Analysis II

LIMITS, CONTINUITY, DIFFERENTIABILITY, AND INTEGRABILITY

CHRISTIAN P. JÄH

SEPTEMBER 28, 2017

Foreword

These note have been written primarily for you, the student. I have tried to make it easy to read and easy to follow.

I do not wish to imply, however, that you will be able to read this text as it were a novel. If you wish to derive any benefit from it, each page slowly and carefully. You must have a pencil and plenty of paper beside you so that you yourself can reproduce each step and equation in an argument. When I say *verify a statement*, *make a substitution*, etc. pp., you yourself must actually perform these operations. If you carry out the explicit and detailed instructions I have given you in remarks, the text, and proofs, I can almost guarantee that you will, with relative ease, reach the conclusions.

One final suggestion. As you come across formulas, record them and their equations/page numbers on a separate sheet of paper for easy reference. You may also find it advantageous to do the same for Definitions and Theorems.

These wise words are borrowed from Morris Tenenbaum and Harry Pollard from the beginning of their book *Ordinary differential equations*. I could not have said it better and it certainly applies to this course.

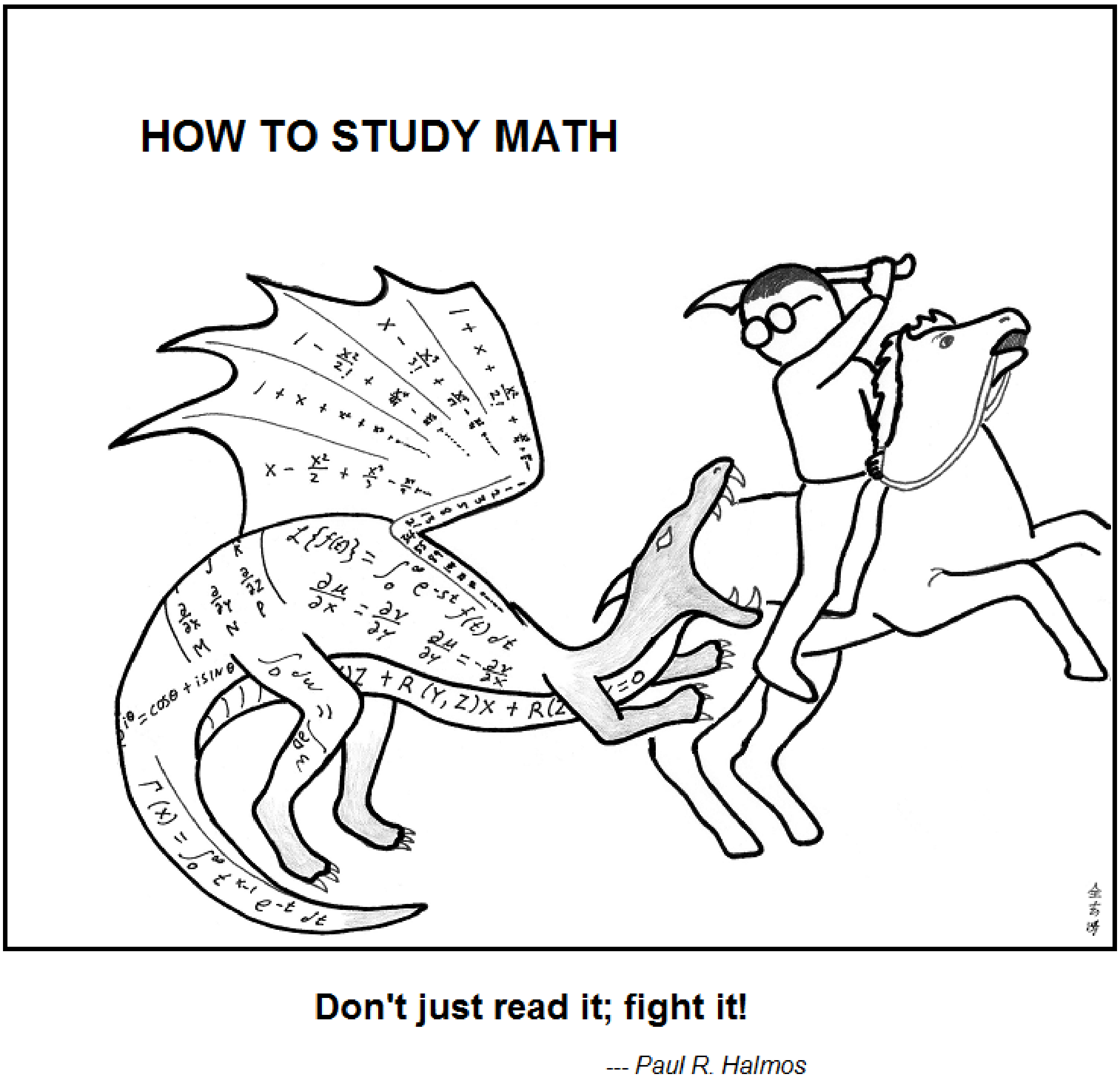


Figure 0.1: Don’t just read it; fight it. – Paul Halmos (The comic is abstrusegoose.com)

Contents

List of symbols

Here I collect a couple of symbols used in the text and reference their definition for you to look up. There are more symbols in Section 1.1.

$\mathbb{R}^{\mathbb{N}_0}$	This denotes the set of all sequences $(a_n)_{n\in\mathbb{N}_0}$ with $a_n \in \mathbb{R}$ for all $n \in \mathbb{N}_0$. $(a_n)_{n\in\mathbb{N}_0} \in \mathbb{R}^{\mathbb{N}_0}$ is equivalent to say $(a_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$. The point of this notation is to stress that (a_n) is a point in a vector space.
$\Omega^{\mathbb{N}_0}$	This denotes the set of all sequences $(a_n)_{n\in\mathbb{N}_0}$ with $a_n \in \Omega$ for all $n \in \mathbb{N}_0$. See also Definition 1.3.
$\lvert\cdot\rvert$	Absolute value of a real number. See Section 2.
$\lVert\cdot\rVert_X$	Is a norm on a linear space X. Some norms have special abbreviations as $\lVert\cdot\rVert_{C^0} = \lVert\cdot\rVert_\infty$. Some norms on $\mathbb{R}^n$ have special notation too. See Sections 2 and 5.5.
$C(\Omega)$	The set of continuous functions $f:\Omega\to\mathbb{R}$, $\Omega\subseteq\mathbb{R}^n$. See also Section 5.5.
$\frac{d}{dx}$	Differentiation operator with respect to x for functions $f:(a,b)\to\mathbb{R}$. See also Section 5.
$\frac{\partial}{\partial x_i}$	Differentiation operator with respect to x_i for functions $f:\Omega\to\mathbb{R}$ with $\Omega\subseteq\mathbb{R}^n$. The ∂ is used to indicate that the function depend on more than one variable. It is called the partial derivative with respect to x_i. See also Section 6.1.2.
$f_n \to f$	If the f_n are functions then this symbol indicates that the sequence converges in a sense to f. It should additional be indicated whether the convergence is pointwise (Definition 8.1) or uniform (Definition 8.2).
$f_n \rightrightarrows f$	This symbol usually means that the sequence $(f_n)_{n\in\mathbb{N}_0}$ converges uniformly to f. In this notes we indicate uniform convergence by specifically stating it. For the definition of uniform convergence see Definition 8.2.
$\mathrm{cl}(\Omega)$	Let $\Omega\subseteq\mathbb{R}^n$ be a set. Then, $\mathrm{cl}(\Omega)$ denotes the closure of the set Ω, i.e. it is the set of all points of Ω union with all limit points of Ω in $\mathbb{R}^n$. See Definition 2.8.
$\mathrm{Sub}(V)$	The set $\mathrm{Sub}(V)$ is the collection of all subs-spaces of the (real) vector space V.

List of (named) theorems

This list is not complete and by no means all theorems you need to know. These are the most important theorems of the class which you should be able to cite with assumptions and conclusions at all times. You should also be familiar with the main ideas of the proofs.

List of important definitions

This list is not a complete list of all definitions you need to know. It is simply a help to quickly navigate around the notes. The text also has an index and some work needs to be done by the reader in preparation of tests and exams.

Glossary

Please find some more here.

Ansatz. An *ansatz* is an assumption about the form of an unknown function which is made in order to facilitate solution of an equation or other problem. Example: find an example fo a function with two extrema. A suitable ansatz is then

$$f'(x) = a(x - x_0)(x - x_1),$$

where x_0 and x_1 are the points in which you want the extrema to be. Another example is partial fractions.

Convexity. See Section 1.5.4. A function is said to be convex on an interval $I = [a, b]$ iff

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

for all $x,\ y \in [a, b]$ and all $\lambda \in [0, 1]$. Another criterion is that the second derivative f'' is non-negative ($f''(x) \geq 0$ for $x \in I$) on the interval. Sometimes, these functions are called *convex downward* or *concave upward*. However, the latter names are uncommon in the academic literature and will not be used in this course. A function f is called concave, if $-f$ is convex. Another way to see convexity is to check weather the graph is always under any secant line one can draw over a given domain. If the function is always above, it ic concave.

Domain. The domain of a function is the set of input values for which the function is defined. The largest possible set of such input values for which a function can be defined is called the natural domain. Example: Let $f : [0, 1] \to \mathbb{R}$ be defined by $f(x) = \sqrt{x}$. The interval $[0, 1]$ is the domain of f. However, given the function $f(x) = \sqrt{x}$, the natural domain is $[0, +\infty)$ since the square root ($\sqrt{\cdot}$) makes sense for all non-negative real numbers.

Extrema. (pl.) The maxima and minima (also pl.) of a function are collectively known as extrema.[1] Sometimes these points are called *turning points*. However, the latter name bears the possibility of being confused with inflexion points.

Function. A functions is a mathematical relationship consisting of a rule linking elements from two sets such that each element from the first set (the domain) links to one and only one element from the second set (the image set or range).

[1] The singular form is extremum. It could be a maximum or minimum.

Graph. Given a function $f : \mathrm{dom}(f) \to \mathbb{R}^m$, then the graph is the set

$$\{(x, f(x)) : x \in \mathrm{dom}(f)\} \subseteq \mathbb{R}^{n+m}.$$

If $n = m = 1$, the graph can also be represented by a picture in a xy-plane showing the curve $y = f(x)$, where x goes through (a part of) the domain.

Image. The image of a function f is the collection of all values that a function can take when the argument goes through the domain of f, i.e. the set

$$\{f(x) : x \text{ in the domain of } f\}.$$

Inflexion point. Inflexion points are the points at which a function changes from convex to concave or from concave to convex. These can be found as the sign changing zeros of f''. Remember that the second derivative may vanish at a point without changing sign. An easy example is $f(x) = x^4$. (Show that!)

Secant. In geometry, a secant of a curve is a line that (locally) intersects two points on the curve.

Stationary point. Given a function f, the stationary points of $f : I \subseteq \mathbb{R} \to \mathbb{R}$ are the points in the domain of f at which $f'(x) = 0$. If the function depends on several variables, i.e. $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$, the stationary points are points in the domain of f for which $\nabla f(x) = 0$.

Such that. A condition used in the definition of a mathematical object, commonly denoted : or |. For example, the rational numbers $\mathbb{Q}$ can be defined as

$$\mathbb{Q} = \{\frac{p}{q} : q \neq 0, p, q \in \mathbb{Z}\}.$$

In sentences, *such that* is sometimes abbreviated by *s.t.*.

CHAPTER

1 Prerequisites

1.1 Some notation used in this notes

Symbols handwritten vs. typed. I tend to use the the hash (#) to indicate the end of a proof when I am writing something by hand. In this notes, the end of a proof will usually be indicated by a □. If there is a short "proof" in a remark, no indication of its end will be given as it is understood that it should be clear. In these notes, I will not use any symbol to indicate contradictions in proofs by contradiction but simply state that we have reached one. In the notes that I make by hand in the lecture, I will mostly use a lightening bolt.

The Greek alphabet. I assume that everyone is familiar with the Greek alphabet and knows how to write the letters:

α	alpha	θ	theta	o	omikron	τ	tau
β	beta	ϑ	theta	π	pi	υ	upsilon
γ	gamma	γ	gamma	ϖ	pi	ϕ	phi
δ	delta	κ	kappa	ρ	rho	φ	phi
ϵ	epsilon	λ	lambda	ϱ	rho	χ	chi
ε	epsilon	μ	mu	σ	sigma	ψ	psi
ζ	zeta	ν	nu	ς	sigma	ω	omega
η	eta	ξ	xi				
Γ	Gamma	Λ	Lambda	Σ	Sigma	Ψ	Psi
Δ	Delta	Ξ	Xi	Υ	Upsilon	Ω	Omega
Θ	Theta	Π	Pi	Φ	Phi		

Table 1.1: Greek Letters

Some more symbols. I assume that you are familiar with the meaning of some symbols described below.

Symbol	Description
$\mathbb{R}$	real numbers
$\mathbb{Z}$	whole numbers, i.e. $\{\dots,-2,-1,0,1,2,\dots\}$
$\mathbb{N}$	natural numbers, i.e. $\{1,2,3,4,\dots\}$
$\mathbb{N}_0$	natural numbers containing 0
$\mathbb{Q}$	rational numbers
	complex numbers

Table 1.2: Notation of certain sets.

We have the following inclusions

$$\mathbb{N} \subset \mathbb{N}_0 \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset .$$

An important class of sets are subsets of the real numbers, called intervals. An interval is a set of numbers characterized by their left and right "boundary". For example

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

which we read as the closed interval a, b. Closed means that it contains a and b. An open interval does not contain the boundary points, i.e.

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

One can also consider the half-open cases

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

and

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

We also denote the real numbers $\mathbb{R}$ by $(-\infty, +\infty)$ at times. All numbers smaller than a would be denoted by $(-\infty, a)$, all numbers smaller or equal to a by $(-\infty, a]$. Similarly, one defines the sets of all numbers larger that or larger or equal to a given number. If we have the situation that we describe x as having either the property $x \geq a$ or $x \leq -a$ for a given $a \geq 0$, then we can write

$$\{x \in \mathbb{R} : \quad x \geq a \quad \text{or} \quad x \leq a\}$$

which is the same as

$$x \in (-\infty, -a] \cup [a, +\infty).$$

1.1.1 Operations on sets

Sets are collections of elements described by some property P. The standard notation for sets is

$$A = \{x : x \, \text{has property} \, P\},$$

where one reads: A consists of all x such that x has property P.

Definition 1.1 (Intersection/Union/Difference)**.**
We denote by $A \cap B$ the intersection of A and B which means that $A \cap B$ contains elements that are in A as well as in B. By $A \cup B$, we denote the union of the two sets A and B which means that $A \cup B$ contains elements that are either in A or in B. With $A \setminus B$, we denote finally the difference of A and B that means that $A \setminus B$ contains all elements in A that are not in B.

Remark 1.1. *Of course the intersection and union is not limited to a finite number. If one has a family of sets $\{A_i : i \in I\}$ indexed by a countable or uncountable set I one can consider the sets $\bigcap_{i \in I} A_i$ and $\bigcup_{i \in I} A_i$. For Example:*

$$\mathbb{R} = \bigcup_{n \in \mathbb{N}} [-n, n], \qquad \{0\} = \bigcap_{n \in \mathbb{N}} \left[-\frac{1}{n}, \frac{1}{n}\right].$$

Definition 1.2 (Subset $A \subseteq B$).
Let A and B be sets. Then, we say that A is a sub-set of B, in symbols, $A \subseteq B$ iff

$$\forall\, x \in A \quad \Rightarrow \quad x \in B.$$

We write $A \subset B$ if we want to signal that A is a proper subset of B, i.e. there are elements in B that do not belong to A.

Definition 1.3 (The set A^B).
Let A and B be two sets. The, we denote the set of all functions $f : B \to A$ as A^B.

Example 1.1. *The set $\mathbb{R}^{\mathbb{R}}$ are all real functions defined on the real line. The set $\mathbb{R}^{\mathbb{N}_0}$ is the set of all real sequences seen as functions from $\mathbb{N}_0$ to $\mathbb{R}$. With the latter we can write $(a_n)_{n\in\mathbb{N}_0} \in \mathbb{R}^{\mathbb{N}_0}$ and vice versa.*

1.2 Some Linear Algebra (of real vector spaces)

We recall the following definition from Linear Algebra. If you do not remember it very clearly, please consult your Linear Algebra notes too.

Definition 1.4 (Vector space).
A real vector space is a set V together with two operations $+ : V \times V \to V$ satisfying $(A1)$ to $(A4)$ and $\cdot : \mathbb{R} \times V \to V$ satisfying $(A5)$ to $(A8)$. The conditions $(A1)$ to $(A4)$ are

(A1) There exists an element $0 \in V$ such that $v + 0 = v$ for all $v \in V$.

(A2) For every $v \in V$ there exists an element $-v \in V$ such that $v + (-v) = 0$.

(A3) For all $u,\ v,\ w \in V$ holds $u + (v + w) = (u + v) + w$.

(A4) For all $u,\ v \in V$ holds $u + v = v + u$.

The conditions $(A5)$ to $(A8)$ are

(A5) For all $v \in V$ holds $1 \cdot = v$, where 1 is the multiplicative identity of $\mathbb{R}$.

(A5) For all $v \in V$ and $\alpha,\ \beta \in \mathbb{R}$ holds $\alpha(\beta \cdot v) = (\alpha\beta) \cdot v$.

(A5) For all $u,\ v \in V$ and $\alpha \in \mathbb{R}$ holds $\alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v$.

(A5) For all $v \in V$ and $\alpha,\ \beta \in \mathbb{R}$ holds $(\alpha + \beta) \cdot v = \alpha \cdot v + \beta \cdot v$.

If we want to emphasise that V is a real vector space, we write $(V, \mathbb{R})$ and if we would like to emphasise the operations as well, we write $(V, \mathbb{R}, +, \cdot)$.

Exercise 1.1. *Convince yourself that $(\mathbb{R}^n, \mathbb{R})$ with*

$$\begin{aligned} &+ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n, \\ &\left(\left[x_1, x_2 \ldots, x_n\right]^T, \left[y_1, y_2 \ldots, y_n\right]^T\right) \mapsto \left[x_1 + y_1, x_2 + y_2 \ldots, x_n + y_n\right] \end{aligned}$$

and

$$\begin{aligned}&\cdot : \mathbb{R}\times\mathbb{R}^n \to \mathbb{R}^n\\ &\left(\lambda, \left[x_1, x_2 \ldots, x_n\right]^T\right) \mapsto \left[\lambda x_1, \lambda x_2 \ldots, \lambda x_n\right]^T\end{aligned}$$

is a real vector space in the sense of Definition 1.4.

Exercise 1.2. *Convince yourself that* $(\mathbb{R}^{\mathbb{N}_0}, \mathbb{R})$ *(see Definition 1.3) with*

$$\begin{aligned}&+ : \mathbb{R}^{\mathbb{N}_0} \times \mathbb{R}^{\mathbb{N}_0} \to \mathbb{R}^{\mathbb{N}_0},\\ &\big((a_n), (b_n)\big) \mapsto (a_n + b_n)\end{aligned}$$

and

$$\begin{aligned}&\cdot : \mathbb{R}\times\mathbb{R}^{\mathbb{N}_0} \to \mathbb{R}^{\mathbb{N}_0}\\ &\big(\lambda, (a_n)\big) \mapsto (\lambda a_n)\end{aligned}$$

is a real vector space in the sense of Definition 1.4. Thus, sequences can be seen as points *in a vector space and we can try to apply intuitive geometric reasoning to this setting.*

Remark 1.2. *The space* $(\mathbb{R}^{\mathbb{N}_0}, \mathbb{R})$ *in the example above is a neat example of an infinite dimensional vector space. For further details on dimension compare your Linear Algebra I notes.*

The following remark is for further study and understanding and not examinable.

Remark 1.3. *The special choice of the field* $\mathbb{R}$ *in Definition 1.4 is only for the purposes of this lecture. It can be replaced by any field* $\mathbb{F}$. *If the notion is familiar to you, you might recognize that* $(V, +)$ *is a group. If you are familiar with these terms you can try to think about* $\mathbb{R}$ *as a vector space over* $\mathbb{Q}$, *i.e. you replace* V *by* $\mathbb{R}$ *and* $\mathbb{R}$ *by* $\mathbb{Q}$ *in the above definition. In this course, we will be concerned with real vector spaces only.*

Definition 1.5 (Scalar product (inner product))**.**
Let V *be a real vector space. A function* $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ *will be called a scalar product if it satisfies the following conditions:*

- *For all* $v \in \mathbb{R}^n$, *we have* $\langle v, v\rangle \geq 0$ *and* $\langle v, v\rangle = 0$ *iff* $v = 0$. *(Positive definiteness)*
- *For all* u, $v \in V$, *we have* $\langle u, v\rangle = \langle v, u\rangle$. *(Symmetry)*
- *For all* u, v, *and* $w \in V$, *and* $\alpha, \beta \in \mathbb{R}$ *we have*

$$\langle u, \alpha v + \beta w\rangle = \alpha\langle u, v\rangle + \beta\langle u, v\rangle. \quad \textit{(Linearity)}$$

Remark 1.4. *In the above definition, tanking property* (ii) *and* (iii) *together, we obtain For all* u, v, *and* $w \in V$, *we have* $\langle \alpha u + \beta v, w\rangle = \alpha\langle u, w\rangle + \beta\langle v, w\rangle$ *as well.*

Example 1.2. *The most important example in our context is*

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i. \tag{1.1}$$

This scalar product induces the Euclid norm

$$\|x\| = \sqrt{\langle x, x\rangle} = \sqrt{\sum_{i=1}^{n} |x_i|^2} = \|x\|_2.$$

This scalar product in, especially in applied mathematics, often denoted by a thick dot and called dot-product: $\langle x|y\rangle = x \cdot y = x^T y$. Since (1.1) is not the only possibility, we introduced $\langle\cdot,\cdot\rangle$ to symbolise a scalar product.

The following remark is for further study and understanding and not examinable.

Remark 1.5. *There are many scalar products that one can establish on $\mathbb{R}^n$. An example for $n = 2$ is*

$$\langle x, y\rangle = x_1 y_1 + 2x_2 y_2 - (x_1 y_2 + y_1 x_2)$$

for all $x, y \in \mathbb{R}^2$. Use the above definition to compute the scalar product $\langle x, y\rangle$ for

$$x = \begin{bmatrix}1, 2\end{bmatrix}^T, \quad y = \begin{bmatrix}-3, 1\end{bmatrix}^T.$$

Proposition 1.1 (Scalar products induce norms)**.**
Let V be a real vector space with scalar product $\langle\cdot|\cdot\rangle$. Then, the scalar product induces a norm by

$$\|x\| = \sqrt{\langle x, x\rangle}.$$

Exercise 1.3. *Use the definition of a scalar product to prove that $\|x\| = \sqrt{\langle x, x\rangle}$ is a norm as stated in the last proposition. For the definition of a norm, 2.1.*

Remark 1.6. *It should be remarked that not all norms on $\mathbb{R}^n$ (or any vector space) are induced by a scalar product. For example, there is no scalar product on $\mathbb{R}^n$ that induces any of the $\|\cdot\|_p$ norms for $p \neq 2$. For the definition of the latter see Example 2.3.*

Remark 1.7. *Scalar products satisfy a very important property for Analysis, the so-called Cauchy-Schwarz inequality. See Theorem 1.7. This inequality says that*

$$|\langle v, w\rangle| \leq \sqrt{\langle v, v\rangle}\sqrt{\langle w, w\rangle}.$$

The Cauchy-Schwarz inequality implies

$$-1 \leq \frac{\langle v, w\rangle}{|v||w|} \leq 1$$

for $v, w \in \mathbb{R}^n \setminus \{0\}$. This allows us to introduce angles $\angle(v, w)$ between vectors v, w by

$$\cos(\angle(v, w)) = \frac{\langle v, w\rangle}{|v||w|}.$$

1.2.1 Basis and dimension

Definition 1.6 (Linear combination)**.**
Let V be a real vector space and $\{v_1, \ldots, v_k\} \subset V$ we call any combination of the type

$$\alpha_1 v_1 + \ldots \alpha_k v_k = \sum_{i=1}^{k} \alpha_i v_i,$$

where the $\alpha_i \in \mathbb{R}$, $i = 1, \dots, k$ a linear combination of $\{v_1, \dots, v_k\}$.

Definition 1.7 (Linear dependence)**.**
A set $\{v_1, \dots, v_k\} \subset V$ is called linearly dependent if it is possible to write

$$0 = \sum_{i=1}^{k} \alpha_i v_i, \tag{1.2}$$

where not all α_i are equal to zero. If (1.2) is only possible $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$, the set is called linearly independent.

Definition 1.8 (Maximal independent set)**.**
A set $B \subseteq V$ is a maximal independent set if $B \cup \{v\}$, for every $v \in V \setminus B$ is a linearly dependent set.

Definition 1.9 (Dimension)**.**
Let V be a vector space and $B \subseteq V$ be a maximally independent set. Then, the dimension of V, is defined to be $|B|$, in symbols: $\dim(V) = |B|$.

Definition 1.10 (Basis)**.**
Let V be a real vector vector space. A set $\mathcal{B} \subseteq V$ is called a basis if it is a maximally linearly independent set.

Definition 1.11 (Linear hull)**.**
Let V be a real vector space and $U \subseteq V$ be a subset. Then, the we denote by $\operatorname{span}(U)$ the set of all finite linear combinations of elements in U. We say $\operatorname{span}(U)$ is the span of U or the linear hull of U.

Proposition 1.2.
Let V be a real vector space and $\mathcal{B} \subseteq V$ be a basis. Then $\operatorname{span}(\mathcal{B}) = V$.

Example 1.3 (Canonical/standard basis)**.** *We define*

$$\delta_{ij} = \begin{cases} 1 & : \quad i = j \\ 0 & : \quad i \neq j \end{cases} .$$

This function is refereed to as Kronecker-δ. We then set

$$e_i = \begin{bmatrix} \delta_{i1} \\ \delta_{i2} \\ \vdots \\ \delta_{in} \end{bmatrix}, \quad i \in \{1, \dots n\}.$$

Then, the set $\{e_i : ni \in \{1, \dots, n\}\}$ is a basis of $\mathbb{R}^n$ and usually referred to as the standard basis or canonical basis of $\mathbb{R}^n$. Since it is a basis, we have for all $x \in \mathbb{R}^n$ that

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n} x_i e_i.$$

1.2.2 Linear maps

Definition 1.12 (Linear map)**.**
Let U and V be two real vector spaces. Then, a map $L : U \to V$ is called linear iff

(i) $L(\lambda u) = \lambda L u$ *for all* $\lambda \in \mathbb{R}$ *and* $u \in U$*, and*

(ii) $L(u + v) = Lu + Lv$ *for all* $u, v \in U$.

1.3 Sequences

Definition 1.13 (Boundedness of sequences)**.**
A sequence $(a_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$ is bounded iff there exits a $C > 0$ such that

$$|a_n| \leq C \quad \forall n \geq 0.$$

Definition 1.14 (Convergence of sequences)**.**
A sequence $(a_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$ converges iff there exists an a such that any $\varepsilon > 0$ there exists an index n_0 such that $|a_n - a| < \varepsilon$ for all $n \geq n_0$. The value a is called the limit of $(a_n)_{n\in\mathbb{N}_0}$. In symbols, we write

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

or

$$a_n \to a \quad \text{for} \quad n \to +\infty.$$

Proposition 1.3 (Uniqueness of limits)**.**
Let $(a_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$ be convergent. Then the limit is unique.

Exercise 1.4. *Prove Proposition 1.3. (Hint: Assume there are two different limits and show that they must be equal.)*

Definition 1.15 (Sub-sequence)**.**
Let $(a_n) \in \mathbb{R}\mathbb{N}_0$. Then $(b_n)_{n\in\mathbb{N}_0}$ is called a subsequence of $(a_n)_{n\in\mathbb{N}_0}$, in symbols $(b_n)_{n\in\mathbb{N}_0} \subseteq (a_n)_{n\in\mathbb{N}_0}$, iff all of the elements of $(a_n)_{n\in\mathbb{N}_0}$ occur amongst the elements of $(a_n)_{n\in\mathbb{N}_0}$ in the same order.

Proposition 1.4 (Convergence of sub-sequences)**.**
Let $(a_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$ be a convergent sequence. Then, all sub-sequences converge every sub-sequence converges to the same limit.

Proof. Let $(a_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$ be convergent, i.e. there exists an $a \in \mathbb{R}$ such that

$$\forall \varepsilon > 0\, \exists n_0 : |a_n - a| < \varepsilon \quad \forall n \geq n_0. \tag{1.3}$$

Let now $(a_{n_k})_{k\in\mathbb{N}_0} \subseteq (a_n)_{n\in\mathbb{N}_0}$ be a subsequence of $(a_n)_{n\in\mathbb{N}_0}$. Since $n_k \geq k$, we have if $k \geq n_0$ that $n_k \geq n_0$. Thus, from (1.3), we get $|a_{n_k} - a| < \varepsilon$ for $k \geq n_0$. Thus, $a_{n_k} \to a$ for $k \to +\infty$. □

The converse of the last proposition is also true and yields

Theorem 1.1.
A sequence $(a_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$ is convergent iff every subsequence converges to the same limit.

Exercise 1.5. *Prove Theorem 1.1.*

Exercise 1.6. *Find a sequence that has three sub-sequences that converge to three different limits. (Hint: it is not always necessary to think in formulas.)*

Proposition 1.5 (Boundedness of convergent sequences)**.**
Let $(a_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$ be convergent. Then, (a_n) is bounded.

Exercise 1.7. *Prove Proposition 1.5.*

Theorem 1.2 (Arithmetic properties of sequences)**.**
Let $(a_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$ and $(b_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$ be sequences with $a_n \to a$ and $b_n \to b$ as $n \to +\infty$. Then, one has

(i) $\lim\limits_{n\to+\infty} \left(a_n + b_n\right) = a + b$,

(ii) $\lim\limits_{n\to+\infty} a_n b_n = ab$, *and*

(iii) $\lim\limits_{n\to+\infty} \dfrac{a_n}{b_n} = \dfrac{a}{b}$ *if $b_n \neq 0$ for all $n \in N_0$ and $b \neq 0$.*

Proof. See your notes from Analysis I or regard it as an exercise. □

Remark 1.8. *The number of different convergent sub-sequences is not limited. The sequence*

$$1,1,2,1,2,3,1,2,3,4,1,2,3,4,5,\ldots$$

has infinitely many such sub-sequences. For every natural number $k \in \mathbb{N}$ there is a subsequence (a_{n_l}) with $a_{n_l} \to k$ as $l \to +\infty$.

Theorem 1.3 (Bolzano–Weierstrass)**.**
Every bounded sequence $(a_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$ has a convergent subsequence.

Proof. See your notes from Analysis I. We will prove a more general version later this semester. □

Definition 1.16 (Cauchy-sequence)**.**
Let $(a_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$ be a sequence. We say that $(a_n)_{n\in\mathbb{N}_0}$ is a Cauchy-sequence (fundamental sequence) iff for any $\varepsilon > 0$ there exists an index n_0 such that $|a_n - a_m| < \varepsilon$ for all $m,\ n \geq n_0$. With qualifiers this reads as

$$\forall \varepsilon > 0 \quad \exists n_0 \in \mathbb{N} \quad \forall m, n \geq n_0 \quad \Rightarrow \quad |a_n - a_m| < \varepsilon$$

Exercise 1.8. *Prove that any subsequence of a Cauchy sequence is a Cauchy sequence.*

Theorem 1.4 (Boundedness of Cauchy sequences)**.**
Let $(a_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$ be a Cauchy sequence. Then, $(a_n)_{n\in\mathbb{N}_0}$ is bounded.

Proof. Fix $\varepsilon > 0$ and choose n_0 such that $|a_m - a_n| < \varepsilon$ for all $m,\ n \geq n_0$. Setting $m = n_0$, we then get $|a_n - a_{n_0}| < \varepsilon$. Thus, we have

$$a_{n_0} - \varepsilon \leq a_n \leq a_{n_0} + \varepsilon$$

for all $n > n_0$. Then, an upper bound for the sequence $(a_n)_{n\in\mathbb{N}_0}$ is given by

$$\max\{a_0, a_1, \ldots, a_{n_0-1}, a_{n_0} + \varepsilon\}$$

and a lower bound by

$$\min\{a_0, a_1, \ldots, a_{n_0-1}, a_{n_0} - \varepsilon\}.$$

□

Proposition 1.6.
Let $(a_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$ be a sequence with limit a. Then $(a_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$ is a Cauchy sequence.

Proof. Let $(a_n)_{n\in\mathbb{N}_0}$ be a sequence with $a_n \to a$. Let $\varepsilon > 0$ and n_0 such that $|a_n - a| < \frac{\varepsilon}{2}$. Then, we get for $m,\ n \geq n_0$ that

$$|a_n - a_m| = |a_n - a + a - a_m| \leq |a_n - a| + |a - a_m| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} < \varepsilon.$$

□

We also have a converse to Proposition 1.6.

Proposition 1.7.
Let $(a_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$ be a Cauchy sequence. Then there exists an $a \in \mathbb{R}$ such that $a_n \to a$ for $n \to +\infty$.

Proof. From Theorem 1.4, we know that $(a_n)_{n\in\mathbb{N}_0}$ is bounded and from Bolzano–Weierstrass we know that there exist a convergent subsequence $(a_{n_k})_{n_k\in\mathbb{N}_0} \subseteq (a_n)_{n\in\mathbb{N}_0}$. Now we show, that $(a_n)_{n\in\mathbb{N}_0}$ has the same limit. Let a be the limit of the convergent subsequence $(a_{n_k})_{n_k\in\mathbb{N}_0}$. Let $\varepsilon > 0$ and choose k_1 such that $|a_{n_k} - a| < \frac{\varepsilon}{2}$ for all $k \geq k_1$. Further, choose k_2 such that $|a_{n_k} - a_m| < \frac{\varepsilon}{2}$ for all $m,\ k \geq k_2$. Now, let $k_0 = \max\{k_1, k_2\}$. Then

$$|a_m - a| = |a_m - a_{n_k} + a_{n_k} - a| \leq |a_m - a_{n_k}| + |a_{n_k} - a| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} < \varepsilon.$$

□

Thus, taking the last two propositions together, we have

Theorem 1.5 (Cauchy criterion for convergence)**.**
Let $(a_n)_{n\in\mathbb{N}_0} \in \mathbb{R}$ be s sequence. There exists an $a \in \mathbb{R}$ such that $a_n \to a$ for $n \to +\infty$ iff $(a_n)_{n\in\mathbb{N}_0} \in \mathbb{R}$ is a Cauchy sequence.

Remark 1.9. *Theorem 1.5 is only valid for sequences in $\mathbb{R}^n$ and ^{n}. In general, as you can learn in the module Metric Spaces, Cauchy sequences do not necessarily have limits. If one considers for example a sequence of rational numbers then there are some, e.g.*

$$x_n = \frac{1}{2}\left(x_{n-1} + \frac{2}{x_{n-1}}\right), \quad n \geq 1$$

with $x_0 = 1$, which converge to irrational numbers. Here, $x_n \to \sqrt{2}$. Thus, the limit exists only if the sequence is considered to be in $\mathbb{R}$. This property is called completeness.

Theorem 1.6 (Limits respect inequalities)**.**
Let $(a_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$ be a sequence and suppose that $a_n \leq C$ for all $n \in N_0$. Then, if $\lim_{n\to+\infty} a_n$ exists, it holds

$$\lim_{n\to\infty} a_n \leq C.$$

Remark 1.10. *The $\leq$ in Theorem 1.6 can not be replaced by $<$ as the example $a_n = \frac{1}{n}$ shows fo which $0 < a_n$ but the limit is equal to 0.*

Exercise 1.9. *Prove Theorem 1.6.*

1.4 Series

Let $(a_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$ be a sequence. The, the infinite sum

$$\sum_{n=0}^{+\infty} a_n$$

is called a series. We call the finite sum

$$S_N = \sum_{n=0}^{N} a_n$$

the Nth partial sum of $\sum_{n=0}^{+\infty} a_n$.

Definition 1.17 (Convergence of series)**.**
We say that $\sum_{n=0}^{+\infty} a_n$ converges if it sequence of partial sums $(S_N)_{N\in\mathbb{N}_0}$ converges. We say that $\sum_{n=0}^{+\infty} a_n$ is absolutely convergent if $\sum_{n=0}^{+\infty} |a_n|$ is convergent.

Remark 1.11. *If a series is not convergent, we say it is divergent.*

Since we have Theorem 1.5, we can conclude

Corollary 1.1.
The series $\sum_{n=0}^{+\infty} a_n$ converges iff the sequence of its partial sums $(S_N)_{N\in\mathbb{N}}$ is a Cauchy sequence.

Remark 1.12. *Writing Corollary 1.1 in other words: The series $\sum_{n=0}^{+\infty} a_n$ converges iff for any $\varepsilon > 0$ there exists an index N_0 such that*

$$\left|\sum_{i=m+1}^{n} a_i\right| < \varepsilon$$

for all $n \geq m \geq N_0$.

Let us conclude a first rather obvious property.

Proposition 1.8.
Let $\sum_{n=0}^{+\infty} a_n$ be a convergent series. Then $a_n \to 0$ for $n \to +\infty$.

Proof. Since $\sum_{n=0}^{+\infty} a_n$ is convergent there exists an S such that

$$\lim_{N\to+\infty} S_N = S.$$

Since $a_n = S_n - S_{n_1}$, using Theorem 1.2, we obtain

$$\lim_{n\to+\infty} a_n = \lim_{n\to+\infty} \left(S_n - S_{n-1}\right) = \lim_{n\to+\infty} S_n - \lim_{n\to+\infty} S_{n-1} = 0.$$

The proof also follows from Remark 1.12 choosing $m = n+1$ which is the just the definition of $\lim_{n\to+\infty} a_n = 0$. □

Remark 1.13. *One can not say much more than $a_n \to 0$ for a convergent sequence. Since the harmonic sequence is divergent, we know that $|a_n| \leq n^{-1}$. No better estimate can be obtained.*

1.5 Elementary properties of functions of one variable

1.5.1 Restrictions of functions

Let A and B be sets and $f : A \to B$ be a function. Now let $C \subseteq A$. We want to define the restriction of the function f to a subset C of its domain. By that we mean the function $g : C \to B$ with $x \mapsto f(x)$. We denote g by $f\Big|_C$ which we speak as f restricted to C.

1.5.2 Monotonic functions

Definition 1.18 (Strictly monotone functions)**.**
Let $I \subseteq \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$. Then, f is called

- *strictly increasing if $f(x) > f(y)$ for all $x > y \in I$, and*
- *strictly decreasing if $f(x) < f(y)$ for all $x > y \in I$.*

If the strict inequalities are replaced by $\geq$ and $\leq$, we speak of non-decreasing and non-increasing functions respectively.

