundedness of f and (iii) from the definition of good kernels we can show

$$\int_{|y|\geq\delta} \underbrace{|f(x-y) - f(x)|}_{\leq 2c} |K_n(y)|\,\mathrm{d}y \leq 2c \underbrace{\int_{|y|\geq\delta} |K_n(y)|\,\mathrm{d}y}_{\epsilon} \leq 2c\epsilon \tag{10.65}$$

We can now use the previous results to show that

$$\begin{aligned} |f * K_n(x) - f(x)| &\leq \int_{\mathbb{R}^d} |f(x-y) - f(x)||K_n(y)|\,\mathrm{d}y \\ &= \underbrace{\int_{|y|<\delta} |f(x-y) - f(x)||K_n(y)|\,\mathrm{d}y}_{\leq M\epsilon} + \underbrace{\int_{|y|\geq\delta} |f(x-y) - f(x)||K_n(y)|\,\mathrm{d}y}_{\leq 2c\epsilon} \\ &\leq \epsilon(M + 2c) \end{aligned} \tag{10.66}$$

Since ϵ can be chosen arbitrarily it follows that

$$f * K_n(x) \xrightarrow{n\to\infty} f(x) \quad \forall x \in \mathbb{R}^d \tag{10.67}$$

With this it is now easy to finish **the proof for Theorem 10.53**. We had seen that

$$\phi(x,y) = \int_{\mathbb{R}} \phi_0(t) P_y(x-t)\,\mathrm{d}t = (\phi_0 * P_y)(x) \tag{10.68}$$

Now let $(\epsilon_n) \subset (0, \infty)$ be a null sequence. Since ϕ_0 is continuous and bounded, and since $(P_{\epsilon_n}) \subset L^1(\mathbb{R}^d)$ is a class of good kernels, it follows from what we have just proven that

$$\lim_{n\to\infty} (\phi_0 * P_{\epsilon_n})(x) = \phi_0(x) \tag{10.69}$$

All in all it follows that

$$\begin{aligned} \nabla^2\phi(x,y) &= 0 \qquad && x \in \mathbb{R},\ y > 0 \\ \phi(x,0) &\xrightarrow{y\to 0} \phi_0(x) \qquad && \forall x \in \mathbb{R} \end{aligned} \tag{10.70}$$

□

Remark 10.60. If $\psi : U \to \mathbb{C}$ is holomorphic, $U \subset \mathbb{C}$ open and $V \subset U$ a domain, then $\psi(V)$ is also a domain with $\psi(\partial V) = \partial\psi(V)$. (Open mapping principle). Then the solution to the Dirichlet problem on V

$$\begin{aligned} \nabla^2\phi &= 0 \qquad && \text{on } V \\ \phi &= \phi_0 \qquad && \text{on } \partial V \end{aligned}$$

can be obtained through a holonomic transformation $\psi : V \to \mathbb{C}_+$ with $\phi(V) = \mathbb{C}_+$ of the Dirichlet problem on the upper half plane.

Chapter 11

Fourier Transform and Basics of Distribution Theory

11.1 Fourier Transform on $L^1(\mathbb{R}^d)$

Definition 11.1. For $f \in L^1(\mathbb{R}^d)$ the function

$$\begin{aligned}\hat{f} : \mathbb{R}^d &\longrightarrow \mathbb{C} \\ k &\longmapsto \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} e^{-ik\cdot x} f(x)\,\mathrm{d}x\end{aligned}$$

is said to be the Fourier transform of f. Sometimes it is written as $(\mathcal{F}f)(k)$.

Remark 11.2. (i) There are several alternative conventions regarding sign and phase of the transform.

$$\hat{f}(k) = \int_{\mathbb{R}^d} e^{-2\pi i k\cdot x} f(x)\,\mathrm{d}x$$

is also a valid definition in other scientific fields, however we will stick to the former definition throughout this script.

(ii) $k \cdot x = \langle k|x\rangle = \sum_{j=1}^d k_j x_j$

(iii) Because $\left|e^{ik\cdot x}\right| = 1$, the integral exists for any $f \in L^1(\mathbb{R}^d)$.

Example 11.3. Consider the function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\alpha)^2}{2\sigma^2}} \quad ,\alpha \in \mathbb{C},\ \sigma > 0$$

The Fourier transform is

$$\begin{aligned}\hat{f}(k) =& \frac{1}{2\pi\sigma} \int_{\mathbb{R}} e^{-ikx} e^{-\frac{(x-\alpha)^2}{2\sigma^2}}\,\mathrm{d}x \\ =& \frac{1}{2\pi\sigma} e^{-ik\alpha} e^{-\frac{\sigma^2}{2}k^2} \int_{\mathbb{R}} e^{-\frac{(x-\alpha+ik\sigma^2)^2}{2\sigma^2}}\,\mathrm{d}x \\ =& \frac{1}{2\pi\sigma} e^{-ik\alpha} e^{-\frac{\sigma^2 k^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}}\,\mathrm{d}x \\ =& \frac{1}{\sqrt{2\pi}} e^{-ik\alpha} e^{-\frac{\sigma^2 k^2}{2}}\end{aligned}$$

Example 11.4. The previous example can be generalized for higher dimensions:

$$f(x) = \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} e^{-\frac{|x|^2}{2\sigma^2}}, \quad x \in \mathbb{R}^d$$

With the Fourier transform

$$\hat{f}(k) = \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-\frac{\sigma^2|k|^2}{2}}, \quad k \in \mathbb{R}^d$$

Example 11.5. Consider the indicator function

$$\chi(x) = \begin{cases} 1, & |x| \leq a \\ 0, & |x| > a \end{cases}, \quad x \in \mathbb{R}$$

It has the Fourier transform

$$\hat{f}(k) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-ikx} \chi(x) \,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \left[\frac{e^{-ikx}}{-ik} \right]_{-a}^{a} = \sqrt{\frac{2}{\pi}} \frac{\sin(ka)}{k}$$

Definition 11.6. For $f \in L^1(\mathbb{R}^d)$ the function

$$\check{f}(x) := \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} f(k) e^{ikx} \,\mathrm{d}k$$

defines the inverse Fourier transform of f.

