Analysis II

Limits, continuity, differentiability, integrability

Christian P. Jäh

December 29, 2020

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, read 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.

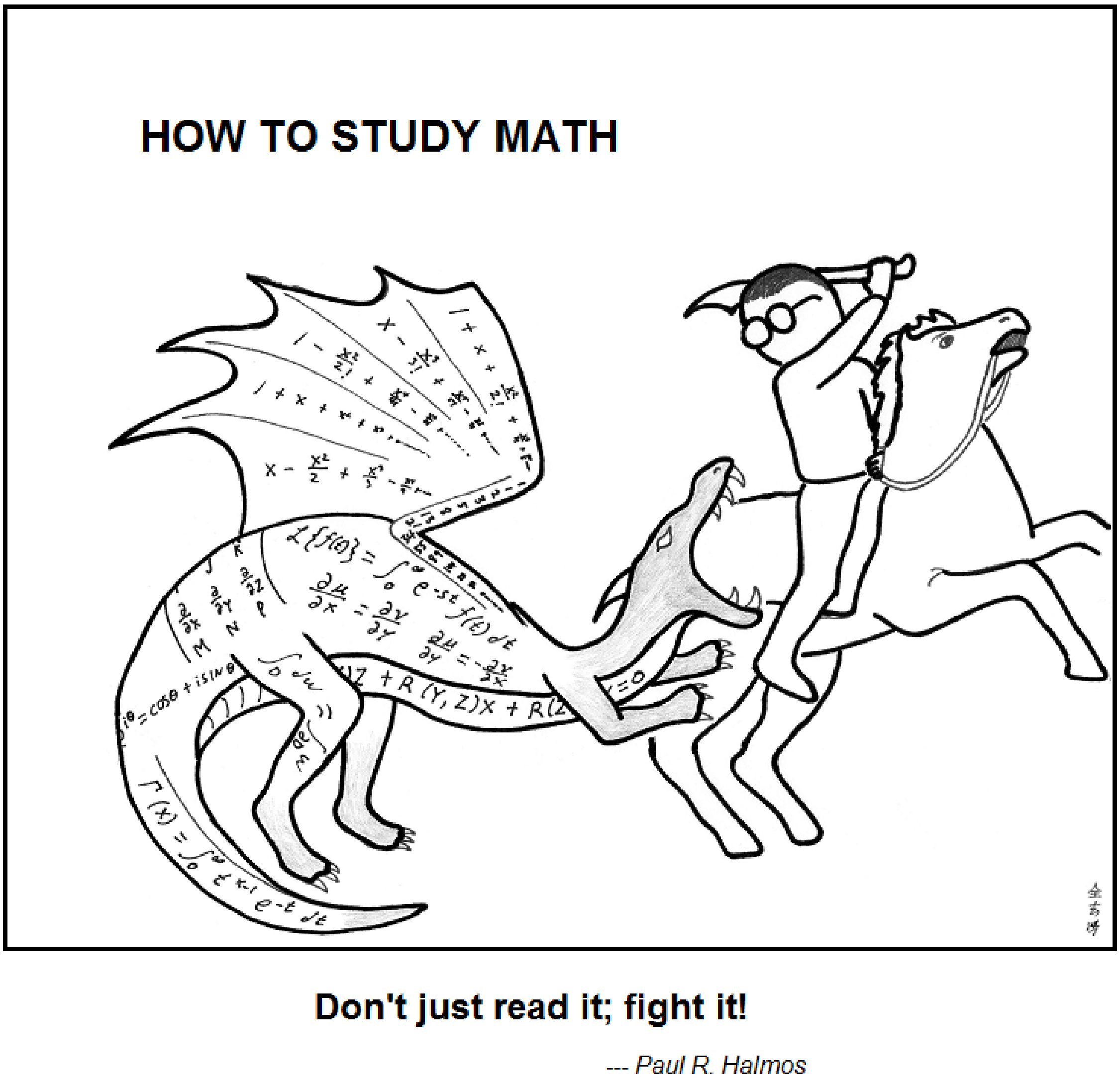


Figure 0.1: Don't just read it; fight it. – Paul Halmos (The comic is abstrusegoose.com)

Contents

How to study for Analysis II

- **Read carefully and deliberately**. As you know, the way one should read in mathematics is quite different from how you may read a history book or a magazine or newspaper. In mathematics you must read slowly, absorbing each phrase. In the first semester, you should all have received a self-explanation training booklet which you should not forget. If you have lost your copy or have never received one, you can contact me and I will bring some to the lecture.