Remark 1.14. *With a slight abuse of words, we say (monotonically) increasing when we mean non-decreasing, i.e. $f(x) \leq f(y)$ if $x \leq y$ and (monotonically) decreasing when we mean non-increasing, i.e. $f(x) \leq f(y)$ if $x \geq y$.*

Remark 1.15. *We say a function is monotone if it is either monotonously increasing or decreasing on it domain. Which one it is depends on the function. For some theorems it is only important that the function is either increasing or decreasing but not which one it is. For example, having a continuous function f which is strictly monotone, is invertible, f^{-1} exists.*

1.5.3 Odd and even functions

Definition 1.19 (Odd/even functions).
Let $f : \mathbb{R} \to \mathbb{R}$. Then, we say that

- *the function f is odd iff $f(-x) = -f(x)$ for all $x \in \mathbb{R}$, and*
- *the function f is said to be even iff $f(-x) = f(x)$ for all $x \in \mathbb{R}$.*

Remark 1.16. *Odd/even functions can be defined on intervals too but one needs to take some care with respect to the symmetry property.*

Remark 1.17. *It is easy to see that every function $f : \mathbb{R} \to \mathbb{R}$ can be written as the sum of an odd function f_{odd} and an even function f_{even}, where*

$$\begin{aligned} f_{odd}(x) &= \frac{1}{2}(f(x) - f(-x)), \\ f_{even}(x) &= \frac{1}{2}(f(x) + f(-x)). \end{aligned}$$

Examples are $\sinh(x)$ and $\cosh(x)$ which are odd and even part of e^x.

1.5.4 Convex and concave functions

Definition 1.20 (Convex/concave function).
Let $f : \mathbb{R} \to \mathbb{R}$ be a function. We say that f is convex on the interval $[a, b]$ iff

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

for all $x, y \in [a, b]$ and all $\lambda \in [0, 1]$. A function f is called concave iff $-f$ is convex.

Proposition 1.9 (Jensen's inequality[a]).
Let $f : \mathbb{R} \to \mathbb{R}$ be convex. Suppose that $x_1, \ldots, x_n \in \mathbb{R}$ and $\lambda_1, \ldots, \lambda_n \in [0, +\infty])$ with $\sum_{i=1}^n \lambda_i = 1$. Then,

$$f(\lambda_1 x_1 + \cdots + \lambda_n x_n) \leq \lambda_1 f(x_1) + \cdots + \lambda_n f(x_n)$$

holds.

[a] Named after the Danish mathematician Johan Jensen (1859–1925).

1.6 Elementary inequalities

Inequalities are one of the most important tools in Analysis. The ones of this and the next section must be in you head at all times.

- Let a, b be non-negative real numbers. Then $\min\{a, b\} \leq a + b \leq 2\max\{a, b\}$.
- Let $a \geq 1$, then $\frac{1}{a} \leq 1$.
- Let a, b be positive real numbers. Then, $\frac{1}{a+b} \leq \frac{1}{a}$ and $\frac{1}{a+b} \leq \frac{1}{b}$.
- Let a, b, and c be non-negative real numbers. Then $a + b - c \leq a + b$ and $a - c \leq a + b - c$ and $b - c \leq a + b - c$ hold true. If $b - c > 0$, one gets also $a \leq a + b - c$ and if $a - c > 0$, one gets $b \leq a + b - c$.

- Let a, b be two real numbers. Then
$$|a+b| \leq |a| + |b| \tag{1.4}$$
holds. This is called triangle inequality.

- Let a, b be two real numbers. Then
$$||a| - |b|| \leq |a-b|$$
holds. This is called the reverse triangle inequality.

Example 1.4. *We need these rather obvious inequalities often in the analysis of sequences. For example, there exists an $n_0 \in \mathbb{N}$ such that*
$$\frac{1}{n^2+n} \leq \frac{1}{n^2} \quad \forall n \geq n_0.$$
If we have a sequences as $\frac{1}{n^2-n}$ the situation is less obvious. However, since there exists a n_0 such that $\frac{n^2}{2} - n \geq 1$ for all $n \geq n_0$, we can estimate
$$\frac{1}{n^2-n} = \frac{1}{\frac{n^2}{2} + \frac{n^2}{2} - n} \leq \frac{1}{n^2} \quad \forall n \geq n_0.$$

Exercise 1.10. *Analyse the convergence of the series*
$$\sum_{i=1}^{+\infty} \frac{n}{n^4 - n^3 + 2n^2 - n + 1}$$
by estimating it against the convergent series $\sum_{n=1}^{+\infty} \frac{1}{n^3}$.

Exercise 1.11. *Show that for any $x \in \mathbb{R}^n$, one has*
$$|x_1| + \cdots + |x_n| \leq n \max_{i=1,\ldots,n} |x_i|.$$
Interpret this in the light of Example 2.3.

1.7 Cauchy–Schwarz, Hölder, and Minkowsky

The inequalities in this chapter are immensely important and need to be memorized.

Theorem 1.7 (Cauchy–Schwarz inequality)**.**
Let $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^n$. Then
$$\sum_{i=1}^{n} a_i b_i \leq \left(\sum_{i=1}^{n} |a_i|^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} |b_i|^2\right)^{\frac{1}{2}}.$$
In shorter notation[a], we can write
$$|\langle a, b\rangle| \leq \|a\|_2 \|b\|_2.$$

[a] See also Definition 1.5 in Section 1.2.

Proof. One proof of this inequality is on the problem sheet for you to find. The solutions contain two different proofs an here I will present another one that one of your class-mates found. To keep the notation simpler, we adopt for a moment the applied mathematicians habit of denoting $\|x\|_2$ by $|x|$. Without loss of generality, we can assume that a and b are in $\mathbb{R}^n \setminus \{0\}$. (Why can we do that?) The define
$$v = |a|b - |b|a. \tag{1.5}$$

Since we have $\langle v, v\rangle \geq 0$, we compute

$$\begin{aligned}
\langle v, v\rangle &= \langle |a|b - |b|a, |a|b - |b|a\rangle \\
&= \langle |a|b, |a|b\rangle + \langle |a|b, -|b|a\rangle + \langle -|b|a, |a|b\rangle + \langle -|b|a, -|b|a\rangle \\
&= 2|a|^2|b|^2 - 2|a||b|\langle a, b\rangle \geq 0.
\end{aligned}$$

From the last line, dividing by $2|a||b|$, we get the Cauchy–Schwarz inequality. (Work out what properties of the scalar product have been used in this proof. See Definition 1.5). ☐

Exercise 1.12. *Try to understand the geometry of the proof above. Use your knowledge from Mathematical Methods II. Show that it can also be proven using the identity* $\|x\|_2^2 = \langle x, x\rangle$ *on the vector* $v = \frac{x}{\|x\|_2} - \frac{y}{\|y\|_2}$.

Exercise 1.13. *As another exercise on the manipulation of scalar products try to show*

$$\langle x, y\rangle \leq \frac{\|x\|_2^2}{2} + \frac{\|y\|_2^2}{2}$$

by using $v = x - y$.

Theorem 1.8 (**Minkowski inequality**).
Suppose $p \in [1, +\infty)$ *and let* $a \in \mathbb{R}^n$ *and* $b \in \mathbb{R}^n$. *Then*

$$\left(\sum_{i=1}^{n} |a_i + b_i|^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{n} |a_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} |b_i|^p\right)^{\frac{1}{p}}.$$

In shorter notation, we may write

$$\|a + b\|_p \leq \|a\|_p + \|b\|_p,$$

where we used $\|\cdot\|_p$ *as defined in Example 2.3. With the appropriate changes, the inequality remains true for* $p = \infty$, *i.e.*

$$\begin{aligned}
&\max_{i=1,\ldots,n} |a_i + b_i| \leq \max_{i=1,\ldots,n} |a_i| + \max_{i=1,\ldots,n} |b_i|, \\
&\|a + b\|_\infty \leq \|a\|_\infty + \|b\|_\infty.
\end{aligned}$$

Remark 1.18. *The case* $p = \infty$ *is very easy to prove. It only relies on the triangle inequality for real numbers (see (1.4)):* $|a_i + b_i| \leq |a_i| + |b_i|$ *for all* $i = 1, \ldots, n$ *implies*

$$\max_{i=1,\ldots,n} |a_i + b_i| \leq \max_{i=1,\ldots,n} |a_i| + \max_{i=1,\ldots,n} |b_i|.$$

CHAPTER

2 Length and distance in $\mathbb{R}^n$

Definition 2.1 (**Norm** (length) on $\mathbb{R}^n$).
A function $\|\cdot\|: \mathbb{R}^n \to \mathbb{R}$ is called a ***norm*** *if*

(P1) $\|x\| \geq 0$ *for all* $x \in \mathbb{R}^n$ *and* $\|x\| = 0$ *iff* $x = 0$. *(Positivity)*

(P2) *For all* $\lambda \in \mathbb{R}$, $x \in \mathbb{R}^n$, $\|\lambda x\| = |\lambda|\|x\|$. *(Homogeneity)*

(P3) *For all* $x, y \in \mathbb{R}^n$, *we have*

$$\|x+y\| \leq \|x\| + \|y\|. \quad \textit{(Triangle inequality)}$$

Remark 2.1. *A vector space which is equipped with a norm, is called a normed vector space. The symbol $\|\cdot\|$ stands for any function $\mathbb{R}^n \to \mathbb{R}$ satisfying the three conditions in the definition above.*

Remark 2.2. *In the definition above, one can replace $\mathbb{R}^n$ by any real/complex vector space if one modifies the properties accordingly. See your notes from Linear Algebra for further details. In Section 5.5, we define norms on function spaces.*

Example 2.1. *The easiest example for a norm is the absolute value function $|\cdot|$ which is defined on $\mathbb{R}$ as*

$$|x| = \begin{cases} x & : \quad x > 0 \\ -x & : \quad x \leq 0 \end{cases}.$$

It is clear that $|x| \geq 0$ for every $x \in \mathbb{R}$ as well as $|x| = 0$ iff $x = 0$. To show $P2$ we take a $\lambda \in \mathbb{R}$ and obtain

$$|\lambda x| = \begin{cases} \lambda x & : \quad \lambda x > 0 \\ -\lambda x & : \quad \lambda x \leq 0 \end{cases} = |\lambda||x|.$$

The triangle inequality $|x+y| \leq |x| + |y|$ for $x, y \in \mathbb{R}$ is a well known fact. See also Section 1.6. Thus, $|\cdot|$ is a norm on $\mathbb{R}$ in the sense of the definition above. The $\cdot$ indicates where the argument goes in the symbol.

Example 2.2. *The second example for such a function is the well known Euclidean length which is defined as*

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{\sum_{i=1}^{n} x_i^2} \tag{2.1}$$

for any $x \in \mathbb{R}^n$. In school and applied mathematics $\|\cdot\|_2$ is often denoted by $|\cdot|$.[1] *If $n = 1$, we get that $\|x\| = \sqrt{x^2} = |x|$ is the usual absolute value.*

[1] Mostly, we do not use this notation since we reserve it for the absolute value function $|\cdot|$ defined on the number line $\mathbb{R}$.

Exercise 2.1. *Prove that $\|\cdot\|_2$ fulfils Properties $P1$ to $P3$ in Definition 2.1.*

Example 2.3. *The 2 in the definition of $\|\cdot\|_2$ in (2.1) plays no special role other than giving the familiar Euclidean distance from the origin of $\mathbb{R}^n$ to x which we call* length *of x. Since there is an everyday meaning to this word, we will generally be speaking about norms since it is sometimes useful to speak about* lengths *that are defined in different terms. We set*

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}},$$

where p may be in $[1,\infty)$. We can extend the definition of $\|\cdot\|_p$ to $p=\infty$ if we set

$$\|x\|_\infty = \max_{i=1,\dots,n} |x_i| = \max\{|x_1|,|x_2|,\dots,|x_n|\}.$$

Let us now introduce the notion of a distance (more generally metric) on $\mathbb{R}^n$.

Definition 2.2 (**Metric** (distance) on $\mathbb{R}^n$)**.**
A function $d:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$ is called a ***metric*** *(distance) if it satisfies*

(H1) For any x, $y\in\mathbb{R}^n$, we have $d(x,y)\geq 0$ and $d(x,y)=0$ iff $x=y$. (Positivity)

(H2) For any x, $y\in\mathbb{R}^n$, we have $d(x,y)=d(y,x)$. (Symmetry)

(H3) For any x, y, and $z\in\mathbb{R}^n$, we have

$$d(x,y)\leq d(x,z)+d(z,y). \qquad \text{(Triangle inequality)}$$

Remark 2.3. *Metrics can not only be defined on vector spaces. This more general theory leads to metric spaces which you can study in the module Matric Spaces in your third year.*

Example 2.4. *A well-known example for a metric is*

$$d_2(x,y)=\|x-y\|_2=\sqrt{\sum_{i=1}^n |x_i-y_i|^2}.$$

This is the so-called Euclidean distance of x and y in $\mathbb{R}^n$. We say this metric is induced by the Euclidean norm $\|\cdot\|_2$. Let us check the properties $(H1)$ to $(H3)$. Since we have $d_2(x,y)=\|x-y\|_2\geq 0$ (see $(P1)$ in Definition 2.1), $(H1)$ follows. By $(P2)$ in Definition 2.1, we get

$$d_2(x,y)=\|x-y\|_2=\|-(y-x)\|_2=|-1|\|y-x\|_2=d_2(y,x).$$

Finally, we have

$$d_2(x,y)=\|x-y\|_2=\|x-z+z-y\|_2\leq\|x-z\|_2+\|z-y\|_2=d_2(x,z)+d_2(z,y)$$

by $(P3)$ in Definition 2.1.

Example 2.5. *The norms defined in Example 2.3 provide another possibility to define distances on $\mathbb{R}^n$. We get*

$$d_p(x,y)=\|x-y\|_p=\left(\sum_{i=1}^n |x_i-y_i|^p\right)^{\frac{1}{p}}. \tag{2.2}$$

Exercise 2.2. *Take* $x = \begin{bmatrix}1 & 1 & 1\end{bmatrix}^T$ *and* $y = \begin{bmatrix}-1 & -2 & 5\end{bmatrix}^T$ *and compute their distances with* d_p *for* $p = 1, 2,$ *and* ∞.

Exercise 2.3. *Using the properties of norms, try to prove that* d_p*, defined in* (2.2)*, is a metric on* $\mathbb{R}^n$ *according to Definition 2.2.*

Remark 2.4. *As you have seen there is some hierarchy here. Scalar products on* $\mathbb{R}^n$ *induce norms (by* $\|x\| = \sqrt{\langle x, x\rangle}$*) on* $\mathbb{R}^n$ *and norms induce metrics (by* $d(x, y) = \|x - y\|$*). However, you can have metrics that are not induced by norms, norms that are not induced by scalar products (e.g.* $\|\cdot\|_p$, $p \neq 2$*). For the notion of scalar products refer to your Lecture Notes of Linear Algebra.*

2.1 Open balls in $\mathbb{R}^n$

Definition 2.3 (Open ball of radius r around $x \in \mathbb{R}^n$).
We define open balls of radius $r > 0$ *around a point* $x \in \mathbb{R}^n$ *by*

$$B_r(x) := \left\{y \in \mathbb{R}^n : d(x, y) < r\right\}.$$

Remark 2.5. *For the moment, the name* open ball *is unjustified. We will introduce the notion of openness as well as prove that the open ball is indeed open in the next section.*

Example 2.6. *Let us consider* $n = 1$, $d(x, y) = |x - y|$*. Then, for* $x \in \mathbb{R}$ *and* $r > 0$*, we have*

$$B_r(x) = (x - r, x + r).$$

By the definition of an open ball, we have

$$\begin{aligned} B_r(x) &= \{y \in \mathbb{R} : d(x, y) < r\} \\ &= \{y : \mathbb{R} : |x - y| < r\}, \end{aligned}$$

i.e. to see what the set $B_r(x)$ *is, we have to solve* $|x - y| < r$ *for* y *and obtain*

$$\begin{aligned} & |x - y| < r \\ \Leftrightarrow\quad & -r < y - x < r \\ \Leftrightarrow\quad & x - r < y < x + r \end{aligned}$$

which, with the notation from Section 1.1, means that $y \in (x - r, x + r)$.

Remark 2.6. *If* $d(x, y) = \|x - y\|_p$*, we get that*

$$B_r(x) := \left\{y \in \mathbb{R}^n : \|x - y\|_p < r\right\}.$$

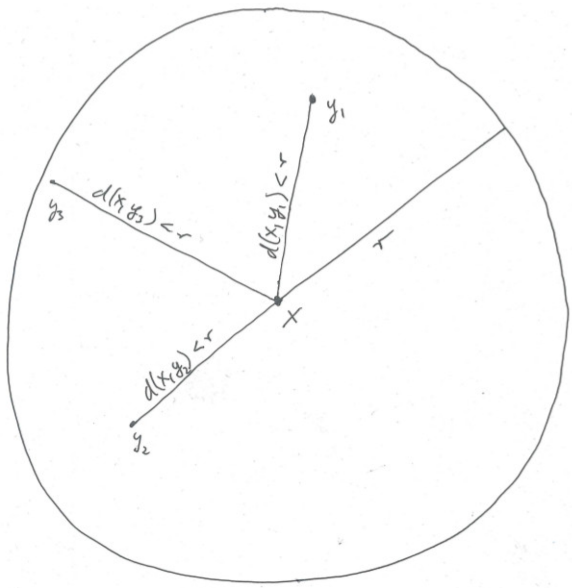


Figure 2.1: Illustration of the definition of $B_r(x)$.

Example 2.7. *Using the definition of d_p by (2.2) we get the open balls $B_r(x)$ associated to d_p as*

$$\begin{aligned} B_r(x) &= \left\{ y \in \mathbb{R}^n : \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}} < r \right\}, \\ &= \left\{ y \in \mathbb{R}^n : \sum_{i=1}^n |x_i - y_i|^p < r^p \right\}. \end{aligned}$$

Draw some pictures of balls with different p values. Can you reproduce the pictures for the values $p = 1$, 2, ∞ that we looked at in class?

Example 2.8. *Let us define the following sets*

$$\begin{aligned} B_1^{(1)}(0) &:= \{ y \in \mathbb{R}^2 : \|y\|_1 = 1 \}, \\ B_1^{(2)}(0) &:= \{ y \in \mathbb{R}^2 : \|y\|_2 = 1 \}, \\ B_1^{(\infty)}(0) &:= \{ y \in \mathbb{R}^2 : \|y\|_\infty = 1 \}. \end{aligned}$$

The sets above are the unit circles in $\mathbb{R}^2$ with respect to the distance measures d_1, d_2, and d_∞. for the definition of those, see (2.2).

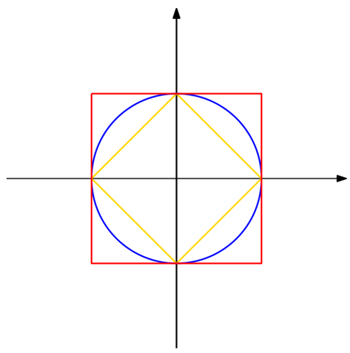


Figure 2.2: The unit-spheres $B_1^{(1)}(0)$ (gold), $B_1^{(2)}(0)$ (blue), and $B_1^{(\infty)}(0)$ (red).

2.2 Three important notions for subsets $\Omega \subseteq \mathbb{R}^n$

Definition 2.4 (Limit point of $\Omega \subseteq \mathbb{R}^n$).
Let $\Omega \subseteq \mathbb{R}^n$. Then, $p \in \mathbb{R}^n$ is called a limit point of Ω iff for all $\varepsilon > 0$, the set $B_\varepsilon(p) \cap \Omega$ is neither equal to $\emptyset$ nor to $\{p\}$. (This means that $B_\varepsilon(p) \cap \Omega$ contains more than one point for all $\varepsilon > 0$.)

Example 2.9. *Let us discuss a list of examples illustrating the last definition:*

- *Let $\Omega = (0,1)$. Then, every point $p \in (0,1)$ and $p = 0$, $p = 1$ are limit points of Ω.*
- *Consider the open ball $B_r(x)$. Then all points $p \in B_r(x)$ are limit points of $B_r(x)$ as well as all points $p \in \{y \in \mathbb{R}^n : d(y,x) = r\}$.*
- *Consider $\{-17, 4, 6, 264, 1034\}$. This set has no limit points. Similarly, $\mathbb{N}$, $\mathbb{N}_0$, and $\mathbb{Z}$ have no limit points.*
- *The set of all limit points of $\mathbb{Q}$ is $\mathbb{R}$. This follows from the density[2] of the rational numbers in the real numbers.*

Proposition 2.1 (Characterization of limit points of $\Omega \subseteq \mathbb{R}^n$).
Let $\Omega \subseteq \mathbb{R}^n$. Then, $p \in \mathbb{R}^n$ is limit point of Ω iff

$$\forall \varepsilon > 0 \quad \text{the set} \quad B_\varepsilon(p) \cap \Omega \quad \text{is infinite.}$$

Proof. The proof of this proposition is on Problem Sheet 3 and thus in the respective solutions. □

From that we get immediately

Corollary 2.1.
Let $\Omega \subseteq \mathbb{R}^n$ and $p \in \mathbb{R}^n$ be a limit point of Ω. Then there exists a sequence $(p_n)_{n \in \mathbb{N}_0} \subseteq \Omega \setminus \{p\}$ such that $p_n \to p$ as $n \to +\infty$.

For later use, let us introduce the notion of an isolated point.

Definition 2.5 (Isolated point of $\Omega \subseteq \mathbb{R}^n$).
Let $\Omega \subseteq \mathbb{R}^n$. Then, $p \in \Omega$ is called an isolated point of Ω iff

$$\exists \varepsilon > 0 : \; B_\varepsilon(p) \cap \Omega = \{p\}.$$

It follows immediately from this definition, that limit points can not be isolated. See also Proposition 2.1.

[2] This means that for every $x \in \mathbb{R}$, the ball $(x - r, x + r)$ contains a rational number however small one chooses $r > 0$.

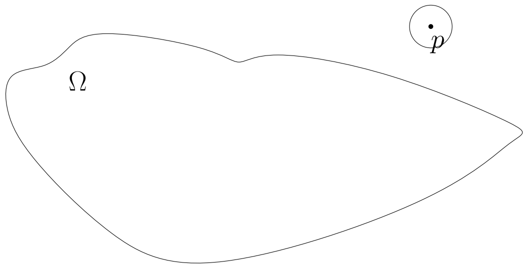


Figure 2.3: Illustration of an Isolated point with neighborhood $B_\varepsilon(p)$.

The next notion is very central and the basis of what is called topology. Sometimes we will refer to the collection of all open sets, in our case generated by a metric or norm, as a topology on the space.

Definition 2.6 (Open sets)**.**
A set $\Omega \subseteq \mathbb{R}^n$ is called open iff

$$\forall x \in \Omega \quad \exists \varepsilon > 0 : \ B_\varepsilon(x) \subseteq \Omega.$$

Figure 2.4: Illustration of an open set.

Remark 2.7. *The definition above can intuitively be read as follows: changing x only ever so slightly does not lead to leaving the set or, since we are in a vector space, at every point of the set we can go a, possibly very small, step in any direction without leaving the set.*

Remark 2.8. *Intuitively, open sets provides a method to distinguish two points. Two points $p_1 \in \mathbb{R}^n$ and $p_2 \in \mathbb{R}^n$ are different iff there exists a $\varepsilon > 0$ such that $B_\varepsilon(p_1) \cap B_\varepsilon(p_2) = \emptyset$. Let us, for instance, consider a convergent sequence $(a_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}^n$ and assume that there are two limits p_1 and p_2. Now, we show that p_1 can not be* distinguished *from p_2 and that, therefore, $p_1 = p_2$. Since $a_n \to p_1$ and $a_n \to p_2$, we get that* *for all* $\varepsilon > 0$ *there must be an* n_0 *such that* $a_n \in B_\varepsilon(p_1)$ *and* $a_n \in B_\varepsilon(p_2)$*, i.e.* $a_n \in B_\varepsilon(p_1) \cap B_\varepsilon(p_2)$*, for all* $n \geq n_0$*. Hence,* $B_\varepsilon(p_1) \cap B_\varepsilon(p_2)$ *is never empty. Thus,* $p_1 = p_2$*.*

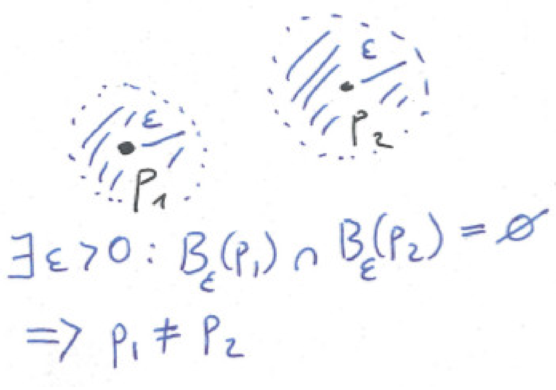


Figure 2.5: Separating two points by open sets.

We can reformulate convergence with the help of open sets.

Theorem 2.1.
A sequence $(a_n)_{n \in \mathbb{N}_0} \subseteq \mathbb{R}^n$ *is convergent to* $a \in \mathbb{R}^n$ *iff for all* $\varepsilon > 0$ *the set* $\{k \in \mathbb{N}_0 : x_k \notin B_\varepsilon(a)\}$ *is finite.*

Proof. We first prove [$\Rightarrow$]: Let $(a_n)_{n \in \mathbb{N}_0} \subseteq \mathbb{R}^n$ be convergent with $a_n \to a$. Then, there exists for all $\varepsilon > 0$ an $n_0 \in \mathbb{N}$ such that $d(a_n, a) < \varepsilon$ for all $n \geq n_0$, i.e. $a_n \in B_\varepsilon(a)$. Thus, for all $\varepsilon > 0$ there are at most $n_0 = n_0(\varepsilon) - 1$ elements a_n for which $a_n \notin B_\varepsilon(a)$. To prove [$\Leftarrow$], we assume that $\{n \in \mathbb{N}_0 : a_n \notin B_\varepsilon(a)\}$ is finite for all $\varepsilon > 0$. Let $n_0 = n_0(\varepsilon)$ be the maximum of this set. Then, for all $\varepsilon > 0$ there exists $\tilde{n}_0$, given by $n_0 + 1$, such that $d(n_n, a) < \varepsilon$ for all $n \geq \tilde{n}_0$, since $a_n \in B_\varepsilon(a)$ for $n \geq \tilde{n}_0$. This concludes the proof. □

Exercise 2.4. *Find an example which shows that the following statement is not true. A sequence* $(a_n)_{n \in \mathbb{N}_0} \subseteq \mathbb{R}^n$ *is convergent to* $a \in \mathbb{R}^n$ *iff for all* $\varepsilon > 0$ *the set* $\{k \in \mathbb{N}_0 : x_k \in B_\varepsilon(a)\}$ *is infinite.*

Example 2.10. *Let us discuss a list of examples illustrating the definition of open sets:*

- *Open intervals* (a, b) *are open. Closed intervals* $[a, b]$ *are not open as are intervals of the form* $[a, b)$ *and* $(a, b]$*.*
- *The open ball* $B_r(p)$ *for* $p \in \mathbb{R}^n$ *is open. To see that we pick an arbitrary* $q \in B_r(p)$ *and show that there exists an* $\delta > 0$ *such that* $B_\delta(q) \subseteq B_r(p)$*. This is the case if we choose* $\varepsilon \in (0, r - d(p, q))$*.*

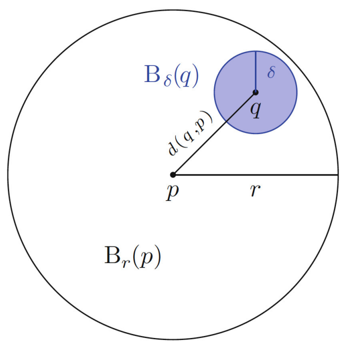


Figure 2.6: The open ball $B_r(p)$ is open.

- *The whole space $\mathbb{R}^n$ is open for any $n \geq 1$ as well as $\emptyset$.*

Remark 2.9. *Let us show that the interval (a,b) is open. If, $a = b$, we have $(a,b) = \emptyset$ and, therefore, (a,b) is open. Now, let $a < b$. Then, for $x\ in(a,b)$, we need to find a ball $B_r(x) = (x-r, x+r)$ with $r > 0$ such that $B_r(x) \subseteq (a,b)$. We can achieve that by choosing $r = \min\{d(x,a), d(x,b)\} = \min\{|x-a|, |x-b|\}$.*

Proposition 2.2.
Let $\Omega \subseteq \mathbb{R}^n$ be an open set. Every $p \in \Omega$ is a limit point of Ω.

Remark 2.10. *Note that this proposition does not imply that all limit points are contained in Ω. It merely states that open sets can not have isolated points. Take $B_1(0)$ as an example. All $x \in B_1(0)$ are limit points of $B_1(0)$ but also the points of $\{y \in \mathbb{R}^n : d(y,0) = 1\}$ are limit points and are not contained in $B_1(0)$.*

Proof. Let $p \in \Omega$. Since Ω is open, there exists an $\varepsilon > 0$ such that $B_\varepsilon(p) \subseteq \Omega$. This remains true for all $\varepsilon' \in (0,\varepsilon)$ since $B_{\varepsilon'}(p) \subseteq B_\varepsilon(p)$. Hence, for all $\varepsilon' \in (0,\varepsilon]$, we have $B_{\varepsilon'}(p) \cap \Omega = B_{\varepsilon'}(p)$ which is never empty not only one point. If $\varepsilon' > \varepsilon$ we can not have $B_{\varepsilon'}(p) \cap \Omega$ since this set always contains $B_\varepsilon(p)$ since $B_\varepsilon(p) \subseteq B_{\varepsilon'}(p)$. The set $B_\varepsilon(p)$ is neither empty not only one point. Thus, we have proven that

$$\forall \varepsilon > 0 : \; B_\varepsilon(p) \cap \Omega \neq \begin{cases} \emptyset \\ \{p\} \end{cases}$$

which is the definition of a limit point. This concludes the proof. □

Exercise 2.5. *Use the statement of Proposition 2.2 to prove that, besides $\emptyset$, no finite set can be open.*

Now, we give a name to sets that contain all their limit points.

Definition 2.7 (Closed set)**.**
A set $\Omega \subseteq \mathbb{R}^n$ is called closed iff every limit point of Ω belongs to Ω.

Example 2.11. *Let us discuss a list of examples illustrating the last definition:*

- *Closed intervals $[a,b]$ are closed. Open intervals (a,b) are not closed as are intervals of the form $[a,b)$ of $(a,b]$.*

- *Let $x \in \mathbb{R}^n$ and $r > 0$. Then, the closed ball $\{y \in \mathbb{R}^n : d(x,y) \leq r\}$ is closed as well. It is the set of al limit points of the open ball $B_r(x)$.*

- *The set*

$$\{y \in \mathbb{R}^n : d(y,x) \leq r\},$$

  called the closed ball of radius r, is closed.

- *The whole space $\mathbb{R}^n$ is closed for any $n \geq 1$ as well as $\emptyset$.*

- *All finite sets $\{x_0, \ldots, x_N\} \subseteq \mathbb{R}^n$, including $\emptyset$, are closed. (In Exercise 2.5 you have proven that the set of limit points of finite sets is $\emptyset$. Since $\emptyset$ is a subset of every set, finite sets contain all their limit points and are, therefore, closed.)*

Example 2.12. *Let us show that the set*

$$\overline{B}_r(x) := \{y \in \mathbb{R}^n : d(y,x) \leq r\},$$

is closed. To show that, we have to show that no point not contained in $\overline{B}_r(x)$ *is limit point of* $\overline{B}_r(x)$*. Let* $y \in \mathbb{R}^n \setminus \overline{B}_r(x)$*. Then there exists* $\varepsilon > 0$*, e.g. given by* $\varepsilon = \frac{d(x,y)-r}{2}$*, such that* $B_\varepsilon(y) \cap \overline{B}_r(x) = \emptyset$*. Thus,* y *can not be limit point of* $\overline{B}_r(x)$*. Thus, all limit points of* $\overline{B}_r(x)$ *must be contained in* $\overline{B}_r(x)$ *which concludes the argument.*

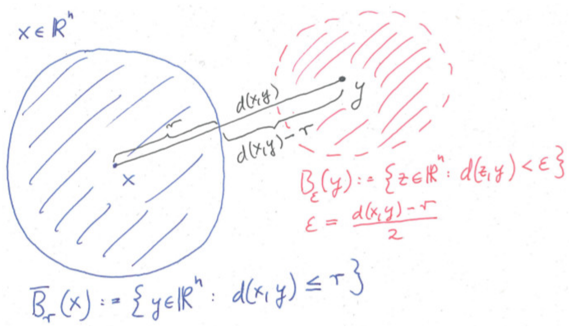


Figure 2.7: Illustration of the argument for the closedness of $\overline{B}_r(x)$.

Remark 2.11 (Characterization of closed sets by limits)**.**
Let us state an alternative characterization of closed sets:
Let $\Omega \subseteq \mathbb{R}^n$*. Then,* Ω *is closed iff for all Cauchy-sequences* $(x_n)_{n \in \mathbb{N}_0} \subseteq \Omega$ *we have that* $\lim_{n \to +\infty} x_n \in \Omega$*.*

Exercise 2.6. *Prove the equivalence stated in Remark 2.11.*

Remark 2.12. *Even though the names might suggest otherwise, open and closed are not mutually exclusive. For example, the set* $\mathbb{R}$ *is open and closed as is* $\mathbb{R}^n$ *in general. Also the set* $\emptyset$ *is open and closed. See also the next proposition.*

Definition 2.8 (Closure)**.**
Let $\Omega \subseteq \mathbb{R}^n$ *and denote by* Ω' *the set of all limit points of* Ω*. Then, we define the closure of* Ω *by*

$$\mathrm{cl}(\Omega) = \Omega \cup \Omega'.$$

Remark 2.13. *By definition we have that* $\mathrm{cl}(\Omega)$ *is closed and that* $\Omega \subseteq \mathrm{cl}(\Omega)$*.*

Proposition 2.3 (Characterization of open/closed sets)**.**
A set $\Omega \subseteq \mathbb{R}^n$ *is*

- *open iff* $\Omega^c = \mathbb{R}^n \setminus \Omega$ *is closed, and*
- *closed iff* $\Omega^c = \mathbb{R}^n \setminus \Omega$ *is open.*

We can prove the following

Proposition 2.4.

Let Ω_1, $\Omega_2 \subseteq \mathbb{R}^n$ be open/closed. Then, the sets $\Omega_1 \cap \Omega_2$ and $\Omega_1 \cup \Omega_2$ are open/closed as well.

Proof. Let us first prove the assertion for open sets. Let Ω_1 and Ω_2 be open. Then $\Omega_1 \cap \Omega_2$ is open. Indeed, let $x \in \Omega_1 \cap \Omega_2$. Then, there exists an $\varepsilon_1 > 0$ such that $B_{\varepsilon_1}(x) \subseteq \Omega_1$ and an $\varepsilon_2 > 0$ such that $B_{\varepsilon_2}(x) \subseteq \Omega_2$. Thus, we have that $B_\varepsilon(x) \in x \in \Omega_1 \cap \Omega_2$ if $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$. Thus, $\Omega_1 \cap \Omega_2$ is open. Let now $x \in \Omega_1 \cup \Omega_2$. Then, there exist $\varepsilon_1, \varepsilon_2 > 0$ such that $B_{\varepsilon_1}(x) \subseteq \Omega_1$ or $B_{\varepsilon_2}(x) \subseteq \Omega_2$. Thus, either $B_{\varepsilon_1}(x) \subseteq \Omega_1 \cup \Omega_2$ or $B_{\varepsilon_2}(x) \subseteq \Omega_1 \cup \Omega_2$. Hence, $\Omega_1 \cup \Omega_2$ is open.
For the proof for the closed sets, we use the De Morgan rules for sets. Let Ω_1 and Ω_2 be closed. We can write $(\Omega_1 \cap \Omega_2)^c = \Omega_1^c \cup \Omega_2^c$ and thus $\Omega_1 \cap \Omega_2 = (\Omega_1^c \cup \Omega_2^c)^c$. Then, by Proposition 2.3, we have that Ω_1^c and Ω_2^c are open which, by the first part of the proof, gives that $\Omega_1^c \cup \Omega_2^c$ is open. Again, by Proposition 2.3, we have that $(\Omega_1^c \cup \Omega_2^c)^c$ is then closed, i.e. $\Omega_1 \cap \Omega_2$ is closed. A similar argument proves the remaining case. □

2.2.1 A list of exercises

All exercises here can be done by using the definitions: We consider all sets to be subsets of $\mathbb{R}^n$ and the metrics d used in the definition of the open ball in Definition 2.3 are one of the d_p introduced in Example 2.3. If nothing else is said, we use

$$d_2(x, y) = \|x - y\|_2 = \left(\sum_{i=1}^{n} |x_i - y_i|^2 \right)^{\frac{1}{2}}.$$

1) Let $\Omega \subseteq \mathbb{R}^n$ be an open set. Prove that every $p \in \Omega$ is a limit point of Ω.

2) Prove that $\emptyset$ is closed and open.

3) Prove that a finite set $\{x_1, \ldots, x_N\} \subseteq \mathbb{R}^n$ has no limit points.

4) Prove that $\mathbb{N}$ has no limit points.

5) Prove that there is no finite open set other than $\emptyset$.

6) Prove that all finite sets $\{x_1, \ldots, x_N\} \subseteq \mathbb{R}^n$ are closed.

CHAPTER

3 Limits of Functions

First let us set up some language to consider functions with values in $\mathbb{R}^m$. Suppose $\Omega \subseteq \mathbb{R}^n$ and

$$\begin{cases} f : \Omega \to \mathbb{R}^m \\ \quad x \mapsto f(x) \end{cases} \tag{3.1}$$

be a map. We recall that Ω is then called the domain of f, in symbols $\mathrm{dom}(f) := \Omega$ and $\mathbb{R}^m$ is called the co-domain. Further, we define the image of f by $\mathrm{im}(f) := \{f(x) : x \in \mathrm{dom}(f)\}$. In general, we have $\mathrm{im}(f) \subseteq \mathbb{R}^m$.

Then, $f(x) = f(x_1, \ldots, x_n)$ is a vector, and thus, also has components, i.e. (3.1) can be understood as

$$\mathbb{R}^n \ni x \mapsto f(x) = f(x_1, \ldots, x_n) = \begin{bmatrix} f_1(x_1, \ldots, x_n) \\ f_2(x_1, \ldots, x_n) \\ \vdots \\ f_m(x_1, \ldots, x_n) \end{bmatrix} \in \mathbb{R}^m.$$

Hence, we are dealing with m functions $f_k : \Omega \to \mathbb{R}$ in n variables.