Example 11.7. Let's revisit Example 11.3, where we found that

$$\hat{f}(k) = \frac{1}{\sqrt{2\pi}} e^{-i\alpha k} e^{-\frac{\sigma^2 k^2}{2}}$$

The inverse Fourier transform of that function is

$$\check{\hat{f}}(k) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\alpha)^2} = f(x)$$

Theorem 11.8 (Fourier inversion theorem)**.** *Let $f, \hat{f} \in L^1(\mathbb{R}^d)$. Then $\check{\hat{f}} = f$.*

Proof. To prove this theorem we will use Theorem 10.58 and the following lemma:

If $f_n \xrightarrow{n\to\infty} f$, then there exists a subsequence $((f_n)_k)_{k\in\mathbb{N}}$ with

$$\lim_{k\to\infty} f_{n_k}(x) = f(x), \quad \forall x \in \mathbb{R}^d$$

Heuristically this theorem can be proven by considering

$$\begin{aligned} \check{\hat{f}}(x) = \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} \hat{f}(k) e^{ikx} \,\mathrm{d}k &= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(y) e^{-iky} \,\mathrm{d}y \right) e^{ikx} \,\mathrm{d}k \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(y) e^{-ik(y-x)} \,\mathrm{d}k \,\mathrm{d}y \end{aligned} \tag{11.1}$$

However, to show this rigorously we should first consider the inversion formula for $f * \delta_l$, with

$$\delta_l(x) = \left(\frac{l^2}{2\pi} \right)^{\frac{d}{2}} e^{-\frac{l^2|x|^2}{2}}, \quad l \in \mathbb{N} \tag{11.2}$$

The Fourier transform is

$$\hat{\delta}_l(k) = \left(\frac{1}{2\pi}\right)^{\frac{d}{2}} e^{-\frac{|k|^2}{2l^2}} \tag{11.3}$$

We've already shown in a previous example that the inversion theorem applies to this function, so we can write

$$\begin{aligned}
(f * \delta_l)(x) &= \int_{\mathbb{R}^d} f(y)\delta_l(x-y)\,\mathrm{d}y = \int_{\mathbb{R}^d} f(y)\check{\hat{\delta}}_l(x-y)\,\mathrm{d}y \\
&= \int_{\mathbb{R}^d} f(y)\left(\frac{1}{(2\pi)^{\frac{d}{2}}}\int_{\mathbb{R}^d} e^{ik(x-y)}\hat{\delta}_l(k)\,\mathrm{d}k\right)\mathrm{d}y \\
&= \int_{\mathbb{R}^d}\left(\frac{1}{(2\pi)^{\frac{d}{2}}}\int_{\mathbb{R}^d} f(y)e^{-iky}\,\mathrm{d}y\right)e^{ikx}\hat{\delta}_l(k)\,\mathrm{d}k \\
&= \int_{\mathbb{R}^d} e^{ikx}\hat{f}(k)e^{-\frac{|k|}{2l^2}}\frac{\mathrm{d}k}{(2\pi)^{\frac{d}{2}}} =: \check{F}_l(x)
\end{aligned} \tag{11.4}$$

Next we want to use the fact that $(\delta_l)_{l\in\mathbb{N}}$ is a class of good kernels. This means that

$$\lim_{l\to\infty}\|\delta_l * f - f\|_{L^1} = 0 \tag{11.5}$$

or in other words

$$\delta_l * f \xrightarrow{l\to\infty} f \tag{11.6}$$

Now using the lemma above we can conclude that there exists a subsequence $(\delta_{l_j} * f)_{j\in\mathbb{N}}$ that converges to $f(x)$ for almost every x. We can apply the dominated convergence theorem to find that

$$\lim_{l\to\infty}\check{F}_l(x) = \int_{\mathbb{R}^d} e^{ikx}\hat{f}(k)\lim_{l\to\infty} e^{-\frac{|k|^2}{2l^2}}\frac{\mathrm{d}k}{(2\pi)^{\frac{d}{2}}} = \check{\hat{f}}(x) \tag{11.7}$$

Finally, this lets us conclude

$$f(x) = \lim_{j\to\infty}\delta_{l_j} * f(x) = \lim_{j\to\infty}\check{F}_{l_j}(x) = \check{\hat{f}}, \quad \text{for a.e. } x \in \mathbb{R}^d \tag{11.8}$$

□

Theorem 11.9 (Algebraisation of the derivative). *Let $f \in C^n(\mathbb{R}^d)$ and $\partial^\alpha f \in L^1(\mathbb{R}^d)$ for all $\alpha \in \mathbb{N}_0^d$ with $|\alpha| < m$. Then*

$$\widehat{\partial^\alpha f}(k) = (ik)^\alpha \hat{f}$$

If $\alpha = (\alpha_1, \cdots, \alpha_d)$ and $|\alpha| = \sum_{j=1}^d \alpha_j \leq m$, then the interpretation of that is

$$\mathcal{F}\left(\frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1}\cdots\partial x_d^{\alpha_d}} f\right)(k) = i^{|\alpha|}\left(k_1^{\alpha_1}\cdots k_d^{\alpha_d}\right)\hat{f}(k)$$

Proof. We will only prove the one-dimensional statement. If $m = 1$ we receive via partial integration

$$\widehat{f'}(k) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f'(x) e^{-ikx} \,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \left[\left(f(x) e^{-ikx} \right)_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f(x) \frac{\mathrm{d}}{\mathrm{d}x} \left(e^{-ikx} \right) \mathrm{d}x \right] \quad (11.9)$$

Since we assumed $f' \in L^1$, the limit $\lim_{x \to \pm\infty} f(x)$ exists. We can write

$$f(x) = f(0) + \int_0^x f'(t) \,\mathrm{d}t \quad (11.10)$$

$$\implies \lim_{x \to \pm\infty} f(x) = f(0) + \lim_{x \to \pm\infty} \int_0^x f'(t) \,\mathrm{d}t \quad (11.11)$$