- **Think with pencil and scratch paper**. Working on mathematics you should always have a pencil and some sheets of paper ready and use them. Test out what is written in the lecture notes, construct examples and counterexamples, draw pictures.
  This will help to clinch the ideas and procedures in your mind before starting the exercises. After you have read and reread a problem carefully, if you still do not see what to do, do not just sit and look at it. Get your pencil going on scratch paper and try to dig it out. Try this "algorithm":

  1. Write down what you want to show.
  2. Write down what you know.
  3. What can you immediately deduce from the known? Apply some known simple inequalities and identities and see what you can get from it.
  4. Think of a strategy. And try to implement it.

5. Have you used all information? Do you have numbers, vectors or functions. Have you used their specific properties?

Another good reason for getting things down is that it is much more easy to help you. Maybe you just overlooked something and with a small hint from me or tutors, you have a light-bulb experience. You should not rob yourself of that by asking to much of the solutions at once.

- **Be independent**. To be clear, being independent does not mean that you should ask no questions at all but to have good judgement over what to ask and when. Sometimes little things will cause considerable confusion or you do not know where even to begin with your studies. Then you should ask for help. Do not be afraid that your question may sound *dumb*. The only *dumb* action is to fail to ask about a topic that you have really tried to grasp and still do not understand. Some people seek help too soon and some wait too long. You will have to use good common sense in this matter.

- **Persevere**. Do not become frustrated if a topic or problem may completely baffle you at first. *Stick with it!* An extremely interesting characteristic of learning mathematics is that at one moment the learner may feel totally at a loss, and then suddenly have a burst of insight that enables her to understand the situation perfectly. If you don't seem to be making any progress after working on a problem for some time, put it aside and attack it again later. Many times you will then see the solution immediately even though you have not been consciously thinking about the problem in the meantime. There is a tremendous sense of satisfaction in having been persistent enough and creative enough to independently solve a problem that had given you a great deal of trouble.

- **Take time to reflect** To learn mathematics well you must take time to do some reflective thinking about the material covered during the last few days or weeks. It takes time for some ideas in mathematics to *soak in*. You may have to live with them a while and do reflective thinking about them before they become a part of you.

- **Concentrate on fundamentals**. Do not try to learn mathematics by memorizing illustrative examples.[1] You will soon become overwhelmed by this approach, and the further you go the less successful you will be. All mathematics is based on a few fundamental principles and definitions. Some of these must be memorized. But if you concentrate on these fundamentals and try to see how each new topic is just a reapplication of them, very little additional memorization will be necessary.[2]

- **Use Heuristics.**[3] If you work on a problem that you can not solve immediately try to consider a problem that is similar but somewhat easier. It is not always easy to find such problems but it will come to you with practice. For example, if you consider a function that depends on a parameter, you can study it by setting the parameter to a specific value. Then do the calculations and, after reaching a satisfying result, follow your calculations step by step with the general parameter. Many more methods may be found in [13] and [10]. Some historical points and further references can be found in [12]. Some evidence of student's benefits from following heuristic methods as well as further references can be found in [5].

[1] Which does not mean you should not have a couple of examples/counterexamples up your sleeve for any concept learned.

[2] For example: If you fully understood how one searches for extrema when one considers functions of one variable, only few additional things have to be memorized in the multi-variable case. (Even less with further knowledge in multi-variable calculus.)

[3] Heuristics is the study of means and methods of problem solving.

Polya's algorithm for solving problems

HOW TO SOLVE IT

UNDERSTANDING THE PROBLEM

First.

You have to *understand* the problem.

What is the unknown? What are the data? What is the condition?
Is it possible to satisfy the condition? Is the condition sufficient to determine the unknown? Or is it insufficient? Or redundant? Or contradictory?

Draw a figure. Introduce suitable notation.

Separate the various parts of the condition. Can you write them down?

DEVISING A PLAN

Second.

Find the connection between the data and the unknown. You may be obliged to consider auxiliary problems if an immediate connection cannot be found. You should obtain eventually a *plan* of the solution.

Have you seen it before? Or have you seen the same problem in a slightly different form?

Do you know a related problem? Do you know a theorem that could be useful?

Look at the unknown! And try to think of a familiar problem having the same or a similar unknown.