Example 3.1. *Change to polar coordinates is a function*

$$\begin{cases} [0, \infty) \times [0, 2\pi) \to \mathbb{R}^2 \\ \begin{bmatrix} r \\ \phi \end{bmatrix} \mapsto \begin{bmatrix} r\cos(\phi) \\ r\sin(\phi) \end{bmatrix} \end{cases}.$$

If we want to talk about limits, we can do this either *component–wise*, or we simply think of points as vectors and use the metrics (distance functions) introduced in Section 2. The latter point of view is slightly harder but better in the long-term.

3.1 Limits of $\mathbb{R}^m$-valued functions

If nothing else is said, we assume in what follows that $d(x, y) = \|x - y\|$ and that $\|\cdot\|$ is the Euclidean norm $\|\cdot\|_2$. If $n = 1$, I will always use $|\cdot|$ to denote the norm which is then called absolute value.

Definition 3.1 (Limit of a $\mathbb{R}^m$–valued function)**.**
Let $\Omega \subseteq \mathbb{R}^n$ and let $p \in \Omega$ be a limit point of Ω. Suppose $f : \Omega \to \mathbb{R}^m$ is a function. Then we write

$$q = \lim_{x \to p} f(x)$$

for a $q \in \mathbb{R}^m$ iff

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad \forall x : \quad 0 < d(x, p) < \delta \quad \Rightarrow \quad d(q, f(x)) < \varepsilon. \tag{3.2}$$

Remark 3.1. *To prove that a function has a limit at a point p/is continuous at p, we have to show that*

$$d(x, p) < \delta \quad \Rightarrow \quad d(q, f(x)) < \varepsilon.$$

This is usually done by showing that there is a constant $C > 0$ such that

$$d(q, f(x)) \leq C d(x, p). \tag{3.3}$$

Setting then $\delta = \frac{\varepsilon}{C}$, we obtain

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad \forall x : \quad d(x, p) < \delta \quad \Rightarrow \quad d(q, f(x)) < \varepsilon.$$

This is the scheme that you will see over and over again in the examples and proofs of theorems involving continuity. It could be that one can not prove (9.10) but something of the kind

$$d(q, f(x)) \leq C g\big(d(x, p)\big)$$

for a "well-behaved" (monotone) function g. Then one can have $d(x, p) < \delta$ implies $d(q, f(x))$ if $\delta = g^{-1}\left(\frac{\varepsilon}{C}\right)$. If $f(p)$ is not defined, as it happens for limits sometimes, one would ask

$$0 < d(x, p) < \delta \quad \Rightarrow \quad d(q, f(x)) < \varepsilon$$

but the strategies are exactly the same.

Remark 3.2. *If the distance function is given by a norm $\|\cdot\|$, i.e. $d(x, y) = \|x - y\|$, then the line (3.2) can be rewritten as*

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad \forall x : 0 < \|x - p\| < \delta \Rightarrow \|f(x) - q\| < \varepsilon.$$

Exercise 3.1. *Set $m = n = 1$ and write Definition 3.1 down explicitly (with the correct distances) in that case.*

Remark 3.3. *Remember that in the following examples $0 < d(x, p) < \delta$ is usually asked since f is not necessary defined at p.*

3.2 Examples

Example 3.2. *Consider $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 2x + 1$. For $\mathbb{R}$, we have $d(x, y) = |x - y|$. We show that $\lim_{x \to 1} f(x) = 3$. What do we have to do? For every $\varepsilon > 0$ we have to find a $\delta > 0$ such that*

$$0 < |x - 1| < \delta \quad \Rightarrow \quad |f(x) - 3| < \varepsilon.$$

Thus, we compute $|f(x) - 3|$ and estimate it in terms of $|x - 1|$. The triangle inequality is our friend. We compute

$$|f(x) - 3| = |2x - 2| = 2|x - 1|.$$

So, if we have $0 < |x - 1| < \delta$, then we have

$$|f(x) - 3| < 2\delta = \varepsilon$$

which leads to the choice $\delta = \frac{\varepsilon}{2}$.

Example 3.3. *Consider* $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = x_1 + x_2$. *We show that* $\lim_{x\to 0} f(x) = 0$. *For that, again, we need to show that for every* $\varepsilon > 0$ *there exists a* $\delta > 0$ *such that* $|f(x)| < \varepsilon$ *whenever* $0 < \|x\|_2 < \delta$. *We have*

$$|f(x)| = |x_1 + x_2| \leq |x_1| + |x_2|.$$

Now, we need to somehow estimate $|x_1| + |x_2|$ *in terms of* $\sqrt{x_1^2 + x_2^2}$. *We have* $\|x\|_1 \leq \sqrt{2}\|x\|_2$.[1] *Thus, if we choose* $\delta = \frac{\varepsilon}{\sqrt{2}}$ *we get*

$$0 < \|x\|_2 < \delta \quad \Rightarrow \quad |f(x)| \leq \sqrt{2}\|x\|_2 < \sqrt{2}\delta = \varepsilon.$$

Example 3.4. *Consider* $f : \mathbb{R}^2 \setminus \{0\} \to \mathbb{R}$, *where* $f(x_1, x_2) = \frac{x_1^2 + 3x_2^2}{\sqrt{x_1^2 + x_2^2}}$. *We prove that* $\lim_{x\to 0} f(x) = 0$. *Hence, we need for every* $\varepsilon > 0$ *a* $\delta > 0$ *such that*

$$0 < \|x\|_2 < \delta \quad \Rightarrow \quad |f(x)| < \varepsilon.$$

We have

$$\begin{aligned} |f(x_1, x_2)| &= \frac{x_1^2 + 3x_2^2}{\sqrt{x_1^2 + x_2^2}} = \frac{x_1^2 + x_2^2 + 2x_2^2}{\sqrt{x_1^2 + x_2^2}} = \sqrt{x_1^2 + x_2^2} + 2\frac{x_2^2}{\sqrt{x_1^2 + x_2^2}} \\ &= \|x\|_2 + 2\frac{x_2^2}{\sqrt{x_1^2 + x_2^2}} \leq \|x\|_2 + 2\frac{x_2^2 + x_1^2}{\sqrt{x_1^2 + x_2^2}} \leq 3\|x\|_2 < 3\delta = \varepsilon \end{aligned}$$

leads to the choice $\delta = \frac{\varepsilon}{3}$.

Example 3.5. *Consider* $f : \mathbb{R}^3 \setminus \{0\} \to \mathbb{R}^2$ *given by*

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto \begin{bmatrix} \frac{x_1 + 2\sqrt{x_2^2 + x_3^2}\sin(x_3)}{(x_1^2 + x_2^2 + 2x_3^2)^{\frac{1}{4}}} \\ x_3 + 1 \end{bmatrix}.$$

Show, using the $\varepsilon - \delta$ *definition of limits, that*

$$\lim_{x\to 0} f(x) = 0.$$

Let us begin by using $\|x\|_2 \leq \|x\|_1$ *and obtain*

$$\begin{aligned} d(f(x), q) &= \left\| \begin{bmatrix} \frac{x_1 + 2\sqrt{x_2^2 + x_3^2}\sin(x_3)}{(x_1^2 + x_2^2 + 2x_3^2)^{\frac{1}{4}}} \\ x_3 + 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\|_2 = \left\| \begin{bmatrix} \frac{x_1 + 2\sqrt{x_2^2 + x_3^2}\sin(x_3)}{(x_1^2 + x_2^2 + 2x_3^2)^{\frac{1}{4}}} \\ x_3 + 1 \end{bmatrix} \right\| \\ &\leq \left| \frac{x_1 + 2\sqrt{x_2^2 + x_3^2}\sin(x_3)}{(x_1^2 + x_2^2 + 2x_3^2)^{\frac{1}{4}}} \right| + |x_3| \end{aligned}$$

Now, we use the triangle inequality, $|\sin(x_3)| \leq 1$, *and the monotonicity of* $(\cdot)^{\frac{1}{4}}$ *with* $\frac{1}{x_1^2 + x_2^2 + 2x_3^2} \leq \frac{1}{x_1^2 + x_2^2 + x_3^2}$ *and we obtain*

$$\left| \frac{x_1 + 2\sqrt{x_2^2 + x_3^2}\sin(x_3)}{(x_1^2 + x_2^2 + 2x_3^2)^{\frac{1}{4}}} \right| \leq \frac{|x_1| + 2\sqrt{x_2^2 + x_3^3}}{(x_1^2 + x_2^2 + x_3^2)^{\frac{1}{4}}}.$$

[1] One can use a slightly worse inequality:

$$|x_1| + |x_2| = \sqrt{|x_1|^2} + \sqrt{|x_2|^2} \leq \sqrt{|x_1|^2 + |x_2|^2} + \sqrt{|x_1|^2 + |x_2|^2} = 2\sqrt{|x_1|^2 + |x_2|^2}.$$

Since we clearly have $|x_i| \le |x_1|+|x_2|+|x_3| = \|x\|_1 \le \sqrt{3}\|x\|_2$ *for any* $i = 1,2,3$ *and* $\sqrt{x_2^2+x_3^3} \le \sqrt{x_1^2+x_2^2+x_3^3} = \|x\|_2$. *This gives us*

$$\begin{aligned}\frac{|x_1|+2\sqrt{x_2^2+x_3^3}}{(x_1^2+x_2^2+x_3^2)^{\frac{1}{4}}}+|x_3| &\le \frac{\sqrt{3}\|x\|_2+2\|x\|_2}{\|x\|^{\frac{1}{2}}}+\sqrt{3}\|x\|_2 \\ &= (2+\sqrt{3})\|x\|_2^{\frac{1}{2}}+\sqrt{3}\|x\|_2 \\ &= \|x\|_2^{\frac{1}{2}}\big(2+\sqrt{3}+\sqrt{3}\|x\|_2^{\frac{1}{2}}\big).\end{aligned}$$

If we assume $\delta \le 1$ *(since* $\|x\|_2 < \delta$*), we then get*

$$\|x\|_2^{\frac{1}{2}}\big(2+\sqrt{3}+\sqrt{3}\|x\|_2^{\frac{1}{2}}\big) \le 6\|x\|_2^{\frac{1}{2}}.$$

Finally, we obtain

$$0 < \|x\|_2 < \delta \quad \Rightarrow \quad d(f(x),q) < \varepsilon$$

if we choose

$$\delta = \min\left\{1, \left(\frac{\varepsilon}{6}\right)^2\right\}.$$

Exercise 3.2. *Some estimates in the Example above could have been made differently. For example*

$$\begin{aligned}&|x_3| = \sqrt{|x_3|^2} \le \|x\|_2,\\ &|x_1|+2\sqrt{x_2^2+x_3^3} \le 3\|x\|_2, \text{and}\\ &|x_1|+2\sqrt{x_2^2+x_3^3} \le 2\|x\|_1 \le 2\sqrt{3}\|x\|_2.\end{aligned}$$

Go through the proof again and see what you can do.

3.3 Characterization of limits via sequences

Theorem 3.1 (Characterization of limits of $\mathbb{R}^m$-valued functions)**.**
Suppose $\Omega \subseteq \mathbb{R}^n$ *and* $p \in \mathbb{R}^n$ *be a limit point of* Ω *and let* $f : \Omega \to \mathbb{R}^m$. *Then*

$$\lim_{x\to p} f(x) = q$$

iff

$$\lim_{n\to+\infty} f(p_n) = q$$

for every sequence $(p_n) \subseteq \Omega \setminus \{p\}$ *with* $p_n \to p$ *as* $n \to +\infty$.

Remark 3.4. *The point* p *is exempt from the possible values of the sequence* (p_n) *in the above theorem since* f *might not be defined at* p.

Proof. First, we prove [⇒]: We have $\lim_{x\to p} f(x) = q$ and $(p_n) \subseteq \Omega \setminus \{p\}$ with $p_n \to p$. For all $\varepsilon > 0$, we need to find $n_0 \in \mathbb{N}$ such that

$$n \ge n_0 \quad \Rightarrow \quad \|f(p_n)-q\|_2 < \varepsilon.$$

We can choose $\delta > 0$ such that

$$0 < \|x-p\|_2 < \delta \quad \Rightarrow \quad \|f(x)-q\|_2 < \varepsilon$$

and then choose n_0 such that for all

$$\forall n \geq n_0 \quad \Rightarrow \quad \|p_n - p\|_2 < \delta \Rightarrow \quad \|f(p_n) - q\|_2 < \varepsilon.$$

Now, we prove [$\Leftarrow$]: This proof is by contrapositive.[2] Thus, we assume that we do not have

$$\forall \varepsilon > 0 \quad \exists \delta > 0 : \quad 0 < \|x - p\|_2 < \delta \quad \Rightarrow \quad \|f(x) - q\|_2 < \varepsilon \tag{3.4}$$

and conclude that we then do not have

$$\lim_{n\to+\infty} f(p_n) = q$$

for every sequence (p_n) in $\Omega \setminus \{p\}$ with $p_n \to p$ as $n \to +\infty$. The negation of (3.4) means there exists an $\varepsilon > 0$ such that for all $\delta > 0$ there exists a x_δ such that $\|x_\delta - p\|_2 < \delta$ and $\|f(x_\delta) - q\|_2 \geq \varepsilon$. Now, we choose $\delta = \frac{1}{n}$ and pick a point x_n. Then $x_n \to p$ for $n \to +\infty$ but $\|f(x_n) - q\|_2 \geq \varepsilon$. □

The last theorem can be used to find out whether limits exist. If we find two sequences with different limits, we can conclude that the function has no limit at that point. Besides the next example, we do not further discuss directional limits aside from the one dimensional case in Section 3.5.

Example 3.6. *Let $f : \mathbb{R}^2 \setminus \{0\} \to \mathbb{R}$ given by*

$$f(x_1, x_2) = \frac{x_1^2 + 3x_2^2}{x_1^2 + x_1^2}.$$

We take the sequences $[p_n, 0]^T$ and $[0, p_n]^T$ with $p_n \in [0, +\infty)$ for all $n \in \mathbb{N}_0$ and $p_n \to 0$. We have

$$\begin{aligned} f(p_n, 0) &= \frac{p_n^2 + 0}{p_n^2 + 0} = 1, \\ f(0, p_n) &= \frac{0 + 3p_n^2}{0 + p_n^2} = 3. \end{aligned}$$

Since the limits are not the same

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

does not exist.

Remark 3.5. *The last example illustrates the fact mentioned in Remark 3.1 that a control of $\|f(x) - [0, 1]^T\|_2$ in terms of $\|x\|_2$ does not always mean that*

$$\|f(x) - [0, 1]^T\|_2 \leq C\|x\|_2$$

but that there could be a function g, in this case $g(x) = \sqrt{x}$, such that

$$\|f(x) - [0, 1]^T\|_2 \leq Cg(\|x\|_2).$$

3.4 Rules for limits

The next theorem follows, as stated in the proof, from results of Analysis I. However, it is a good exercise to prove the result directly by using the definition.

[2] Note that a proof by contrapositive is not a proof by contradiction. A proof of $A \Rightarrow B$ by contradiction assumes A and $\neg B$ and derives a contradiction. A proof of $A \Rightarrow B$ by contrapositive proofs the equivalent statement $\neg B \Rightarrow \neg A$. A good introduction can be found here in Chapter 5.

Theorem 3.2 (Arithmetic rules for limits).
Let $\Omega \subseteq \mathbb{R}^n$, $p \in \mathbb{R}^n$ be a limit point of Ω and

$$f : \Omega \to \mathbb{R}, \quad g : \Omega \to \mathbb{R}.$$

Suppose

$$\lim_{x\to p} f(x) = q, \quad \lim_{x\to p} g(x) = r.$$

Then

- $\lim_{x\to p} (\lambda \cdot f)(x) = \lambda \cdot q$, *for all* $\lambda \in \mathbb{R}$,
- $\lim_{x\to p} (f+g)(x) = q + r$,
- $\lim_{x\to p} (f \cdot g)(x) = q \cdot r$, *and*
- $\lim_{x\to p} (\frac{f}{g})(x) = \frac{q}{r}$, *provided* $r \neq 0$.

Proof. This theorem follows immediately from Theorem 1.2 using Theorem 3.1. □

3.5 One sided limits of functions on $\mathbb{R}$

Definition 3.2 (Right-sided limit of f).
Let $f : (a,b) \to \mathbb{R}$ and $p \in [a,b)$. Then, we write $f(p+) = q$ if

$$\lim_{n\to+\infty} f(p_n) = q$$

for any sequence $(p_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$ with $p_n \in (p,b)$ for all $n \in \mathbb{N}_0$ and $p_n \to p$.

In the same way, we define

Definition 3.3 (Left-sided limit of f).
Let $f : (a,b) \to \mathbb{R}$ and $p \in (a,b]$. Then, we write $f(p-) = q$ if

$$\lim_{n\to+\infty} f(p_n) = q$$

for any sequence $(p_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$ with $p_n \in (a,p)$ for all $n \in \mathbb{N}_0$ and $p_n \to p$.

We may also write

$$\begin{aligned} f(p+) &= \lim_{x\to p+} f(x) = \lim_{x\to p+0} f(x), \\ f(p-) &= \lim_{x\to p-} f(x) = \lim_{x\to p-0} f(x). \end{aligned}$$

Remark 3.6. *Definition 3.2 can be rephrased as (ε,δ)-criterion in the following way: a function $f : (a,b) \to \mathbb{R}$ has a right limit at $p \in [a,b)$ iff*

$$\exists q \in \mathbb{R} \text{ s.t. } \forall \varepsilon > 0\, \exists \delta > 0 : \; 0 < h < \delta \;\Rightarrow\; |f(p+h) - q| < \varepsilon.$$

Similarly, one rephrases Definition 3.3.

We state the following result without proof.

Theorem 3.3 (Limits by one-sided limits)**.**
Let I be an open interval and $f : I \to \mathbb{R}$. Then,

$$\lim_{x\to p} f(x) = q$$

iff

$$\lim_{x\to p-} f(x) = q \quad \textit{and} \quad \lim_{x\to p+} f(x) = q.$$

Example 3.7. *Let $f : \mathbb{R}\setminus\{0\} \to \mathbb{R}$ be defined by*

$$f(x) = \frac{x}{|x|}.$$

Then $f(x) = 1$ for $x > 0$ and $f(x) = -1$ for $x < 0$. Thus,

$$\begin{aligned} f(0+) &= \lim_{x\to 0+} f(x) = 1, \\ f(0-) &= \lim_{x\to 0-} f(x) = -1. \end{aligned}$$

Example 3.8. *The function $f : \mathbb{R} \to \mathbb{R}$ given as*

$$f(x) = \begin{cases} \frac{1}{x} & : \quad x > 0 \\ 0 & : \quad x \le 0 \end{cases}.$$

The limit $f(0-)$ exists and is equal to 0 and the limit $f(0+)$ does not exist.

CHAPTER

4 Continuity

For notational remarks, please refer to the very beginning of Chapter 3.

4.1 Definition of continuity

We are now ready to state another central notion of Analysis:

Definition 4.1 (Continuity at a point of $\mathbb{R}^m$ valued functions).
Let d be a metric[a] on $\mathbb{R}^n$ and let $\Omega \subseteq \mathbb{R}^n$ with $p \in \Omega$. Then, $f : \Omega \to \mathbb{R}^m$ is called continuous at p iff

$$\forall \varepsilon > 0 \quad \exists \delta > 0 : \quad \big(x \in \Omega : d(x,p) < \delta\big) \quad \Rightarrow \quad d(f(x), f(p)) < \varepsilon. \tag{4.1}$$

[a]See Definition 2.2.

One immediate consequence of the definition is that functions defined on some $\Omega \subseteq \mathbb{R}^n$ are continuous at isolated points. See Definition 2.5. To see that let $\Omega \in \mathbb{R}^n$ have at least one isolated point (for example $\Omega = \{p\}$, $p \in \mathbb{R}^n$) and $f : \Omega \to \mathbb{R}^m$ a function. Pick an isolated point p of Ω. Then, there exists $\delta > 0$ such that $B_\delta(p) \cap \Omega = \{p\}$. Thus, for every $\varepsilon > 0$, one can choose this δ so that $\|x - p\|_2 < \delta$ is only satisfied for $x = x_0$. Thus, we obtain

$$\|f(x) - f(p)\|_2 = \|f(x_0) - f(x_0)\|_2 = 0 < \varepsilon.$$

Also remember the following important remark.

Remark 4.1. *A function f can only be continuous where it is defined, i.e it makes no sense to ask about the continuity of a function at points p where $f(p)$ does not exist. For example, the function*

$$f(x) = \begin{cases} -1 & : \quad x < 0 \\ 1 & : \quad x > 0 \end{cases}$$

is continuous on $\mathbb{R} \setminus \{0\}$.

Remark 4.2. *If $d(x,y) = \|x - y\|$, then we can rewrite (9.13) as*

$$\forall \varepsilon > 0 \quad \exists \delta > 0 : \quad \|x - p\| < \delta \quad \Rightarrow \quad \|f(p) - f(x)\| < \varepsilon.$$

The d or $\|\cdot\|$ may be any of the metrics/norms in Example 2.5/2.3 respectively.

Exercise 4.1. *Set $m = n = 1$ and write Definition 4.1 down explicitly (with the correct distances) in that case.*

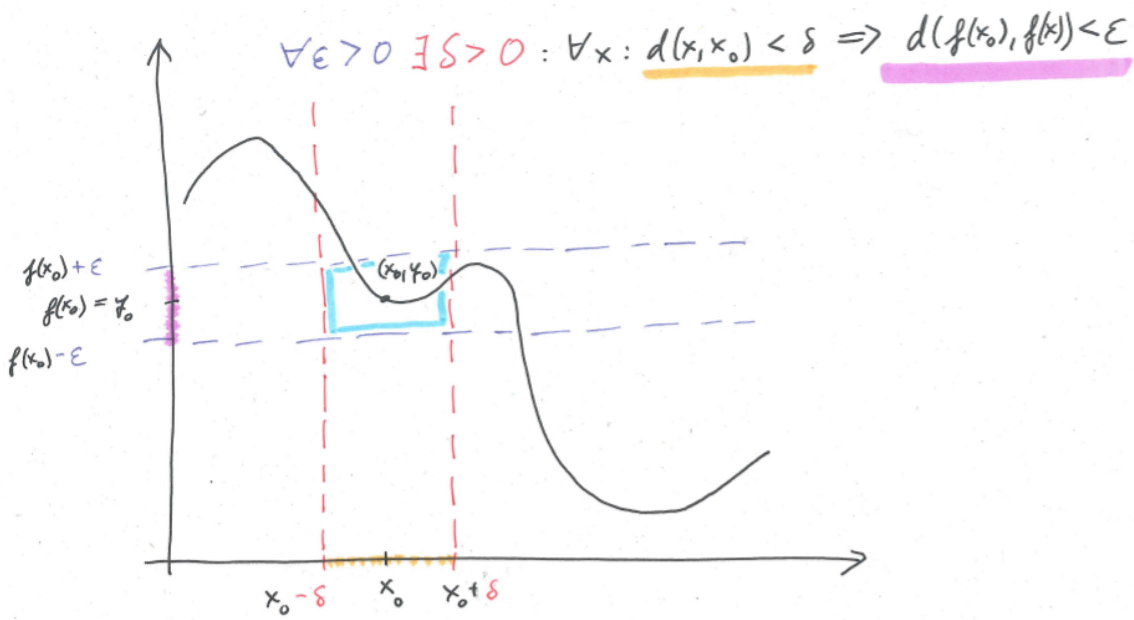


Figure 4.1: Illustration of the definition of continuity for $m = n = 1$.

Remark 4.3. *We can rewrite*

$$d(x, p) < \delta \quad \Rightarrow \quad d(f(x), f(p)) < \varepsilon$$

as

$$x \in B_\delta(p) \quad \Rightarrow \quad f(x) \in B_\varepsilon(f(p))$$

or

$$f(B_\delta(p)) \subseteq B_\varepsilon(f(p)). \tag{4.2}$$

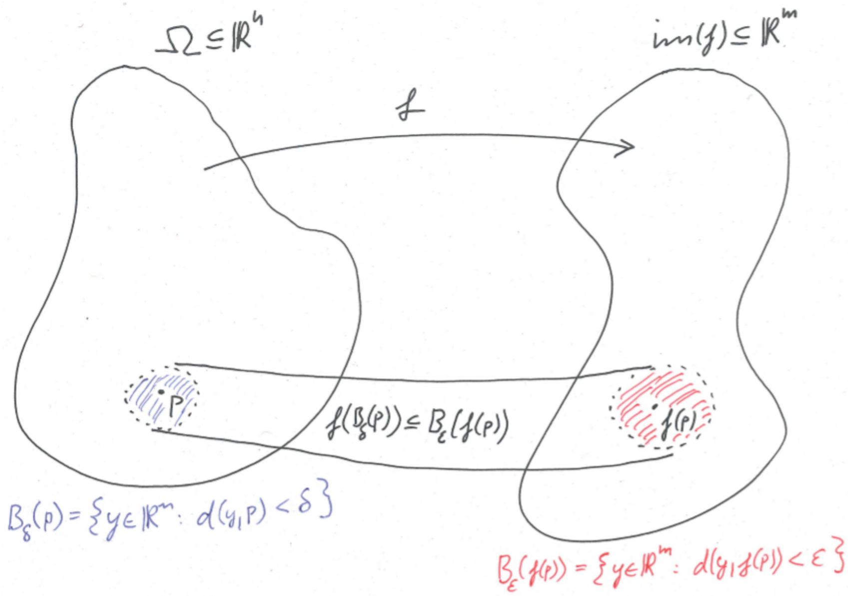


Figure 4.2: Illustration of continuity. Especially (4.2).

Definition 4.2 (Continuity of $\mathbb{R}^m$ valued functions)**.**
Let $\Omega \subseteq \mathbb{R}^n$. Then, the function $f : \Omega \to \mathbb{R}^m$ is said to be continuous[a] *iff it is continuous at all $p \in \Omega$.*[b]

[a] Sometimes we will also say f is continuous on Ω.
[b] See Definition 4.1.

Using Definition 3.1 and Theorem 3.1, it holds immediately

Theorem 4.1 (Limit characterization of continuity)**.**
Let $\Omega \subseteq \mathbb{R}^n$, $f : \Omega \to \mathbb{R}^m$, and $p \in \Omega$ limit point of Ω. Then, f is continuous at p iff

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

and iff for all $(p_n)_{n \in \mathbb{N}_0} \subseteq \Omega \setminus \{p\}$ with $p_n \to p$ holds

$$\lim_{n \to +\infty} f(p_n) = f(p).$$

4.2 Examples

Example 4.1. *Consider $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$. We show that this function is continuous at all $x_0 \in \mathbb{R}$. We need to show that*

$$\forall x_0 \in \mathbb{R}\ \forall \varepsilon > 0\ \exists \delta > 0 : \big(\forall x \in \mathbb{R},\ d(x, x_0) < \delta\big) \ \Rightarrow\ d(f(x), f(x_0)) < \varepsilon$$

which means that

$$\forall x_0 \in \mathbb{R}\ \forall \varepsilon > 0\ \exists \delta > 0 : \big(\forall x \in \mathbb{R},\ |x - x_0| < \delta\big) \ \Rightarrow\ |f(x_0) - f(x)| < \varepsilon.$$

We have

$$|x^2 - x_0^2| = |(x - x_0)||(x + x_0)|.$$

If we choose $\delta < 1$, we have $|x - x_0| < 1$ and therefore $|x| < 1 + |x_0|$. Thus, $|x + x_0| < 2|x_0| + 1$ by the triangle inequality and, hence,

$$|x^2 - x_0^2| \leq (2|x_0| + 1)|x - x_0|.$$

Finally, $\delta := \min\left\{1, \frac{1}{2|x_0|+1}\right\}$ concludes the argument.

The continuity of $f(x) = x^2$ in the last example could be proved slightly differently. See the next example.

Example 4.2. *Consider $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$. We show that this function is continuous at all $x_0 \in \mathbb{R}$. Let us set $x = x_0 + \delta'$ with $\delta' \in (-\delta, \delta)$. Then, $d(x_0, x) = |\delta'| < |\delta|$. Now, we get*

$$\begin{aligned} d(f(x), f(x_0)) &= |f(x) - f(x_0)| = |(x_0 + \delta')^2 - x_0^2| \\ &= |2x_0\delta' + \delta'^2| = |\delta'||2x_0 + \delta'| \\ &\leq |2x_0 + 1||\delta'|, \end{aligned}$$

where we require $\delta \leq \frac{1}{2}$ and finally have to choose

$$\delta = \min\left\{1, \frac{1}{2|x_0| + 1}\right\}.$$

Example 4.3. *Consider $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 2x^2 - 8x + 6$. Let us prove continuity everywhere. Thus, let $p \in \mathbb{R}$ and $\varepsilon > 0$. Then, we get*

$$\begin{aligned} d(f(x), f(p)) = |f(x) - f(p)| &= |2x^2 - 8x + 6 - (2p^2 - 8p + 6)| = |2(x^2 - p^2) + 8(p - x)| \\ &\leq 2|x^2 - p^2| + 8|p - x| = 2|x + p||x - p| + 8|p - x| \\ &\leq 2(|x| + |p|)|x - p| + 8|p - x,| \end{aligned}$$

where we used $|x + p| \leq |x| + |p|$. Now, setting $\delta \leq 1$, we get $|x + p| \leq 1 + |p|$. With that, we have

$$d(f(x), f(p)) = |f(x) - f(p)| \leq (2(1 + |p|) + 8)|x - p| < \varepsilon$$

if we choose

$$\delta = \min\left\{1, \frac{\varepsilon}{2(1 + |p|) + 8}\right\}.$$

Example 4.4. *Let $f : [0, \infty) \to \mathbb{R}$ with $f(x) = \sqrt{x}$. We prove that f is continuous on $\mathbb{R}$. Let $p = 0$ and $\varepsilon > 0$. We get*

$$|x| < \delta \quad \Rightarrow \quad |\sqrt{x}| = \sqrt{|x|} < \varepsilon$$

if we choose $\delta = \sqrt{\varepsilon}$. Now let $p \in (0, +\infty)$ and $\varepsilon > 0$. We estimate

$$\begin{aligned} |\sqrt{x} - \sqrt{p}| &= \left|(\sqrt{x} - \sqrt{p})\frac{\sqrt{x} + \sqrt{p}}{\sqrt{x} + \sqrt{p}}\right| \leq \frac{|x - p|}{\sqrt{x} + \sqrt{p}} \\ &\leq \frac{|x - p|}{\sqrt{p}}. \end{aligned}$$

Thus, we get

$$|x - p| < \delta \quad \Rightarrow \quad |\sqrt{x} - \sqrt{p}| < \varepsilon$$

for $\delta = \sqrt{p}\varepsilon$.

Exercise 4.2. *Consider $f : \mathbb{R}^3 \to \mathbb{R}^2$ given by*

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto \begin{bmatrix} \frac{x_1 + 2\sqrt{x_2^2 + x_3^2}\sin(x_3)}{(x_1^2 + x_2^2 + 2x_3^2)^{\frac{1}{4}}} \\ x_3 + 1 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \mapsto \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Show that f is continuous at $x = 0$.

4.3 Discontinuity

Definition 4.3 (Discontinuous at a point of $\mathbb{R}^m$ valued functions)**.**
Let $\Omega \subseteq \mathbb{R}^n$ and $p \in \Omega$. Then, $f : \Omega \to \mathbb{R}^m$ is called discontinuous at p iff f is not continuous at p.

Example 4.5. *The function $f : R \to \mathbb{R}$, given by*

$$f(x) = \begin{cases} 1 & : \quad x \neq 0 \\ 0 & : \quad x = 0 \end{cases}$$

is not continuous at $x = 0$ since $f(0) = 0$ but

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

The following example uses one-sided limits from Section 3.5.

Example 4.6. *The function* $\mathrm{sgn}:\mathbb{R}\to\mathbb{R}$*, defined by*

$$\mathrm{sgn}(x)=\begin{cases}1 & : & x>0\\ 0 & : & x=0\\ -1 & : & x<0\end{cases}$$

is discontinuous at $x=0$ *since* $\mathrm{sgn}(0)=0$ *but*

$$\lim_{x\to 0-}\mathrm{sgn}(x)=-1 \quad \textit{and} \quad \lim_{x\to 0+}\mathrm{sgn}(x)=1.$$

Example 4.7. *The Dirichlet-function (which is the characteristic function*[1] *of* $\mathbb{Q}$*), given by*

$$\chi_{\mathbb{Q}}(x)=\begin{cases}1 & : & x\in\mathbb{Q}\\ 0 & : & x\in\mathbb{R}\setminus\mathbb{Q}\end{cases},$$

is nowhere continuous.

Exercise 4.3. *Prove that the function* $\chi_{\mathbb{Q}}$ *from the example above is nowhere continuouss.*

Example 4.8. *The function* $f:\mathbb{R}\to\mathbb{R}$*, given by*

$$f(x)=\begin{cases}\sin\left(\frac{1}{x}\right) & : & x\neq 0\\ 0 & : & x=0\end{cases}$$

is not continuous at $x=0$ *as*

$$\lim_{x\to 0}f(x)$$

does not exist.

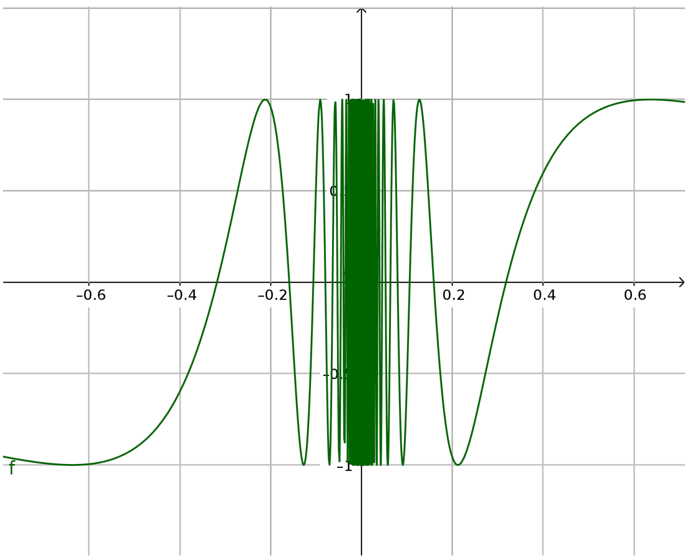


Figure 4.3: The function $f(x)=\sin\left(\frac{1}{x}\right)$.

[1] Also called indicator function.

4.3.1 Further counterexamples in continuity

Example 4.9 (Dirichlet function)**.**
We consider the function $\chi_{\mathbb{Q}} : \mathbb{R} \to \mathbb{R}$ *defined by*

$$\chi_{\mathbb{Q}}(x) = \begin{cases} 1 & : \quad x \in \mathbb{Q} \\ 0 & : \quad x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

This function is at no point continuous. If we let $f : \mathbb{R} \to \mathbb{R}$ *be a function which is continuous and has zeros at* $x_1, \dots, x_n$*, then* $f(x)\chi_{\mathbb{Q}}(x)$ *is continuous at the points* $x_1, \dots, x_n$*.*

Example 4.10. *We consider the function* $f : [0,1] \to \mathbb{R}$ *defined by*

$$f(x) = \begin{cases} 1 & : \quad x = 0 \\ 1/n & : \quad x = \frac{m}{n} \in \mathbb{Q} \\ 0 & : \quad x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

where the $\frac{m}{n}$ *is always considered to be in lowest terms. This function is discontinuous at* $x \in \mathbb{Q}$ *and continuous at* $x \in \mathbb{R} \setminus \mathbb{Q}$*.*

Example 4.11. *We consider the function* $f : \mathbb{R} \to \mathbb{R}$ *defined by*

$$f(x) = \begin{cases} x & : \quad x \in \mathbb{Q} \\ -x & : \quad x \in \mathbb{R} \setminus \mathbb{Q} \end{cases} .$$

This function is only at $x = 0$ *continuous.*

Example 4.12. *We consider the function* $f : \mathbb{R} \to \mathbb{R}$ *with*

$$f(x) = \begin{cases} 1 & : \quad x \in \mathbb{Q} \\ -1 & : \quad x \in \mathbb{R} \setminus \mathbb{Q} \end{cases} .$$

This function is nowhere differentiable but its absolute value $|f|(x) \equiv 1$ *is.*

Example 4.13. *For a function* $f : \mathbb{R}^2 \to \mathbb{R}$ *it is not enough to be continuous in each variable to be continuous. We consider*

$$f(x,y) = \begin{cases} \frac{xy}{x^2+y^2} & : \quad x^2 + y^2 \neq 0 \\ 0 & : \quad x = y = 0 \end{cases} .$$

In every disc $B_\varepsilon(0)$ *exist points of the form* (a,a) *at which* f *has the value* $\frac{1}{2}$*. For every fixed value of* y*, say* $y_0 \in \mathbb{R}$*, the function* $g(x) := f(x, y_0)$ *is continuous. Similarly, the function* $h(y) := f(x_0, y)$ *is continuous for every fixed* $x_0 \in \mathbb{R}$*.*

Exercise 4.4. *Prove the claimed properties in the above examples where the details have been left open.*

4.4 Continuity and component-wise continuity

We show now a result connecting the continuity of f with the component functions f_k. Remember that $f : \Omega \to \mathbb{R}^m$ means that

$$\mathbb{R}^n \supseteq \Omega \ni x \mapsto f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix} = \begin{bmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_m(x_1, \dots, x_n) \end{bmatrix} \in \mathbb{R}^m.$$

We have

Theorem 4.2 (Component-wise continuity at a point).
Let $\Omega \subseteq \mathbb{R}^n$, $f : \Omega \to \mathbb{R}^m$ *with*

$$f(x) = \Big[f_1(x), f_2(x), \dots, f_m(x)\Big]^T.$$

Then, f *is continuous at* $p \in \Omega$ *iff the function* $f_k(x)$ *is continuous at* p *for all* $k \in \{1, \dots, m\}$.

Proof. This proof is an exercise (or see your lecture notes from class). All you need are the definitions and the inequalities

$$\|x\|_2 \le \|x\|_1 \le \sqrt{n}\|x\|_2, \quad x \in \mathbb{R}^n.$$

□

Remark 4.4. *Theorem 4.2 is useful to decide whether functions like*

$$f(x) = \begin{bmatrix} |x_1 - 1| \\ |x_2 - 2| + |3x_3 - 5| \\ |x_3 - 3| \end{bmatrix}$$

is continuous at a point $x_0 \in \mathbb{R}^3$ *as one has only to check the the component functions*

$$\begin{aligned} f_1(x_1, x_2, x_3) &= |x_1 - 1|, \\ f_2(x_1, x_2, x_3) &= |x_2 - 2| + |3x_3 - 5|, \text{ and} \\ f_3(x_1, x_2, x_3) &= |x_3 - 3| \end{aligned}$$

which might be easier.