Furthermore this limit has to be equal to 0, so

$$\lim_{|x| \to \infty} f(x) = 0 \quad (11.12)$$

$$\implies \left[f(x) e^{-ikx} \right]_{-\infty}^{\infty} = \lim_{x \to \infty} e^{-ikx} f(x) - \lim_{x \to \infty} f(x) e^{ikx} = 0 \quad (11.13)$$

This leads us to

$$\widehat{f'}(k) = ik \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-ikx} \,\mathrm{d}x = ik \hat{f}(k) \quad (11.14)$$

The proof for $m > 1$ can be found via induction. □

Theorem 11.10. *Let $f \in L^1(\mathbb{R}^d)$ and $m \in \mathbb{N}_0$. If*

$$x \longmapsto x^\alpha f(x) \in L^1(\mathbb{R}^d), \quad \forall \alpha \in \mathbb{N}_0^d,\ |\alpha| \leq m$$

then $\hat{f} \in C^m(\mathbb{R}^d)$ and

$$\partial^\alpha \hat{f}(k) = \mathcal{F}\left[(-ix)^\alpha f(x) \right](k)$$

Proof. Again we will only consider the one-dimensional case. Assume $m = 1$ (the proof for $m > 1$ follows from induction). We can write out the difference quotient for $\hat{f}$ at $k \in \mathbb{R}$

$$\frac{\hat{f}(k+h) - \hat{f}(k)}{h} = \frac{1}{h} \int_{\mathbb{R}} f(x) \left(e^{-i(k+h)x} - e^{-ikx} \right) \frac{\mathrm{d}x}{\sqrt{2\pi}}, \quad h \in \mathbb{R} \setminus \{0\} \quad (11.15)$$

However, because

$$\left| \frac{e^{-ixh} - 1}{h} \right| \leq |x|, \quad \forall x \in \mathbb{R},\ h \neq 0 \quad (11.16)$$

and because we assumed that $xf \in L^1$, we can use the dominated convergence theorem to conclude

$$\begin{aligned} \lim_{h \to 0} \frac{\hat{f}(k+h) - \hat{f}(k)}{h} &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-ikx} \underbrace{\lim_{h \to 0} \left(\frac{e^{-ixk} - 1}{h} \right)}_{\frac{\mathrm{d}}{\mathrm{d}x} \left(e^{-ikx} \right) \big|_{x=0}} \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (-ixf(x)) e^{-ikx} \,\mathrm{d}x = \widehat{-ixf}(k) \end{aligned} \quad (11.17)$$

□

Theorem 11.11. *Let $f, g \in L^1(\mathbb{R}^d)$. Then*

$$\widehat{f * g}(k) = (2\pi)^{\frac{d}{2}} \hat{f}(k)\hat{g}(k)$$

Proof. Without proof. □

Example 11.12 (Solving inhomogeneous, linear ODEs). We want to find the general solution of

$$\ddot{x} - x = f, \quad f, \hat{f} \in L^1(\mathbb{R})$$

The solution space of this equation is

$$\mathcal{L} = \mathcal{L}_{\text{hom}} + x_s$$

The space of homogeneous solutions $\mathcal{L}_{\text{hom}}$ is equal to span $\{e^x, e^{-x}\}$, and x_s is *one* solution of the inhomogeneous equation. Let ϕ denote the Fourier transform of x_s, so

$$\phi(k) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x_s(t) e^{-ikt} \,\mathrm{d}t$$

Then ϕ satisfies the equation

$$-k^2\phi(k) - \phi(k) = \hat{f}(k)$$

Or rearraranged to solve for ϕ

$$\phi(k) = -\frac{1}{1+k^2}\hat{f}(k), \quad k \in \mathbb{R}$$

We can then rewrite $\phi(k)$ as

$$\phi(k) = -\hat{g}(k)\hat{f}(k) \qquad\qquad \text{with } \hat{g} = \frac{1}{1+k^2} \in L^1(\mathbb{R})$$

and then use the previous theorems to conclude

$$x_s(t) = \check{\phi}(t) = (2\pi)^{\frac{1}{2}} \mathcal{F}^{-1} \underbrace{\left[\hat{f}\hat{g}(2\pi)^{\frac{1}{2}}\right]}_{\widehat{f*g}}(t) = -\frac{1}{\sqrt{2\pi}}(f * g)(t)$$

11.2 Fourier Transform on $L^2(\mathbb{R}^d)$

Definition 11.13 (Hilbert space)**.** For this section we introduce the Hilbert space of Lebesgue square-integrable functions

$$L^2(\mathbb{R}^d) := \left\{ f : \mathbb{R}^d \to \mathbb{C} \text{ measurable} \,\middle|\, \|f\|_{L^2}^2 = \int_{\mathbb{R}^d} |f(x)|^2 \,\mathrm{d}x < \infty \right\}$$

This space is also important in quantum mechanics, as wave functions are elements of L^2.

Definition 11.14. The space $L^2(\mathbb{R}^d)$ is a Hilbert space, i.e. a complete, normed vector space with an inner product

$$\langle f|g\rangle := \int_{\mathbb{R}^d} \overline{f(x)} g(x)\,\mathrm{d}x$$

that has the following properties:

(i) $\langle f|f\rangle \geq 0$ and $\langle f|f\rangle = 0 \iff f = 0$

(ii) $\langle f|g\rangle = \overline{\langle g|f\rangle}$

(iii) $\langle f|g+\lambda h\rangle = \langle f|g\rangle + \lambda \langle f|h\rangle$

(ii) and (iii) imply

$$\langle \lambda f|g\rangle = \overline{\lambda}\,\langle f|g\rangle$$

The inner product induces a norm

$$\|f\|_{L^2}^2 = \langle f|f\rangle \left(= \int_{\mathbb{R}} \underbrace{\overline{f(x)} f(x)}_{|f(x)|^2} \,\mathrm{d}x \right)$$

Since the Fourier transform cannot directly be defined for $L^2(\mathbb{R}^d)$, we will first consider the space of rapidly decreasing functions, the so called *Schwartz space* $\mathcal{S}(\mathbb{R}^d)$.