Here is a problem related to yours and solved before. Could you use it? Could you use its result? Could you use its method? Should you introduce some auxiliary element in order to make its use possible?

Could you restate the problem? Could you restate it still differently? Go back to definitions.

If you cannot solve the proposed problem try to solve first some related problem. Could you imagine a more accessible related problem? A more general problem? A more special problem? An analogous problem? Could you solve a part of the problem? Keep only a part of the condition, drop the other part; how far is the unknown then determined, how can it vary? Could you derive something useful from the data? Could you think of other data appropriate to determine the unknown? Could you change the unknown or the data, or both if necessary, so that the new unknown and the new data are nearer to each other?

Did you use all the data? Did you use the whole condition? Have you taken into account all essential notions involved in the problem?

CARRYING OUT THE PLAN

Third.

Carry out your plan.

Carrying out your plan of the solution, *check each step*. Can you see clearly that the step is correct? Can you prove that it is correct?

LOOKING BACK

Fourth.

Examine the solution obtained.

Can you *check the result?* Can you check the argument?

Can you derive the result differently? Can you see it at a glance?

Can you use the result, or the method, for some other problem?

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 A.1.

$\mathbb{R}^{\mathbb{N}_0}$	This denotes the set of all sequences (a_n) with $a_n \in \mathbb{R}$ for all $n \in \mathbb{N}$. $(a_n) \in \mathbb{R}^{\mathbb{N}_0}$ is equivalent to say $(a_n) \subseteq \mathbb{R}$. The point of this notation is to stress that (a_n) is a point in a vector space.
$\Omega^{\mathbb{N}}$	This denotes the set of all sequences (a_n) with $a_n \in \Omega$ for all $n \in \mathbb{N}$. See also Definition A.3.
$\lim_{n\to\infty} a_n$	This denotes the limit of a sequence. For example in $\mathbb{R}$ or $\mathbb{R}^d$. However, the sequence may also be a sequence of functions. See Definitions 1.11 and 6.2.
$\lvert \cdot \rvert$	Absolute value of a real number. See Section 1.3.
$\lVert \cdot \rVert_X$	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 $\mathbb{R}^d$ have special notation too. See Sections 1.3 and 5.7.
$C(\Omega)$	The set of continuous functions $f : \Omega \to \mathbb{R}$, $\Omega \subseteq \mathbb{R}^d$. See also Section 5.7.
$C[a,b]$	The set of continuous functions $f : [a,b] \to \mathbb{R}$, $\Omega \subseteq \mathbb{R}^d$. See also Section 5.7.
$PC[a,b]$	The set of all piecewise continuous functions on $[a,b]$. See Definition 7.6.

$C^k[a,b]$	The set of k-times differentiable functions $f : [a,b] \to \mathbb{R}$ such that $f^{(i)} \in C[a,b]$ for $i = 0,\dots,k$. See also Section 5.7.
$\mathcal{S}[a,b]$	The real vector space of all step-functions $f : [a,b] \to \mathbb{R}$. See Definitions 7.2 and 7.3.
$\mathcal{R}[a,b]$	The set of all regulated functions on the interval $[a,b]$. See Definition 7.5 and Theorem 7.4.
$\mathcal{I}(f)$	Step function integral. See Definition 7.2.
$\int\limits_a^b f(x)\mathrm{d}x$	Can stand for the step function integral $\mathcal{I}(f)$ if $f \in \mathcal{S}[a,b]$, the regulated Integral (Definition 7.8), or the Riemann integral (Definition 7.10).
$\overline{\int_a^b} f(x)\mathrm{d}x$	Stands for the upper integral. See Definition 7.9.
$\underline{\int_a^b} f(x)\mathrm{d}x$	Stands for the lower integral. See Definition 7.9.
$\frac{d}{dx}$	Differentiation operator with respect to x for functions $f : (a,b) \to \mathbb{R}$. See also Section 5.3.
$\frac{\partial}{\partial x_i}$	Differentiation operator with respect to x_i for functions $f : \Omega \to \mathbb{R}$ with $\Omega \subseteq \mathbb{R}^d$. 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 9.1.2.
$f_n \to f$	If the f_n are functions then this symbol indicates that the sequence (f_n) converges in a sense to f. It should additionally be indicated whether the convergence is pointwise (Definition 6.1) or uniform (Definition 6.2).
$f_n \rightrightarrows f$	This symbol usually means that the sequence (f_n) converges uniformly to f. In this notes we indicate uniform convergence by specifically stating it. For the definition of uniform convergence see Definition 6.2.