4.5 Operations with continuous functions

Theorem 4.3 (Arithmetic properties of continuous functions).
Let $\Omega \subseteq \mathbb{R}^n$, $p \in \Omega$. *Consider* $f, g : \Omega \to \mathbb{R}$ *which are continuous (at* p*). Then*

- $f + g$ *is continuous (at* p*),*
- $f \cdot g$ *is continuous (at* p*), and*
- $\frac{f}{g}$ *is continuous (at* p*).*

Proof. Using limits, the proof follows from results of Analysis I (Theorems 3.1 together with Theorems 4.1 and 1.2) whenever p is a limit point of Ω. In isolated points, the result is easy to prove. □

Exercise 4.5. *Let* $\Omega \subseteq \mathbb{R}^n$, $p \in \Omega$. *Consider* $f, g : \Omega \to \mathbb{R}^m$ *which are continuous (at* p*). Prove that* $f + g$ *is continuous (at* p*).*

Theorem 4.4 (Composition of continuous functions).
Let $\Omega_1 \subseteq \mathbb{R}^n$, $\Omega_2 \subseteq \mathbb{R}^m$ *and*

$$\begin{aligned} f &: \Omega_1 \to \Omega_2, \\ g &: \Omega_2 \to \mathbb{R}^k. \end{aligned}$$

Further suppose that f is continuous at $p \in \Omega_1$ and g is continuous at $f(p) \in \Omega_2$. Then, $g \circ f$ is continuous at p.

Exercise 4.6. *Before you read the proof have a go yourself. You only need the definition of continuity. Draw a picture similar to Figure 4.2 to get an idea what to do.*

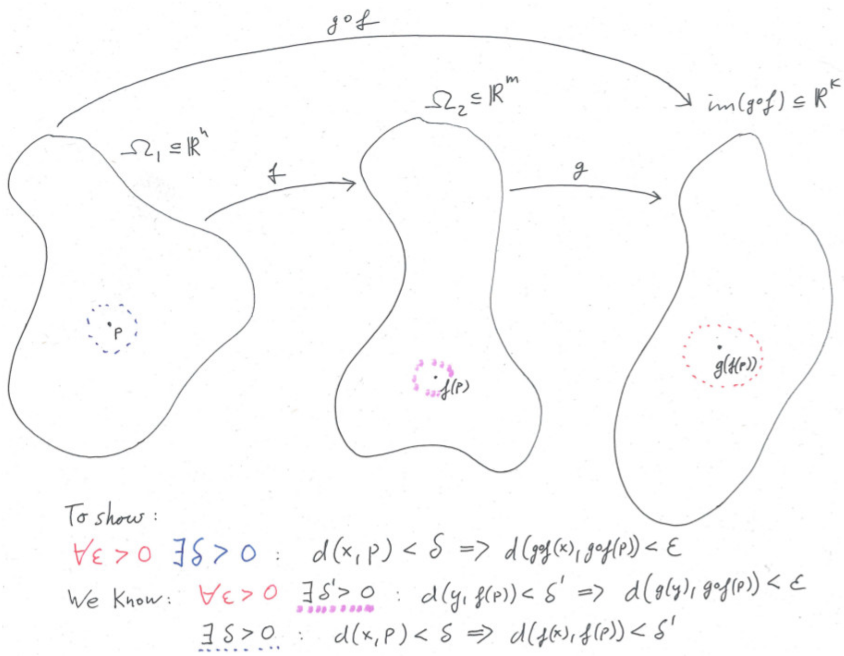


Figure 4.4: A graphical analysis of the composition theorem Theorem 4.4 and ist proof.

Proof. We need to show that for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\|x - p\|_2 < \delta \quad \Rightarrow \quad \|g(f(x)) - g(f(p))\|_2 < \varepsilon.$$

Let $\varepsilon > 0$. Since g is continuous at $f(p)$, we can find a $\delta' > 0$ such that

$$\|y - f(p)\|_2 < \delta' \quad \Rightarrow \quad \|g(y) - g(f(p))\|_2 < \varepsilon.$$

Since f is continuous at p, we can now find a $\delta > 0$ such that

$$\|x - p\|_2 < \delta \quad \Rightarrow \quad \|f(x) - f(p)\|_2 < \delta'.$$

Hence, we have that

$$\|x - p\|_2 < \delta \quad \Rightarrow \quad \|g(f(x)) - g(f(p))\|_2 < \varepsilon.$$

□

4.6 Continuous functions $f : [a, b] \to \mathbb{R}$

The purpose of this section is to introduce and prove two important theorems for continuous functions

$$f : [a, b] \to \mathbb{R}$$

of one variable. These theorems are called the *Intermediate value theorem* (IVT) and the *Extreme value theorem* (EVT) which is also commonly called *Theorem of Weierstrass*.

4.6.1 The Intermediate Value Theorem (IVT)

Theorem 4.5 (Intermediate value theorem (IVT))**.**
Suppose $f : [a,b] \to \mathbb{R}$ is continuous and $y_0 \in \mathbb{R}$ is between $f(a)$ and $f(b)$. Then there exists a $x_0 \in [a,b]$ such that $f(x_0) = y_0$.

Remark 4.5. *The converse of the IVT is not true. There are discontinuous functions $f : [a,b] \to \mathbb{R}$ which have the property that for any value y_0 between $f(a)$ and $f(b)$ there exists an $x_0 \in [a,b]$ such that $f(x_0) = y_0$. An example is given by $f : [0,1] \to \mathbb{R}$ defined by*

$$f(x) = \begin{cases} 2x & : \quad x \in [0, \frac{1}{2}] \\ x & : \quad x \in (\frac{1}{2}, 1] \end{cases}.$$

The function clearly takes any value between $f(0) = 0$ and $f(1) = 1$ but is not continuous. (Draw a picture.) See also the next remark.

Remark 4.6. *The assumption of continuity can in general not be dropped from Theorem 4.5. An example is the function $f : [0,1] \to \mathbb{R}$ defined by*

$$f(x) = \begin{cases} 0 & : \quad x \in [0, \frac{1}{2}] \\ 1 & : \quad x \in (\frac{1}{2}, 1] \end{cases}.$$

There is no value $y \in (0,1)$ for which there is a $x \in [0,1]$ such that $f(x) = y$.

Remark 4.7. *As we will see in the proof, the fact that $[a,b]$ is a closed interval of real numbers is important as well since the real numbers are complete[2]. If we look at continuous functions on $\mathbb{Q}$, i.e. $f : [a,b] \cap \mathbb{Q} \to \mathbb{R}$ which are continuous, then we do not have the intermediate value property. For example, the function $f(x) = x^2$ on $[1,2] \cap \mathbb{Q}$ does not take all values between 1 and 2 as there is no $x \in [1,2] \cap \mathbb{Q}$ such that $f(x) = 2$ since $\sqrt{2}$ is irrational.*

To prove the Intermediate Value Theorem, we need a Lemma.

Lemma 4.1.
Suppose $f : [a,b] \to \mathbb{R}$ is continuous at x_0, and $f(x_0) > 0$. Then there exists a $\delta > 0$ such that $f(x) > 0$ for $x \in (x_0 - \delta, x_0 + \delta)$.

Proof. Since f is continuous at x_0, we have that for all $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|x - x_0| < \delta \quad \Rightarrow \quad -\varepsilon < f(x) - f(x_0) < \varepsilon.$$

Thus, since $f(x_0) > 0$, we get, choosing $\varepsilon = \frac{f(x_0)}{2}$ that

$$-\frac{f(x_0)}{2} < f(x) - f(x_0) < \frac{f(x_0)}{2} \quad \Leftrightarrow \quad \frac{f(x_0)}{2} < f(x) < \frac{3}{2} f(x_0)$$

and hence $f(x) > 0$ for $x \in (x_0 - \delta, x_0 + \delta)$. ☐

Now, we are ready to prove the IVT.

[2] Remember that means that Cauchy sequences are convergent.

Proof of Theorem 4.5. First, without loss f generality, we can assume that $f(a) < y_0 < f(b)$. Then, we define the function h by

$$h(x) = f(x) - y_0.$$

From that, we get $h(a) < 0$ and $h(b) > 0$ and we need to show that there is a $x_0 \in [a, b]$ such that $h(x_0) = 0$.

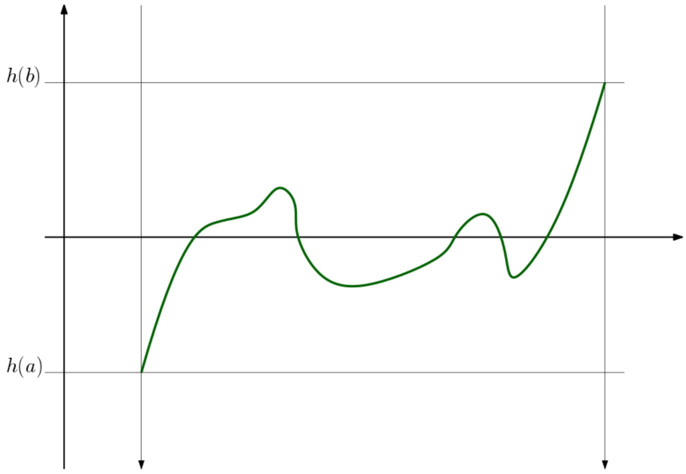


Figure 4.5: Illustration of the proof. Possible picture of h. Compare with the construction described afterwards.

We define the following partition of $[a, b]$:

$$[a, b] = X_- \cup X_0 \cup X_+,$$

where

$$\begin{aligned} X_- &= \{y \in [a, b] : h(y) < 0\}, \\ X_+ &= \{y \in [a, b] : h(y) > 0\}, \quad \text{and} \\ X_0 &= \{y \in [a, b] : h(y) = 0\}. \end{aligned}$$

We have $a \in X_-$ and $b \in X_+$ and are left to show that $X_0 \neq \emptyset$. Since $\mathbb{R}$ is complete, we know that there is a x_0 such that $x_0 = \sup X_-$. Since $X_- \subseteq [a, b]$, we have that $x_0 \in [a, b]$. We will now show that x_0 belongs neither to X_- nor to X_+ which means it belongs to X_0 which concludes the proof. First, from Lemma 4.1, we have that $x_0 \notin X_-$. If it were, there would be $\delta > 0$ such that $(x_0, x_0 + \delta) \subseteq X_-$, i.e. $x_0 + \frac{\delta}{2} \in X_-$ and this contradicts that $x_0 = \sup X_-$. Now, $x_0 \notin X_+$ by Lemma 4.1 since if it were, there would be a $\delta > 0$ such that $(x_0 - \delta, x_0) \subseteq X_+$. This would imply that, for instance, $x_0 - \frac{\delta}{2}$ is an upper bound of X_-. This contradicts that $x_0 = \sup X_-$ is the lowest upper bound of X_-. □

4.6.2 Applications of the IVT

Example 4.14. *Prove that $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = x^5 - x^4 + x^3 - x^2 + x + 1$ has a zero between -1 and 0. To prove that we compute $f(-1) = -4$ and $f(0) = 1$. By the IVT (see Theorem 4.5), we obtain that there must be a $x_0 \in [-1, 0]$ such that $f(x_0) = 0$.*

Example 4.15. *We show that $f : \mathbb{R} \to \mathbb{R}$ with*

$$f(x) = x^3 + \frac{2}{1 + x^2}$$

is subjective.[3] *To show that, we compute the limits for $x \to \pm\infty$:*

$$\lim_{x\to-\infty} f(x) = -\infty, \quad \lim_{x\to\infty} f(x) = \infty.$$

This means that

$$\forall c < 0\, \exists M < 0 : x < M \quad \Rightarrow \quad f(x) < C$$

and

$$\forall C > 0\, \exists M > 0 : x > M \quad \Rightarrow \quad f(x) > C.$$

Thus, given any $y_0 \in \mathbb{R}$, we can choose $M > 0$ such that $f(-M) < y_0$ and $f(M) > y_0$. Then, by the IVT (see Theorem 4.5), there exists $x_0 \in [-M, M]$ such that $f(x_0) = y_0$.

Theorem 4.6 (Brouwer's Fixed Point Theorem)**.**
Let $f : [a, b] \to [a, b]$ be continuous. Then there exists a $x_0 \in [a, b]$ such that

$$f(x_0) = x_0.$$

Definition 4.4 (Fixed point)**.**
Let $I \subseteq \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$ be a function. Then, $x_0 \in I$ is called a fixed point of f iff $f(x_0) = x_0$.

Remark 4.8. *Brouwer's fixed point theorem asserts that a continuous function $f : [a, b] \to [a, b]$ possesses at least one fixed point.*

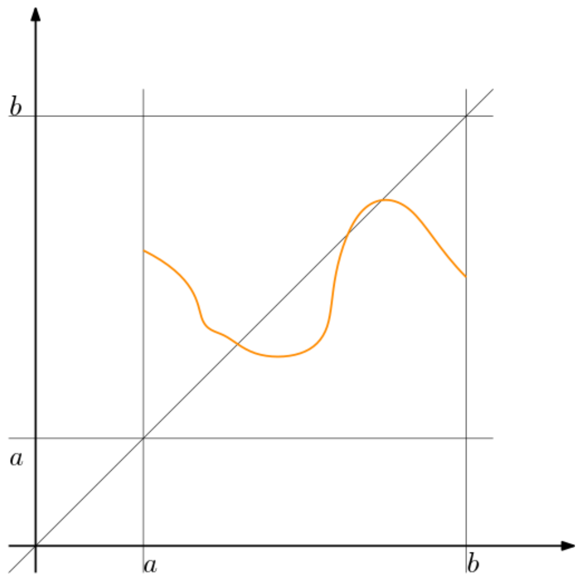


Figure 4.6: Illustration of the Brouwer's Fixed Point Theorem.

Proof. If $f(a) = a$ or $f(b) = b$, we have nothing to prove. Thus, we assume that $f(a) > a$ and $f(b) < b$. We want to use the intermediate value theorem and define the function

$$h(x) = x - f(x)$$

[3]That means that for all y in the co-domain $\mathbb{R}$ there exists an x in the domain $\mathbb{R}$ such that $f(x) = y$.

which has the properties $h(a) < 0$ and $h(b) > 0$. Thus, by the IVT, there exists a $x_0 \in [a, b]$ such that $h(x_0) = 0$ which implies

$$f(x_0) = x_0.$$

□

4.6.3 Weierstrass' Extremal Value Theorem

We introduce the notions of local and global maximum.

Definition 4.5 (Local maximum).
Let $I \subseteq \mathbb{R}$ and $f : I \to \mathbb{R}$ be a function. Then, $x_0 \in I$ is called a local maximum of f if there exists a $\delta > 0$ such that

$$f(x) \leq f(x_0) \quad x \in (x_0 - \delta, x_0 + \delta).$$

Definition 4.6 (Global maximum).
Let $I \subseteq \mathbb{R}$ and $f : I \to \mathbb{R}$ be a function. Then, $x_0 \in I$ is called a global minimum if

$$f(x) \leq f(x_0) \quad x \in I.$$

Remark 4.9. *Local/global minima are defined similarly.*

Exercise 4.7. *Write down the according definition for a local minimum and a global minimum for a function $f : I \to \mathbb{R}$, where $I \subseteq \mathbb{R}$ is an interval.*

Exercise 4.8. *Draw pictures and illustrate the above defined notions of local and global maximum/minimum. Convince yourself that neither local not global maxima/minima must be unique.*

We also introduce the notion of a bounded function. This is essentially a recap of results of Section 3.3.1 of Analysis I (Bounded sets).

Definition 4.7 (Bounded from above).
Let $I \subseteq \mathbb{R}$. Then, a function $f : I \to \mathbb{R}$ is called bounded from above if there exists a constant $C > 0$ such that for all $f(x) \leq C$ for all $x \in I$.

Definition 4.8 (Bounded from below).
Let $I \subseteq \mathbb{R}$. Then, a function $f : I \to \mathbb{R}$ is called bounded from below if $-f$ is bounded from above.

Definition 4.9 (Bounded).
Let $I \subseteq \mathbb{R}$. Then, a function $f : I \to \mathbb{R}$ is called bounded if f is bounded from above and bounded from below.

Remark 4.10. *Let us set $f(I) := \{f(x) : x \in [a, b]\}$. Then, the boundedness of f is equivalent to the boundedness of the set $f(I)$, i.e.*

$$\sup(f(I)) < +\infty \quad \textit{and} \quad \inf(f(I)) > -\infty.$$

If only one of the two holds, we have the function bounded above or below respectively.

The next important theorem is Weierstrass' Extreme Value Theorem.

Theorem 4.7 ((Weierstrass') Extreme Value Theorem (EVT))**.**
Suppose that $f : [a,b] \to \mathbb{R}$ *is continuous. Then,*

(i) f *is bounded.*

(ii) f *attains its maximal/minimal values, i.e. there exists a global maximum/minimum.*

Remark 4.11. *The converse of the EVT is not true. There exist functions* $f : [a,b] \to \mathbb{R}$ *which are bounded and attain their maximum and minimum but are not continuous. An example is given by* $f : [0,1] \to \mathbb{R}$ *defined by*

$$f(x) = \begin{cases} 2x & : \quad x \in [0, \frac{1}{2}] \\ \frac{1}{2} & : \quad x \in (\frac{1}{2}, 1] \end{cases}.$$

The minimum value 0 *and the maximum value* 1 *are attained at* $x = 0$ *and* $x = \frac{1}{2}$ *respectively but* f *is not continuous. (Draw a picture!) See also Remarks 4.13, 4.14, and 4.15.*

To prove that theorem, we need to prepare ourselves with another Lemma.

Lemma 4.2.
Suppose $f : [a,b] \to \mathbb{R}$ *is continuous at* x_0*. Then there exists a* $\delta > 0$ *and a constant* $C > 0$ *such that* $|f(x)| \leq C$ *for all* $x \in [x_0 - \delta, x_0 + \delta]$*.*

Proof. Since f is continuous at x_0, we have that for all $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|x - x_0| < \delta' \quad \Rightarrow \quad -\varepsilon < f(x) - f(x_0) < \varepsilon.$$

Setting $\varepsilon = 1$, we get that

$$f(x_0) - 1 < f(x) < 1 + f(x_0) \quad \forall x \in (x_0 - \delta, x_0 + \delta).$$

Now, we choose $\delta = \frac{\delta'}{2}$ and get that $[x_0 - \delta, x_0 + \delta] \subseteq (x_0 - \delta, x_0 + \delta)$. Thus, we have that there exists a $\delta > 0$ such that f is bounded on $[x_0 - \delta, x_0 + \delta]$ as claimed. □

Now we prove the Extremal value theorem.

Proof. First, we prove (i). We define

$$X = \{x \in [a,b] : f \text{ is bounded on } [a,x]\}.$$

Clearly $a \in X$, i.e. $X \neq \emptyset$. Also, X is clearly bounded by b. By the completeness of $\mathbb{R}$, we have $s = \sup X \in [a,b]$. By Lemma 4.2, we get $s > a$ and that $s \notin [a,b)$. To see that, we assume that $s \in [a,b)$. Then, by Lemma 4.2, we can find a $\delta > 0$ such that f is bounded on $[a,s]$ and also on $[a, s+\delta]$ since f is then bounded on $[s-\delta, s+\delta]$. This contradicts the fact that $s = \sup X$ is the least upper bound of X. Thus, $s = b$ and it remains to show that $b \in X$. We have shown so far that f is bounded on all $[a, b-\delta]$ for all $\delta \in (0, b-a)$. Since f is continuous at b, we have that, again by Lemma 4.2 there is $\delta' > 0$ such that f is bounded on $[b-\delta', b]$. Hence there is a $\delta > 0$ such that $[a, b-\delta] \cap [b-\delta', b] \neq \emptyset$; for $\delta < \delta'$. Thus, $b \in X$.

Now, we attend to part (ii). Since, by part (i), f is bounded on $[a,b]$ there exists, by completeness of $\mathbb{R}$ a $y_0 = \sup\{f(x) : x \in [a,b]\} =: X$. We have to show that there exists a $x_0 \in [a,b]$ such that $f(x_0) = y_0$. We assume that such an x_0 does not exist. Then, the function

$$g(x) := \frac{1}{y_0 - f(x)}$$

is everywhere defined and continuous. Thus, g is bounded, i.e. there exists an $M > 0$ such that

$$|g(x)| \leq M.$$

This means, by the definition of g that

$$\begin{aligned} \tfrac{1}{y_0 - f(x)} \leq M \quad &\Leftrightarrow \\ y_0 - f(x) \geq \tfrac{1}{M} \quad &\Leftrightarrow \\ f(x) \leq y_0 - \tfrac{1}{M} & \end{aligned}$$

which contradicts that $y_0 = \sup X$. For the minimum, we argue similarly. This concludes the proof. □

Remark 4.12. *One can prove the boundedness of a continuous function $f : [a,b] \to \mathbb{R}$ with the help of Bolzano–Weierstrass (see Theorem 1.3). Let us assume that f is not bounded from above on $[a,b]$. Then there exists a x_1 such that $f(x_1) > 1$ and x_2 such that $f(x_2)$, etc. The sequence $(x_n)_{n\in\mathbb{N}} \subseteq [a,b]$ is bounded. Thus, by Bolzano–Weierstrass, there exists a convergent subsequence $(x_{n_k})_{k\in\mathbb{N}} \subseteq (x_n)_{n\in\mathbb{N}}$ with limit $x_0 \in [a,b]$. By construction, we have $f(x_{n_k}) \to +\infty$. This contradicts continuity as the limit should be $f(x_0)$. The proof that f is bounded below proceeds similarly.*

Remark 4.13. *The closedness of the interval in the statement of the EVT is essential. For instance, consider $f : (0,1) \to \mathbb{R}$ with $f(x) = \frac{1}{x}$. The function is continuous but not bounded on $(0,1)$.*

Remark 4.14. *The continuity of the function f in the statement of the EVT is essential. For instance, consider $f : [0,1] \to \mathbb{R}$ with*

$$f(x) = \begin{cases} \frac{(-1)^n n}{n+1} & : \quad x = \frac{m}{n} \in \mathbb{Q} \\ 0 & : \quad x \in \mathbb{R} \setminus \mathbb{Q} \end{cases},$$

where $\frac{m}{n}$ in the definition of the function is regarded to be in lowest terms. In every neighborhood of every point in $[0,1]$, the values of f come arbitrarily close to the numbers -1 and 1 but always stay strictly between them.

Remark 4.15. *Consider $f : [0,1] \to \mathbb{R}$ given by*

$$f(x) = \begin{cases} n & : \quad x = \frac{m}{n} \in [0,1] \cap \mathbb{Q} \\ 0 & : \quad x \in [0,1] \setminus \mathbb{Q} \end{cases},$$

where $\frac{m}{n}$ is in lowest terms. This function is finite in every point but not bounded on $[0,1]$ and thus shows again that continuity can not be dropped from Theorem 4.7. To see that assume that there is a $x_0 \in [0,1]$ such that f is bounded on $[x_0 - \delta, x_0 + \delta]$. (If $x = a$ or $x = b$ we consider the appropriate "half"-interval.) Then, the denominators of $x \in [x_0 - \delta, x_0 + \delta] \cap \mathbb{Q}$ must be bounded as well as the numerators. However, this means there are only finitely many rational elements in $[x_0 - \delta, x_0 + \delta]$ which is not true. Thus, f is not bounded on any Interval $I \subseteq [0,1]$.

4.7 Continuity of linear maps

We convince ourselves that a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ is continuous. For the definition of Linear Map see Definition 1.12. Since $Lx - Lx_0 = L(x - x_0)$, it is sufficient to prove that L is continuous at $x_0 = 0$.
Let us denote

$$L = \begin{bmatrix} l_{11} & \dots & l_{1n} \\ \vdots & & \vdots \\ l_{m1} & \dots & l_{mn} \end{bmatrix}.$$

With that, we obtain

$$L = \begin{bmatrix} l_{11} & \dots & l_{1n} \\ \vdots & & \vdots \\ l_{m1} & \dots & l_{mn} \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n l_{1j} h_j \\ \vdots \\ \sum_{j=1}^n l_{mj} h_j \end{bmatrix}.$$

We want to show that for all $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\|h\|_2 < \delta \quad \Rightarrow \quad \|Lh\|_2 < \varepsilon. \tag{4.3}$$

We have

$$\|Lh\|_2 = \left\| \begin{bmatrix} \sum_{j=1}^n l_{1j} h_j \\ \vdots \\ \sum_{j=1}^n l_{mj} h_j \end{bmatrix} \right\|_2 \le \left\| \begin{bmatrix} \sum_{j=1}^n l_{1j} h_j \\ \vdots \\ \sum_{j=1}^n l_{mj} h_j \end{bmatrix} \right\|_1 \le \sum_{i=1}^m \left| \sum_{j=1}^n l_{ij} h_j \right|.$$

We further obtain for $i = 1, \dots, m$

$$\left| \sum_{j=1}^n l_{ij} h_j \right| \quad \le \quad \sum_{j=1}^n |l_{ij}||h_j|$$

Now, we get for all $i = 1, \dots, m$ that

$$|l_{ij}| \le \max_{\substack{i=1,\dots,m \\ j=1,\dots,n}} |l_{ij}| =: L.$$

Note that the right hand side of the last inequality does not depend on i any-more and we get

$$\left| \sum_{j=1}^n l_{ij} h_j \right| \le L \sum_{j=1}^n |h_j|.$$

Hence, we obtain

$$\sum_{i=1}^m \left| \sum_{j=1}^n l_{ij} h_j \right| \le \sum_{i=1}^m \left(L \sum_{j=1}^n |h_j| \right) \le mL \sum_{j=1}^n |h_j| \le m\sqrt{n} L \|h\|_2.$$

Let us set $C_L = m\sqrt{n}L$. Finally, we get

$$\|Lh\|_2 \le C_l \|h\|_2$$

and then (4.3) by choosing $\delta = \frac{\varepsilon}{C_L}$.

CHAPTER

5 Differentiation in one variable

5.1 Definition for one variable

Definition 5.1 (Differentiability at a point)**.**
Let $f : (a, b) \to \mathbb{R}$ be a function. Then, f is said to be differentiable at $x_0 \in (a, b)$ iff

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} \tag{5.1}$$

exists. If the limit exists, we denote it by $f'(x_0)$, i.e

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}.^{a} \tag{5.2}$$

[a] We say that $f'(x_0)$ is the derivative of f at x_0. Thus one can say that f is differentiable at x_0 if the derivative of f exists at x_0.

Graphically, this definition says that the derivative of f at x_0 is the slope of the tangent line to the graph $y = f(x)$ at x_0, which is the limit as $h \to 0$ of the slopes of the lines through the points $(x_0, f(x_0))$ and $(x_0 + h, f(x_0 + h))$, which are called secants. Thinking about the derivative graphically, one always should keep in mind, that $h \to 0$ does not only mean $h \to 0+$ or $h \to 0-$ but that they both must exist and agree. See the next remark.

Remark 5.1. *Differentiability for $f : (a, b) \to \mathbb{R}$ can be defined using the one sided limits from Section 3.5. The function f is then called differentiable at x_0 if the two one-sided limits*

$$\lim_{h \to 0-} \frac{f(x_0 + h) - f(x_0)}{h}, \quad \lim_{h \to 0+} \frac{f(x_0 + h) - f(x_0)}{h}$$

exist and are the same. If they exist but are not the same, then the function is not differentiable at x_0 but has a right-derivative

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

and a left-derivative *at x_0*

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

at x_0. One can also have the case that only one of the two exists.

Remark 5.2. *Let us discuss a couple of ways to write the statement of the differentiability of f in different ways:*

1. *The differential quotient* (5.1) *can be written as*
$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}. \tag{5.3}$$

2. *Definition 5.1 can be restated as: $f : (a, b) \to \mathbb{R}$ is differentiable at $x_0 \in (a, b)$ iff there exists a function $\varphi : (a - x_0, b - x_0) \to \mathbb{R}$ such that $\lim_{h \to 0} \frac{\varphi(h)}{h} = 0$ and a number A such that*
$$f(x_0 + h) = f(x_0) + Ah + \varphi(h).$$
The number A is given by $f'(x_0)$ by (5.2)*. This means that f is, close to x_0, well approximated by a linear function with a very small; by very small we mean that it gets to 0 faster then the distance h from $x_0 + h$ to x_0.*

3. *The last statement is often rephrased as*
$$f(x + h) = f(x) + Ah + o(h)$$
or, in the spirit of (5.3)*, as*
$$f(x) = f(x_0) + A(x - x_0) + o(x - x_0).$$
We say that ϕ is little-o of h. Intuitively, that means that one can approximate $f(x)$ by $f(x_0) + f'(x_0)(x - x_0)$ if x is close enough to x_0, i.e. in a (small) neighborhood of f, one can replace f by its tangent $T(x) = f(x_0) + f'(x_0)(x - x_0)$.

Exercise 5.1. *Prove the claimed equivalence of Definition 5.1 with the second statement in Remark 5.2.*

Remark 5.3. *Using the definition of a limit from the previous Chapter, we can rewrite the definition of a derivative as follows: We say that $f : (a, b) \to \mathbb{R}$ is differentiable at $x_0 \in (a, b)$ iff*
$$\exists A \in \mathbb{R} \text{ s.t. } \forall \varepsilon > 0\ \exists \delta > 0 :\ 0 < |h| < \delta \ \Rightarrow \left| \frac{f(x_0 + h) - f(x_0)}{h} - A \right| < \varepsilon.$$

We finish the section on the definition of derivatives by stating

Definition 5.2 (Differentiability & Derivative)**.**
Let $f : (a, b) \to \mathbb{R}$ be a function. Then, f is said to be differentiable on (a, b) iff f is differentiable at x_0 for all $x_0 \in (a, b)$. The derivative $f' : (a, b) \to \mathbb{R}$ is given by $x \mapsto f'(x)$ as defined in (5.2).

Remark 5.4. *By saying that $f : [a, b] \to \mathbb{R}$ is differentiable on $[a, b]$, we mean that f is differentiable on (a, b) according to Definition 5.2 and that*
$$\lim_{h \to 0+} \frac{f(a + h) - f(a)}{h} \quad \text{and} \quad \lim_{h \to 0-} \frac{f(b + h) - f(b)}{h}$$
exist.

Remark 5.5. *Sometimes we might just say that a function $f : I \to \mathbb{R}$ be differentiable by which we mean that is is supposed to be differentiable on its domain I taking into account one-sided limits at boundary points.*

5.1.1 Examples of derivatives

Example 5.1. *Let $I \subseteq \mathbb{R}$ be an interval and let us compute the derivative of $f : I \to \mathbb{R}$, $f(x) = c$, where $c \in \mathbb{R}$ is a fixed constant. We get*

$$\frac{f(x+h) - f(x)}{h} = \frac{c - c}{h} = \frac{0}{h} = 0.$$

Thus,

$$\frac{f(x+h) - f(x)}{h} = 0,$$

i.e. $f'(x) = 0$ for all $x \in I$.

Remark 5.6. *The converse of Example 5.1 is also true. For that see Lemma 5.1.*

Example 5.2. *Let us compute the derivative of $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x$ with the differential quotient. We have*

$$\frac{f(x+h) - f(x)}{h} = \frac{(x+h) - x}{h} = \frac{h}{h} = 1. \tag{5.4}$$

Thus,

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

Example 5.3. *Let us compute the derivative of $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$ with the differential quotient. We have*

$$\begin{aligned}
\frac{f(x+h) - f(x)}{h} &= \frac{(x+h)^2 - x^2}{h} = \frac{x^2 + 2hx + h^2 - x^2}{h} && (5.5)\\
&= 2x + h. && (5.6)
\end{aligned}$$

Thus,

$$\lim_{n\to 0} \frac{f(x+h) - f(x)}{h} = 2x.$$

Example 5.4. *Let us compute the derivative of $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \sqrt{x}$ with the differential quotient. We have*

$$\begin{aligned}
\frac{f(x+h) - f(x)}{h} &= \frac{\sqrt{x+h} - \sqrt{x}}{h} = \frac{\sqrt{x+h} - \sqrt{x}}{h} \cdot \frac{\sqrt{x+h} + \sqrt{x}}{\sqrt{x+h} + \sqrt{x}}\\
&= \frac{1}{h} \frac{(x+h) - x}{\sqrt{x+h} + \sqrt{x}} = \frac{1}{\sqrt{x+h} + \sqrt{x}}.
\end{aligned}$$

This leads to

$$\lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = \frac{1}{2\sqrt{x}},$$

where we use the continuity of $\sqrt{\cdot}$ and limit calculus. See also Example 4.4.

Example 5.5. *We consider $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = |x|$. The function f is not differentiable at $x = 0$ but only on $(-\infty, 0)$ and $(0, \infty)$. On The first interval we have $f(x) = -x$ and on the second $f(x) = x$. At $x = 0$, we have*

$$\begin{aligned}
\lim_{h\to 0+} \frac{f(0+h) - f(0)}{h} &= \lim_{h\to 0+} \frac{h}{h} = 1, \quad \textit{and}\\
\lim_{h\to 0-} \frac{f(0+h) - f(0)}{h} &= \lim_{h\to 0-} \frac{-h}{h} = -1.
\end{aligned}$$

See also Example 3.7.

Example 5.6. *There exist also function which are differentiable at one point but are not continuous anywhere else. Let*

$$\chi_{\mathbb{Q}}(x) = \begin{cases} 1 & : \quad x \in \mathbb{Q} \\ 0 & : \quad x \in \mathbb{R} \setminus \mathbb{Q} \end{cases} .$$

Then, the function

$$f(x) = x^2 \chi_{\mathbb{Q}}(x)$$

is differentiable at $x = 0$ but for no $x \neq 0$ continuous. The continuity part is clear from Example 4.9. Let us investigate the existence of

$$\lim_{h \to 0} \frac{f(0+h) - f(0)}{h}. \tag{5.7}$$

We have that $|f(h)| \leq h^2 = |h|^2$ and thus,

$$0 < |h| < \varepsilon \quad \Rightarrow \quad \left| \frac{f(h)}{h} \right| < \varepsilon$$

which implies that (5.7) exists and is equal to 0.

Remark 5.7. *As we have seen, e.g. with Example 5.5, there are more differentiable than continuous functions. However, the examples so far are differentiable at most points and only problematic at very few. Until quite late, it was widely believed that almost all functions possess even infinitely many derivatives. It was first Bolzano [2] and then Weierstrass [17] who showed that there are* Monster functions *which are everywhere continuous but at no point differentiable. An example is*

$$f(x) = \sum_{n=1}^{+\infty} \frac{1}{2^n} \cos(10^n \pi x).$$

We have not yet all the tools to understand why the assertion is true but can still admire that such a function exists and that one can write down an example as explicit as this.

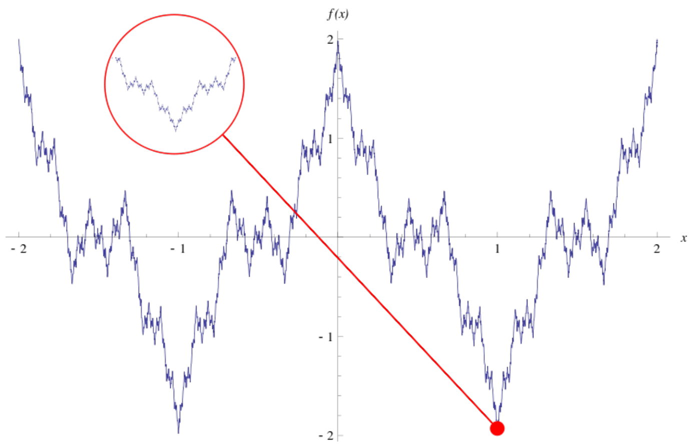


Figure 5.1: A glimpse into the Weierstrass Monster Function.

For further information on this function and related ones see also [8], [1], and [9].

5.1.2 Differentiability ⇒ Continuity

Theorem 5.1 (Differentiable functions are continuous)**.**

Suppose $f:(a,b)\to\mathbb{R}$ is differentiable at $x_0\in(a,b)$. Then f is continuous at x_0.

Remark 5.8. *The converse of theorem 5.1 is not true as we have already noted in Remark 5.7. Another simple function which is continuous on $\mathbb{R}$ but not differentiable at $x=0$ is $f(x)=|x|$.*

Proof. Since (a,b) contains only limit points, we can use Theorem 4.1. We show that $\lim_{x\to x_0} f(x) = f(x_0)$. We compute, using the limit calculus,

$$\begin{aligned}\lim_{x\to x_0} f(x) &= \lim_{x\to x_0}\left(f(x_0)+\frac{f(x)-f(x_0)}{x-x_0}(x-x_0)\right) \\ &= f(x_0)+\lim_{x\to x_0}\left(\frac{f(x)-f(x_0)}{x-x_0}(x-x_0)\right) \\ &= f(x_0)+f'(x_0)\lim_{x\to x_0}(x-x_0) \\ &= f(x_0).\end{aligned} \tag{5.8}$$

The step from the second to the third line is possible since f is differentiable at x_0 and we have Theorem 3.2. □

Remark 5.9. *The idea of the first line (5.8) is simply that one needs to introduce the differential quotient in some way. Thus, one first writes*

$$f(x)=f(x_0)+f(x)-f(x_0)$$

and then multiplies a 1 suitably to get

$$f(x)=f(x_0)+(f(x)-f(x_0))\cdot\frac{x-x_0}{x-x_0}=f(x_0)+\frac{f(x)-f(x_0)}{x-x_0}(x-x_0),\quad x\neq x_0.$$

Remark 5.10. *The result of Theorem 5.1 holds true for $f:[a,b]\to\mathbb{R}$ is one uses the appropriate one-sided limits in the definition of continuity and differentiability at the boundary points.*

Remark 5.11. *Another proof of Theorem 5.1, using the (ε,δ)-definition of continuity, proceeds as follows. Let $f:(a,b)\to\mathbb{R}$ be differentiable on (a,b) and let $x_0\in(a,b)$. Then, for all $\varepsilon>0$ we have to show that there is a $\delta>0$ such that*

$$|x-x_0|<\delta \quad\Rightarrow\quad |f(x)-f(x_0)|<\varepsilon.$$

Since we want to link this to the derivative of f, we need to introduce the differential quotient:

$$|f(x)-f(x_0)|=\left|\frac{f(x)-f(x_0)}{x-x_0}(x-x_0)\right|\le\left|\frac{f(x)-f(x_0)}{x-x_0}\right||x-x_0|.$$

Since

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

we have that there exists $\delta'>0$ such that

$$0<|x-x_0|<\delta' \quad\Rightarrow\quad \left|\frac{f(x)-f(x_0)}{x-x_0}-f'(x_0)\right|<1.$$

which gives

$$\sup_{0<|x-x_0|<\delta'}\left|\frac{f(x)-f(x_0)}{x-x_0}\right|\le 1+|f'(x_0)|.$$

Thus, we get

$$0<|x-x_0|<\delta \quad\Rightarrow\quad |f(x)-f(x_0)|\le\big(1+|f'(x_0)|\big)|x-x_0|<\varepsilon$$

choosing $\delta=\min\left\{\delta',\frac{\varepsilon}{1+|f'(x_0)|}\right\}$. This concludes the proof.