Definition 11.15 (Schwartz space)**.** The Schwartz space $\mathcal{S}(\mathbb{R}^d)$ is defined as the function space

$$\mathcal{S}(\mathbb{R}^d) := \left\{ f \in C^\infty(\mathbb{R}^d) \,\middle|\, x \mapsto x^\beta \partial^\alpha f \text{ bounded}, \quad \forall \alpha, \beta \in \mathbb{N}_0^d \right\}$$

Example 11.16. (i) Smooth functions with compact support $f \in C^\infty(\mathbb{R}^d)$ are also elements of $\mathcal{S}(\mathbb{R}^d)$, for example

$$f(x) = \begin{cases} \exp\left(-\sum_{j=1}^d \frac{1}{1-|x_j|^2}\right), & |x_j| < 1 \\ 0, & \text{else} \end{cases}$$

(ii) For every polynomial $p(x)$, the function

$$f(x) = p(x) e^{-|x|^2}$$

defines a function in $\mathcal{S}(\mathbb{R}^d)$.

Remark 11.17. Because of the continuity and the rapid decrease towards infinity we can find that

$$\mathcal{S}(\mathbb{R}^d) \subset L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$$

and one can show that $\mathcal{S}(\mathbb{R}^d)$ is dense in $L^2(\mathbb{R}^d)$, i.e.

$$\forall f \in L^2(\mathbb{R}^d)\; \exists\, (f_n) \subset \mathcal{S}(\mathbb{R}^d) : \quad \|f_n - f\|_{L^2} \xrightarrow{n\to\infty} 0$$

Theorem 11.18. *Let $f \in \mathcal{S}(\mathbb{R}^d)$. Then $\hat{f} \in \mathcal{S}(\mathbb{R}^d)$ and the restriction of the Fourier transform to $\mathcal{S}(\mathbb{R}^d)$*

$$\mathcal{F}_{\mathcal{S}} : \mathcal{S}(\mathbb{R}^d) \longrightarrow \mathcal{S}(\mathbb{R}^d)$$

is an isomorphism. Furthermore

$$\left\langle \hat{f} \middle| \hat{g} \right\rangle = \langle f|g\rangle , \quad \forall f, g \in \mathcal{S}(\mathbb{R}^d)$$

with the inner product

$$\langle f|g\rangle = \int_{\mathbb{R}^d} \overline{f(x)} g(x) \, \mathrm{d}x$$

Proof. To prove that $\hat{f} \in \mathcal{S}$ we use the fact that

$$k^\beta \partial^\alpha \hat{f}(k) = (-i)^{|\alpha|+|\beta|} \mathcal{F}_{\mathcal{S}} \left[\partial^\beta x^\alpha f(k) \right] , \quad k \in \mathbb{R}^d, \ \forall \alpha, \beta \in \mathbb{N}_0^d \tag{11.18}$$

Next we want to prove that $\mathcal{F}_{\mathcal{S}}$ is an isomorphism. This is trivial however since

$$\forall f \in \mathcal{S}(\mathbb{R}^d) : \quad {\mathcal{F}_{\mathcal{S}}}^{-1} \mathcal{F}_{\mathcal{S}}(f) = f \tag{11.19}$$

To prove the final statement we can explicitly calculate

$$\begin{aligned}
\left\langle \hat{f} \middle| \hat{g} \right\rangle &= \int_{\mathbb{R}^d} \overline{\hat{f}(k)} \hat{g}(k) \, \mathrm{d}k \\
&= \int_{\mathbb{R}^d} \overline{\left(\int_{\mathbb{R}^d} f(x) e^{-ikx} \, \mathrm{d}x \, \frac{\mathrm{d}x}{(2\pi)^{\frac{d}{2}}} \right) \hat{g}(x) \, \mathrm{d}k} \\
&= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \overline{f(x)} e^{ikx} \frac{\mathrm{d}x}{(2\pi)^{\frac{d}{2}}} \right) \hat{g}(k) \, \mathrm{d}k \\
&= \int_{\mathbb{R}^d} \overline{f(x)} \underbrace{\left(\int_{\mathbb{R}^d} \hat{g}(k) e^{ikx} \frac{\mathrm{d}k}{(2\pi)^{\frac{d}{2}}} \right)}_{\check{\hat{g}}(x) = g(x)} \mathrm{d}x \\
&= \int_{\mathbb{R}^d} \overline{f(x)} g(x) \, \mathrm{d}x = \langle f|g\rangle
\end{aligned} \tag{11.20}$$

□

Remark 11.19. Since not all functions $f \in L^2(\mathbb{R}^d)$ are integrable, the limit

$$\lim_{R \to \infty} \int_{|x|<R} f(x) e^{-ikx} \frac{\mathrm{d}x}{(2\pi)^{\frac{d}{2}}} = \hat{f}(k)$$

doesn't converge for every $k \in \mathbb{R}^d$, only for almost every.

Theorem 11.20. *The Fourier transform $\mathcal{F}_{\mathcal{S}}$ can be uniquely and continuously continued on $L^2(\mathbb{R}^d)$. The resulting mapping*

$$\mathcal{F} : L^2(\mathbb{R}^d) \longrightarrow L^2(\mathbb{R}^d)$$

is linear and unitary, i.e. $\forall f, g \in L^2(\mathbb{R}^d)$ we have

$$\langle \mathcal{F}(f) | \mathcal{F}(g) \rangle = \langle f | g \rangle$$

which is also known as the Plancherel identity.