$\mathrm{cl}(\Omega)$	Let $\Omega \subseteq \mathbb{R}^d$ 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 $\mathbb{R}^d$. See Definition 8.5.
$\mathrm{Sub}(V)$	The set $\mathrm{Sub}(V)$ is the collection of all subs-spaces of the (real) vector space V.
$f\big\|_X$	Restriction of a function to a sub-set of its domain. Let $f : A \to B$ be a function and $X \subseteq A$. Then, the function $g : C \to B$ with $x \mapsto f(x)$ is denoted by $f\big\|_C$.
χ_A	Denotes the characteristic function (or indicator function) of the set A. See Definition 7.1.

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 without which you can not work.
For the starred theorems you need to be able to give detailed proofs and for the rest you should know the main ideas. Other theorems, not listed here, might still be asked in the exam if their proof is straight forward and you might need them in arguing other results.

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 A.5.4. A function is said to be convex on an 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 $\lambda \in [0, 1]$. 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.[4] 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).

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

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

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.

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

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. $f : \Omega \subseteq \mathbb{R}^d \to \mathbb{R}$, 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 Sequences and series in $\mathbb{R}^d$

This chapter serves as a gentle introduction to Analysis 2. We will look at some definitions we made in Analysis 1 and will transfer them from $\mathbb{R}^1$ to $\mathbb{R}^d$ with arbitrary $d \geq 1$.

At the start, this will leads us to think about distance and length in $\mathbb{R}^d$. The first section is intended to motivate these investigations and the second and third will introduce the precise notions of length and distance that we will use in this course.

1.1 Some notational remarks

Throughout the notes, I will continue to denote sequences by (a_n) regardless of whether they are sequences in $\mathbb{R}$ or in $\mathbb{R}^d$, $d \geq 1$.
If we consider a sequence in $\mathbb{R}^d$, it is understood that the elements a_n of (a_n) are of the type

$$\mathbb{R}^d \ni a_n = \begin{bmatrix} a_1^{(n)} \\ \vdots \\ a_d^{(n)} \end{bmatrix} = \begin{bmatrix} a_1^{(n)} & \dots & a_d^{(n)} \end{bmatrix}^T.$$

I will usually denote the ith component by a lower index and the counting index of the sequence in parenthesis in an upper index. This notation may take some time to get used to. I suggest you do the exercises in this section on an extra sheet of paper to get the necessary acquaintance.

In special cases as $\mathbb{R}^2$, like in the next section, we may simplify notation by the choice of extra letters like x_n and y_n. However, for the general case we use the above described notation. To get a better feeling of the statements, you may always specialise them to $d = 2$ to see better what the notation is trying to say.

1.2 Convergence of sequences $(a_n) \subseteq \mathbb{R}^2$

We start with an repetition of

Definition 1.1 (Convergence of sequences in $\mathbb{R}$)**.**
A sequence $(a_n) \subseteq \mathbb{R}$ is ***convergent*** *if and only if there exists an $a \in \mathbb{R}$ such that*

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

Let us write one more times in plain English, what that means:

Now, let us ask ourselves the following question. Let $(x_n), (y_n) \subseteq \mathbb{R}$ be two sequences and let,

$$a_n := \begin{bmatrix} x_n \\ y_n \end{bmatrix} \in \mathbb{R}^2$$

be a sequence in $\mathbb{R}^2$. How would you define convergence $(a_n) \to a$ in this case?

How do you get the right definition?

1. How do you calculate $|x - y|$ for $x, y \in \mathbb{R}^2$? (A-level maths)

2. Read Definition 1.1 out loud.

3. Now, do read the meaning of Definition 1.1 out loud. What does $|a_n - a| < \varepsilon$ mean?

Give it a go and write down some thoughts. Why not start with a verbal description, which you thought about on the last page, which you den transform in more and more formal terms?

Definition 1.2 (Convergence of $(a_n) \subseteq \mathbb{R}^2$ (Student version I))**.**

Definition 1.3 (Convergence of $(a_n) \subseteq \mathbb{R}^2$ (Student version II))**.**

Definition 1.4 (Convergence of $(a_n) \subseteq \mathbb{R}^2$ (Student version III))**.**

Let us now write down the precise definitions and investigate their relationships.