Remark 5.12. *The (ε,δ)-proof in Remark 5.11 proves something more general than intended: Let $I \subseteq \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$ be a function such that there exist constants $\delta > 0$ and $C > 0$ such that*

$$\sup_{0<|x-x_0|\le\delta} \left| \frac{f(x)-f(x_0)}{x-x_0} \right| \le C. \tag{5.9}$$

Those function are called Lipschitz continuous and an equivalent definition is to ask that they satisfy

$$|f(x)-f(y)| \le C|x-y| \quad \forall x,y \in I \tag{5.10}$$

for a constant $C > 0$. An example is $f(x) = |x|$ which is not differentiable at $x = 0$ but still satisfies (5.9) with $C = 1$. By the Mean Value Theorem, see Theorem 5.6, we get that all continuously differentiable[1] *functions are also Lipschitz-continuous: for all $x, y \in I$, we have that there exists $\xi \in (x,y)$ such that*

$$|f(x)-f(y)| = |f'(\xi)||x-y|.$$

This yields (5.10) by estimating $|f'(\xi)| \le \sup_{x\in I}|f'(x)| = C$.

Remark 5.13. *To seal this section let us streamline the argument from Remark 5.11 a little bit. Let $f : (a,b) \to \mathbb{R}$ be differentiable at $x_0 \in (a,b)$. That means that there exists $A \in \mathbb{R}$ such that for all $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$0 < |x-x_0| < \delta \quad \Rightarrow \quad \left| \frac{f(x)-f(x_0)}{x-x_0} - A \right| < \varepsilon.$$

Setting $\varepsilon = 1$, we obtain, using the triangle inequality, that there exists $\delta' > 0$ such that

$$\left| \frac{f(x)-f(x_0)}{x-x_0} \right| \le 1 + |A|$$

for all $x \in (a,b)$ with $0 < |x-x_0| < \delta'$. Thus, we get

$$|f(x)-f(x_0)| \le (1+|A|)|x-x_0|$$

for $0 < |x-x_0| < \delta'$. Taking

$$\delta = \min\left\{\delta', \frac{\varepsilon}{1+|A|}\right\},$$

we get for all $\varepsilon > 0$ that

$$0 < |x-x_0| < \delta \quad \Rightarrow \quad |f(x)-f(x_0)| < \varepsilon. \tag{5.11}$$

This implies that f is continuous at x_0 by Theorem 4.1 as (5.11) means $\lim_{x\to x_0} f(x) = f(x_0)$.

5.1.3 Are derivatives continuous?

Let $f : (a,b) \to \mathbb{R}$ be a differentiable function. Then, the derivative $x \mapsto f'(x)$ is not necessary continuous. A standard example is

$$f(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & : \quad x \in [-1,1] \setminus \{0\} \\ 0 & : \quad x = 0 \end{cases}.$$

This function is continuous on $[-1,1]$ and differentiable on $[-1,1]$ but the derivative

$$f'(x) = 2x\left(\sin\left(\frac{1}{x}\right)\right) - \cos\left(\frac{1}{x}\right)$$

[1] That means that the derivative of f not only exits on the entire domain of f but f' is itself a continuous function.

is not continuous at $x=0$ as the limit

$$\lim_{x\to 0}\cos\left(\frac{1}{x}\right)$$

does not exist.

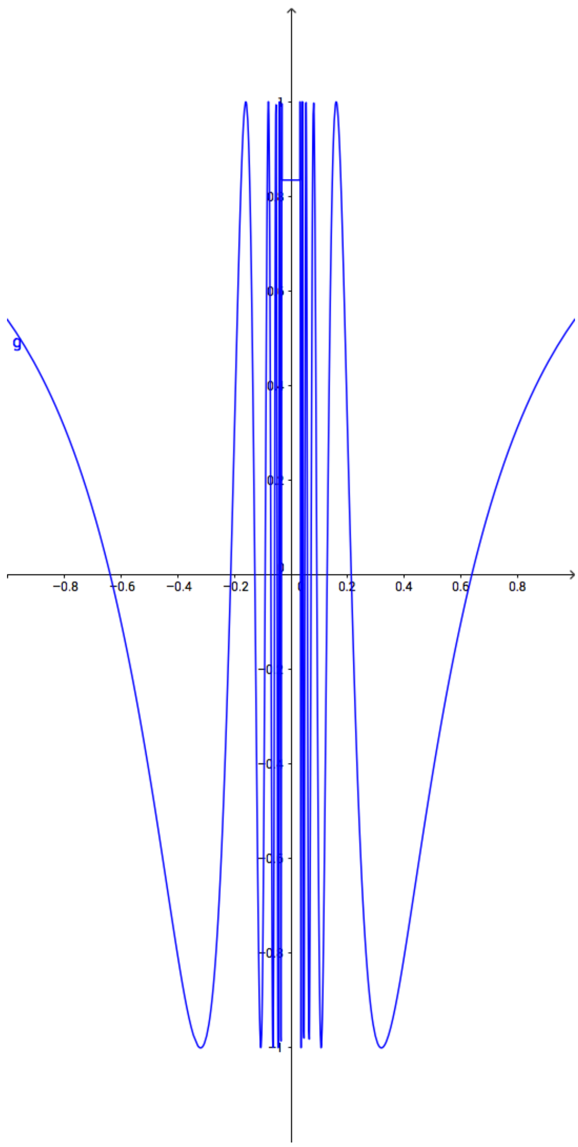


Figure 5.2: The function $\cos(\frac{1}{x})$. As x approaches zero, the functions tries to take all values between -1 and 1 and can therefore not have a limit.

Below, we plot f. The reader should use GeoGebra or an equivalent tool to get better pictures as we are here limited to the inanimate nature of paper.

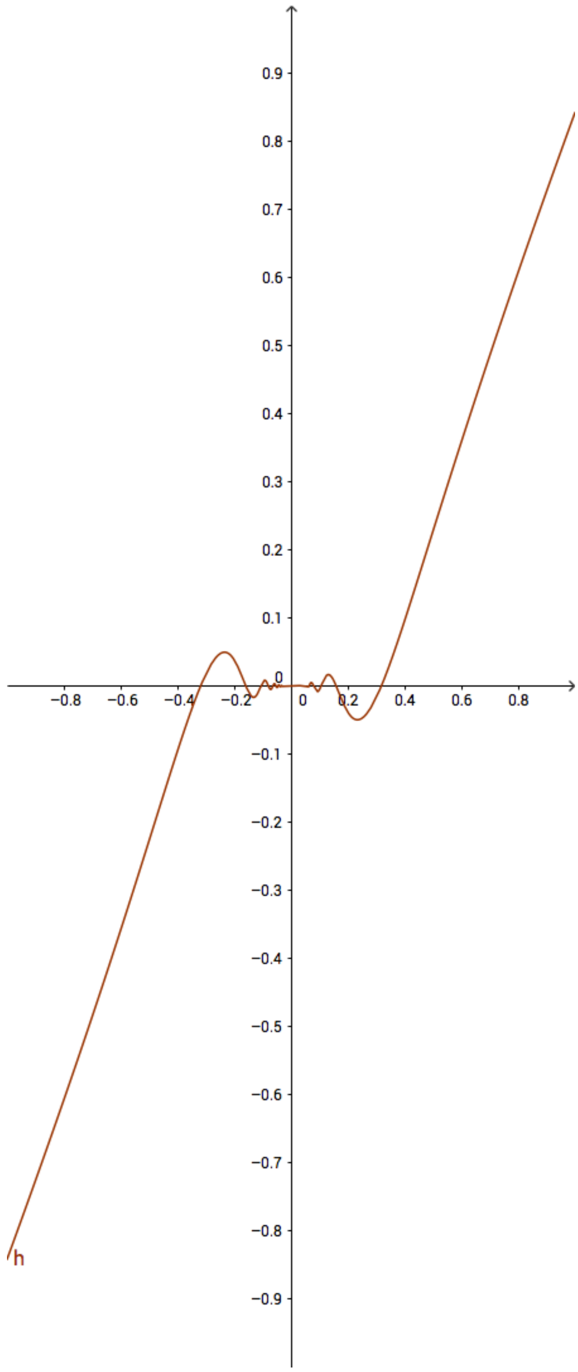


Figure 5.3: A plot of $f(x) = x^2 \sin\left(\frac{1}{x}\right)$.

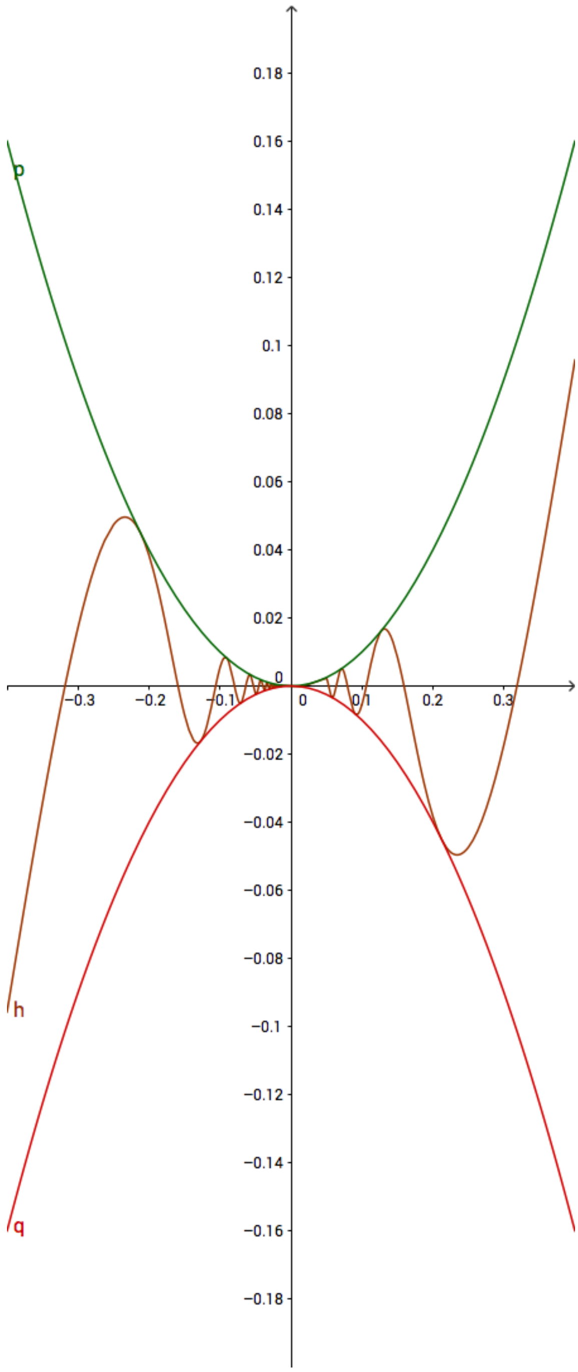


Figure 5.4: A more detailed plot of $f(x) = x^2 \sin\left(\frac{1}{x}\right)$ around $x = 0$ with the enveloppe $\pm x^2$.

5.2 Operations with differentiable functions

As we have done in Section 4.5 for continuous functions, we investigate now what operations we are allowed to do with differential functions and how the resulting derivatives are computed.

Theorem 5.2 (Chain rule)**.**
Let $g : [a, b] \to \mathbb{R}$ and let $f : I \to \mathbb{R}$, where $I \subseteq \mathbb{R}$ is an interval containing the range of g, so that

$f \circ g$ is defined. Suppose that g is differentiable at $x_0 \in [a,b]$ and f is differentiable at $g(x_0)$. Then, $h = f \circ g$ is differentiable at x_0 and

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

Remark 5.14. *In the composition $f(g(x))$ we refer to f as the outer function, and g as the inner function. We can describe the basic mechanism of the chain rule as follows: differentiate the outer function holding the inner function as a constant. Then, multiply the result by the derivative of the inner function. If there is a composition of more than two functions, e.g. $f(g(h(x)))$, the above process is simply repeated as many times as necessary. We leave it as an exercise to write down a precise statement for that case.*

Remark 5.15. *A proof that can often be found in text books works as follows:*

$$\begin{aligned}
\lim_{h\to 0} \frac{f(g(x_0+h)) - f(g(x_0))}{h} &= \lim_{h\to 0} \frac{f(g(x_0+h)) - f(g(x_0))}{h} \frac{g(x_0+h) - g(x_0)}{g(x_0+h) - g(x_0)} \\
&= \lim_{h\to 0} \frac{f(g(x_0+h)) - f(g(x_0))}{g(x_0+h) - g(x_0)} \frac{g(x_0+h) - g(x_0)}{h} \\
&= \lim_{h\to 0} \frac{f(g(x_0+h)) - f(g(x_0))}{g(x_0+h) - g(x_0)} \lim_{h\to 0} \frac{g(x_0+h) - g(x_0)}{h} \\
&= \left(\lim_{h\to 0} \frac{f(g(x_0+h)) - f(g(x_0))}{g(x_0+h) - g(x_0)} \right) g'(x_0) \\
&= \left(\lim_{h\to 0} \frac{f(g(x_0) + g(x_0+h) - g(x_0)) - f(g(x_0))}{g(x_0+h) - g(x_0)} \right) g'(x_0) \\
&= \left(\lim_{t\to 0} \frac{f(g(x_0) + t) - f(g(x_0))}{t} \right) g'(x_0) \\
&= f'(g(x_0))g'(x_0),
\end{aligned}$$

where we set $t = g(x_0+h) - g(x_0)$. There are several sins in this proof. Can you spot them? First, the proof does not apply to constant functions g since you then commit the deadly sin of dividing by zero. Also, the function $g(x_0+h) - g(x_0)$ might be 0 for a sequence of h due to oscillations of g. Furthermore, the limits $t \to 0$ and $h \to 0$ are not equivalent as $h \to 0$ implies $t \to 0$ but, again due to possible oscillations, $t \to 0$ does not imply $h \to 0$. Also, the step of computing the product of limits (3rd equal sign) is only justified if we can ensure the existence of both (see Theorem 3.2) and the existence of

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

is unclear due to the limits $t \to 0$ and $h \to 0$ not being equivalent.
As a final remark let me make clear that the proof works if one excludes *the following situation: g has the following property:*

$$\exists \varepsilon_0 > 0 \forall \varepsilon \in (0, \varepsilon_0] \exists x \in (x_0 - \varepsilon, x_0 + \varepsilon) \setminus \{x_0\} \quad \textit{such that} \quad g(x) = g(x_0).$$

A way to fix the problem is given in the proof below.

Proof. Orthodox proofs of the chain rule are somewhat technical and often opaque to students. However, the proof presented in the slides can be extended into a mathematically rigorous argument. We follow Peter F. McLoughlin who presented this proof in the American Mathematical Monthly as a page filler. See January Issue of 2013 p. 94 or here.

First, when $g(x) - g(x_0) \neq 0$, we can write

$$\frac{f(g(x)) - f(g(x_0))}{x - x_0} = \frac{f(g(x)) - f(g(x_0))}{g(x) - g(x_0)} \frac{g(x) - g(x_0)}{x - x_0}. \tag{5.12}$$

Case I: We assume that for every $\varepsilon > 0$ there exists an $x \in (x_0 - \varepsilon, x_0 + \varepsilon) \setminus \{x_0\}$ such that $g(x) = g(x_0)$. Picking such a sequence, we get

$$\lim_{n \to +\infty} \frac{g(x_n) - g(x_0)}{x_n - x_0} = 0.$$

Since g is differentiable, i.e.

$$\lim_{x \to x_0} \frac{g(x) - g(x_0)}{x - x_0}$$

exists and the uniqueness of limits, we get

$$\lim_{x \to x_0} \frac{g(x) - g(x_0)}{x - x_0} = 0. \tag{5.13}$$

Let now $(x_n)_{n \in \mathbb{N}} \subseteq (a, b)$ be a sequence with $x_n \to x_0$ as $n \to +\infty$. We can partition $(x_n)_{n \in \mathbb{N}}$ into two subsequences $(x_{n_l})_{l \in \mathbb{N}} \subseteq (x_n)_{n \in \mathbb{N}}$ and $(x_{n_k})_{k \in \mathbb{N}} \subseteq (x_n)_{n \in \mathbb{N}}$ such that $g(x_{n_l}) = g(x_0)$ and $g(x_{n_k}) \neq g(x_0)$.[2] Thus, we get

$$\lim_{l \to +\infty} \frac{f(g(x_{n_l})) - f(g(x_0))}{x_{n_l} - x_0} = 0.$$

For $(x_{n_k})_{k \in \mathbb{N}}$, we can use (5.12) to get

$$\begin{aligned} \lim_{k \to +\infty} \frac{f(g(x_{n_k})) - f(g(x_0))}{x_{n_k} - x_0} &= \lim_{k \to +\infty} \frac{f(g(x_{n_k})) - f(g(x_0))}{g(x_{n_k}) - g(x_0)} \frac{g(x_{n_k}) - g(x_0)}{x_{n_k} - x_0} \\ &= \lim_{k \to +\infty} \frac{f(g(x_{n_k})) - f(g(x_0))}{g(x_{n_k}) - g(x_0)} \cdot \lim_{k \to +\infty} \frac{g(x_{n_k}) - g(x_0)}{x_{n_k} - x_0} = 0 \end{aligned}$$

by (5.13). Thus, we have

$$\lim_{n \to +\infty} \frac{f(g(x_n)) - f(g(x_0))}{x_n - x_0} = 0.$$

This gives the result in case I.

Case II Let us assume that there is an $\varepsilon > 0$ such that $g(x) \neq g(x_0)$ for all $x \in (x_0 - \varepsilon, x_0 + \varepsilon)$. Then the result follows by taking limits $x \to x_0$ on both sides of (5.12) and basic limit calculus. □

Example 5.7. *Let us compute a couple of examples with the chain rule:*

1. $h(x) = \sin(x^2)$, $h'(x) = \cos(x^2) \cdot 2x$, $f(x) = \sin(x)$, $g(x) = x^2$
2. $h(x) = (1 + x)^{-\frac{1}{2}}$, $h'(x) = -\frac{1}{2}(1 + x)^{-\frac{3}{2}} \cdot 1$, $f(x) = \frac{1}{\sqrt{x}}$, $g(x) = 1 + x$
3. $h(x) = e^{\sin(x^2)}$, $h'(x) = e^{\sin(x^2)} \cdot (2x\cos(x^2))$, $f(x) = e^x$, $g(x) = \sin(x^2)$.

Theorem 5.3 (Arithmetic operations with differentiable functions)**.**
Let $f, g : [a, b] \to \mathbb{R}$ differentiable at $x_0 \in [a, b]$. Then

1) $(\lambda \cdot f)'(x_0) = \lambda \cdot f'(x_0)$, *for all* $\lambda \in \mathbb{R}$,

2) $(f + g)'(x_0) = f'(x_0) + g'(x_0)$,

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

[2] One of them could be empty, the argument proceeds nevertheless. Do you see that?

4) $\left(\frac{f}{g}\right)'(x_0) = \dfrac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g^2(x_0)}$*, provided* $g(x_0) \neq 0$*.*

Proof. Property (i) is clear by the arithmetic rules for limits. Let us proof property (ii). We compute

$$\begin{aligned} \frac{(f+g)(x_0+h)-(f+g)(x_0)}{h} &= \frac{f(x_0+h)+g(x_0+h)-f(x_0)-g(x_0)}{h} \\ &= \frac{f(x_0+h)-f(x_0)}{h} + \frac{g(x_0+h)-g(x_0)}{h}. \end{aligned}$$

Taking the limit $h \to 0$ on both sides yields, since

$$\lim_{h\to 0} \frac{f(x_0+h)-f(x_0)}{h} \quad \text{and} \quad \lim_{h\to 0} \frac{g(x_0+h)-g(x_0)}{h} \tag{5.14}$$

exist, the result in (ii). For (iii), we calculate

$$\begin{aligned} \frac{(fg)(x_0+h)-(fg)(x_0)}{h} &= \frac{f(x_0+h)g(x_0+h)-f(x_0)g(x_0)}{h} \\ &= \frac{f(x_0+h)g(x_0+h)-f(x_0)g(x_0)+f(x_0)g(x_0+h)-f(x_0)g(x_0+h)}{h} \\ &= \frac{(f(x_0+h)-f(x_0))g(x_0+h)+f(x_0)(g(x_0+h)-g(x_0))}{h}. \end{aligned}$$

Taking the limit $h \to 0$ on both sides gives (iii) since g is continuous at x_0 and the limits in (5.14) exist. Property (iv) follows by combining (iii) with the chain rule since $\left(\frac{1}{g}\right)'(x) = -\frac{g'(x)}{g^2(x)}$. We can prove the rule also directly by

$$\begin{aligned} \frac{\left(\frac{f}{g}\right)(x_0+h)-\left(\frac{f}{g}\right)(x_0)}{h} &= \frac{\frac{f(x_0+h)}{g(x_0+h)}-\frac{f(x_0)}{g(x_0)}}{h} = \frac{1}{h}\frac{f(x_0+h)g(x_0)-f(x_0)g(x_0+h)}{g(x_0+h)g(x_0)} \\ &= \frac{1}{h}\frac{f(x_0+h)g(x_0)-f(x_0)g(x_0+h)+f(x_0)g(x_0)-f(x_0)g(x_0)}{g(x_0+h)g(x_0)} \\ &= \frac{1}{h}\frac{(f(x_0+h)-f(x_0))g(x_0)}{g(x_0+h)g(x_0)} - \frac{1}{h}\frac{(g(x_0+h)-g(x_0))f(x_0)}{g(x_0+h)g(x_0)}. \end{aligned}$$

Taking the limit $h \to 0$ and basic limit calculus provides the result due to (5.14) and the fact that g is continuous at x_0 (see Thm. 5.1). $\square$

Exercise 5.2. *Use Theorem 5.3 to prove that polynomials of any degree are differentiable on* $\mathbb{R}$*. Remember, a polynomial of degree* $n \in \mathbb{N}_0$ *is given by*

$$p(x) = \sum_{k=1}^{n} a_k x^k,$$

where the a_k *are real numbers. (Hint: Start with* $p(x) = a_0$ *and* $p(x) = a_1 x$ *and work your way up from there. Induction is your friend.)*

5.3 Properties of differentiable functions

We introduce the notion of a stationary point by

Definition 5.3 (Stationary point)**.**
Let $f : [a,b] \to \mathbb{R}$ *and* $x_0 \in (a,b)$*. If* f *is differentiable at* x_0 *and* $f'(x_0) = 0$*, then we call* x_0 *a stationary point of* f*.*

We prove now

Theorem 5.4 (Fermat's Theorem)**.**
Let $f : [a,b] \to \mathbb{R}$ and suppose f is differentiable at $x_0 \in (a,b)$. Suppose that f has a local maximum at x_0. Then $f'(x_0) = 0$, i.e. x_0 is stationary point of f.

Proof. We have, by the definition of local maximum, that there exists a $\delta > 0$ such that

$$f(x_0+h) - f(x_0) \leq 0 \quad \forall\, |h| < \delta$$

which yields

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

and

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

Since $\lim_{h\to 0} \frac{f(x_0+h)-f(x_0)}{h}$ exists, we must have

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

since the left and right limit must be equal. □

From Fermat's theorem, we get immediately

Corollary 5.1 (Classification of extrema)**.**
Let $f : [a,b] \to \mathbb{R}$. Then the (local/global) maxima and minima of f can only be at points x_0 of one of three types:

(i) stationary point of f,

(ii) a point in (a,b) at which f is not differentiable,

(iii) at the boundary of $[a,b]$, i.e. at $x = a$ or $x = b$.

Example 5.8. *Consider $f(x) = x^2$ and $g(x) = |x|$ on $[-1,1]$. Then, we have that the global minimum of f is a stationary point, $x_0 = 0$ and the global maxima are on the boundary at $x_1 = -1$ and $x_2 = 1$. For g, we have that the global minimum is at a point where the function is not differentiable, namely $x_0 = 0$. The global maximuma are at $x_1 = -1$ and $x_2 = 1$.*

The next theorem is quite intuitive and has many applications in Analysis.

Theorem 5.5 (Rolle's Theorem)**.**
Let $f : [a,b] \to \mathbb{R}$ be continuous and differentiable on (a,b) and $f(a) = f(b)$. Then there exists an $x_0 \in (a,b)$ such that $f'(x_0) = 0$.

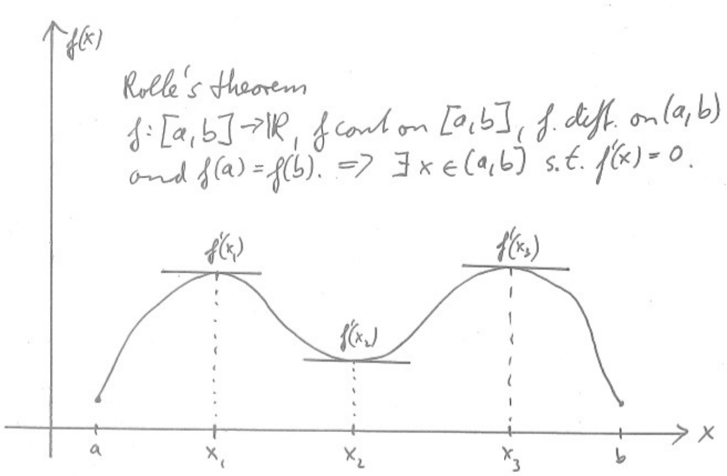


Figure 5.5: Illustration of Rolle's Theorem.

Proof. First, if f is constant, the result is clear. Now, let f not be constant, i.e. there exists a $c \in [a,b]$ such that $f(c) \neq f(a)$. Since f is continuous on $[a,b]$, by the EVT (see Thm. 4.7) there exist $x_{\max}$ and $x_{\min}$ in $[a,b]$ where the function f attains its maximum and minimum respectively. Since $f(a) = f(b)$, at least one of the two is in (a,b), denote it by x_0. By the theorem of Fermat, we have $f'(x_0) = 0$. $\square$

Note that for Rolle's theorem, we require continuity on the closed interval $[a,b]$ but differentiability only on the open interval (a,b). The proof is deceptively simple, but the result is nevertheless non-trivial since it relies on the Extreme Value Theorem (see Thm. 4.7), which, in turn, relies on the completeness of $\mathbb{R}$. The theorem would not be true if we restricted attention to functions defined on the rationals $\mathbb{Q}$.

The mean value theorem is an immediate consequence of Rolle's theorem.

Theorem 5.6 (Mean Value Theorem)**.**
Let $f : [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and differentiable on (a,b). Then there exists $x_0 \in (a,b)$ such that

$$f'(x_0) = \frac{f(b) - f(a)}{b - a}.$$

Graphically, this result says that there is point $x_0 \in (a,b)$ at which the slope of the graph, i.e. the slope of the tangent at $y = f(x)$ at x_0, is equal to the slope of the secant through the endpoints $(a, f(a))$ and $(b, f(b))$.

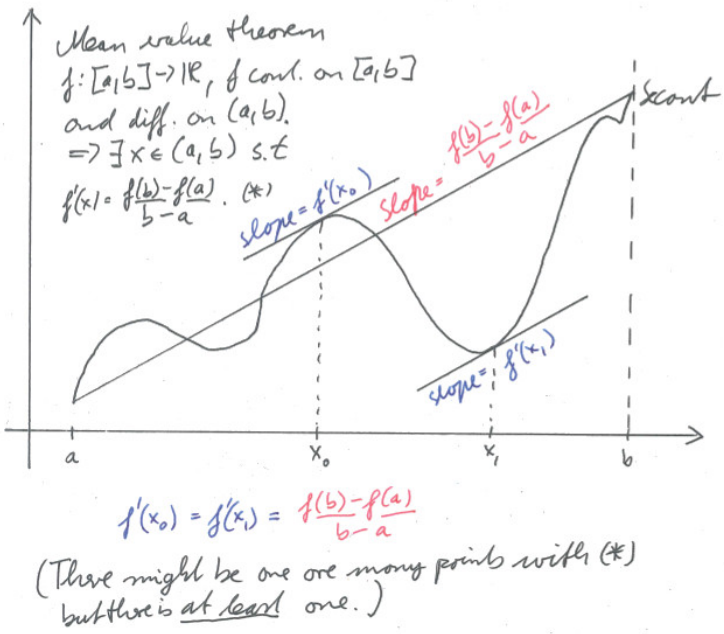


Figure 5.6: Illustration od the Mean Value Theorem.

Proof. We define an auxiliary function g by

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

This function satisfies $g(a) = f(a) = g(b)$. Thus, the assumptions of Rolle's theorem are satisfied since g is continuous on $[a, b]$ and differentiable on (a, b). Thus, there exists an $x_0 \in (a, b)$ such that $g'(x_0) = 0$ which implies

$$f'(x_0) - \frac{f(b) - f(a)}{b - a} = 0$$

The proof is complete. □

The next lemma is simple yet of some use to us in the remainder of these notes.

Lemma 5.1.
Suppose $f : [a, b] \to \mathbb{R}$ and differentiable on $[a, b]$. Further, let $f'(x) = 0$ for all $x \in [a, b]$. Then there exists $c \in \mathbb{R}$ such that $f(x) = c$ for all $x \in [a, b]$.

Proof. For all $x \in (a, b]$, there exists, by Theorem 5.6, a $y \in [a, x]$ such that

$$\frac{f(x) - f(a)}{x - a} = f'(y) = 0.$$

Thus, $f(x) = f(a)$ for all $x \in [a, b]$. □

5.4 Derivatives of higher order

If a function $f : (a, b) \to \mathbb{R}$ has a derivative $x \mapsto f'(x)$ on (a, b) and f' is itself differentiable, then we denote the derivative of f' by f'' and call it the second derivative of f. Continuing, we obtain

$$f, f', f'', f^{(3)}, f^{(4)}, \ldots, f^{(k)}$$

each of which is the derivative of the preceding. The function $f^{(k)}$ is called the kth derivative of f. It is also called the derivative of order n of f. It is also denoted by

$$\frac{d^k f}{dx^k}.$$

For $f^{(k)}(x_0)$ to exist, $f^{(k-1)}(x)$ must exist in a neighborhood[3] $(x_0-\delta, x_0+\delta)$ of x_0, and $f^{(k-1)}$ must be differentiable at x. Since $f^{(k-1)}$ must exist in neighborhood of x_0, f^{n-2} must be differentiable in that neighborhood.

5.5 Function spaces

We consider now the collection of all real functions which are continuous. Thus, functions are considered as points in an appropriate space which allows us to carry some geometric intuition over to much more complicated situations than $\mathbb{R}^n$.

Definition 5.4 (The space $C[a,b]$)**.**
Let $\emptyset \neq [a,b] \subseteq \mathbb{R}$ be an interval. Then, by $C[a,b] = C^0[a,b]$, we denote the set of all functions $f:[a,b] \to \mathbb{R}$ which are continuous on $[a,b]$.

Exercise 5.3. *Convince yourself that $C[a,b]$ is a real vector space, where $+$ is the usual pointwise definition*

$$+ : C[a,b] \times C[a,b] \to C[a,b]$$
$$(f,g) \mapsto f+g,$$

where $(f+g)(x) := f(x)+g(x)$ for all $x \in [a,b]$, and

$$\cdot : \mathbb{R} \times C[a,b] \to C[a,b]$$
$$(\lambda, f) \mapsto \lambda f,$$

where $(\lambda f)(x) := \lambda \cdot f(x)$ for all $x \in [a,b]$. The $\cdot$ in $\lambda \cdot f(x)$ is the product of $\mathbb{R}$.

Let us generalize Definition 2.1 from $\mathbb{R}^n$ to a general real vector space.

Definition 5.5 (Norm)**.**
Let V be a real vector space. Then a function $\|\cdot\| : V \to \mathbb{R}$ is called a norm if

(P1) $\|x\| \geq 0$ for all $x \in V$ and $\|x\| = 0$ iff $x = 0$. (Positivity)

(P2) For all $\lambda \in \mathbb{R}$, and all $x \in V$, $\|\lambda x\| = |\lambda|\|x\|$. (Homogeneity)

(P3) For all x, $y \in V$, we have

$$\|x+y\| \leq \|x\| + \|y\|. \quad \textit{(Triangle inequality)}$$

Now, we introduce a norm on $C[a,b]$: for $f \in C[a,b]$, we set

$$\|f\|_\infty := \sup_{x \in [a,b]} |f(x)|.$$

Since we have that $|\cdot|$ is a norm on $\mathbb{R}$, we get that $\|\cdot\|_\infty$ is a norm on $C^0(\Omega)$ with the properties

[3] or in a one-sided neighborhood $(x_0-\delta, x_0]$, $[x_0, x_0+\delta)$ if x_0 is a boundary point of an interval on which f is defined.

(P1) $\|f\|_\infty \geq 0$ and $\|f\|_\infty = 0$ iff $f = 0$,

(P2) $\|\lambda f\|_\infty = |\lambda| \|f\|_\infty$, and

(P3) $\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$.

Definition 5.6 (The space $C^k[a,b]$).
We say $f \in C^k[a,b]$ iff

$$\forall j \in \{1, \dots, k\} : \qquad f^{(j)} \in C^{k-j}[a,b].$$

Exercise 5.4. *Convince yourself that all $C^k[a,b]$ are real vector spaces and that, for $k \geq 1$, the derivative $\frac{d}{dx} : C^k[0,1] \to C^{k-1}[a,b]$ is a linear map from $C^k[0,1]$ to $C^{k-1}[a,b]$.*

Also on the space $C^k[a,b]$, $k \geq 1$, we can introduce a norm:

$$\begin{aligned} \|f\|_{C^k} &:= \|f\|_\infty + \|f^{(1)}\|_\infty + \cdots + \|f^{(k)}\|_\infty \\ &= \sum_{i=0}^{k} \|f^{(i)}\|_\infty. \end{aligned}$$

Definition 5.7 (The space $C^\infty[a,b]$).
We define the space $C^\infty[a,b]$ by

$$C^\infty[a,b] = \bigcap_{k \geq 0} C^k[a,b].$$

Remark 5.16. *It is clear, that one can not simply extend the norm-definitions from $C^k[a,b]$ to $C^\infty[a,b]$ as we need to involve series.*

Remark 5.17. *The above spaces can easily be defined on open intervals (a,b). However, then we can not easily introduce a norm as we have no Extreme Value Theorem on open sets which guarantees the boundedness of continuous functions. Clearly, $f \in C(0,1)$ for $f(x) = \frac{1}{x}$ but $\|f\|_\infty$ is not finite.*

CHAPTER

6 Differentiation of functions on $\mathbb{R}^n$

In this chapter, we generalize the notion of differentiability from function of one variable to functions of several variables. Further, we will allow the co-domain of f to be $\mathbb{R}^m$ for $m \geq 1$.

6.1 Definition

The definition we are going to start with, is a multi-dimensional generalization of the characterisation of differentiation at a point by approximation by a linear function as described in point 2 of Remark 5.2.

Definition 6.1 (Total Derivative)**.**
Let $\Omega \subseteq \mathbb{R}^n$ be an open set and let $f : \Omega \to \mathbb{R}^m$ be a function. Then, f is called differentiable at $x_0 \in \Omega$ if there exists a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ and a function $r : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$f(x_0 + h) = f(x_0) + L \cdot h + r(h) \tag{6.1}$$

and

$$\lim_{h \to 0} \frac{\|r(h)\|_2}{\|h\|_2} = 0. \tag{6.2}$$

Remark 6.1. *It is worth noting, that the map L depends on x_0, i.e. $L = L(x_0)$. Also, the dependence on x_0 has not to be linear but only $L(x_0)$ has to be a linear map for every fixed x_0. See also the next example.*

Example 6.1. *We consider*

$$f : \mathbb{R}^2 \to \mathbb{R}^2, \quad f(x_1, x_2) = \begin{bmatrix} x_1^3 + x_2 \\ x_1 + x_2 \end{bmatrix}.$$

We set $x_0 = [x_1\, x_2]^T$, $h = [h_1\, h_2]^T$ and calculate

$$\begin{aligned} f(x_1 + h_1, x_2 + h_2) &= \begin{bmatrix} (x_1 + h_1)^3 + x_2 + h_2 \\ x_1 + h_1 + x_2 + h_2 \end{bmatrix} \\ &= \begin{bmatrix} x_1^3 + x_2 \\ x_1 + x_2 \end{bmatrix} + \begin{bmatrix} 3h_1 x_1^2 + 3h_1^2 x_1 + h_1^3 + h_2 \\ h_1 + h_2 \end{bmatrix} \\ &= \begin{bmatrix} x_1^3 + x_2 \\ x_1 + x_2 \end{bmatrix} + \begin{bmatrix} 3x_1^2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} 3h_1^2 x_1 + h_1^3 \\ 0 \end{bmatrix}. \end{aligned}$$

By that we have that f is differentiable for all x_0 if we can show

$$\lim_{h \to 0} \frac{\|r(h)\|_2}{\|h\|_2} = 0 \tag{6.3}$$

for

$$r(h) = \begin{bmatrix} 3h_1^2 x_1 + h_1^3 \\ 0 \end{bmatrix}.$$

The matrix $\begin{bmatrix} 3x_1^2 & 1 \\ 1 & 1 \end{bmatrix}$ *which is the total derivative* $Df(x_0)$ *for the* f *in the example, depends on the point where the derivative is computed and the dependence is not linear. However, the matrix defines a linear map from* $\mathbb{R}^2$ *to* $\mathbb{R}^2$ *by*

$$\begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \mapsto \begin{bmatrix} 3x_1^2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}.$$

Exercise 6.1. *Prove (6.3) in the previous example.*

One can give an alternative definition of differentiability by

Definition 6.2 (Total Derivative II)**.**
Let $\Omega \subset \mathbb{R}^n$ *be an open set and let* $f : \Omega \to \mathbb{R}^m$ *be a function. Then,* f *is called differentiable at* $x_0 \in \Omega$ *if there exists a linear map* $L : \mathbb{R}^n \to \mathbb{R}^m$ *and a function* $r : \mathbb{R}^n \to \mathbb{R}^m$ *such that*

$$\lim_{h \to 0} \frac{\|f(x_0 + h) - f(x_0) - Lh\|_2}{\|h\|_2} = 0. \tag{6.4}$$

Lemma 6.1.
Definitions 6.1 and 6.2 are equivalent.

Proof. [$\Rightarrow$] Let $\Omega \subseteq \mathbb{R}^n$ be an open set and let $f : \Omega \to \mathbb{R}^m$ be a function. Let f be differentiable by Definition 6.1. From (6.1), we get

$$f(x_0 + h) - f(x_0) - Lh = r(h).$$

Then, taking norms on both sides, dividing by $\|h\|_2$ and the limit $h \to 0$, we get (6.4) by (9.13). Thus, f is differentiable at x_0 by Definition 6.2.
[$\Leftarrow$]: Let $\Omega \subseteq \mathbb{R}^n$ be an open set and let $f : \Omega \to \mathbb{R}^m$ be a function. Let f be differentiable by Definition 6.2. Take the L from (6.1); then we have

$$\begin{aligned} f(x_0 + h) &= f(x_0) + f(x_0 + h) - f(x_0) \\ &= f(x_0) + L \cdot h + f(x_0 + h) - f(x_0) - Lh. \end{aligned}$$

Setting $r(h) = f(x_0 + h) - f(x_0) - Lh$, we are left to prove that

$$\lim_{h \to 0} \frac{\|r(h)\|_2}{\|h\|_2} = 0.$$

This follows from (6.4) by

$$\lim_{h \to 0} \frac{\|r(h)\|_2}{\|h\|_2} = \lim_{h \to 0} \frac{\|f(x_0 + h) - f(x_0) - Lh\|_2}{\|h\|_2} = 0.$$

□

Lemma 6.2 (Uniqueness of the total derivative)**.**
Let $\Omega \subseteq \mathbb{R}^n$ *be open and* $f : \Omega \to \mathbb{R}^m$ *be differentiable at* x_0*. Then,* L *in Definition 6.2 (and by the last*

Lemma in Definition 6.1) is uniquely determined.