Proof. Without proof. □

Remark 11.21. The continuiuty of $\mathcal{F}$ doesn't imply that $\hat{f}$ is continuous. Secondly, the Plancherel identity also yields

$$\|f\|_{L^2} = \sqrt{\langle f|f \rangle} = \sqrt{\left\langle \hat{f} \middle| \hat{f} \right\rangle} = \left\| \hat{f} \right\|_{L^2}$$

11.3 Outlook: Tempered Distributions

Definition 11.22. A tempered distribution f is a continuous, linear mapping

$$\begin{aligned} f : \mathcal{S}(\mathbb{R}^d) &\longrightarrow \mathbb{C} \\ \phi &\longmapsto f(\phi) = (f, \phi) \left(= \int f(x)\phi(x) \, \mathrm{d}x \right) \end{aligned}$$

Theorem 11.23. *Tempered distributions are linear, continuous mappings.*

Proof. To prove linearity, let $\phi, \psi \in \mathcal{S}(\mathbb{R}^d)$ and $\lambda \in \mathbb{C}$. Then

$$(f, \phi + \lambda\psi) = (f, \phi) + \lambda(f, \psi) \tag{11.21}$$

For the continuity, we want to consider any sequence $(\phi_n) \subset \mathcal{S}(\mathbb{R}^d)$ that converges to $\phi \in \mathcal{S}(\mathbb{R}^d)$. I.e.

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}^d} \left| x^\beta \partial^\alpha (\phi_n(x) - \phi(x)) \right| = 0, \quad \forall \alpha, \beta \in \mathbb{N}_0^d \tag{11.22}$$

Then we can conclude that

$$\lim_{n\to\infty} |(f, \phi_n) - (f, \phi)| = 0 \tag{11.23}$$

□

Remark 11.24. The space of all tempered distributions is denoted as $\mathcal{S}'(\mathbb{R}^d)$.

Example 11.25. One important example is the Dirac deltra distribution:

$$\delta : \mathcal{S}(\mathbb{R}^d) \longrightarrow \mathbb{C}$$

It maps a function to its value at 0.

$$(\delta, \phi) = \int \delta(x)\phi(x) \, \mathrm{d}x = \phi(0) \in \mathbb{C}$$

Chapter 12

Operator Theory

12.1 Linear Operators

Throughout this chapter X and Y will denote vector spaces over the same scalar field $\mathbb{K}$. Also, I want to quickly recap some normed vector spaces that we will use from here on out.

(i) The real numbers

$$X = \mathbb{R}$$
$$\|x\| = |x|, \quad x \in \mathbb{R}$$

(ii) The euclidian space

$$X = \mathbb{R}^n$$
$$\|x\| = \left(\sum_{k=1}^{n} \xi_k^2 \right)^{\frac{1}{2}}, \quad x = (\xi_1, \cdots, \xi_n) \in \mathbb{R}^n$$

(iii) The space of bounded sequences l^∞

$$X = l^\infty := \{(\xi_k) \subset \mathbb{R} \,|\, (\xi_k) \text{ bounded}\}$$
$$\|x\|_{l^\infty} = \sup_{k \in \mathbb{N}} |\xi_k|, \quad x = (\xi_n) \in l^\infty$$

(iv) The space of converging sequences c

$$X = c = \{(\xi_k) \subset \mathbb{R} \,|\, (\xi_k) \text{ is convergent}\}$$
$$\|x\|_c = \sup_{k \in \mathbb{N}} |\xi_k|, \quad x = (\xi_n) \in c$$

c can be considered a subspace of l^∞ because it is a subset of l^∞ and its norm is just a restriction of $\|\cdot\|_{l^\infty}$.

(v) The space of bounded functions $B(A)$

$$X = B(A) = \{f : A \subset \mathbb{R} \to \mathbb{R} \,|\, f \text{ bounded}\}$$
$$\|x\|_\infty = \sup_{t \in A} |f(t)|, \quad f \in B(A)$$

(vi) The space of continuous functions $C(A)$

$$X = C(A) = \{f : A \to \mathbb{R} \,|\, f \text{ continuous}\}$$
$$\|x\|_C = \max_{t \in A} |f(t)|, \quad f \in C(A)$$

(vii) Sequence spaces $l^p,\ p \geq 1$

$$X = l^p = \left\{ (\xi_n) \subset \mathbb{R} \,\middle|\, \sum_{k=1}^{\infty} |\xi_k|^p < \infty \right\}$$

$$\|x\|_{l^p} = \left(\sum_{k=1}^{\infty} |\xi_k|^p \right)^{\frac{1}{p}}, \quad x = (\xi_n) \in l^p$$

(viii) The space of Lebesgue measurable functions $L^p(A),\ p \geq 1$

$$X = L^p(A) = \left\{ f : A \to \mathbb{R} \,\middle|\, \int_A |f(t)|^p \,\mathrm{d}t < \infty \right\}$$

$$\|x\|_{L^p} = \left(\int_A |f(t)|^p \,\mathrm{d}t \right)^{\frac{1}{p}}, \quad f \in L^p(A)$$

Definition 12.1. A linear operator T is a mapping

$$T : \mathcal{D}(T) \subset X \longrightarrow Y$$

such that

(i) The domain $\mathcal{D}(T)$ is a subspace of X

(ii) $\forall x, y \in \mathcal{D}(T),\ \forall \alpha \in \mathbb{K} : \quad T(x + \alpha y) = Tx + \alpha Ty$

If $Y = \mathbb{K}$, then T is said to be a linear functional.

Example 12.2. (i) Let $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$. If $A \in \mathbb{R}^{m \times n}$ then we can define

$$Tx = Ax, \quad x \in \mathbb{R}^n$$

such that for $x = (\xi_1, \cdots, \xi_n)$ we have

$$Tx = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix} = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_m \end{pmatrix}$$

Then $\mathcal{D}(T) = \mathbb{R}^n$ and T is a linear operator.

(ii) Let $X = C([a, b])$ and $Y = C([a, b])$. Then

$$(Tx)(t) = \int_{\alpha}^{t} x(s) \,\mathrm{d}s\,, \quad t \in [a, b]$$

defines a linear operator with $\mathcal{D}(T) = C([a, b])$.