Definition 1.5 (Convergence of $(a_n) \subseteq \mathbb{R}^2$, Version I)**.**

Definition 1.6 (Convergence of $(a_n) \subseteq \mathbb{R}^2$, Version II)**.**

Definition 1.7 (Convergence of $(a_n) \subseteq \mathbb{R}^2$, Version III)**.**

Question:

What theorem(s) should we expect to be true?

Theorem 1.1.

Proof. The full proof will be on the problem sheet for you to carry out. However, let us discuss the main points and difficulties.

1.3 Length and distance in $\mathbb{R}^d$

1.3.1 The notion of a norm

Recall the calculation of

1. Length $|x|$ of x for an $x \in \mathbb{R}^2$.

2. Length $|x|$ of x for an $x \in \mathbb{R}^3$.

Recall properties of length $\mathbb{R}^2$ and $\mathbb{R}^3$:

1. How is the length of x and λx related if $x \in \mathbb{R}^2$ or $\mathbb{R}^3$ and $\lambda \in \mathbb{R}$.

2. Can length be negative?

3. How is the length of $x + y$ related to the length of x and y?

With that, we introduce the notions of a **norm** (length) and **metric** (distance) on $\mathbb{R}^d$. Pay attention to the notation. Usually, students find it difficult to use at first. The way to overcome it and get really familiar, is to work with it.

Definition 1.8 (**Norm** (length) on $\mathbb{R}^d$).
A function $\|\cdot\| : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ is called a ***norm*** *if and only if*

(P1) For all $x \in \mathbb{R}^d$, it holds

$$\|x\| \geq 0 \quad \textit{(Positivity)}$$

and $\|x\| = 0$ if and only if $x = 0$.

(P2) For all $\lambda \in \mathbb{R}$, $x \in \mathbb{R}^d$, it holds

$$\|\lambda x\| = |\lambda|\|x\|. \quad \textit{(Homogeneity)}$$

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

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

Remark 1.1. *A vector space which is equipped with a norm, is called a normed vector space. As always, the $\cdot$ in $\|\cdot\|$ indicates where the argument goes, e.g. $\|x\|$ is the norm of $x \in \mathbb{R}^d$ for a given norm. The notation $\|\cdot\|$ is used if we want to talk about the norm as a function on $\mathbb{R}^d$ and the notation $\|x\|$ means the concrete norm of x.*[1,2]

Remark 1.2. *The definition of a norm is an abstract one. On some spaces, we have many norms as we will soon see. However, on some vector spaces, $\mathbb{R}^d$ being an example, all norms are related.*

[1] That is the same difference as between a function f and its value at x which is denoted by $f(x)$.
[2] Again, $\|\cdot\|$ stands for any function $\mathbb{R}^d \to \mathbb{R}_{\geq 0}$ satisfying the three conditions in the definition above.

Example 1.1. *The easiest example for a norm is the absolute value function* $|\cdot|$ *which is defined on* $\mathbb{R}$ *as*

$$\begin{cases} |\cdot| : \mathbb{R} \to \mathbb{R}_{\geq 0} \\ x \mapsto \begin{cases} x & : \ x > 0 \\ 0 & : \ x = 0 \\ -x & : \ x < 0 \end{cases} \end{cases}.$$

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

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

The triangle inequality $|x + y| \leq |x| + |y|$ *for* $x, y \in \mathbb{R}$ *is a well known fact from Analysis 1. Thus,* $|\cdot|$ *is a norm on* $\mathbb{R}$ *in the sense of the definition above. Again, the* $\cdot$ *indicates where the argument goes.*

Example 1.2. *The second example for such a function is the well known* ***Euclidean length (or norm)*** *which is defined as*

$$\|x\|_2 = \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_d|^2} = \sqrt{\sum_{i=1}^{d} |x_i|^2} \tag{1.3.1}$$

for any $x = [x_1, \ldots, x_d]^T \in \mathbb{R}^d$.

Exercise 1.1. *Calculate the Euclidean norm of the vectors* $\begin{bmatrix} 2 \\ -3 \end{bmatrix}$ *and* $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$.