Proof. Let $L, M : \mathbb{R}^n \to \mathbb{R}^m$ be linear maps satisfying at x_0 (6.1). Then

$$\begin{aligned}
\lim_{h\to 0} \frac{\|Lh - Mh\|_2}{\|h\|_2} &= \lim_{h\to 0} \frac{\|-f(x_0+h)+f(x_0)+Lh-Mh+f(x_0+h)-f(x_0)\|_2}{\|h\|_2} \\
&\le \lim_{h\to 0} \frac{\|-f(x_0+h)+f(x_0)+Lh\|_2}{\|h\|_2} \\
&\qquad + \lim_{h\to 0} \frac{\|-Mh+f(x_0+h)-f(x_0)\|_2}{\|h\|_2} = 0.
\end{aligned}$$

Now, let $v \in \mathbb{R}^n \setminus \{0\}$ and set $h = tv$. Then, by the linearity of M and L, we get $Mh = tMv$ and $Lh = tLv$. Thus, we obtain

$$\begin{aligned}
\lim_{h\to 0} \frac{\|Lh - Mh\|_2}{\|h\|_2} &= 0 \\
&= \lim_{t\to 0} \frac{|t|\|Lv - Mv\|_2}{|t|\|v\|_2} = \frac{\|Lv - Mv\|_2}{\|v\|_2}.
\end{aligned}$$

Hence, $(L-M)v$ must be zero. This concludes the proof. □

Remark 6.2. *The total derivative has different symbols in the literature. Sometimes it is just denoted by $f'(x_0)$ as in the 1D case or by df_{x_0}.*

Remark 6.3. *As in the 1D case, we can write* (6.1) *as*

$$f(x) = f(x_0) + L(x - x_0) + r(x - x_0),$$

where

$$\lim_{x\to x_0} \frac{\|r(x-x_0)\|_2}{\|x - x_0\|_2} = 0.$$

We can also write this with the small-o notation:

$$f(x_0 + h) = f(x_0) + Lh + o(h).$$

6.1.1 Directional Derivative

Restricting the directions of h in Definition 6.1, we can give

Definition 6.3 (Directional Derivative)**.**
Let $v \in \mathbb{R}^n \setminus \{0\}$, $\Omega \subseteq \mathbb{R}^n$ open, $x_0 \in \Omega$, and $f : \Omega \to \mathbb{R}^m$. Then, if

$$\lim_{t\to 0} t^{-1}\big(f(x_0 + tv) - f(x_0)\big) \tag{6.5}$$

exists, it is called the directional derivative of f in direction v at x_0 and denoted by $D_v f(x_0)$.

Theorem 6.1.
Let $\Omega \subseteq \mathbb{R}^n$ open, $x_0 \in \Omega$, and $f : \Omega \to \mathbb{R}^m$. Then, if f is differentiable at x_0, the directional derivative of f at x_0 exists for all $v \in \mathbb{R}^n\{0\}$ and we have $D_v f(x_0) = Df(x_0)v$.

6.1.2 Partial Derivatives

An important special case of directional derivatives, according to Definition 6.3, are the so-called partial derivatives of a function $f:\Omega\to\mathbb{R}^m$. We set $v=e_k$, where e_k is the kth vector of the canonical basis of $\mathbb{R}^n$, i.e.

$$e_k=\begin{bmatrix} e_{k1}\\ e_{k2}\\ \vdots\\ e_{kn}\end{bmatrix},\quad e_{kj}=\delta_{kj},$$

where δ_{kj} is the Kronecker function given by

$$\delta_{kj}=\begin{cases} 0 & : \quad j\neq k\\ 1 & : \quad j=k\end{cases}.$$

Then, letting $\tilde{e}_j$ be the canonical basis of $\mathbb{R}^m$, we can write

$$f(x)=\begin{bmatrix} f_1(x)\\ \vdots\\ f_m(x)\end{bmatrix}=\sum_{j=1}^m f_j(x)\tilde{e}_j.$$

If f is differentiable (see Def. 6.1), we obtain from (6.5) that

$$\begin{aligned}\lim_{t\to 0} t^{-1}\big(f(x_0+te_k)-f(x_0)\big) &= \sum_{j=1}^m \lim_{t\to 0} t^{-1}\big(f_j(x_0+te_k)-f_j(x_0)\big)\tilde{e}_j\\ &= \begin{bmatrix} \lim_{t\to 0} t^{-1}\big(f_1(x_0+te_k)-f_1(x_0)\big)\\ \vdots\\ \lim_{t\to 0} t^{-1}\big(f_m(x_0+te_k)-f_m(x_0)\big)\end{bmatrix}.\end{aligned}$$

We set

$$\frac{\partial f_j}{\partial x_k}(x_0)=\lim_{t\to 0} t^{-1}\big(f_j(x_0+te_k)-f_j(x_0)\big).$$

We give these special directional derivatives a name by

Definition 6.4 (Partial Derivatives)**.**
Let $\Omega\subseteq\mathbb{R}^n$ open, $p\in\Omega$, and $f:\Omega\to\mathbb{R}^m$. If the limit

$$\lim_{h\to 0}\frac{f_j(p_1,\dots,p_{k-1},p_k+h,p_{k+1},\dots,p_n)-f_j(p_1,\dots,p_n)}{h} \tag{6.6}$$

exists, we denote it by $\frac{\partial f_j}{\partial x_k}(p)$ and call it the kth partial derivative of f_j at p.

Remark 6.4. *As you can see in (6.6), that means that the partial derivative $\frac{\partial f_j}{\partial x_k}(x_0)$ is given by the derivative of the component function f_j with respect to the variable x_k. All other variables are left constant. In the special case $m=1$, we can write*

$$\frac{\partial f}{\partial x_k}(p)=\lim_{h\to 0}\frac{f(p_1,\dots,p_{k-1},p_k+h,p_{k+1},\dots,p_n)-f(p_1,\dots,p_n)}{h},$$

where $p=[p_1,\dots p_n]^T\in\Omega$. Compare Definition 5.1.

Example 6.2. *Consider $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = 2x^2 + 3y$, $(x_0, y_0) \in \mathbb{R}^2$. First, we compute the partial derivative with respect to x:*

$$\begin{aligned}\frac{\partial f}{\partial x}(x_0, y_0) &= \lim_{h\to 0} \frac{f(x_0+h, y_0) - f(x_0, y_0)}{h} \\ &= \lim_{h\to 0} \frac{2(x_0+h)^2 + 3y_0 - 2x_0^2 - 3y_0}{h} = \lim_{h\to 0} \frac{4x_0 h + 2h^2}{h} \\ &= 4x_0.\end{aligned}$$

Now, we compute the partial derivative of f with respect to y:

$$\begin{aligned}\frac{\partial f}{\partial y}(x_0, y_0) &= \lim_{h\to 0} \frac{f(x_0, y_0+h) - f(x_0, y_0)}{h} \\ &= \lim_{h\to 0} \frac{2x_0^2 + 3(y_0+h) - 2x_0^2 - 3y_0}{h} = \lim_{h\to 0} \frac{3h}{h} \\ &= 3.\end{aligned}$$

Theorem 6.2.
Let $\Omega \subseteq \mathbb{R}^n$ open, $x_0 \in \Omega$, and $f : \Omega \to \mathbb{R}^m$. If f is differentiable at x_0, then all partial derivatives $\frac{\partial f_j}{\partial x_k}(x_0)$, $j = 1, \ldots, m$, $k = 1, \ldots, n$ exist and the total derivative $Df(x_0)$ has the matrix representation

$$Df(x_0) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \frac{\partial f_1}{\partial x_2}(x_0) & \ldots & \frac{\partial f_1}{\partial x_n}(x_0) \\ \vdots & \vdots & \ldots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x_0) & \frac{\partial f_m}{\partial x_2}(x_0) & \ldots & \frac{\partial f_m}{\partial x_n}(x_0) \end{bmatrix}.$$

Remark 6.5. *The converse of the last theorem is not true. The existence of all partial derivatives does not imply that f is differentiable. In fact it does not even imply that f is continuous. As an example consider*

$$f(x, y) = \begin{cases} \frac{x^2 y}{x^4 + y^2} & : \quad x^2 + y^2 \neq 0 \\ 0 & : \quad x = y = 0 \end{cases}.$$

To show that this function is not continuous, one has to choose curves other than straight lines going through the origin. In fact, if you choose the coordinate-axis or a line $y = L(x) = mx$ and consider $f(x, mx)$, you find that the limit is 0 as x tends to 0. However, the "symmetry" of the function suggest to look at points (a, a^2) which lead to

$$f(a, a^2) = \frac{a^2 \cdot a^2}{a^4 + a^4} = \frac{1}{2}.$$

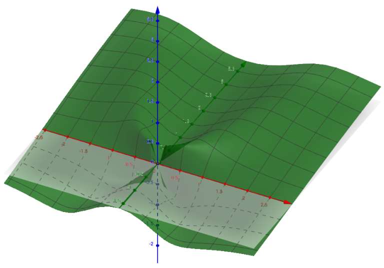
Since we can choose (a, a^2) arbitrarily close to $(0, 0)$, we get the discontinuity of f at $(0, 0)$.

6.2 Meaning of Differentiability in $\mathbb{R}^n$

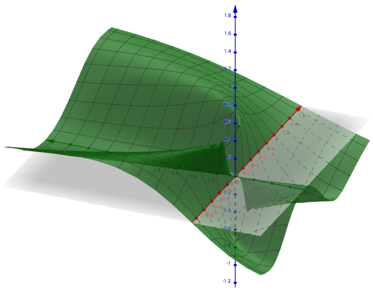
If $f : \Omega \to \mathbb{R}$, then the function graph

$$\left\{(x, f(x)) \in \mathbb{R}^{n+1} : x \in \mathbb{R}^n\right\}$$

is a hypersurface in $\mathbb{R}^{n+1}$. The word hypersurface means that the dimension of the surface is one less than the surrounding space. For example, the sphere $\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$ is a 2-dimensional hypersurface in $\mathbb{R}^3$ as is any plane. In $\mathbb{R}^3$, the hyperplanes are just called planes but in $\mathbb{R}^n$, we distinguish the different planes. in $\mathbb{R}^4$ we have the hyperplanes which are $4 - 1 = 3$-dimensional, we have



(a) The function f from one perspective



(b) and from another respective.

Figure 6.1: GeoGebra plots of f from different perspectives. Use GeoGebra to get a better impression of the function and to see the non-differentiability at $(x, y) = (0, 0)$.

2-dimensional planes and 1-dimensional planes which we call lines.

The function f is then differentiable at x_0 if there exists a tangent plane at x_0, which is a hyperplane in $\mathbb{R}^{n+1}$ (since the graph lives there), which is the graph of

$$g(x) = f(x_0) + L(x - x_0).$$

The set of points $V = \{(x, Lx) : x \in \mathbb{R}^n\}$ is a sub-space of $\mathbb{R}^{n+1}$ and the graph of g is given by $V + (x_0, f(x_0))$. See Analytic Geometry/Linear Algebra. This can be interpreted as small changes in x lead to only small changes in $f(x)$ (in a linear way).

Consider

$$f(x, y) = -\left[(x-1)^2 + (y-1)^2\right] + 5$$

which is a function $f : \mathbb{R}^2 \to \mathbb{R}$. Thus, the graph $\{(x, f(x)) : x \in \mathbb{R}^2\} \subseteq \mathbb{R}^3$ is a 2-dimensional (hyper)surface in $\mathbb{R}^3$.

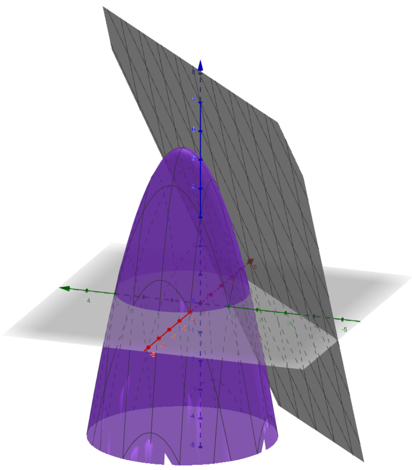


Figure 6.2: The function f with its tangent plane $E(x, y) = 7 - 2x + 2y$ at $(x_0, y_0) = (2, 0)$.

For (x, y) close to $(x_0, y_0) = (2,0)$, one could then write

$$\begin{aligned}
f(x,y) &\approx f(x_0,y_0) + L\begin{bmatrix} x-x_0 \\ y-y_0 \end{bmatrix} \\
&= f(x_0,y_0) + \begin{bmatrix} \frac{\partial f}{\partial x}(x_0,y_0) & \frac{\partial f}{\partial y}(x,y) \end{bmatrix} \begin{bmatrix} x-x_0 \\ y-y_0 \end{bmatrix} \\
&= 7 - 2x + 2y.
\end{aligned}$$

For a function $f : \Omega \to \mathbb{R}$, $\Omega \subseteq \mathbb{R}^n$ open, one may say

$$\begin{aligned}
f(x_1+\Delta x_1, \ldots x_n + \Delta x_n) &\approx f(x_1,\ldots,x_n) + \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) & \frac{\partial f}{\partial x_2}(x) & \ldots & \frac{\partial f}{\partial x_n}(x) \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \\ \vdots \\ \Delta x_n \end{bmatrix} \\
&= f(x_1,\ldots,x_n) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_1,\ldots,x_n)\Delta x_i.
\end{aligned}$$

if f is differentiable at $x \in \Omega$. The $x_i + \Delta x_i$ are just another way of writing $x_i + h_i$, which are the components of $x + h$. Engineers and scientist like this notation and you may see it here and there.
See also your lecture notes from Mathematical Methods II.

This finds applications in error considerations in physics and engineering as one can calculate how much measuring errors propagate into final results. This way one can determine how certain results are and what quantities need to be measures with greater care. See also the Wikipedia article on Propagation of uncertainty or the Weppage Uncertainty as Applied to Measurements and Calculations from John Denker. You can also find some information

CHAPTER

7 Compactness

7.1 Definition

Definition 7.1 (Open cover)**.**
Let $\Omega \subseteq \mathbb{R}^n$ be a subset and $(\mathcal{O}_\alpha)_{\alpha\in I}$ be a collection of open sets $\mathcal{O}_\alpha \subseteq \mathbb{R}^n$ indexed by some (possible uncountable) index set I. Then we say that $(\mathcal{O}_\alpha)_{\alpha\in I}$ is an open cover of Ω iff

$$\Omega \subseteq \bigcup_{\alpha\in I} \mathcal{O}_\alpha,$$

i.e. for all $x \in \Omega$ there exists an $\alpha \in I$ such that $x \in \mathcal{O}_\alpha$.

Example 7.1. *Let $\Omega = \{1,2,3\}$. Then, the family $\{\mathcal{O}_1, \mathcal{O}_2, \mathcal{O}_3\} = (\mathcal{O}_\alpha)_{\alpha\in\{1,2,3\}}$ with $\mathcal{O}_1 = (\frac{1}{2}, \frac{3}{2})$, $\mathcal{O}_2 = (\frac{3}{2}, \frac{5}{2})$, and $\mathcal{O}_3 = (\frac{5}{2}, \frac{7}{2})$ is an open cover.*

Example 7.2. *The set $\Omega = [0,1]$ is covered by $\{\mathcal{O}_1, \mathcal{O}_2\} = (\mathcal{O}_\alpha)_{\alpha\in\{1,2\}}$ with $\mathcal{O}_1 = (-1,1)$, and $\mathcal{O}_2 = (0,2)$.*

Example 7.3. *The closed square $[0,1]^2$ is covered by the 4 open balls of radius 1 around the corner points, i.e. $\{\mathcal{O}_1, \ldots, \mathcal{O}_4\} = (\mathcal{O}_\alpha)_{\alpha\in\{1,2,3,4\}}$ is an open cover with $\mathcal{O}_1 = B_1((0,0))$, $\mathcal{O}_2 = B_1((1,0))$, $\mathcal{O}_3 = B_1((0,1))$, and $\mathcal{O}_4 = B_1((1,1))$.*

Definition 7.2 (Finite sub-cover)**.**
Let $\Omega \subseteq \mathbb{R}^n$ be a subset and $(\mathcal{O}_\alpha)_{\alpha\in I}$ be an open cover of Ω. We say that $(\mathcal{O}_\alpha)_{\alpha\in I}$ admits a finite sub-cover iff there exists a finite subset $J \subseteq I$ such that

$$\Omega \subseteq \bigcup_{\alpha\in J} \mathcal{O}_\alpha.$$

Definition 7.3 (Compact set)**.**
A set $\Omega \subseteq \mathbb{R}^n$ is called compact iff every open cover $(\mathcal{O}_\alpha)_{\alpha\in I}$ of Ω admits a finite sub-cover.

Remark 7.1. *One also says that $\Omega \subseteq \mathbb{R}^n$ is compact iff it has the Heine-Borel property. The Heine-Borel property is that any open cover of Ω admits a finite sub-cover.*

Example 7.4. *First we consider a couple of non-examples and the simplest examples and then we give criteria that identify the sub-sets of $\mathbb{R}^n$ which are compact.*

1. *The set* $(0,1)$ *is not compact as the cover* $\bigcup_{\varepsilon>0}(\varepsilon,1)$ *does not admit a finite sub-cover. Notice that* $(\varepsilon_2,1)\supset(\varepsilon_1,1)$ *if* $0<\varepsilon_2<\varepsilon_1$ *but never* $(0,1)=(\varepsilon,1)$ *however small one chooses* $\varepsilon>0$.

2. $\mathbb{R}$ *is not compact as the open cover*
$$\mathbb{R}=\bigcup_{R>0}(-R,R)$$
does not have a finite sub-cover.

3. $\mathbb{R}^n$ *is not compact as*
$$\mathbb{R}^n=\bigcup_{R>0}(-R,R)^n$$
does not admit a finite sub-cover.

4. *A point* $\{x\}$ *(and with that any finite set) is a compact set since from any open cover* $(U_\alpha)_{\alpha\in I}$*, we can choose always one* U_α *that covers* $\{x\}$*. Compare the Definition of open sets Definition 2.6.*

Theorem 7.1.
The closed interval $[a,b]$ *is compact.*

Proof. Assume that there exists an open cover $(\mathcal{O}_\alpha)_{\alpha\in I}$ of $[a,b]$ which does not admit a finite sub-cover. We bisect the interval into
$$[a,b]=[a,c_1]\cup[c_1,b],$$
where $c_1=\frac{a+b}{2}$. Now, $(\mathcal{O}_\alpha)_{\alpha\in I}$ also covers $[a,c_1]$ and $[c_1,b]$. At least one of them does not admit a finite sub-cover, since otherwise the union of the two finite sub-covers would e a finite sub-cover of $[a,b]$. We call the non-coverable interval I_1 and repeat the bisection and the same argument. We then obtain a sequence of intervals with
$$I_1\supset I_2\supset I_3\subset\ldots$$
such that the length of I_n is given by $|b-a|2^{-n}$. Moreover, by construction, none of the I_n can be covered by a finite sub-cover of $(\mathcal{O}_\alpha)_{\alpha\in I}$. We pick a sequence $(x_k)_{k\in\mathbb{N}}$ by picking $x_k\in I_k$. By construction, for all $m,k\geq k_0$, we have $x_m,\ x_k\in I_{k_0}$ and, hence,
$$|x_m-x_k|\leq|b-a|2^{-k_0}<|b-a|2^{-k_0+1}.$$
Thus, the sequence $(x_k)_{k\in\mathbb{N}}$ is a Cauchy sequence and, by the completeness of $\mathbb{R}$, it converges to a point $x\in[a,b]$. Since $(\mathcal{O}_\alpha)_{\alpha\in I}$ covers $[a,b]$ there must exist an $\alpha\in I$ such that $x\in\mathcal{O}_\alpha$. Since the $\mathcal{O}_\alpha$ is open there exists an $\varepsilon>0$ such that $B_\varepsilon(x)\subseteq\mathcal{O}_\alpha$ and, since $x_k\to x$ as $k\to+\infty$, we can choose $K>0$ such that
$$k\geq K\quad\Rightarrow\quad|x_k-x|<\frac{\varepsilon}{2}$$
and such that $2^{-K+1}|b-a|<\frac{\varepsilon}{2}$. Then, $y\in I_K$ implies
$$|x-y|\leq|x-x_K|+|x_K-y|<\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.$$
This means $I_N\subseteq B_\varepsilon(x)\subseteq\mathcal{O}_\alpha$ and, hence, I_N is covered by a single element of $(\mathcal{O}_\alpha)_{\alpha\in I}$. This contradicts our assumption since by construction, non of the I_n can be covered by a finite sub-cover of $(\mathcal{O}_\alpha)_{\alpha\in I}$. □

Using the same technique, one can show

Theorem 7.2.
Let $n \in \mathbb{N}$ and $a_i \leq b_i$ real numbers for $i = 1, \dots, n$. The cuboid

$$[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n] \subseteq \mathbb{R}^n$$

is compact.

7.2 Properties of compact sets

Theorem 7.3 (Compact sets are closed)**.**
Let $\Omega \subseteq \mathbb{R}^n$ be compact. Then, Ω is closed, i.e. contains all its limit points.

Proof. Let $x \in \mathbb{R}^n$ be a limit point of Ω with $x \neq \Omega$. We show that Ω can then not be compact. We do that by constructing an open cover of Ω which has no finite sub-cover. Let

$$\begin{aligned} X_\varepsilon &:= \{y \in \mathbb{R}^n : d(y,x) > \varepsilon\} \\ &= \mathbb{R}^n \setminus \{y \in \mathbb{R}^n : d(y,x) \leq \varepsilon\}. \end{aligned}$$

The set X_ε is open (see Proposition 2.3) and $(X_\varepsilon)_{\varepsilon \in (0,+\infty)}$ covers Ω since

$$\mathbb{R}^n \setminus \{x\} = \bigcup_{\varepsilon > 0} X_\varepsilon.$$

Since x is a limit point of Ω, we have that for every $\varepsilon > 0$, the set X_ε contains at least one more point of Ω which implies that $(X_\varepsilon)_{\varepsilon \in (0,+\infty)}$ does not admit a finite sub-cover. $\square$

Theorem 7.4 (Compact sets are bounded)**.**
Let $\Omega \subseteq \mathbb{R}^n$ be compact. Then, Ω is a bounded set, i.e. there exists an $R > 0$ such that $\Omega \subseteq B_R(0)$.

Proof. The family $(B_R(0))_{R \in (0,+\infty)}$ is an open cover of Ω and, since Ω is compact, there exists a finite sub-cover $B_{R_1}(0), \dots, B_{R_n}(0)$. Then, Ω is contained in $B_R(0)$ with $R = \max\{R_1, \dots, R_n\}$. $\square$

Theorem 7.5.
If $\Omega \subseteq \mathbb{R}^n$ is compact, then every sequence $(x_n)_{n \in \mathbb{N}_0} \subseteq \Omega$ has a convergent subsequence with limit in Ω.

To prove that we first prove the following generalisation of Theorem 1.3.

Theorem 7.6 (Bolzano-Weierstrass)**.**
Every bounded sequence $(x_k)_{k \in \mathbb{N}_0} \subseteq \mathbb{R}^n$ has a convergent subsequence.

Proof. Let $(x_k)_{k \in \mathbb{N}_0} \subseteq \mathbb{R}^n$ be a bounded sequence, i.e. there exists a constant $C > 0$ such that $\|x_n\|_2 \leq C$ for all $n \geq 0$. We denote $x_k = [x_1^{(k)}, \dots, x_n^{(k)}]^T$. The sequence $(x_1^{(k)})_{k \in \mathbb{N}_0}$ of first components of the elements of $(x_k)_{k \in \mathbb{N}_0}$ is a bounded real sequence. Thus, by Theorem 7.6, there exists a convergent sub-sequence $(x_1^{(k_l)})_{l \in \mathbb{N}_0} \subseteq (x_1^{(k)})_{k \in \mathbb{N}_0}$. Now we consider $(x_{k_l})_{l \in \mathbb{N}_0}$ the corresponding subsequence of $(x_k)_{k \in \mathbb{N}_0}$. By the same arguments, we can now pick a convergent subsequence of the $(x_2^{(k_l)})_{l \in \mathbb{N}_0}$ and thus a corresponding sub-sequence of $(x_{k_l})_{l\in\mathbb{N}_0}$ which is then a subsequence of $(x_k)_{k\in\mathbb{N}_0}$. Continuing like this for the remaining $n-2$ components, we obtain a sub-sequence of $(x_k)_{k\in\mathbb{N}_0}$ in which the sequences of all components converge and, therefore, the sequence itself converges in $\mathbb{R}^n$. □

Now to the proof of Theorem 7.5.

Proof of Theorem 7.5. By Theorems 7.3 and 7.4, we have that Ω is closed and bounded. Thus, any sequence $(x_n)_{n\in\mathbb{N}_0} \subseteq \Omega$ is bounded and, by Theorem 7.6, has a convergent subsequence $(x_{k_l})_{l\in\mathbb{N}_0}$. Since $(x_{k_l})_{l\in\mathbb{N}_0} \subseteq \Omega$, we have, by Remark 2.11, that the limit of $(x_{k_l})_{l\in\mathbb{N}_0}$ belongs to Ω. □

We now prove a property of closed sub-sets of compact sets.

Theorem 7.7.
Let $\Omega \subseteq \mathbb{R}^n$ be compact and $C \subseteq \Omega$ be closed. Then, C is compact.

Proof. Let $(\mathcal{O}_\alpha)_{\alpha\in I}$ be an open cover of C. By Proposition 2.3, we have that $\mathbb{R}^n \setminus C$ is open since C is closed. Thus, $(\mathcal{O}_\alpha)_{\alpha\in I} \cup \mathbb{R}^n \setminus C$ covers Ω and, since Ω is compact, there exists a finite sub-cover that covers Ω. Omitting $\mathbb{R}^n \setminus C$, we get a finite sub-cover of C. □

Remark 7.2. *A mistake that one may make at the beginning is the following: Since Ω is compact, we have that every open cover $(\mathcal{O}_\alpha)_{\alpha\in I}$ contains a finite sub-cover with $\Omega \subseteq \mathcal{O}_{\alpha_1} \cup \cdots \cup \mathcal{O}_{\alpha_k}$. Since $C \subseteq \Omega$, we have $C \subseteq \mathcal{O}_{\alpha_1} \cup \cdots \cup \mathcal{O}_{\alpha_k}$. This does not prove the claim in Theorem 7.7 since this only concerns the covers which also cover Ω. However, one has to prove that* every *open cover of C admits a finite sub-cover.*

The main theorem of this section is

Theorem 7.8 (Heine–Borel)**.**
A set $\Omega \subseteq \mathbb{R}^n$ is compact iff it is closed and bounded.

Proof. $[\Rightarrow]$: Follows from Theorems 7.4 and 7.3.
$[\Leftarrow]$: Let Ω be closed and bounded. Then, there is a $L > 0$ such that

$$\Omega \subseteq [-L, L]^n.$$

By Theorems 7.7 and 7.2, we have that Ω is compact. □

7.3 Uniform continuity – Theorem of Heine

Let us recall the definition of a function $f : (a,b) \to \mathbb{R}$ being continuous on (a,b): For all $x_0 \in (a,b)$ and all $\varepsilon > 0$ there exists $\delta = \delta(x_0, \varepsilon) > 0$ such that

$$|x - x_0| < \delta \quad \Rightarrow \quad |f(x) - f(x_0)| < \varepsilon.$$

Important is, that we can not avoid, in general, that δ depends not only on ε but on the point x_0 as well. An example, with the appropriate interpretations at the boundary, is $f : [0, +\infty) \to \mathbb{R}$, $f(x) = \sqrt{x}$, where we have to choose $\delta = \sqrt{x_0}\varepsilon$. Can we choose δ in a different way such that it does not depend on x_0? Let us first consider f on $[1, +\infty)$:

$$|f(x) - f(x_0)| = \left|(\sqrt{x} - \sqrt{x_0})\frac{\sqrt{x}+\sqrt{x_0}}{\sqrt{x}+\sqrt{x_0}}\right| \le \frac{1}{1+\sqrt{x_0}}|x - x_0| \le \frac{1}{2}|x - x_0| < |x - x_0|.$$

Which gives us

$$|x-x_0|<\delta \quad \Rightarrow \quad |f(x)-f(x_0)|<\varepsilon$$

for $\delta=\varepsilon$. Now let us look at f on $[0,1]$. By the triangle inequality, we get

$$|\sqrt{x}-\sqrt{x_0}| \le |\sqrt{x}|+|\sqrt{x_0}| = |\sqrt{x}+\sqrt{x_0}|$$

which gives

$$\begin{aligned} |\sqrt{x}-\sqrt{x_0}|^2 &= |\sqrt{x}-\sqrt{x_0}||\sqrt{x}-\sqrt{x_0}| \\ &\le |\sqrt{x}-\sqrt{x_0}||\sqrt{x}+\sqrt{x_0}| = |x-x_0| \end{aligned}$$

which yields

$$|x-x_0|<\delta \quad \Rightarrow \quad |f(x)-f(x_0)|<\varepsilon$$

for $\delta=\varepsilon^2$. Thus, we can choose $\delta=\varepsilon^2$ on $[0,+\infty)$.

For which functions can we do this in general? To investigate this question further, let us first be more precise. We introduce

Definition 7.4 (Uniform continuity)**.**
Let I be an interval. A function $f: I \to R$ said to be uniformly continuous iff for all $\varepsilon>0$ there exists $\delta>0$ such that

$$\forall x,y \in I: \; |x-y|<\delta \quad \Rightarrow \quad |f(x)-f(y)|<\varepsilon.$$

Remark 7.3. *Let us contrast the definition of continuity again against uniform continuity: for the first, we have*

$$\forall x\in I \quad \forall \varepsilon>0 \quad \exists \delta>0 \; : \big(\forall y \in I: \quad |x-y|<\delta\big) \quad \Rightarrow \quad |f(x)-f(y)|<\varepsilon$$

while uniform continuity means

$$\forall \varepsilon>0 \quad \exists \delta>0 \quad \big(\forall x,y \in I \; : \quad |x-y|<\delta\big) \quad \Rightarrow \quad |f(x)-f(y)|<\varepsilon.$$

Example 7.5. *Let us consider $f:(0,1)\to \mathbb{R}$, $f(x)=\frac{1}{x}$. We compute*

$$|f(x_0)-f(x)| = \left|\frac{1}{x}-\frac{1}{x_0}\right| = \left|\frac{x_0-x}{x_0 x}\right| \le \frac{1}{|x_0||x|}|x-x_0| = \frac{1}{x_0}\cdot\frac{1}{x}\cdot|x-x_0|.$$

Now, from $|x-x_0|<\delta$, we obtain

$$x_0-\delta<x<x_0+\delta \quad \Rightarrow \quad \frac{1}{x}<\frac{1}{x_0-\delta}.$$

Assume $\delta \le \frac{x_0}{2}$, which gives us $\frac{1}{x}<\frac{1}{x_0-\delta}\le\frac{2}{x_0}$. Thus, we get

$$|f(x_0)-f(x)| \le \frac{1}{x_0 x}|x-x_0| \le \frac{1}{x_0}\frac{2}{x_0}|x-x_0|.$$

Hence, the choice $\delta=\min\left\{\frac{x_0}{2},\frac{x_0^2\varepsilon}{2}\right\}$ yields

$$|x-x_0|<\delta \quad \Rightarrow \quad |f(x)-f(x_0)|<\varepsilon.$$

Thus, f is continuous on $(0,1)$. However, we will see that it is not uniformly continuous. To prove that, let us assume that f is uniformly continuous, i.e. for all $\varepsilon>0$ there exists a $\delta>0$ such that $|x-y|<\delta$ implies

$|f(x)-f(y)|<\varepsilon$. Now, we set $\varepsilon=\frac{1}{2}$, i.e. we have a $\delta>0$ such that $|x-y|<\delta$ implies $|f(x)-f(y)|<\frac{1}{2}$. Thus, we can choose a number $\eta>0$ such that $\eta<\delta$ and set $x=\eta$ and $y=\frac{\eta}{2}$. Thus,

$$|x-y|=\left|\frac{\eta}{2}\right|<\eta<\delta.$$

Now, we get

$$|f(x)-f(y)|=\left|\frac{1}{\eta}-\frac{2}{\eta}\right|=\frac{1}{\eta}>\frac{1}{2}$$

if $\eta>0$ is small enough. Thus, we reached a contradiction and f can not be uniformly continuous.

From the next theorem, we learn that continuous functions on compact sets are always uniformly continuous on that set.

Theorem 7.9 (Heine's theorem)**.**
Let $f:[a,b]\to\mathbb{R}$ be continuous. Then, f is uniformly continuous on $[a,b]$.

Remark 7.4. *Let me recall the main ideas of the proof first. The first part is to divide the interval $[a,b]$ in pieces on which we can control f. By compactness, we can write*

$$[a,b]=B_{\frac{\delta(x_1,\varepsilon)}{2}}(x_1)\cup\cdots\cup B_{\frac{\delta(x_k,\varepsilon)}{2}}(x_k).$$

We have that for all $i\in\{1,\dots,k\}$ it holds

$$x,y\in B_{\frac{\delta(x_i,\varepsilon)}{2}}(x_i)\quad\Rightarrow\quad|f(x)-f(y)|<\varepsilon. \tag{7.1}$$

If we choose now

$$\delta:=\frac{1}{2}\min\{\delta(x_1,\varepsilon),\dots,\delta(x_k,\varepsilon)\}$$

e get a δ which depends only on ε.We would like to have that whenever $x,y\in[a,b]$ with $|x-y|<\delta$, we get $|f(x)-f(y)|<\varepsilon$. However, we only have (7.1). (See that it is not quite what we need?) What can we do? We could try to estimate

$$|f(x)-f(y)|=|f(x)-f(x_j)+f(x_j)-f(y)|\le|f(x)-f(x_j)|+|f(x_j)-f(y)|,$$

where $j\in\{1,\dots,k\}$ is such that $x\in B_{\delta(x_j,\varepsilon)}(x_j)$. Now we can use (7.1) to get

$$|f(x)-f(y)|\le|f(x)-f(x_j)|+|f(x_j)-f(y)|<\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon$$

as $|x-x_j|<\delta(x_j,\varepsilon)$ for $x\in B_{\frac{\delta(x_j,\varepsilon)}{2}}(x_j)$. The details are provided below where I dropped the dependence on ε in the notation.

Proof. Proof of Theorem 7.9. Since f is continuous at $x\in[a,b]$, there exists $\delta_x>0$ such that

$$|x-y|<\delta_x\quad\Rightarrow\quad|f(x)-f(y)|<\frac{\varepsilon}{2}.$$

We define

$$B_{\frac{\delta_x}{2}}(x)=\left\{y\in[a,b]:|x-y|<\frac{\delta_x}{2}\right\}.$$

The $B_{\frac{\delta_x}{2}}(x)$ form an open cover of $[a,b]$. Since $[a,b]$ is compact, there exists a finite sub-cover $B_{\frac{\delta_{x_1}}{2}}(x)$, $\dots$, $B_{\frac{\delta_{x_k}}{2}}(x)$ still covering $[a,b]$. Let now

$$\delta:=\frac{1}{2}\min\{\delta_{x_1},\dots,\delta_{x_k}\}.$$

If now $|x-y|<\delta$, we claim that we have $|f(x)-f(y)|<\varepsilon$. Since the $B_{\frac{\delta_{x_1}}{2}}(x), \ldots, B_{\frac{\delta_{x_k}}{2}}(x)$ cover $[a,b]$, we have that for all $x \in [a,b]$, there exists a $j \in \{1,\ldots,k\}$ such that $x \in B_{\frac{\delta_{x_j}}{2}}(x)$, i.e. $|x-x_j| < \frac{\delta_{x_j}}{2} < \delta_{x_j}$ and, therefore, $|f(x)-f(x_j)| < \frac{\varepsilon}{2}$ (that is how we choose the δ_x in the beginning.) Moreover, we have

$$|y-x_j| \leq |y-x| + |x-x_j| < \frac{\delta_{x_j}}{2} + \frac{\delta_{x_j}}{2} = \delta_{x_j}$$

which then gives $|f(y)-f(x_j)| < \frac{\varepsilon}{2}$. Thus,

$$|f(y)-f(x)| \leq |f(x)-f(x_j)| + |f(y)-f(x_j)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

$\square$

Chapter

8 Sequences of functions

8.1 Notions of convergence

Consider $f_n : \mathbb{R} \to \mathbb{R}$ with $f_n(x) = e^{-nx^2}$. If $x = 0$, we have $f_n(x) = 1$. If $x \neq 0$, we have $f_n(x) = e^{-nx^2} = \left(e^{-x^2}\right)^n$. This implies $f_n(x) \to 0$ as $n \to +\infty$. Hence, $(f_n)_{n\in\mathbb{N}_0}$ for every fixed x to

$$f(x) = \begin{cases} 1 & : \quad x = 0 \\ 0 & : \quad x \neq 0 \end{cases}.$$

Note that the limit function f is not continuous even though all f_n are continuous functions. Let us make this notion of convergence more precise with

Definition 8.1 (Pointwise convergence)**.**
Let $\Omega \subseteq \mathbb{R}^n$ and $(f_n)_{n\in\mathbb{N}_0}$ be a sequence of functions $f_n : \Omega \to \mathbb{R}$. We say $f_n \to f$ pointwise as $n \to +\infty$ iff

$$\lim_{n\to+\infty} f_n(x) = f(x), \tag{8.1}$$

where $(f_n(x))_{n\in\mathbb{N}_0} \subseteq \mathbb{R}$.