(iii) Consider $X = C([a,b])$ and $Y = C([a,b])$. We can define

$$(Tx)(t) = x'(t), \quad t \in [a,b]$$

T is a linear operator with $C([a,b]) \supset \mathcal{D}(T) = C^1([a,b])$.

(iv) Let $X = L^p([a,b])$ and $Y = L^q([a,b])$. Choose a fixed measurable function $\phi : [a,b] \to \mathbb{R}$. Then

$$(Tx)(t) = \phi(t)x(t), \quad t \in [a,b]$$

defines a linear operator. The domain in this case is

$$\mathcal{D}(T) = \left\{ x \in L^p([a,b]) \,\middle|\, \int_a^b |\phi(t)x(t)|^q \, \mathrm{d}t < \infty \right\}$$

(v) Consider $X = l^\infty$ and $Y = \mathbb{R}$. Then

$$Tx = \lim_{k \to \infty} \xi_k, \quad x = (\xi_k) \in l^\infty$$

is a linear functional with $\mathcal{D}(T) = c$.

Definition 12.3. Let $T : \mathcal{D}(T) \to Y,\ \mathcal{D}(T) \subset X$ be a linear operator. If $\exists C > 0$ such that

$$\|Tx\| \leq C\|x\|$$

then T is said to be bounded. The number

$$\|T\| = \sup_{\substack{x \in \mathcal{D}(T) \\ x \neq 0}} \frac{\|Tx\|}{\|x\|}$$

is the operator norm of T.

Example 12.4. Consider $X = Y = C([0,1])$. We can define the operator T as

$$(Tf)(t) = \int_0^t f(s) \, \mathrm{d}s \,, \quad f \in C([0,1]) = \mathcal{D}(T)$$

T is a bounded operator. This can be shown by explicitely calculating the norm

$$\begin{aligned}
\|Tf\| &= \max_{t \in [0,1]} \left| \int_0^t f(s) \, \mathrm{d}s \right| \\
&\leq \max_{t \in [0,1]} \int_0^t |f(s)| \, \mathrm{d}s \\
&\leq \max_{t \in [0,1]} \int_0^t \max_{s \in [0,1]} |f(s)| \, \mathrm{d}s \\
&= \|f\| \max_{t \in [0,1]} \int_0^t \mathrm{d}s = \|f\| \max_{t \in [0,1]} t = \|f\|
\end{aligned}$$

Thus we have shown that $\|T\| \leq 1$. We can further show that $\|T\| = 1$. To do that, assume $f = 1$. Trivially, this results in $\|f\| = 1$ and further

$$(Tf)(t) = \int_0^t 1 \,\mathrm{d}s = t$$

This gives us

$$\|Tf\| = 1 \implies \|T\| \geq \frac{\|Tf\|}{\|f\|} = 1$$

This implies $\|T\| = 1$.

Example 12.5. Again, consider $X = Y = C([0,1])$. This time we look at

$$(Tf)(t) = f'(t), \quad \mathcal{D}(T) = C^1([0,1])$$

T is an unbounded operator. To prove this take $f_n(t) = t^n \in C([0,1]),\ n \geq 1$. We compute

$$\|f_n\| = \max_{t\in[0,1]} |t^n| = 1, \quad \|Tf_n\| = \max_{t\in[0,1]} \left|nt^{n-1}\right| = n$$

Then

$$\|T\| \geq \frac{\|Tf_n\|}{\|f_n\|} = n, \quad \forall n \geq 1$$

So there doesn't exist a $C > 0$ such that $n \leq C$, thus T cannot be bounded.

Theorem 12.6. *Let X be a finite-dimensional normed space. If T is a linear operator on X, then T is bounded.*

Proof. Without proof. □

Definition 12.7. Let $T : \mathcal{D}(T) \to Y$ be a linear operator. T is said to be continuous in $x_0 \in \mathcal{D}(T)$ if

$$\forall \epsilon > 0,\ \exists \delta > 0 : \quad \|x - x_0\| < \delta \implies \|Tx - Tx_0\| < \epsilon, \quad \forall x \in \mathcal{D}(T)$$

Theorem 12.8. *Let $T : \mathcal{D}(T) \to Y$ be a linear operator. Then*

(i) T is continuous $\iff$ T is bounded

(ii) If T is continuous in a single point, then it is continuous everwhere

Proof. To prove the first statement, we want to consider $T \neq 0$ (since $T = 0$ is trivial). This implies that $\|T\| \neq 0$. Assume T is bounded, and take $x_0 \in \mathcal{D}(T)$. Now let $\epsilon > 0$ and $\delta = \frac{\epsilon}{\|T\|}$ such that $\|x - x_0\| < \delta,\ x \in \mathcal{D}(T)$. Then

$$\|Tx - Tx_0\| = \|T(x - x_0)\| \leq \|T\|\|x - x_0\| < \|T\|\delta = \|T\|\frac{\epsilon}{\|T\|} = \epsilon \tag{12.1}$$

Thus proving that T is continuous. Now inversely, let T be continuous in $x_0 \in \mathcal{D}(T)$. If we choose $\epsilon = 1$, then we can find a δ such that

$$\|x - x_0\| < \delta \implies \|Tx - Tx_0\| < \epsilon = 1 \tag{12.2}$$

If we now take any $y \neq 0$ from $\mathcal{D}(T)$ and set $x = x_0 + \frac{\delta}{2\|y\|}y$, then we can show

$$\|x - x_0\| = \frac{\delta}{2} < \delta \implies \|Tx - Tx_0\| < \epsilon = 1 \tag{12.3}$$

Therefore we have

$$1 > \|Tx - Tx_0\| = \|T(x - x_0)\| = \left\|T\frac{\delta}{2\|y\|}y\right\| = \frac{\delta}{2\|y\|}\|Ty\| \tag{12.4}$$

Thus

$$\frac{\delta}{2\|y\|}\|Ty\| < 1 \implies \|Ty\| < \frac{2}{\delta}\|y\| \tag{12.5}$$

Since $y \in \mathcal{D}(T)$ was chosen arbitrarily, this implies that T is bounded. The second statement follows trivially from the first one, as we have shown that if T is continuous in one point x_0, it is bounded and if it is bounded then it is continuous everywhere. □