Remark 1.3. *In school and applied mathematics, also in the module Mathematical Methods 2,* $\|\cdot\|_2$ *is often denoted by* $|\cdot|$. *We do not use this notation since we reserve it for the absolute value function* $|\cdot|$ *defined on the number line* $\mathbb{R}$. *If* $d = 1$, *we get that* $\|x\|_2 = \sqrt{x^2} = \max\{x, -x\} = |x|$ *is the usual absolute value.*

Exercise 1.2. *Prove that $\|\cdot\|_2$ fulfils Properties $(P1)$ to $(P3)$ in Definition 1.8 for $d = 2$.*

Example 1.3. *The 2 in the definition of $\|\cdot\|_2$ in (1.3.1) plays no special role other than giving the familiar Euclidean distance from the origin of $\mathbb{R}^d$ 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*

$$\|x\|_p = (|x_1|^p + \ldots + |x_d|^p)^{\frac{1}{p}} = \left(\sum_{i=1}^{d} |x_i|^p\right)^{\frac{1}{p}}, \tag{1.3.2}$$

where p may be in $[1, +\infty)$.

We can extend the definition of $\|\cdot\|_p$ *to* $p = +\infty$ *if we set*

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

For any $p \in [0, +\infty]$, *we call* $\|\cdot\|_p$ *the* ℓ^p*-norm.*
If $d = 1$, *we have* $\|x\|_p = |x|$, *where* $|x| = \max\{x, -x\}$.

Exercise 1.3. *Calculate the* $\|\cdot\|_3$*-norm of the vectors* $\begin{bmatrix} 2 & , & -3 \end{bmatrix}^T$ *and* $\begin{bmatrix} -1 & , & 2 \end{bmatrix}^T$.

For us, the three norms

- $\|x\|_1 =$
- $\|x\|_2 =$
- $\|x\|_\infty =$

are the most important and therefore, we shall investigate some relations between them. They are called the ℓ^1, ℓ^2, and ℓ^∞ norm respectively. We also call $\|x\|_\infty$ the **supremum norm** (maximum norm) of x.

We have

Theorem 1.2.

Let $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$ be the norms on $\mathbb{R}^d$ as defined above. Then, we have the following inequalities

1. $\|x\|_\infty \leq \|x\|_1 \leq d\,\|x\|_\infty$,
2. $\|x\|_\infty \leq \|x\|_2 \leq \sqrt{d}\,\|x\|_\infty$, *and*
3. $\|x\|_2 \leq \|x\|_1 \leq d\,\|x\|_2$

for all $x \in \mathbb{R}^d$.

Remark 1.4. *The situation of Theorem 1.2 is typical for norms on $\mathbb{R}^d$. In fact, for two arbitrary norms $\|\cdot\|_a$ and $\|\cdot\|_b$ we can always find positive constants c and C such that*

$$c\|x\|_a \leq \|x\|_b \leq C\|x\|_a. \tag{1.3.3}$$

If two norms satisfy an estimate of the type (1.3.3), we say that the norms are equivalent. See also the article Norm in the Encyclopedia of Mathematics.

Proof of Theorem 1.2.

The inequality $\|x\|_\infty \leq \|x\|_1$ is clear since $\|x\|_1$ contains $\max\limits_{i=1,\ldots,d} |x_i|$.
Further, we have

$$\begin{aligned} \|x\|_1 &= |x_1| + \ldots + |x_d| \leq d \max_{i=1,\ldots,d} |x_i| \\ &= d\|x\|_\infty. \end{aligned}$$

since $|x_i| \leq \max\limits_{i=1,\ldots,d} |x_i|$ for all $i = 1, \ldots, d$.
Thus, we proved the first statement of the theorem.

For the first inequality of the second statement, let $j \in \{1, \ldots, d\}$ be such that $|x_j| = \max_{i=1,\ldots,d} |x_i|$.
Then,

$$\begin{aligned} \|x\|_\infty &= \sqrt{|x_j|^2} \leq \\ &= \|x\|_2. \end{aligned}$$

For the second case, we obtain

$$\|x\|_2 = \sqrt{|x_1|^2 + \ldots + |x_d|^2} \leq$$

since $|x_i|^2 \leq \max_{i=1,\ldots,d} |x_i|^2$ for all $i = 1, \ldots, d$.
Further, we have $\max |x_i|^2 \leq (\max |x_i|)^2$.
Hence, we obtain $\|x\|_2 \leq \sqrt{d}\|x\|_\infty$.