The line (8.1) can be rewritten as

$$\forall x \in \Omega \quad \forall \varepsilon > 0 \quad \exists n_0 = n_0(x, \varepsilon) \in \mathbb{N}_0 : \ n \geq n_0 \quad \Rightarrow \quad |f_n(x) - f| < \varepsilon.$$

Note the dependence of n_0 not only on ε but also on x. In the above example, one can see this clearly as n_0 must be larger and larger as x is closer and closer to 0. The result is, that the limit function f is not continuous. Could we get better convergence behavior if we ask $|f_n(x) - f(x)| < \varepsilon$ not for one point but over all $x \in \Omega$, i.e. we define convergence by requiring that there exists an index n_0 such that $f_n(x)$ differs from $f(x)$ no more than ε for all x. To be more precise, let us introduce

Definition 8.2 (Uniform convergence)**.**
Suppose $\Omega \subseteq \mathbb{R}^n$ and $(f_n)_{n\in\mathbb{N}_0}$ is a sequence of functions $f_n : \Omega \to \mathbb{R}$. We then say $f_n \to f$ uniformly as $n \to +\infty$ iff

$$\forall \varepsilon > 0 \quad \exists n_0 = n_0(\varepsilon) \in \mathbb{N} : \ n \geq n_0 \quad \Rightarrow \quad \sup_{x\in\Omega} |f(x) - f_n(x)| < \varepsilon. \tag{8.2}$$

Generalizing the considerations in Section 5.5, we denote by $C(\Omega)$ the set of all continuous functions $\Omega \to \mathbb{R}$. The reader may convince herself that $C(\Omega)$ is a real vector space. As before, we can introduce a norm by

$$\|f\|_\Omega := \sup_{x\in\Omega} |f(x)|$$

As in the case of $\Omega = [a,b]$, we can only expect the norm to be finite if Ω is a compact set. See Theorem **??**. When necessary, we will make the appropriate remarks.

We can restate (8.2) as

$$\forall \varepsilon > 0 \quad \exists n_0 \in \mathbb{N} : \ n \geq n_0 \quad \Rightarrow \quad \|f_n - f\| < \varepsilon$$

or

$$\lim_{n\to+\infty} \|f_n - f\|_\infty = 0.$$

Proposition 8.1 (Uniform ⇒ pointwise)**.**
Let $\Omega \subseteq \mathbb{R}^n$ and $(f_n)_{n\in\mathbb{N}_0}$ be a sequence of functions such that there exists a function $f : \Omega \to \mathbb{R}$ with $f_n \to f$ uniformly. Then $f_n \to f$ pointwise.

Exercise 8.1. *Prove Proposition 8.1.*

We can of course formulate a Cauchy-criterion

Definition 8.3 (Cauchy criterion)**.**
Suppose $\Omega \subseteq \mathbb{R}^n$ and $(f_n)_{n\in\mathbb{N}_0}$ is a sequence of functions $f_n : \Omega \to \mathbb{R}$ is uniformly Cauchy on Ω of for every $\epsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that

$$\forall m, n \geq n_0 \ : \ \sup_{x\in\Omega} |f_n(x) - f_m(x)| < \varepsilon.$$

One can prove the following

Theorem 8.1.
Suppose $\Omega \subseteq \mathbb{R}^n$ and $(f_n)_{n\in\mathbb{N}_0}$ is a sequence of functions $f_n : \Omega \to \mathbb{R}$. Then, (f_n) is uniformly convergent iff it is uniformly Cauchy.

Exercise 8.2. *Prove Theorem 8.1. First assume that $(f_n)_{n\in\mathbb{N}_0}$ is uniformly convergent and deduce that it is Cauchy. Review the case for real sequences. For the converse prove first that $(f_n(x)_{n\in\mathbb{N}_0}) \subseteq \mathbb{R}$ is Cauchy in $\mathbb{R}$ and then set $f(x) = \lim_{n\to+\infty} f_n(x)$ for all $x \in \Omega$. Then show the uniform convergence $f_n \to f$.*

The central result of this section is

Theorem 8.2.
Suppose $\Omega \subseteq \mathbb{R}^n$ and $(f_n)_{n\in\mathbb{N}_0}$ is a sequence of continuous functions $f_n : \Omega \to \mathbb{R}$. Assume that $f_n \to f$ uniformly for a function $f : \Omega \to \mathbb{R}$. Then, f is continuous.

Proof. We assume that there exists a function f such that $(f_n)_{n\in\mathbb{N}}$ is uniformly convergent to f. We show that f is continuous, i.e. that for all $x_0 \in \Omega$ and all $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\|x - x_0\|_2 < \delta \quad \Rightarrow \quad |f(x) - f(x_0)| < \varepsilon.$$

We compute

$$\begin{aligned} |f(x)-f(x_0)| &\le \sup_{x\in\Omega}|f(x)-f(x_0)| \\ &= \sup_{x\in\Omega}|f(x)-f_n(x)+f_n(x)-f(x_0)| \\ &\le \sup_{x\in\Omega}|f(x)-f_n(x)|+\sup_{x\in\Omega}|f_n(x)-f(x_0)|. \end{aligned}$$

Since $f_n \to f$ uniformly, there exists $n_1 \in \mathbb{N}$ such that $\sup_{x\in\Omega}|f(x)-f_n(x)| < \frac{\varepsilon}{3}$. Thus, we have

$$\begin{aligned} |f(x)-f(x_0)| &< \frac{\varepsilon}{3}+\sup_{x\in\Omega}|f_n(x)-f(x_0)| \\ &= \frac{\varepsilon}{3}+\sup_{x\in\Omega}|f_n(x)-f_n(x_0)+f_n(x_0)-f(x_0)| \\ &\le \frac{\varepsilon}{3}+\sup_{x\in\Omega}|f_n(x)-f_n(x_0)|+\sup_{x\in\Omega}|f_n(x_0)-f(x_0)|. \end{aligned}$$

By Proposition 8.1, there exists $n_2 \in \mathbb{N}$ such that $\sup_{x\in\Omega}|f_n(x_0)-f(x_0)| = |f_n(x_0)-f(x_0)| < \frac{\varepsilon}{3}$ for all $n \ge n_2$. Further, the f_n are continuous on Ω, hence, there exists for all $n \ge n_0 = \max\{n_1,n_2\}$ a $\delta = \delta(n,\varepsilon) > 0$ such that $\sup_{x\in\Omega}|f_n(x)-f_n(x_0)| < \frac{\varepsilon}{3}$. Thus, we have that

$$\forall x_0 \in \Omega \quad \forall \varepsilon > 0 \quad \exists \delta > 0 : \|x-x_0\|_2 < \delta \quad \Rightarrow \quad |f(x)-f(x_0)| < \varepsilon.$$

□

8.2 Examples & counterexamples

Example 8.1 (Do pointwise limits preserve boundedness?)**.**
Consider $(f_n)_{n\in\mathbb{N}}$ with $f_n : (0,1) \to \mathbb{R}$ given by

$$f_n(x) = \frac{n}{nx+1}.$$

We have

$$\lim_{n\to+\infty} f_n(x) = \lim_{n\to+\infty} \frac{1}{x+\frac{1}{n}} = \frac{1}{x}.$$

Thus, $f_n \to f$ pointwise as $n \to +\infty$. We have $|f_n(x)| < n$ for $x \in (0,1)$, i.e. f_n is bounded on $(0,1)$ for any n but the pointwise limit is not.

Example 8.2. *One can make the last example stronger. A pointwise convergent sequence need not to be bounded even if it converges pointwise to 0. Consider $(f_n)_{n\in\mathbb{N}}$ with $f_n : [0,1] \to \mathbb{R}$ defined by*

$$f_n(x) = \begin{cases} 2n^2x & : \quad 0 \le x \le \frac{1}{2n} \\ 2n^2\left(\frac{1}{n}-x\right) & : \quad \frac{1}{2n} < x < \frac{1}{n} \\ 0 & : \quad \frac{1}{n} \le x \le 1 \end{cases}.$$

If $x \in (0,1]$, we get $f_n(x) = 0$ for all $n \ge \frac{1}{x}$, i.e. $f_n(x) \to 0$ as $n \to +\infty$. If $x = 0$, we have $f_n(x) = 0$ for all $n \ge 1$. Thus, $f_n \to 0$ pointwise on $[0,1]$ as $n \to +\infty$. We have $\max_{x\in[0,1]} f_n(x) = n \to +\infty$ as $n \to +\infty$.

Exercise 8.3. *Consider $(f_n)_{n\in\mathbb{N}}$ with $f_n : \mathbb{R} \to \mathbb{R}$ defined by*

$$f_n(x) = \frac{\sin(nx)}{n}.$$

The sequence f_n converges pointwise to $f \equiv 0$ as well as uniformly. For all $\varepsilon > 0$, we have $\|f_n\|_\infty < \varepsilon$ for $n \ge \lfloor \frac{1}{\varepsilon} \rfloor + 1$, where $\lfloor \frac{1}{\varepsilon} \rfloor$ means the largest integer smaller or equal to $\frac{1}{\varepsilon}$.

8.3 (Real) Power series

For this section, you may also review your Analysis I notes. You will learn much more about power series in the complex setting in the module Complex Variables.

Theorem 8.3 ((Real) Power series)**.**
Let $(a_n) \subseteq \mathbb{R}$ be a sequence of numbers and suppose that

$$\sum_{k=0}^{+\infty} |a_n| R^n < +\infty. \tag{8.3}$$

Then, the sequence $(f_N)_{N\in\mathbb{N}_0}$ of partial sums

$$f_N(x) = \sum_{k=0}^{N} a_n x^n$$

of functions converges uniformly on $[-R, R]$.

Proof. We show that for every $\varepsilon > 0$ there exists an index $n_0 \in \mathbb{N}$ such that

$$\forall n \geq n_0 \quad \Rightarrow \quad \|f_n - f\|_\infty < \varepsilon.$$

Since $|x^n| \leq R^n$ for $x \in [-R, R]$, we have that

$$\sum_{k=0}^{+\infty} a_n x^n$$

converges for every $x \in [-R, R]$ by (8.3). We have

$$\sup_{x\in[-R,R]} |f_n(x) - f(x)| = \sup_{x\in[-R,R]} \left| \sum_{k=n+1}^{+\infty} a_k x^k \right| \leq \sum_{k=n+1}^{+\infty} |a_k| \|x^k\|_\infty = \sum_{k=n+1}^{+\infty} |a_k| R^k.$$

Since (8.3), we have that

$$\lim_{n\to+\infty} \sum_{k=n}^{+\infty} |a_k| R^k = 0.$$

Thus, we have that there exists n_0 such that

$$\sum_{k=n}^{+\infty} |a_k| R^k < \varepsilon \quad \forall n \geq n_0.$$

Hence, we obtain

$$\sup_{x\in[-R,R]} |f_n(x) - f(x)| < \varepsilon$$

for $n \geq n_0$. This concludes the proof. □

Example 8.3. *The series*

$$\sum_{k=0}^{+\infty} \frac{x^k}{k!} \tag{8.4}$$

converges uniformly. We define the exponential function $e^x = \exp(x)$ by

$$e^x := \sum_{k=0}^{+\infty} \frac{x^k}{k!} \tag{8.5}$$

.

Exercise 8.4. *Check that (8.4) converges uniformly. Use the Definition (8.5) to prove $e^{x+y} = e^x e^y$.*

Example 8.4. *We define (see Taylor's Theorem and the Mathematical Methods II module) the following functions*

$$\sin(x) := \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1}, \tag{8.6}$$

$$\cos(x) := \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k)!} x^{2k}. \tag{8.7}$$

Exercise 8.5. *Check that the series in the definitions* (8.6) *and* (8.7) *converge uniformly and deduce*

$$\frac{d}{dx}\sin(x) = \cos(x), \quad \frac{d}{dx}\cos(x) = -\sin(x).$$

You may for the moment ignore the issue of when one is allowed to exchange the differentiation and summation.[1]

[1] One can do that exactly when the series is uniformly convergent.

CHAPTER

9 Integration

9.1 Step functions and their integrals

first, we remind ourselves about a definition from set theory and introduce the notion of an indicator function.

Definition 9.1 (Indicator function)**.**
Let $A \subseteq \mathbb{R}^n$. Then, the function

$$\chi_A(x) = \begin{cases} 1 & : \quad x \in A \\ 0 & : \quad x \neq A \end{cases}.$$

is called indicator function of A.

Remark 9.1. *The letter χ is a the Greek letter* chi*.*

The following properties of characteristic functions are useful.

Proposition 9.1 (Properties of characteristic functions)**.**
Let $\mathcal{U}$ be a set and $A, B \subseteq \mathcal{U}$.

(i) If $A \subseteq B$ then $\chi_B(x) \geq \chi_A(x)$.

(ii) The characteristic function of A^c is given by

$$\chi_{A^c}(x) = 1 - \chi_A(x).$$

(iii) For $A \cap B$, we have $\chi_{A\cap B}(x) = \chi_A(x)\chi_B(x)$.

(iv) For $A \cup B$, we have

$$\chi_{A\cup B}(x) = \chi_A(x) + \chi_B(x) - \chi_A(x)\chi_B(x).$$

(v) The characteristic function of $A \setminus B$ is given by

$$\chi_{A\setminus B}(x) = \chi_A(x) - \chi_A(x)\chi_B(x).$$

Exercise 9.1. *Prove Proposition 9.1 in detail.*

The next notion is a class of simple function which will help us to construct integrals by means of an approximating procedure.

Definition 9.2 (Step function).
Let $I \subseteq \mathbb{R}$ be an interval. Then, a function $g : I \to \mathbb{R}$ is called a step-function if g can be written as

$$g(x) = \sum_{i=1}^{n} c_i \chi_{A_i}(x) \tag{9.1}$$

for $c_i \in \mathbb{R}$ and $A_i \subseteq I$ intervals for $i = 1, \ldots, n$.

Example 9.1. *The function $f : [0,10] \to \mathbb{R}$, given by*

$$g(x) = \begin{cases} 1 & : \quad 0 \leq x \leq 3 \\ -3 & : \quad 3 < x < 4 \\ 8 & : \quad 4 \leq x \leq 9 \\ 2 & : \quad 9 < x \leq 10 \end{cases}$$

is a step function as we can write

$$g(x) = \chi_{[0,3]}(x) - 3\chi_{(3,4)}(x) + 8\chi_{[4,9]}(x) + 2\chi_{(9,10]}(x).$$

Example 9.2. *The intervals A_i in Definition 9.2 can always assumed to be disjoint. We have*

$$g : [0,3] \to \mathbb{R}, \quad g(x) = \chi_{[0,2]}(x) + 3\chi_{[1,3]}(x)$$

which can be written as

$$g(x) = \chi_{[0,1)}(x) + 4\chi_{[1,2]}(x) + 3\chi_{(2,3]}(x).$$

Remark 9.2. *From the last example, we get the alternative characterization of step functions. Namely, $g : I \to \mathbb{R}$ is a step-function iff there exists a partition $(x_1, \ldots, x_n)$, where $a = x_1 < x_2 < \cdots < x_{n-1} < x_n = b$ such that g is constant on each of the intervals (x_i, x_{i+1}) for $i = 1, \ldots, n-1$.*

Remark 9.3. *Suppose g is a step function by Definition 9.2. Then, the representation in (9.1) is not unique. Indeed, we may find $\tilde{c}_i \in \mathbb{R}$, $\tilde{A}_i \subseteq I$ for $i = 1, \ldots, m$ such that*

$$g(x) = \sum_{i=1}^{n} c_i \chi_{A_i}(x) = \sum_{i=1}^{m} \tilde{c}_i \chi_{\tilde{A}_i}(x).$$

Theorem 9.1 (The space of step functions $\mathcal{S}(I)$).
Let $I \subseteq \mathbb{R}$ be an interval and denote by $\mathcal{S}(I)$ the set of all step functions $I \to \mathbb{R}$. Then, $\mathcal{S}(I)$ is a real vector space, i.e. for all $f, g \in \mathcal{S}(I)$, we have

$$\lambda f + \mu g \in \mathcal{S}(I) \quad \forall \lambda, \mu \in \mathbb{R}.$$

Moreover, $\mathcal{S}(I)$ is an algebra, i.e.

$$f \cdot g \in \mathcal{S}(I).$$

Proof. This follows immediately from Definition 9.2 taking into account

$$\chi_A(x)\chi_B(x) = \chi_{A \cap B}(x).$$

□

Definition 9.3 (Length $|I|$ of an interval)**.**
Let $I \subseteq \mathbb{R}$ be an interval. Then, its length $|I|$ is defined to be

$$|[a,b]| = |[a,b)| = |(a,b)| = |(a,b]| = |b-a|.$$

Next, we define the integral of a step function. Thinking intuitively, we would like to set

$$\mathcal{I}(g) := \sum_{i=1}^{n} c_i |A_i|. \tag{9.2}$$

However, as said in Remark 9.3, we need to know that

$$\sum_{i=1}^{n} c_i |A_i| = \sum_{i=1}^{m} \tilde{c}_i |\tilde{A}_i|,$$

where

$$g(x) = \sum_{i=1}^{n} c_i \chi_{A_i}(x), \quad g(x) = \sum_{i=1}^{m} \tilde{c}_i \chi_{\tilde{A}_i}(x).$$

Intuitively that is clear but intuition alone is not a proof. We show

Theorem 9.2 (Step function integral, Independence of representation)**.**
Let $I \subseteq \mathbb{R}$ be an interval and $g : I \to \mathbb{R}$ be a step function. Then, the integral of g, denoted by $\mathcal{I}(g)$, is defined as

$$\mathcal{I}(g) = \sum_{i=1}^{n} c_i |A_i|,$$

where

$$g(x) = \sum_{i=1}^{n} c_i \chi_{A_i}(x). \tag{9.3}$$

The value of $\mathcal{I}(g)$ does not depend on the particular representation (9.3).

Let $I = [a,b] \subseteq \mathbb{R}$. Then, we introduce the following notation for a step function $f : I \to \mathbb{R}$:

$$\mathcal{I}(g) = \int_I g(x)dx = \int_a^b g(x)dx.$$

Sketch of the proof of Theorem 9.2. The proof proceed in several steps. We will leave out a couple of technical details.

Step 1: Suppose, we have two representations of $g : I \to \mathbb{R}$ as

$$g(x) = \sum_{i=1}^{n} c_i \chi_{A_i}(x) = \sum_{i=1}^{m} \tilde{c}_i \chi_{\tilde{A}_i}(x).$$

Then, by Theorem 9.1, we have that

$$h(x) = \sum_{i=1}^{n} c_i \chi_{A_i}(x) - \sum_{i=1}^{m} \tilde{c}_i \chi_{\tilde{A}_i}(x)$$

is a step-function and that $h(x) \equiv 0$. By Definition (9.2), we would have that

$$\sum_{i=1}^{n} c_i |A_i| - \sum_{i=1}^{m} \tilde{c}_i |\tilde{A}_i|$$

would be the integral of h. Thus, if we show that it vanishes, we will have

$$\sum_{i=1}^{n} c_i|A_i| = \sum_{i=1}^{m} \tilde{c}_i|\tilde{A}_i|.$$

Now, without loss of generality, we assume that $g(x) \equiv 0$ and we need to show that for any representation

$$g(x) = \sum_{i=1}^{n} c_i \chi_{A_i}(x) = 0, \tag{9.4}$$

we have

$$\sum_{i=1}^{n} c_i|A_i| = 0. \tag{9.5}$$

Step 2: We show, using (9.4), that

$$\left|\sum_{i=1}^{n} c_i|A_i|\right| \le \|g\|_\infty |I|$$

which then implies (9.5) if $g(x) \equiv 0$.

To do this, we subdivide the A_i as follows: Find disjoint sets $B_1, \ldots, B_N$ such that $\bigcup_{j=1}^{N} B_j = I$ and such that $B_i \cap A_j$ is either B_i or empty. Then we can write

$$|A_i| = \sum_{j: B_j \cap A_i \neq} |B_j|$$

and, therefore,

$$\sum_{i=1}^{n} c_i|A_i| = \sum_{i=1}^{n} \sum_{j: B_j \cap A_i \neq \emptyset} c_i|B_j| = \sum_{j=1}^{N} \sum_{i: B_j \cap A_i \neq \emptyset} c_i|B_j| = \sum_{j=1}^{N} a_j|B_j|,$$

where we set

$$a_j = \sum_{i: B_j \cap A_i \neq \emptyset} c_i$$

which are the values of g on B_j. We have $\|g\|_\infty = \max_{j=1,\ldots,N} |a_i|$ and, therefore,

$$\begin{aligned} \left|\sum_{i=1}^{n} c_i|A_i|\right| &= \left|\sum_{j=1}^{N} a_j|B_j|\right| \le \sum_{j=1}^{N} |a_j||B_j| \\ &\le \|g\|_\infty \sum_{j=1}^{N} |B_j| = \|g\|_\infty |I|. \end{aligned}$$

This concludes the proof. □

From the previous proof, we get

Corollary 9.1.
Let $I \subseteq \mathbb{R}$ be a bounded interval and $f: I \to \mathbb{R}$ a step function. Then

$$\mathcal{I}(f) \le |I| \|f\|_\infty.$$

Remark 9.4. *We can also prove the last corollary directly if we assume that we have shown that a step function g can always be written as*

$$g(x) = \sum_{i=1}^{n} c_i \chi_{A_i}(x)$$

where the A_i are such that $A_i \cap A_j = \emptyset$ if $i \neq j$. Then we have

$$\begin{aligned} |\mathcal{I}(g)| &= \left|\sum_{i=1}^{n} c_i |A_i|\right| \leq \sum_{i=1}^{n} |c_i||A_i| \\ &\leq \max_{i=1,\dots,n} |c_i| \sum_{i=1}^{n} |A_i| = \max_{i=1,\dots,n} |c_i||I|. \end{aligned}$$

9.1.1 Properties of the step function integral $\mathcal{I}(f)$

Proposition 9.2 (Properties of $\mathcal{I}(f)$)**.**
Let $I \subseteq \mathbb{R}$ be a compact interval. Then, the following properties hold for $f, g \in \mathcal{S}(I)$:

(i) $|f| \in \mathcal{S}(I)$ *and*

$$\mathcal{I}(f) \leq \mathcal{I}(|f|).$$

(ii) The map[a] $\mathcal{I}(f)$ *is a linear map from* $\mathcal{S}(I)$ *to* $\mathbb{R}$ *such that*

$$\left|\int_I f(x)dx\right| \leq |I|\|f\|_\infty.$$

(iii) If $f(x) \geq 0$ for all $x \in I$, then

$$\int_I f(x)dx \geq 0.$$

(iv) If $f(x) \geq g(x)$ for all $x \in I$, then

$$\int_I f(x)dx \geq \int_I g(x)dx.$$

[a] Also operator, which is a technical term for functions that one uses if the arguments of the function are functions themselves. If the range of the operator is $\mathbb{R}$ or , we call the operator a functional.

Proof. (i) If f is written as

$$f(x) = \sum_{i=1}^{n} c_i \chi_{A_i}(x)$$

then $|f|$ can be written as

$$|f|(x) = \sum_{i=1}^{n} |c_i|\chi_{A_i}(x).$$

Using Theorem 9.2 and Corollary 9.1, we get the result.

(ii) Follows from Corollary 9.1 or from (i) by using that $f(x) \leq \|f\|_\infty < +\infty$ for $f \in \mathcal{S}(I)$.

(iii) If $f \in \mathcal{S}(I)$ and $f(x) \geq 0$, we have that all c_i in

$$f(x) = \sum_{i=1}^{n} c_i \chi_{A_i}(x)$$

satisfy $c_i \geq 0$. Thus, by Definition 9.2, we get $\mathcal{I}(f) \geq 0$ as the sum of non-negative numbers is non-negative.

(iv) Follows from (iii) realising that $f - g \in \mathcal{S}(I)$ and $f(x) - g(x) \geq 0$.

□

9.2 Regulated functions and their integrals

Definition 9.4 (Regulated function)**.**
Let $I = [a, b]$ be a compact interval. Then $f : I \to \mathbb{R}$ is called regulated if there exists a sequence $(f_n)_{n\in\mathbb{N}}$ of step functions $f_n \in \mathcal{S}(I)$ such that $f_n \to f$ uniformly as $n \to +\infty$. The set of all regulated functions is denoted by $\mathcal{R}(I)$.

Remark 9.5. *One can characterize regulated functions as follows: a function $f : [a, b] \to \mathbb{R}$ is regulated iff both the limit*

$$\lim_{x\to c-} f(x) \quad \textit{and} \quad \lim_{x\to c+} f(x)$$

exist for all $c \in (a, b)$ as well as $f(a+)$ and $f(b-)$.

Exercise 9.2. *Prove the assertion in Remark 9.5.*

Remark 9.6. *The reader may convince himself that $\mathcal{R}(I)$ is a real vector space. An immediate consequence of the definition is that $\mathcal{S}(I) \subseteq \mathcal{R}(I)$ as a linear sub-space of $\mathcal{R}(I)$. since products of uniformly convergent sequences converge uniformly, and $\mathcal{S}(I)$ is an algebra, we get that also $\mathcal{R}(I)$ is an algebra. Thus, products of regulated functions are regulated.*

Readers unfamiliar with the terminology of a piecewise continuous function may ignore the definition in the first reading and replace piecewise continuous simply by continuous in the remainder of the text.

Definition 9.5 (Piecewise continuous function)**.**
Let $I = [a, b] \subseteq \mathbb{R}$ be a compact interval. Then $f : I \to \mathbb{R}$ is called piecewise continuous iff there exist finitely many points $x_1, \dots, x_n \in I$ such that

(i) f is continuous on $I \setminus \{x_1, \dots, x_n\}$ and

(ii) the limits

$$\lim_{x\to x_k+} f(x) \quad \textit{and} \quad \lim_{x\to x_k-} f(x)$$

exist for all $k = 1, \dots, n$.

We denote the set of all piecewise continuous functions $f : [a, b] \to \mathbb{R}$ by $PC[a, b]$.

Theorem 9.3 ($PC[a, b] \subseteq \mathcal{R}([a, b])$)**.**
Let $f \in PC[a, b]$. Then $f \in \mathcal{R}(I)$.

Proof. Since $f \in PC[a, b]$, there exists a partition $a = x_1 < x_2 < \cdots < x_{n-1} < x_n = b$ of $[a, b]$ such that f is continuous on (x_i, x_{i+1}) and the limits $\lim_{x\to x_i+} f(x)$, $\lim_{x\to x_i-} f(x)$ exist for $i = 1, \dots, n-1$. We prove that the restriction of f to the interval (x_i, x_{i+1}), denoted by $f\big|_{(x_i,x_{i+1})}$, can be uniformly approximated by step functions on $[x_i, x_{i+1}]$. We can extend $f\big|_{(x_i,x_{i+1})}$ to a function $f\big|_{[x_i,x_{i+1}]}$ since e have the left and right limits at the boundary. Then, we glue those together to get a uniform approximation of f on $[a, b]$. By this discussion, it is enough to focus on the interval $[x_i, x_{i+1}]$, i.e. without loss of generality, we can assume that f is continuous on $[a, b]$. So, for every $\varepsilon > 0$ we have to construct a step-function $g \in \mathcal{S}([a, b])$ such that $\|f - g\|_\infty < \varepsilon$. By Heine's Theorem, we have that there exists a $\delta > 0$ such that

$$|x - y| < \delta \quad \Rightarrow \quad |f(x) - f(y)| < \varepsilon.$$

Thus, we divide $[a,b]$ into sub-intervals

$$[a,b] = \left[a, a+\frac{\delta}{2}\right) \cup \left[a+\frac{\delta}{2}, a+2\frac{\delta}{2}\right) \cup \cdots \cup \left[a+k\frac{\delta}{2}, a+(k+1)\frac{\delta}{2}\right) \cup \cdots \cup [a+N\delta, b]\,.$$

Now, we can define a step function $g : I \to \mathbb{R}$ by

$$g(x) = f\left(a+k\frac{\delta}{2}\right) \quad \forall x \in \left[a+k\frac{\delta}{2}, a+(k+1)\frac{\delta}{2}\right).$$

Then, by construction,

$$|f(x) - g(x)| = |f(x) - f(y)|,$$

for a y with $|x-y| \le \frac{\delta}{2} < \delta$. Thus, $|f(x) - g(x)| < \varepsilon$ and, hence,

$$\|f-g\|_\infty = \sup_{x\in[a,b]} |f(x) - g(x)| < \varepsilon.$$

This concludes the proof. □

Corollary 9.2 ($C[a,b] \subseteq \mathcal{R}[a,b]$)**.**
Continuous functions $f : [a,b] \to \mathbb{R}$ are regulated on $[a,b]$.

Remark 9.7. *We say that a set $A \subseteq \mathbb{R}^n$ is dense in $\mathbb{R}^n$ if for every $x \in \mathbb{R}^n$ there exists a sequence $(x_n)_{n\in\mathbb{N}_0} \subseteq A$ such that $x_n \to x$ as $n \to +\infty$. An example is $\mathbb{Q}$ in $\mathbb{R}$ or $\mathbb{Q}^n$ in $\mathbb{R}^n$.*
Here, we have a similar situation. The space $\mathcal{S}(I)$ is dense in $PC[a,b]$ since $\mathcal{S}(I) \subseteq PC[a,b]$ and for every $f \in PC[a,b]$ there exists a sequence $(f_n)_{n\in\mathbb{N}_0} \subseteq \mathcal{S}(I)$ with $f_n \to f$ uniformly, i.e. using $d(f,g) = \|f-g\|_\infty$ as a distance. This is the content of Theorem 9.3.

Theorem 9.4 (Construction of the Cauchy–Riemann Integral)**.**
Let $I \subseteq \mathbb{R}$ be a compact interval. Suppose $f \in \mathcal{R}(I)$ and let $(f_n)_{n\in\mathbb{N}_0} \subseteq \mathcal{S}(I)$ be a sequence of step-functions such that $\|f_n - f\|_\infty \to 0$ for $n \to +\infty$. Then $\mathcal{I}(f_n)$ converges and the limit is independent of the particular sequence $(f_n)_{n\in\mathbb{N}_0}$.

Proof. First let us prove the convergence. Since $f_n \to f$ uniformly, we have that for all $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}_0$ such that for all $n, m \ge n_0$ follows that $\|f_n - f_m\|_\infty < \frac{\varepsilon}{|I|}$. Therefore, by Corollary 9.1,

$$|\mathcal{I}(f_n) - \mathcal{I}(f_m)| = |\mathcal{I}(f_n - f_m)| \le |I| \|f_m - f_n\|_\infty < \varepsilon.$$

Hence, $(\mathcal{I}(f_n))_{n\in\mathbb{N}_0}$ is a Cauchy sequence in $\mathbb{R}$ and it converges by the completeness of $\mathbb{R}$.
Now let us turn to the independence from the particular sequence $(f_n)_{n\in\mathbb{N}_0}$. Let $(f_n)_{n\in\mathbb{N}_0}$ and $(g_n)_{n\in\mathbb{N}_0}$ be two sequences in $\mathcal{S}(I)$ that converge uniformly to f. Let $\varepsilon > 0$. Then, there exists $n_0 \in \mathbb{N}$ such that $\|f_n - f\|_\infty < \frac{\varepsilon}{2|I|}$ for and $\|g_n - f\|_\infty < \frac{\varepsilon}{2|I|}$ for $n \ge n_0$. Hence, we have

$$\|f_n - g_n\|_\infty = \|f_n - f + f - g_n\|_\infty \le \|f_n - f\|_\infty + \|g_n - f\|_\infty < \frac{\varepsilon}{|I|}.$$

Thus, we get

$$\|\mathcal{I}(f_n - g_n)\|_\infty \le |I| \|f_n - g_n\|_\infty < \varepsilon$$

for all $n \ge n_0$. We also used Corollary 9.1. □

Definition 9.6 (Cauchy–Riemann Integral)**.**
Let $I \subseteq \mathbb{R}$ be a compact interval. Let $f \in \mathcal{R}(I)$. Then, the Cauchy–Riemann Integral

$$\int_I f(x)dx := \lim_{n\to+\infty} \mathcal{I}(f_n)$$

where $(f_n)_{n\in\mathbb{N}_0} \subseteq \mathcal{S}(I)$ is such that $f_n \to f$ uniformly as $n \to +\infty$.

We agree on the following conventions

- If $a > b$, we define
$$\int_a^b f(x)dx := -\int_b^a f(x)dx.$$
- We also set
$$\int_a^a f(x)dx = 0.$$

Proposition 9.3 (Properties of the Cauchy–Riemann Integral)**.**
Let $I = [a,b] \subseteq \mathbb{R}$ be a compact interval, $c \in [a,b]$ and $f, g \in \mathcal{R}(I)$. Then, the following properties hold

(i) For $\lambda, \mu \in \mathbb{R}$, we have $\displaystyle\int_a^b \big(\lambda f(x) + \mu g(x)\big)dx = \lambda \int_a^b f(x)dx + \mu \int_a^b g(x)dx,$

(ii) $\displaystyle\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx,$

(iii) $\displaystyle\int_a^b f(x)dx \geq 0$ *if* $f(x) \geq 0$ *for all* $x \in I$,

(iv) $\displaystyle\int_a^b f(x)dx \geq \int_a^b g(x)dx$ *if* $f(x) \geq g(x)$ *for all* $x \in I$,

(v) $\displaystyle\left|\int_a^b f(x)dx\right| \leq \int_a^b |f(x)|dx,$

(vi) $\displaystyle\left|\int_a^b f(x)dx\right| \leq |b-a|\|f\|_\infty,$ *and*

(vii) $\displaystyle\left|\int_a^b f(x)g(x)dx\right| \leq \int_a^b |g(x)||f(x)|dx \leq \|g\|_\infty \int_a^b |f(x)|dx.$

Remark 9.8. *The last Proposition implies in particular that the Cauchy–Riemann Integral is a linear map from $PC[a,b]$ to $\mathbb{R}$ in the sense of Definition 1.12.*

9.3 The Riemann-Integral

Suppose $I = [a,b] \subseteq \mathbb{R}$ is a compact interval, or, more generally, $I = [a_1,b_1] \times \cdots \times [a_n,b_n]$ is a compact cuboid in $\mathbb{R}^n$.
Suppose $f : I \to \mathbb{R}$ is a function and define

$$\overline{\mathcal{I}(f)} = \inf_{\substack{h\in\mathcal{S}(I)\\ h\geq f}} \mathcal{I}(h)$$

and

$$\underline{\mathcal{I}(f)} = \sup_{\substack{h\in\mathcal{S}(I)\\ h\leq f}} \mathcal{I}(h).$$

Since $h_1 \geq h_2$ implies $\mathcal{I}(h_1) \geq \mathcal{I}(h_2)$, we have

$$\underline{\mathcal{I}(f)} \leq \overline{\mathcal{I}(f)}. \tag{9.6}$$

Definition 9.7 (Riemann integrability)**.**
Let I be as above, $f : I \to \mathbb{R}$. Then, f is said to be Riemann integrable if $\underline{\mathcal{I}(f)} = \overline{\mathcal{I}(f)}$, $\underline{\mathcal{I}(f)} = \overline{\mathcal{I}(f)}$. The Riemann integral $R - \int_I f(x)dx$ is defined as the value of $\underline{\mathcal{I}(f)} = \overline{\mathcal{I}(f)}$.

Remark 9.9. *Not all functions are Riemann integrable. The Dirichlet function $\chi_{\mathbb{Q}}$ is not Riemann-integrable on any compact interval $[a,b] \subseteq \mathbb{R}$.*

Exercise 9.3. *Prove the assertion in Remark 9.9 by showing $\underline{\mathcal{I}(f)} \neq \overline{\mathcal{I}(f)}$.*

Theorem 9.5 (Riemann-Integral)**.**
Let I be as above and $f : I \to \mathbb{R}$. Then, if f is regulated, it is Riemann integrable and

$$R - \int_I f(x)dx = \int_I f(x)dx.$$

Remark 9.10. *The R in-front of the integral is just to remind us that it is a different integral. As long as we know what we are talking about, we may drop the R if convenient. In the above theorem however, we state that two definitions of an integral agree under certain conditions and therefore should use different symbols.*

Remark 9.11. *There are functions which are not regulated but Riemann integrable. Thus, the Riemann integral is more general than the integral of regulated functions. An example is given by $f : [0,1] \to \mathbb{R}$, defined by*

$$f(x) = \begin{cases} 1 & : \quad x = \frac{1}{n},\ n \in \mathbb{N} \\ 0 & : \quad x \neq \frac{1}{n},\ n \in \mathbb{N} \end{cases}.$$

Though f is not regulated, the Riemann integral exists

$$\int_0^1 f(x)dx = 0.$$

First we discuss that the Riemann integral os equal to 0*. To this end, we define*

$$h_n(x) = \begin{cases} 1 & : \quad x \in \left[0, \frac{1}{n}\right] \\ f(x) & : \quad x \in \left(\frac{1}{n}, 1\right] \end{cases}.$$

An appropriate partition is given by $\{x_0, x_1, \ldots, x_n\}$ *with*

$$0 = x_0 < x_1 = \frac{1}{n} < x_2 = \frac{1}{n-1} < \cdots < x_n = 1$$

and we have $h_n \geq f(x)$*. Further, since* h_n *is constant on* $[0, \frac{1}{n}]$ *and only at finitely many points different from constant equal to* 0 *on* $[\frac{1}{n}, 1]$*, we get*

$$\int_0^1 h_n(x)dx = \frac{1}{n}.$$

We further have

$$\inf_{\substack{h \in \mathcal{S}([a,b]) \\ h \geq f}} \int_0^1 h(x)dx \leq \inf_{n \in \mathbb{N}} \int_0^1 h_n(x)dx = \inf_{n \in \mathbb{N}} \frac{1}{n} = 0.$$

Since $0 \leq f(x)$*, we also have*

$$0 \leq \sup_{\substack{h \in \mathcal{S}([a,b]) \\ h \leq f}} \int_0^1 h(x)dx.$$

By (9.6)*, we get*

$$0 \leq \sup_{\substack{h \in \mathcal{S}([a,b]) \\ h \leq f}} \int_0^1 h(x)dx \leq \inf_{\substack{h \in \mathcal{S}([a,b]) \\ h \geq f}} \int_0^1 h(x)dx \leq 0.$$

Thus they are all equal to 0 *which is the definition of* f *having* 0*-Riemann integral.*
Now, we see that f *is not regulated. To this end we will assume that it is. Thus, let* g *be a step function with partition* $\{x_0 < x_1 < \cdots < x_n\}$ *such that* $\|f - g\|_\infty < \varepsilon$*. Then,* g *is constant on* $(0, x_1)$*, say equal to* c_1*. By the definition of* f*, there exist* $x_1, x_2 \in (0, x_1)$ *such that* $f(x_1) = 0$ *and* $f(x_2) = 1$*. Next, we have*

$$|f(x_1) - g(x_1)| = |c_1| \quad \text{and} \quad |f(x_2) - g(x_2)| = |1 - c_1|. \tag{9.7}$$

and $|c_1| + |1 - c_1| \geq 1$ *which implies* $|c_1| \geq \frac{1}{2}$ *or* $|1 - c_1| \geq \frac{1}{2}$*. By* (9.7) *that implies* $\varepsilon > \frac{1}{2}$*. Thus,* f *can not be uniformly approximated by a step function.*

Now we prove that, if the regulated integral of f exists it agrees with its Riemann-Integral as stated in Theorem 9.5.