Corollary 12.9. *Let T be a bounded linear operator. Then*

(i) For $x_n, x \in \mathcal{D}(T)$ we have $x_n \to x \implies Tx_n \to Tx$

(ii) The set $\ker(T) = \{x \in \mathcal{D}(T) \,|\, Tx = 0\}$ is a null set and closed in X

Proof. Left as an exercise for the reader. □

Theorem 12.10. *Let $T : \mathcal{D}(T) \to Y$ be a bounded linear operator, with Y a Banach space. Then T has an extension $\tilde{T} : \overline{\mathcal{D}(T)} \to Y$ where $\tilde{T}$ is a bounded linear operator and $\|\tilde{T}\| = \|T\|$.*

Proof. In this proof we only want to show how such a $\tilde{T}$ can be constructed. Let $x \in \overline{\mathcal{D}(T)}$. Then there is a sequence $x_n \in \mathcal{D}(T)$ such that $x_n \to x$. Since T is linear and bounded, we can find

$$\|Tx_n - Tx_m\| \leq \|T(x_n - x_m)\| \leq \|T\|\|x_n - x_m\| \xrightarrow{n,m\to\infty} 0 \tag{12.6}$$

So $(Tx_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in Y. Because Y is a Banach space there exists some $y \in Y$ such that Tx_n converges to y. Now we define $\tilde{T}x := y$, and show that $\tilde{T}x$ is well-defined. If $(z_n)_{n\in\mathbb{N}} \subset \mathcal{D}(T)$ is another sequence converging to x, then $Tz_n \to y'$. Now consider the sequence

$$(v_n)_{n\in\mathbb{N}} = (x_1, z_1, x_2, z_2, x_3, z_3, \cdots) \tag{12.7}$$

This sequence also converges to x, and $Tv_n \to y''$. However we can also find

$$Tv_{2k+1} \longrightarrow y = y'' \qquad\qquad Tv_{2k} \longrightarrow y' = y'' \tag{12.8}$$

Thus $y = y'$. □

12.2 Dual Spaces

Definition 12.11 (Normed spaces of Operators)**.** Let X, Y be normed spaces and $T : X \to Y$ a bounded linear operator. Then $B(X, Y)$ is the set of all such bounded linear operators. If we define for $x \in X,\ \alpha \in \mathbb{K}$

$$\begin{aligned} (T_1 + T_2)(x) &= T_1 x + T_2 x, && T_1, T_2 \in B(X,Y) \\ (\alpha T)(x) &= \alpha T x, && T \in B(X,Y) \end{aligned}$$

then $B(X, Y)$ is a vector space.

Theorem 12.12. *The vector space $B(X, Y)$ is a normed space with the operator norm*

$$\|T\| = \sup_{x \neq 0} \frac{\|Tx\|}{\|x\|} = \sup_{\|x\|=1} \|Tx\|$$

Proof. Left as an exercise for the reader. □

Theorem 12.13. *If Y is a Banach space, then $B(X, Y)$ is also a Banach space.*

Proof. Let $(T_n) \subset B(X, Y)$ be a Cauchy sequence. We need to show that there exists some $T \in B(X, Y)$ such that $T_n \to T$. Let $x \in X$ and define

$$Tx = \lim_{n\to\infty} T_n x \tag{12.9}$$

Consider the sequence $T_n x$. It is possible to show that this is a Cauchy sequence

$$\|T_n x - T_m x\| = \|(T_n - T_m)x\| = \|T_n - T_m\|\|x\| \xrightarrow{n,m\to\infty} 0 \tag{12.10}$$

Since Y is complete, there exists some $y \in Y$ such that $T_n x \to y := Tx$. Thus we have shown that T is indeed mapping X to Y. We now need to prove that T is linear and bounded (and thus element of $B(X, Y)$).

$$\begin{aligned} T(\alpha x + \beta z) &= \lim_{n\to\infty} T_n(\alpha x + \beta z) \\ &= \lim_{n\to\infty} (\alpha T_n x + \beta T_n z) \\ &= \alpha \lim_{n\to\infty} T_n x + \beta \lim_{n\to\infty} T_n z = \alpha T x + \beta T z \end{aligned} \tag{12.11}$$

This shows that T is linear. Now let $\epsilon > 0$. Then

$$\exists N \in \mathbb{N} : \quad \|T_n - T_m\| < \frac{\epsilon}{2}, \quad \forall n, m \geq N \tag{12.12}$$

If we let $n \geq N$ we can use this to show

$$\begin{aligned} \|T_n x - Tx\| &= \left\|T_n x - \lim_{m\to\infty} T_m x\right\| \\ &= \lim_{m\to\infty} \|T_n x - T_m x\| \\ &\leq \lim_{m\to\infty} \|T_n - T_m\|\|x\| \leq \frac{\epsilon}{2}\|x\| < \epsilon\|x\| \end{aligned} \tag{12.13}$$

Thus showing that T is bounded. This also implies that $T_n \to T$, proving that $B(X, Y)$ is a Banach space. $\square$

Definition 12.14 (Dual Spaces)**.** The set of all bounded linear functionals $f : X \to \mathbb{K}$ with the norm

$$\|f\| = \sup_{x \neq 0} \frac{|f(x)|}{\|x\|} = \sup_{\|x\|=1} |f(x)|$$

is said to be the dual space of X, and is written as $X' = B(X, \mathbb{K})$.

Theorem 12.15. *The dual space X' of a normed space X is a Banach space.*

Proof. Without proof. $\square$

Definition 12.16. Let $X, \tilde{X}$ be normed spaces. A bijective linear operator $T : X \to \tilde{X}$ that perserves the norm (i.e. $\|Tx\| = \|x\|, \ \forall x \in X$) is said to be an isomorphism. If such an isomorphism exists, then X and $\tilde{X}$ are called isomorphic normed spaces.