For the third statement, we remember that

$$x = x_1 e_1 + \ldots + x_d e_d,$$

where $\{e_i : i = 1, \ldots, d\}$ is the standard basis of $\mathbb{R}^d$.
By the triangle inequality for $\| \cdot \|_2$, we get

$$\begin{aligned} \|x\|_2 &= \|x_1 e_1 + \ldots + x_d e_d\|_2 \leq |x_1|\|e_1\| + |x_d|\|e_d\|, \\ &\leq |x_1| + \ldots + |x_d| = \|x\|_1, \end{aligned}$$

since $\|e_i\|_2 = 1$ for all $i \in \{1, \ldots, d\}$.
To prove the second inequality of the third statement, we estimate

$$\|x\|_1 = |x_1| + \ldots + |x_d| \leq d\|x\|_2,$$

where we used $|x_j| \leq \|x\|_2$, $j = 1, \ldots, d$.[3]
This concludes the proof. □

[3]We could have combined estimates 1 and 2 to get the right hand side of the third estimate. However, it is worthwhile to prove it in its own right.

Remark 1.5. *The estimate in the third statement in Theorem 1.2 is not the best possible. However, it will suffice for everything we want to do in this module. On the problem sheet, you will find questions leading to the proof of a sharper version*

$$\forall x \in \mathbb{R}^d \, : \; \|x\|_2 \leq \|x\|_1 \leq \sqrt{d}\, \|x\|_2.$$

Let us draw some pictures to get a better grasp of Theorem 1.2.

Figure 1.1: Unit balls in different norms.

1.3.2 The notion of a metric

Recall the calculation of

1. the distance $|x-y|$ between x and y for $x, y \in \mathbb{R}^2$.

2. the distance $|x-y|$ between x and y for $x, y \in \mathbb{R}^3$.

Recall properties of distance in $\mathbb{R}^2$ and $\mathbb{R}^3$:

1. How is the distance of x to y related to the distance from y to x.

2. Can distance be negative?

3. What is the relation of distances if you walk from a point x to a point y when you take a detour to a point z?

Let us now introduce the notion of a **distance**, in more general situations called **metric**, on $\mathbb{R}^d$, $d \geq 1$.

Definition 1.9 (Metric (distance) on $\mathbb{R}^d$**).**
*A function $\rho : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ is called a **metric** (distance) if it satisfies*

(H1) For any x, $y \in \mathbb{R}^d$, we have

$$\rho(x, y) \geq 0 \qquad \text{(Positivity)}$$

and $\rho(x, y) = 0$ if and only if $x = y$.

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

$$\rho(x, y) = \rho(y, x). \qquad \text{(Symmetry)}$$

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

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

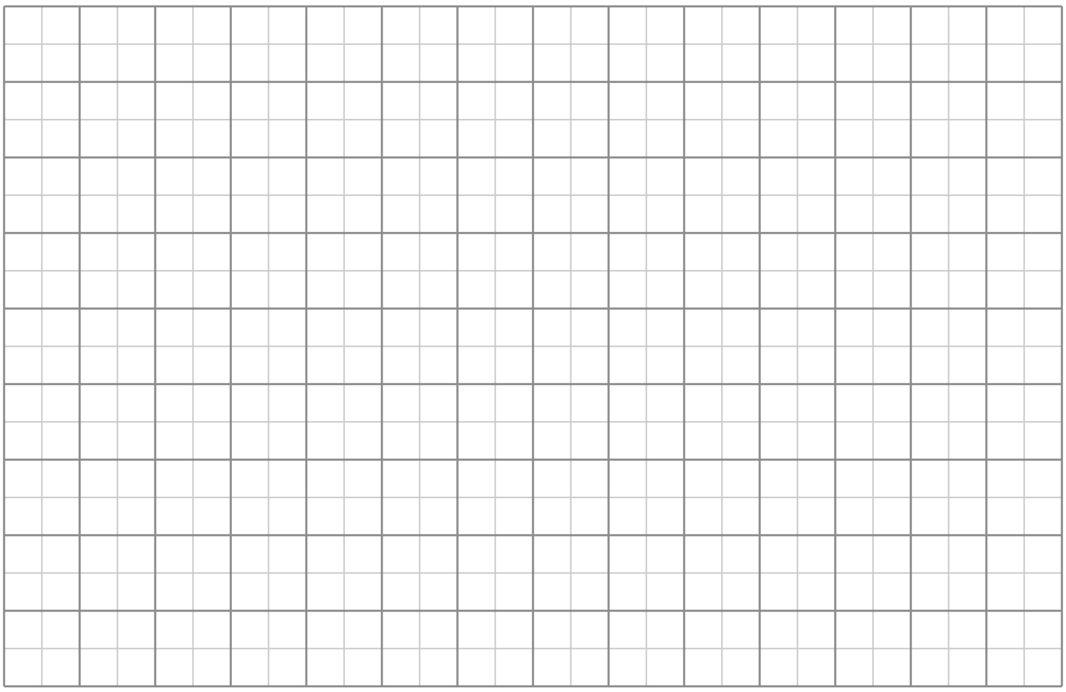


Figure 1.2: Calculating the distance between two points in $\mathbb{R}^2$ using Pythagoras' theorem.