Proof of Theorem 9.5. Let $(h_n)_{n \in \mathbb{N}_0} \subseteq \mathcal{S}(I)$ such that $\|f - h_n\|_\infty < \frac{1}{n}$. Then

$$\begin{aligned} \overline{h}_n &:= h_n + \frac{1}{2n} \geq f \\ \underline{h}_n &:= h_n - \frac{1}{2n} \leq f \end{aligned}$$

and $\overline{h}_n \to f$ and $\underline{h}_n \to f$ uniformly. Indeed, fo instance

$$\|f - \overline{h}_n\|_\infty = \left\| f - h_n - \frac{1}{2n} \right\|_\infty < \frac{1}{2n} + \frac{1}{2n} = \frac{1}{n}.$$

Therefore, we get

$$\overline{\mathcal{I}(f)} \leq \int_I f(x)dx,$$

$$\underline{\mathcal{I}(f)} \geq \int_I f(x)dx.$$

The last two inequalities follow from

$$\overline{\mathcal{I}(f)} \leq \int_I \overline{h}_n(x)dx \quad \text{and} \quad \underline{\mathcal{I}(f)} \geq \int_I \underline{h}_n(x)dx.$$

Hence, we get

$$\int_I f(x)dx \leq \underline{\mathcal{I}(f)} \leq \overline{\mathcal{I}(f)} \leq \int_I f(x)dx.$$

□

Corollary 9.3.
If $f : [a, b] \to \mathbb{R}$ is piecewise continuous then f is Riemann integrable.

The next is a n-dimensional version of the last corollary which one can get from a straight forward generalisation of Definition 9.7 and Theorem 9.5.

Corollary 9.4.
Suppose $Q = [a_1, b_1] \times \cdots \times [a_n, b_n] \subseteq \mathbb{R}^n$ is a n-dimensional compact cuboid and $f : Q \to \mathbb{R}$ continuous. Then, f is Riemann integrable.

9.3.1 Properties of the Riemann Integral

As for the Cauchy–Riemann–Integral, one can prove

Theorem 9.6 (The space of Riemann integrable functions)**.**
Let $I \subseteq \mathbb{R}$ be a compact interval and $f : [a, b] \to \mathbb{R}$, $g : [a, b] \to \mathbb{R}$ be Riemann integrable. Then, for all $\lambda, \mu \in \mathbb{R}$, the function $\lambda f + \mu g$ is Riemann integrable, i.e. the set of Riemann integrable functions is a real vector space. Furthermore, fg is a Riemann integrable function, i.e. the set of Riemann integrable functions is an algebra.

and

Proposition 9.4 (Properties of the Riemann Integral)**.**
Let $I = [a, b] \subseteq \mathbb{R}$ be a compact interval, $c \in [a, b]$ and $f, g : I \to \mathbb{R}$ be Riemann integrable. Then, the following properties hold

(i) For $\lambda, \mu \in \mathbb{R}$, we have $\displaystyle\int_a^b \lambda f(x) + \mu g(x)dx = \lambda \int_a^b f(x)dx + \mu \int_a^b g(x)dx,$

(ii) $\displaystyle\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx,$

(iii) $\int_a^b f(x)dx \geq 0$ *if* $f(x) \geq 0$ *for all* $x \in I$*,*

(iv) $\int_a^b f(x)dx \geq \int_a^b g(x)dx$ *if* $f(x) \geq g(x)$ *for all* $x \in I$*,*

(v) $\left|\int_a^b f(x)dx\right| \leq \int_a^b |f(x)|dx,$

(vi) $\left|\int_a^b f(x)dx\right| \leq |b-a|\|f\|_\infty$*, and*

(vii) $\left|\int_a^b f(x)g(x)dx\right| \leq \int_a^b |g(x)||f(x)|dx \leq \|g\|_\infty \int_a^b |f(x)|dx.$

9.4 Improper Integrals on $\mathbb{R}$

Improper integrals are integrals over non-compact sets that are written as limits of integrals over compact sets. We use the notion of regulated integrals but everything can be done for Riemann integrals as well.

Example 9.3. *Consider*

$$\int_{(0,1]} \frac{1}{\sqrt{x}} dx$$

is an improper integral since $\mathrm{im}_{x\to 0+} \frac{1}{\sqrt{x}} = +\infty$*. The integral should be understood as*

$$\int_0^1 \frac{1}{\sqrt{x}} dx = \lim_{\varepsilon\to 0+} \int_\varepsilon^1 \frac{1}{\sqrt{x}} dx. \tag{9.8}$$

The problem is that since $(0,1]$ *is not compact and* $\frac{1}{\sqrt{x}}$ *is not bounded,* $\frac{1}{\sqrt{x}}$ *is not a uniform limit of step functions and, hence, the definition of the Cauchy-Riemann Integral does not work. However, considering the right hand side of (9.8), we have the function* $f(x) = \frac{1}{\sqrt{x}}$ *over the compact interval* $[\varepsilon, 1]$*. Since* f *is continuous there, it is, by Heine's theorem, also uniformly continuous. Thus, we can define its Cauchy-Riemann integral which would of course depend on* ε*. So, (9.8) says that we say that the integral over* $(0,1]$ *exists if the limit* $\lim_{\varepsilon\to 0+} \mathcal{I}_\varepsilon(f)$ *exists.*

Similarly, we would like to interpret

$$\int_0^\infty f(x)dx \quad \text{as} \quad \lim_{R\to+\infty} \int_0^R f(x)dx$$

if the limit exists and

$$\int_{-\infty}^{+\infty} f(x)dx = \lim_{R\to+\infty} \int_{-R}^0 f(x)dx + \lim_{R\to+\infty} \int_0^R f(x)dx$$

if both limits exist. The improper integral may depend on the way we take the limit and we will need some care in defining the integrals.

To make the above a bit more precise, we say that $\int_I f(x)dx$ is an improper integral of $f : I \to \mathbb{R}$ if $|I| = +\infty$ or there exists an $x_0 \in \mathrm{cl}(I)$ such that $\lim_{x\to x_0} |f(x)| = +\infty$. Examples are

$$\int_0^1 \frac{1}{x} dx, \quad \int_1^{+\infty} \frac{1}{x} dx.$$

Definition 9.8 (Convergence of improper integrals)**.**
Let $I \subseteq \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$. Further, $\int_I f(x)dx$ be an improper integral and $(I_n)_{n\in\mathbb{N}_0} \subseteq I$ be a sequence of compact intervals with $I_{n+1} \supseteq I_n$ such that $\mathrm{cl}(I) = \bigcup_{n=0}^{\infty} I_n$. If $f\big|_{I_n}$ is regulated for every $n \geq 0$, then $\int_I f(x)dx$ is said to be convergent (or f is integrable over I) if the sequence $(A_n)_{n\in\mathbb{N}_0}$, defined by

$$A_n := \int_{I_n} f(x)dx,$$

converges.

Example 9.4. *We consider the integral*

$$\int_{\pi}^{+\infty} \frac{\sin(x)}{x} dx.$$

as the function f is clearly continuous on the intervals $[k\pi, (k+1)\pi]$, $k \geq 1$ and

$$\bigcup_{k\in\mathbb{N}} [k\pi, (k+1)\pi] = [\pi, +\infty),$$

we only have to check whether there is a constant $C > 0$ such that

$$\int_{k\pi}^{(k+1)\pi} \frac{\sin(x)}{x} dx \leq C \quad \forall n \geq 1.$$

We have

$$\left| \int_{k\pi}^{(k+1)\pi} \frac{\sin(x)}{x} dx \right| \leq \int_{k\pi}^{(k+1)\pi} \left| \frac{\sin(x)}{x} \right| dx \leq \int_{k\pi}^{(k+1)\pi} \frac{1}{x} dx \leq \frac{1}{k}$$

and

$$\int_{k\pi}^{(k+1)\pi} \frac{\sin(x)}{x} dx = \int_{k\pi}^{(k+1)\pi} (-1)^k \left| \frac{\sin(x)}{x} \right| dx, \quad \forall k \geq 1.$$

Hence, by Leibniz' criterion, we have that $\int_{I_n} f(x)dx$ converges with

$$I_n := \bigcup_{1\leq k\leq n} [k\pi, (k+1)\pi]$$

and

$$\int_{I_n} f(x)dx = \sum_{k=1}^{n} (-1)^k \int_{k\pi}^{(k+1)\pi} \left| \frac{\sin(x)}{x} \right| dx.$$

We further introduce the notion of absolute integrability.

Definition 9.9 (Absolute Integrability on $\mathbb{R}$)**.**
Let $I \subseteq \mathbb{R}$ be an interval and let $(I_n)_{n\in\mathbb{N}_0} \subseteq I$, be a sequence of compact intervals such that $\mathrm{cl}(I) = \bigcup_{n=0}^{\infty} I_n$, $I_{n+1} \supseteq I_n$. Then, a function $f : I \to \mathbb{R}$ with the property that $f\big|_{I_n}$ is regulated for any $n \in \mathbb{N}_0$

is said to be absolutely integrable if the sequence $(A_n)_{n\in\mathbb{N}_0}$, defined by

$$A_n := \int_{I_n} |f(x)| dx,$$

converges.

Remark 9.12. *By $\mathrm{cl}(I)$, we denote the closure of I which means it adds the set of all limit points of I to I. See also Definition 2.8.*

Example 9.5. *A function $f:\mathbb{R}\to\mathbb{R}$ is called absolutely integrable if*

$$\lim_{R\to+\infty}\int_{-R}^{R} |f(x)| dx < +\infty.$$

Example 9.6. *Let us consider*

$$\int_1^{+\infty} \frac{dx}{x^2}.$$

With $I_n := [1,n]$, we get

$$\int_{I_n} \frac{dx}{x^2} = \int_1^n \frac{dx}{x^2} = 1 - \frac{1}{n}$$

which converges to 1. Thus, we also have that

$$\int_1^{+\infty} \frac{\sin(x)}{x^2} dx$$

is absolutely convergent since

$$\int_1^{+\infty} \left|\frac{\sin(x)}{x^2}\right| dx \le \int_1^{+\infty} \frac{dx}{x^2} < +\infty$$

as seen above.

Example 9.7. *The function $\frac{\sin(x)}{x}$ is not absolutely integrable. Thus,*

$$\int_{-\infty}^{+\infty} \left|\frac{\sin(x)}{x}\right| dx = +\infty.$$

However, $\frac{\sin(x)}{x}$ is integrable as we have seen, i.e.

$$\int_{-\infty}^{+\infty} \frac{\sin(x)}{x} dx$$

converges.

9.5 The fundamental theorem of calculus and primitives

Theorem 9.7 (Fundamental Theorem of Calculus)**.**

Let $f \in C[a,b]$ and define

$$F(x) = \int_a^x f(y)\,dy. \tag{9.9}$$

Then, F is (uniformly) continuous on $[a,b]$ and differentiable on $[a,b]$ and $F'(x) = f(x)$ for all $x \in [a,b]$.

Remark 9.13. *In (9.9), the choice of the lower limit does not matter. In fact, one can prove the theorem for*

$$F(x) = \int_{x_0}^x f(y)\,dy,$$

where $x_0 \in [a,b]$ fixed and arbitrary.

From the fundamental theorem, we get the following *reconstruction* formula for f from its derivative.

Corollary 9.5.
Let $f \in C^1[a,b]$, then

$$f(x) = f(a) + \int_a^x f'(y)\,dy.$$

Remark 9.14. *The function F in Theorem 9.7 is called a primitive of $f : [a,b] \to \mathbb{R}$. The Corollary 9.5 is the reason why the name anti-derivative that some people use is misleading. One can only reconstruct the function f from its derivative if one knows at least one function value. Otherwise, one can only say that there exists a constant $C \in \mathbb{R}$ such that*

$$f(x) + C = \int_a^x f'(y)\,dy.$$

See also the discussion at the end of this section.

Remark 9.15. *I would like to draw attention to the fact that not all Riemann integrable functions have primitives. The signum function*

$$f(x) = \operatorname{sgn}(x) = \begin{cases} -1 & : \quad x < 0 \\ 0 & : \quad x = 0 \\ 1 & : \quad x > 0 \end{cases}$$

is integrable on $[-1,1]$ but there exists no function F such that $F'(x) = \operatorname{sgn}(x)$. In general, one can say that a function with jumps, as a general step function for example, can not have a primitive as derivatives can not have jumps. However, if f is continuous, Theorem 9.7 guarantees the existence of a primitive.

Proof of Theorem 9.7. Since $f \in C[a,b]$, the function $F(x)$ is well defined for all $x \in [a,b]$ by Theorem 9.5. First we show that $F(x)$ is continuous. To this end, let $\varepsilon > 0$ and we compute

$$\begin{aligned} |F(x+h) - F(x)| &= \left| \int_a^{x+h} f(y)\,dy - \int_a^x f(y)\,dy \right| = \left| \int_x^{x+h} f(y)\,dy \right| \\ &\leq \int_x^{x+h} |f(y)|\,dy \leq M|h|, \end{aligned}$$

where $M := \max_{x\in[a,b]} |f(x)|$. Thus, with $\delta = \frac{\varepsilon}{M}$, we get that

$$|h| < \delta \quad \Rightarrow \quad |F(x+h) - F(x)| < \varepsilon$$

for $x \in (a,b)$ with obvious adaptations for $x = a$ and $x = b$. Now, to show differentiability, we calculate

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{h}\left(\int_a^{x+h} f(y)dy - \int_a^{x} f(y)dy\right) = \frac{1}{h}\int_x^{x+h} f(y)dy. \tag{9.10}$$

Next, we have

$$\int_x^{x+h} f(y)dy = \int_x^{x+h} f(x)dy + \int_x^{x+h} (f(y) - f(x))dy$$

and estimate

$$\left|\int_x^{x+h} (f(y) - f(x))dy\right| \leq h \sup_{y\in[x,x+h]} |f(x) - f(y)|.$$

Since f is continuous on $[a,b]$, we can, for a given $\varepsilon > 0$, choose h sufficiently small such that

$$\sup_{y\in[x,x+h]} |f(x) - f(y)| < \varepsilon.$$

From (9.10), we get, using

$$\int_x^{x+h} f(x)dy = f(x)h,$$

that for all $\varepsilon > 0$ there exists $\delta > 0$ such that

$$0 < |h| < \delta \quad \Rightarrow \quad \left|\frac{F(x+h) - F(x)}{h} - f(x)\right| < \varepsilon.$$

Modification at the boundary are clear. This concludes the proof. □

The next theorem is a more common formulation of Corollary 9.5.

Theorem 9.8.
Let $F \in C^1[a,b]$ with $F'(x) = f(x)$ for all $x \in [a,b]$. Then

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

Remark 9.16. *We also sometimes write*

$$\int_a^b f(x)dx = F(x)\Big|_a^b = \Big[F(x)\Big]_a^b = F(b) - F(a).$$

Proof. We define an auxiliary function $G : [a,b] \to \mathbb{R}$ by

$$G(x) = F(x) - F(a) - \int_a^x f(y)dy.$$

Then,

$$G'(x) = f(x) - f(x) = 0.$$

by Theorem 9.7. Thus, by Lemma 5.1, G is constant and since $G(a) = 0$ is identically zero. This concludes the proof. □

The set of all primitives (sometimes called anti-derivatives) of a function $f \in C[a,b]$, i.e.

$$\mathcal{P}(f) := \{f \in C^1[a,b] : F' = f\}$$

is denoted by

$$\int f dx.$$

All functions in $\mathcal{P}(f)$ differ by an additive constant. We write

$$\int f dx = g(x) + C$$

for a particular choice $g \in \mathcal{P}(f)$. For example, we always have

$$\int f dx = \int_a^x f(y) dy + C$$

and for all $g \in \mathcal{P}(f)$ there exists a constant C such that

$$g(x) = \int_a^x f(y) dy + C.$$

Remark 9.17. *Sometimes, primitives are called anti-derivatives. As primitives are not unique, this is a quite bad name. As a mathematician you should stick to primitive or indefinite integral the latter being the best terminology.*

9.6 Some Rules for Integration

Differentiation is mechanics, integration is art.

The next theorem is in some sense the inverse version of the product-rule for differentiation. See Theorem 5.3. Given f and g are differentiable, then we get

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

and, slightly handwavingly, we obtain

$$\int (f \cdot g)' dx = f \cdot g = \int f' \cdot g dx + \int f \cdot g' dx$$

which leads to

$$\int f \cdot g' dx = f \cdot g - \int f' \cdot g dx.$$

Using Theorem 9.7, we can prove a version for definite integrals.

Theorem 9.9 (Integration by parts)**.**
Suppose $f, g \in C^1[a,b]$. Then

$$\int_a^b f'(x) g(x) dx = f(b)g(b) - f(a)g(a) - \int_a^b f(x) g'(x) dx.$$

Proof. The proof follows from the product rule in Theorem 5.3 and integration over $[a,b]$ using the Fundamental Theorem of Calculus, Theorem 9.7. □

The correct use of Theorem 9.9 takes some exercise. With experience, the reader will learn what kinds of integrals are to be attacked by integration by parts.

Example 9.8. *Let us suppose that we want to compute the integral*

$$\int_0^\pi \cos^2(x)\,dx.$$

We obtain

$$\begin{aligned}\int_0^\pi \cos^2(x)\,dx &= \Big[\sin(x)\cos(x)\Big]_0^\pi + \int_0^\pi \sin^2(x)\,dx \\ &= \int_0^\pi \sin^2(x)\,dx. \qquad (9.11)\end{aligned}$$

At this point it seems that we have not really won much. However, the reader might remember that $\cos^2(x)+\sin^2(x)=1$. *Therefore, we obtain*

$$\begin{aligned}\int_0^\pi \sin^2(x)\,dx &= \int_0^\pi \big(1-\cos^2(x)\big)\,dx = \int_0^\pi 1\,dx - \int_0^\pi \cos^2(x)\,dx \\ &= \pi - \int_0^\pi \cos^2(x)\,dx.\end{aligned}$$

With that, from (9.11), we finally obtain

$$\int_0^\pi \cos^2(x)\,dx = \frac{\pi}{2}$$

and in fact we also showed that

$$\int_0^\pi \sin^2(x)\,dx = \frac{\pi}{2}.$$

Other very typical examples where integration by parts often leads to results are functions of the type $\sin(x)e^{ax+b}$, $\cos(x)e^{ax+b}$ and $x^k e^{ax+b}$. For more examples and material to calculate see the module Mathematical Methods.

Theorem 9.10 (Integration by Substitution)**.**
Let $[\alpha,\beta]\subseteq\mathbb{R}$, $[a,b]\subseteq\mathbb{R}$, $f:[\alpha,\beta]\to\mathbb{R}$ *be continuous on* $[a,b]$, *and let* $g:[a,b]\to[\alpha,\beta]$ *be differentiable on* $[a,b]$. *Further, suppose that* g' *is Riemann integrable. Then, we have*

$$\int_{g(a)}^{g(b)} f(x)\,dx = \int_a^b f(g(x))g'(x)\,dx. \qquad (9.12)$$

Proof. Since f is continuous on $[a,b]$ and g is differentiable on $[\alpha,\beta]$, we have that $f\circ g:[\alpha,\beta]\to\mathbb{R}$ is well-defined and continuous on $[a,b]$. Thus $f\circ g$ is Riemann integrable. Since g' is Riemann integrable and the product of Riemann integrable functions is Riemann integrable (see Theorem 9.6), we have that the right hand side of (9.12) is well defined as is the left hand side. Let F be a primitive of f. Since F is differentiable (Theorem 9.7), we obtain that $F\circ g:[a,b]\to\mathbb{R}$ is differentiable and by the chain rule (Theorem 5.2), we get

$$\frac{d}{dx}F(g(x)) = f(g(x))g'(x)\,dx.$$

Thus, we have that

$$\int_a^b f(g(x))g'(x)dx = \int_a^b \frac{d}{dx}F(g(x))dx.$$

By Corollary 9.5 (or Theorem 9.8), we obtain

$$\int_a^b \frac{d}{dx}F(g(x))dx = F(g(b)) - F(g(a)) = \int_{g(a)}^{g(b)} f(x)dx.$$

□

As for integration by parts, the knowledge of the best choice of substitutions comes with experience. The reader is encouraged to take the time to simply calculate some integrals. The reader should not give up at the slightest problem as integration is really an art. A fantastic book on the matters of this section is [11].

Example 9.9. *Simple examples if the integrand is a function of the type $e^{f(x)}$, $\cos(f(x))$, or $\sin(f(x))$ with f being a linear function. For instance, for $A \in \mathbb{R} \setminus \{0\}$, $B \in \mathbb{R}$, we have*

$$\begin{aligned} \int_a^b e^{Ax+B} dx &= \frac{1}{A}\int_a^b Ae^{Ax+B}dx = \frac{1}{A}\int_a^b g'(x)e^{g(x)}dx \\ &= \frac{1}{A}\int_{Aa+B}^{Ab+B} e^y dy = \frac{1}{A}\left(e^{Ab+B} - e^{Aa+B}\right). \end{aligned}$$

For such a simple case substitution is a bit of n overkill as the integral of e^{Ax+B} is obvious. See also the discussion at the end of this section.

An interesting integral that students get often asked and that is surprisingly difficult for them is something like

$$\int x^2 e^{-x^3} dx, \quad \text{or} \quad 2\int xe^{x^2} dx.$$

Do you see how to solve it? The "trick" is that one realises that these integrals are essentially of the form $\int f'(x)e^{f(x)}dx$ which makes the solution obvious:

$$\int f'(x)e^{f(x)}dx = e^{f(x)} + C.$$

Compare also Theorem 9.10 and the subsequent notes. To complete the Lecture Notes let us compute the first example and leave a more general result as an exercise to the reader:

$$\int x^2 e^{-x^3} dx = -\frac{1}{3}\int (-3x^2)e^{-x^3} dx = -\frac{1}{3}e^{-x^3} + C.$$

The same "tricks" work obviously for

$$\int f'(x)\sin(f(x))dx, \quad \text{and} \quad \int f'(x)\cos(f(x))dx.$$

We leave the details also here to the inclined reader. For instance the reader may formulate and prove a general result that contains the above examples.

9.7 Uniform convergence and integration

The theorems of this chapter are not examinable in the sense that you do not have to state them. However, you should understand the examples in the last four questions of Problem Sheet 10 and that is what this section relates to. See also the practice exam.

In this section we consider the questions under which conditions we can exchange integrals with limits, i.e. given a sequence $(f_n)_{n\in\mathbb{N}_0}$ of functions we ask: When is

$$\lim_{n\to+\infty}\int_a^b f_n(x)\,dx = \int_a^b \lim_{n\to+\infty} f_n(x)\,dx \tag{9.13}$$

true?
In general, the answer depends on the space where the sequences come from and the type of convergence. A general result is that it usually does not work if one has only pointwise convergence.

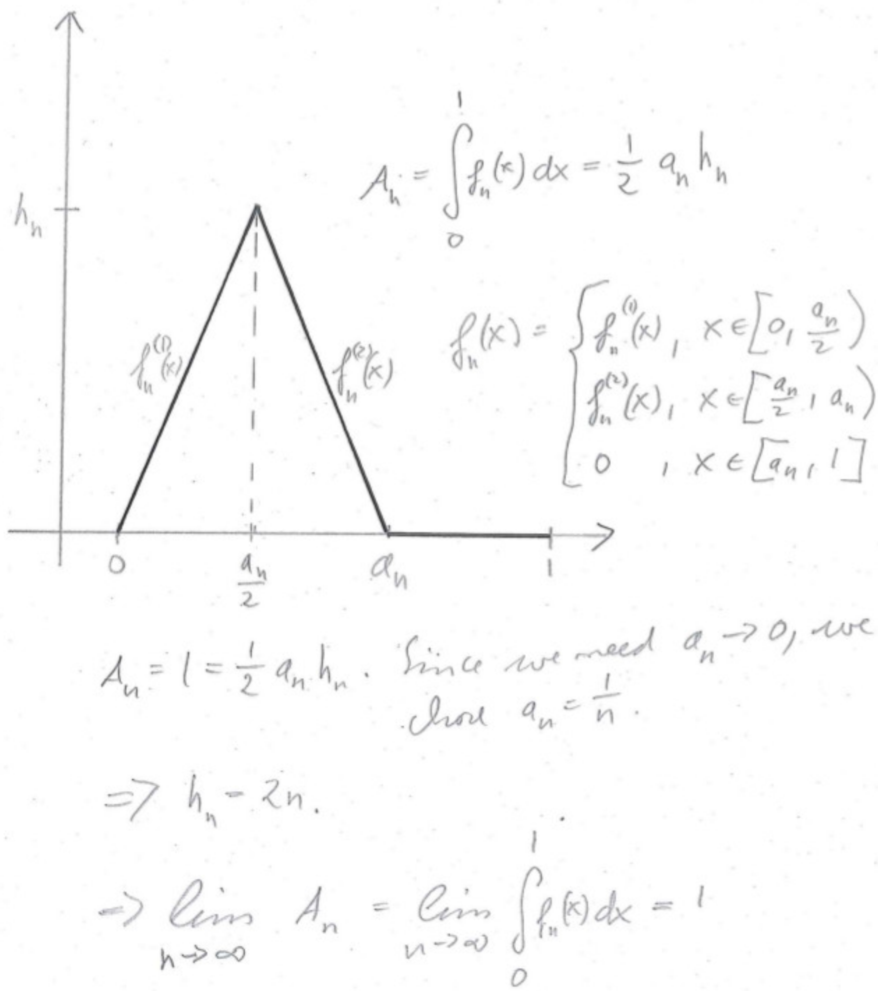


Figure 9.1: A sequence of continuous f_n with $f_n \to 0$ pointswise but $\int_0^1 f_n(x)\,dx = 1$.

In fact, one can even construct a sequence of continuous functions $(f_n)_{n\in\mathbb{N}_0}$ which converge pointwise to $f \equiv 0$ but the limit of the integrals grows unbounded.

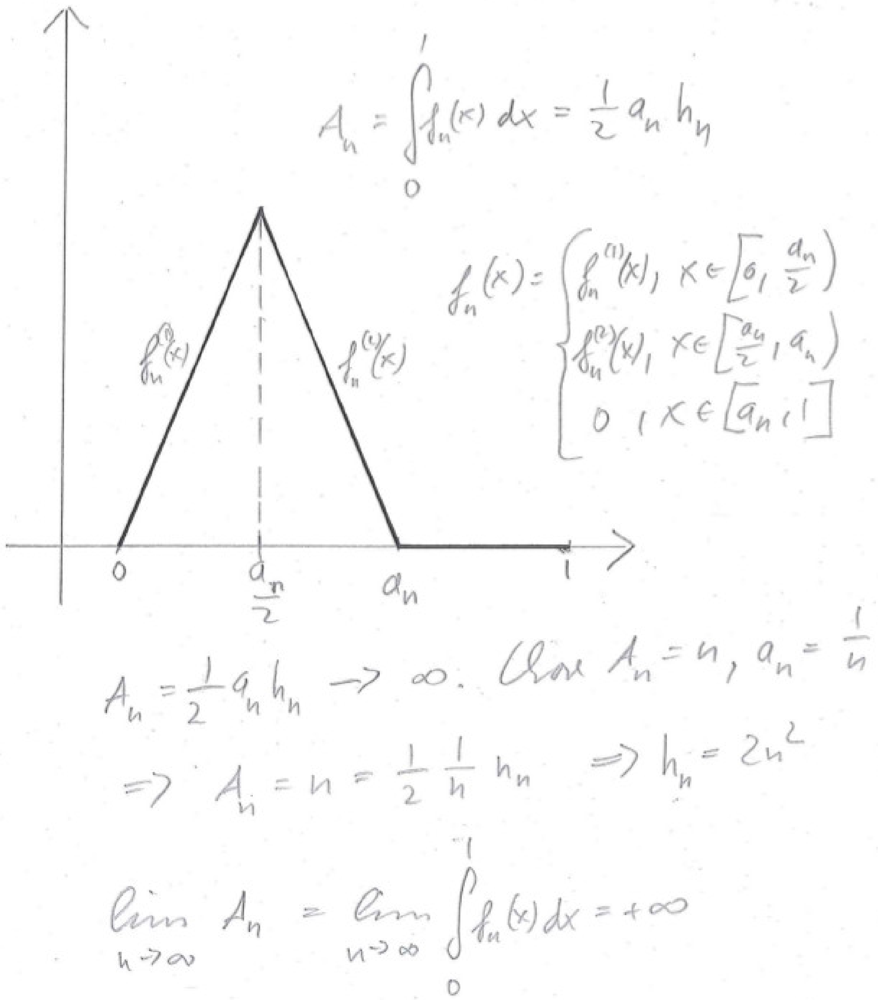


Figure 9.2: A sequence of continuous f_n with $f_n \to 0$ pointswise but $\int_0^1 f_n(x)dx = +\infty$.

Thus, in the subsequent theorems, we will be concerned with uniform convergence.

Theorem 9.11.
Let $[a,b]$ be a compact interval and let $(f_n)_{n\in\mathbb{N}_0} \subseteq \mathcal{R}[a,b]$. Suppose that there exists $f : [a,b] \to \mathbb{R}$ such that $f_n \to f$ uniformly. Then

$$\lim_{n\to+\infty} \int_a^b f_n(x)dx = \int_a^b \lim_{n\to+\infty} f_n(x)dx. \tag{9.14}$$

Remark 9.18. *Since regulated functions are Riemann integrable, we get, under the hypothesis of Theorem 9.11, that*

$$\lim_{n\to+\infty} R-\int_a^b f_n(x)dx = R-\int_a^b \lim_{n\to+\infty} f_n(x)dx.$$

To prove Theorem 9.11, we first need another theorem.

Theorem 9.12.
Let I be an interval and $(f_n)_{n\in\mathbb{N}_0} \subseteq \mathcal{R}[a,b]$. Suppose that $f_n \to f$ uniformly for a function $f : [a,b] \to \mathbb{R}$. Then $f \in \mathcal{R}[a,b]$.

Proof. Let $\varepsilon > 0$. Then there exists $f_{n,\varepsilon} \in \mathcal{S}[a,b]$ such that $\|f_n - f_{n,\varepsilon}\|_\infty < \frac{\varepsilon}{2}$. Now there exists an $n_0 \in \mathbb{N}_0$ such that $\|f - f_n\|_\infty < \frac{\varepsilon}{2}$ for all $n \geq n_0$. Then

$$\|f - f_{n,\varepsilon}\|_\infty \leq \|f - f_n\|_\infty + \|f_n - f_{n,\varepsilon}\|_\infty < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus, f is regulated. □

Remark 9.19. *Theorem 9.12 can be improved by saying that $\mathcal{R}[a,b]$ is a complete metric space with respect to $d_\infty(f,g) = \|f - g\|_\infty$. All that remains to be shown is that if $(f_n)_{n\in\mathbb{N}_0} \subseteq \mathcal{R}[a,b]$ is a Cauchy sequence then there exists $f \in \mathcal{R}[a,b]$ such that $f_n \to f$ uniformly. Indeed, for all $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}_0$ such that*

$$\forall m,n \geq n_0 : \sup_{x\in I} |f_n(x) - f_m(x)| < \varepsilon.$$

Thus, for all $x \in I$, we have that $(f_n(x))_{n\in\mathbb{N}_0}$ is a Cauchy sequence in $\mathbb{R}$. Since $\mathbb{R}$ is a complete space, it is convergent. Thus, we can set

$$f(x) = \lim_{n\to+\infty} f_n(x).$$

By the argument in the proof of Theorem 9.12, we have that $f \in \mathcal{R}[a,b]$.

Proof of Theorem 9.11. By Theorem 9.12, we have that $f \in \mathcal{R}[a,b]$. Then, we get

$$\int_a^b f_n(x)dx - \int_a^b f(x)dx = \int_a^b (f_n(x) - f(x))dx$$

which are all well-defined since $f - f_n \in \mathcal{R}[a,b]$. Thus, we can estimate

$$\left|\int_a^b (f_n(x) - f(x))dx\right| \leq \int_a^b |f_n(x) - f(x)|dx \leq |b-a| \sup_{x\in[a,b]} |f_n(x) - f(x)|.$$

Since $f_n \to f$ uniformly, we have that for all $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that

$$\forall n \geq n_0 \quad \Rightarrow \quad \|f_n - f\|_\infty < \frac{\varepsilon}{|b-a|}.$$

Thus, for all $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that

$$\forall n \geq n_0 : \quad \left|\int_a^b f_n(x)dx - \int_a^b f(x)dx\right| < \varepsilon.$$

Thus, (9.15) holds. (See Definition 1.14.) □

Following the similar arguments, we can show

Theorem 9.13.
Let $[a,b]$ be a compact interval and let $(f_n)_{n\in\mathbb{N}_0} \subseteq C[a,b]$. Suppose that there exists $f : [a,b] \to \mathbb{R}$ such that $f_n \to f$ uniformly. Then

$$\lim_{n\to+\infty} \int_a^b f_n(x)dx = \int_a^b \lim_{n\to+\infty} f_n(x)dx. \tag{9.15}$$

The integral can be interpreted as Cauchy-Riemann or Riemann integral.

Exercise 9.4. *Theorem 9.13 follows from Theorem 9.11 by Theorem 9.3. However, use the proof of Theorem 9.11 above as a blue print to prove Theorem 9.13 again in detail by bare hands.*

APPENDIX

A Relations of function spaces

The relations between the function spaces introduced in this course, are not trivial and will take a while to sink in. First, let us recall the spaces we have considered

Name	Type of functions (Real vector space of ...)
$C[a,b]$	all functions $f:[a,b]\to\mathbb{R}$ which are continuous on $[a,b]$.
$C^k[a,b]$	all functions $f:[a,b]\to\mathbb{R}$ which are k-times differentiable and $f^{(i)}\in C[a,b]$ for $i=0,\dots,k$.
$C^\infty[a,b]$	all functions $f:[a,b]\to\mathbb{R}$ which are ∞-times differentiable and $f^{(i)}\in C[a,b]$ for $i\in\mathbb{N}_0$.
$PC[a,b]$	all functions $f:[a,b]\to\mathbb{R}$ which are picewise continuous on $[a,b]$. See Definition 9.5.
$\mathcal{S}[a,b]$	all step-functions $g:[a,b]\to\mathbb{R}$
$\mathcal{R}[a,b]$	all functions $f:[a,b]\to\mathbb{R}$ such that there exists a sequence of functions $(f_n)_{n\in\mathbb{N}_0}\subseteq\mathcal{S}[a,b]$ with $f_n\to f$ uniformly.

Table A.1: A list of the function spaces introduced in the course.

Of course, we could also consider $C(a,b)$ instead of $C[a,b]$. However, the first space is not complete with respect to $d_\infty(f,g)=\|f-g\|_\infty$ while the latter one is.

What we can see immediately is that

$$C^\infty[a,b]\subseteq C^k[a,b]\subseteq C[a,b],$$

where the inclusions are strict, i.e. there exits a function f in $C^k[a,b]$ which is not in $C^\infty[a,b]$ and a function g in $C[a,b]$ which does not belong to $C^k[a,b]$. See also Remark 5.7.

What else can we say? Since all spaces above are vector spaces, they all contain the function identically zero. Also, since $\mathcal{S}[a,b]$ contains constant functions, we have that $C[a,b]\cap\mathcal{S}[a,b]\neq\emptyset$. As vector spaces, the intersection is isomorphic to $\mathbb{R}$ as only the constant step functions are continuous functions.

By definition of $\mathcal{R}[a,b]$, we have $\mathcal{S}[a,b]\subseteq\mathcal{R}[a,b]$ and, by Corollary 9.2, $C[a,b]\subseteq\mathcal{R}[a,b]$. Again, the containment is strict. The relation between $\mathcal{S}[a,b]$ and $C[a,b]$ is like the relation between $\mathbb{Q}$ and $\mathbb{R}\setminus\mathbb{Q}$, i.e. $\mathcal{S}[a,b]$ is dense in $\mathcal{R}[a,b]$ and for all $f\in C[a,b]$ there exists $g\in\mathcal{S}[a,b]$ such that $d_\infty(f,g)=\|f-g\|_\infty<\varepsilon$. See also Remark 9.7.

Figure A.1: Schematic depiction of the relation between the function spaces introduced. Please takeinto account that this is only depicting the relation of being contained. The more complictaed nature of their density is more difficult to depict.

List of Figures

Bibliography

[1] Pieter C. Allaart and Kiko Kawamura *The Takagi Function: a Survey*, Real Analysis Exchange 37(1) (2011), pp. 1–54. ProjectEuclid

[2] Bernard Bolzano *Funktionenlehre*, Eds.: K. Rychlik, Prague 1831.

[3] Carl B. Boyer *The history of the calculus and its conceptual development*, Dover Books on Mathematics, Mineola NY, 1959. ISBN 9780486605098.

[4] Robert A. Bjork John Dunlosky, and Nate Kornell *Self-Regulated Learning: Beliefs, Techniques, and Illusions*, Annual Review of Psychology, 64(1) (2013). doi:10.1146/annurev-psych-113011-143823.

[5] Gerald A. Goldin *Problem Solving Heuristics, Affect, and Discrete Mathematics: A Representational Discussion* in Theories of Mathematical Education, Springer, New York (2010). doi:10.1007/978-3-642-00742-2_24

[6] A. Rupert Hall *Philosophers at War: The Quarrel between Newton and Leibniz*, Cambridge University Press, 1980, ISBN 9780521227322

[7] Paul Halmos *Finite-Dimensional Vector Spaces*, Springer, Undergraduate Texts in Mathematics, Corr. 2nd printing 1993 edition. ISBN: 978-0387900933.

[8] Godfrey H. Hardy *Weierstrass's Non-Differentiable Function*, Transactions of the American Mathematical Society 17(3), (1916), pp. 301–325. doi:10.2307/1989005

[9] Merek Jarnicki and Peter Pflug *Continuous Nowhere Differentiable Functions*, Springer, 2015, ISBN 978-3-319-12669-2, doi:10.1007/978-3-319-12670-8.

[10] Zbigniew Michalewicz and David B. Fogel *How to Solve It: Modern Heuristics*, Springer, New York, 2nd edition, 2004. ISBN 9783540224945

[11] Paul J. Nahin *Inside Interesting Integrals*, Springer, New York, Undergraduate Lecture Notes in Physics, ISBN 978-1-4939-1276-6, doi:10.1007/978-1-4939-1277-3.

[12] Nicholas Mousoulides and Bharath Sriraman *Heuristics in Mathematics Education* in Encyclopedia of Mathematics Education, Springer, New York, pp. 253–255 (2014). doi:10.1007/978-94-007-4978-8_172

[13] George Polya *How to Solve It*, Princeton University Press, Princeton, Notable Centenary Edition, 2014 (original 1945). ISBN 9780691164076

[14] Arthur Rosenthal *The History of Calculus*, The American Mathematical Monthly, 58(2) (1951), pp. 75–86. doi:10.2307/2308368.

[15] Karl Randall *Improving study habits in mathematics*, The mathematics teacher, 55(7) (1962). Jstor.

[16] Michael Spivak *Calculus*, Addison-Wesley, World Student Series, 1986.

[17] Karl Weierstrass Über continuirliche Functionen eines reellen Arguments, die für keinen Werth des letzteren einen bestimmten Differentialquotienten besitzen. Read in the Royal Academy of Sciences on 18 July 1872. doi:10.1007/978-3-322-91273-2_5