Example 12.17. (i) The dual space of l_n^p is

$$(l_n^p)' = l_n^q, \quad \frac{1}{p} + \frac{1}{q} = 1, \ 1 < p < \infty$$

So let $f \in (l_n^p)'$ be a bounded linear functional. We can define a basis

$$\left.\begin{aligned} e_1 &= (1, 0, \cdots, 0) \\ &\vdots \\ e_n &= (0, \cdots, 0, 1) \end{aligned}\right\} \in l_n^p$$

This lets us express elements of l_n^p in the following way

$$x = \sum_{k=1}^{n} \xi_k e_k \in l_n^p$$

We can then write out f as

$$f(x) = f\left(\sum_{k=1}^{n} \xi_k e_k\right) = \sum_{k=1}^{n} \xi_k f(e_k) = \sum_{k=1}^{n} \gamma_k \xi_k = \langle u | x \rangle$$

where $u = (\gamma_1, \cdots, \gamma_n), \ \gamma_k = f(e_k), \ k = 1, \cdots, n$. To compute the norm of f we want to use Hölder's inequality

$$\sum_{k=1}^{n} |x_k y_k| \leq \left(\sum_{k=1}^{n} |x_k|^q\right)^{\frac{1}{q}} \left(\sum_{k=1}^{n} |y_k|^p\right)^{\frac{1}{p}}, \quad \frac{1}{p} + \frac{1}{q} = 1$$

With this we can write

$$|f(x)| = \left|\sum_{k=1}^{n} \gamma_k \xi_k\right| \le \sum_{k=1}^{n} |\gamma_k \xi_k| \le \left(\sum_{k=1}^{n} |\gamma_k|^q\right)^{\frac{1}{q}} \left(\sum_{k=1}^{n} |\xi_k|^p\right)^{\frac{1}{p}} = \|u\|_{l^q} \|x\|_{l^p}, \quad \forall x \in l_n^p$$

This implies $\|f\| \le \|u\|_{l^q}$. Now let $x = (\pm|\gamma_1|^{q-1}, \cdots, \pm|\gamma_n|^{q-1})$ where we use $+$ is $\gamma_k \ge 0$, and $-$ otherwise. Then

$$|f(x)| = \sum_{k=1}^{n} \gamma_k(\pm|\gamma_k|^{q-1}) = \sum_{k=1}^{n} |\gamma_k|^q$$

and

$$\|x\|_{l^p} = \left(\sum_{k=1}^{n} |\gamma_k|^{(q-1)p}\right)^{\frac{1}{p}} = \left(\sum_{k=1}^{n} |\gamma_k|^q\right)^{1-\frac{1}{q}}$$

Using these two steps we can write

$$|f(x)| = \sum_{k=1}^{n} |\gamma_k|^q = \left(\sum_{k=1}^{n} |\gamma_k|^q\right)^{\frac{1}{q}} \left(\sum_{k=1}^{n} |\gamma_k|^q\right)^{1-\frac{1}{q}} = \|u\|_{l^q} \|x\|_{l^p}$$

thus proving $\|f\| = \|u\|$. As a result, this shows that f is an isomorphism of $(l_n^p)'$ to l_n^q. In other words, any bounded linear function f canm be written as

$$f(x) = \sum_{k=1}^{n} \gamma_k \xi_k =: \langle u|x\rangle\,, \quad u = (\gamma_n) \subset l_n^p$$

(ii) $(l_n^1)' = l_n^\infty$ and $(l_n^\infty)' = l_n^1$

(iii) $(l^p)' = l^q, \quad \frac{1}{p} + \frac{1}{q} = 1$

(iv) $(l^1)' = l^\infty$

(v) $c' = (c_0)' = l^1$

(vi) $(L^p(A))' = L^q(A)$ and $(L^1(A))' = L^\infty(A)$

(vii) $(C(A))' =$ "functions of bounded variation"

Definition 12.18. Let $w : [a,b] \to \mathbb{R}$ be a function. w is said to be of bounded variation on $[a,b]$ if its total variation

$$\mathrm{Var}(w) = \sup \sum_{j=1}^{n} |w(t_j) - w(t_{j-1})|$$

is finite. The supremum is taken over all partitions $a = t_0 < t_1 < \cdots < t_n = b$.

Example 12.19. If w is non-decreasing, then w has bounded variation. This can be explicitly shown

$$\mathrm{Var}(w) = \sup \sum_{j=1}^{n} |w(t_j) - w(t_{j-1})| = \sup \sum_{j=1}^{n} (w(t_j) - w(t_{j-1})) = w(b) - w(a)$$

Remark 12.20. A functio w has bounded variation if it can be written as a difference of two non-decreasing functions. I.e. $\exists w_1, w_2 : [a, b] \to \mathbb{R}$ non-decreasing, such that $w = w_1 - w_2$.

Lemma 12.21. *Let $BV([a, b])$ be the set of all functions on $[a, b]$ that have bounded variation. It is obvious that $BV([a, b])$ is a vector space over $\mathbb{R}$, if we define the norm*

$$\|w\| = |w(a)| + \mathrm{Var}(w), \quad w \in BV([a, b])$$

$BV([a, b])$ is a Banach space

Remark 12.22. Let $x \in C([a, b])$ and $w \in BV([a, b])$. Then one can see that the Riemann-Stieltjes integral

$$\int_a^b \mathrm{d}w(t) = \lim_{\lambda \to \infty} \sum_{k=1}^{n} x(\xi_k)(w(t_k) - w(t_{k-1}))$$

exists, with $\lambda = \max |t_k - t_{k-1}|,\ \xi_k \in [t_{k-1}, t_k]$. If $w \in C^1([a, b])$ then

$$\int_a^b x(t)\,\mathrm{d}w(t) = \int_a^b x(t)w'(t)\,\mathrm{d}t$$

Theorem 12.23. *Every $f \in (C([a, b]))'$ can be expressed as a Riemann-Stieltjes integral*

$$f(x) = \int_a^b x(t)\,\mathrm{d}w(t)$$

with $\|f\| = \mathrm{Var}(w)$