Example 1.4. *A well-known example for a metric is the well-known Euclidean distance*

$$\rho_2(x,y) = \sqrt{\sum_{i=1}^{d} |x_i - y_i|^2}.$$

We say this metric is induced by the Euclidean norm $\|\cdot\|_2$ as

$$\rho_2(x,y) = \|x-y\|_2. \tag{1.3.4}$$

Let us check the properties $(H1)$ to $(H3)$. Since we have $\rho_2(x,y) = \|x-y\|_2 \geq 0$ (see $(P1)$ in Definition 1.8), $(H1)$ follows. By $(P2)$ in Definition 1.8, we get

$$\begin{aligned}
\rho_2(x,y) &= \|x-y\|_2 = \|-(y-x)\|_2 \\
&= |-1|\|y-x\|_2 \\
&= \rho_2(y,x).
\end{aligned}$$

Finally, we have

$$\begin{aligned}
\rho_2(x,y) = \|x-y\|_2 &= \|x-z+z-y\|_2 \\
&\leq \|x-z\|_2 + \|z-y\|_2 = \rho_2(x,z) + \rho_2(z,y)
\end{aligned}$$

by $(P3)$ in Definition 1.8.

Example 1.5. *The norms defined in Example 1.3 provide another possibility to define distances on $\mathbb{R}^d$. We get*

$$\rho_p(x,y) = \|x-y\|_p = \left(\sum_{i=1}^{d} |x_i - y_i|^p\right)^{\frac{1}{p}} \tag{1.3.5}$$

and

$$\rho_\infty(x,y) = \max_{i=1,\dots,d} |x_i - y_i|.$$

Exercise 1.4. *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* ρ_p *for* $p = 1$, 2, *and* ∞.

Exercise 1.5. *Using the properties of norms, try to prove that* ρ_p, *defined in* (1.3.5), *is a metric on* $\mathbb{R}^d$ *according to Definition 1.9.*

Remark 1.6. *In this module, we will only consider special classes of metrics which are, in fact, induced by norms as on* (1.3.4). *However, metrics can not only be defined on vector spaces and do even then not necessarily come from norms; the more general theory leads to metric spaces which is just a non-empty set equipped with a metric, i.e. a function with the properties stated in Definition 1.9.*

1.4 Conclusions

Reading 1. *This Sections 1.4 and 1.5 form this week's reading. Please take the time to work though it with care and keep a list of questions which you can either ask me, your tutor or the MLSC staff.*

How should you go through the reading?

1. *First read through skipping proofs.*
2. *Read the definitions and theorems again and write down examples in simple cases i.e.* $d = 1$ *(Analysis 1) and* $d = 2$*.*
3. *Work through the proofs. If you do not understand certain steps, write down the questions with a useful reference and ask me, your tutor, or the MLSC staff.*

With the discussion of norm and metric, we are prepared, to make the following definitions regarding the boundedness and convergence of sequences in $\mathbb{R}^d$ with respect to (w.r.t.) a norm $\|\cdot\|$ or a metric ρ.

Definition 1.10 (Boundedness of $(a_n) \subseteq \mathbb{R}^d$)**.**
Let $\|\cdot\|$ *be a norm on* $\mathbb{R}^d$*. Let* $(a_n) \subseteq \mathbb{R}^d$ *be a sequence. We say that* (a_n) *is* ***bounded*** w.r.t. $\|\cdot\|$ *if and only if there exists* $C > 0$ *such that*

$$\forall n \in \mathbb{N} \; : \; \|a_n\| \leq C.$$

Remark 1.7. *The choice of the norm in Definition 1.10 is of no matter. In light of Remark 1.4, we have that for any two norms* $\|\cdot\|_a$ *and* $\|\cdot\|_b$*, that if* (a_n) *is bounded w.r.t.* $\|\cdot\|_a$ *then it is bounded w.r.t. to* $\|\cdot\|_b$ *and vice versa.*

Exercise 1.6 (Skip this exercise in a first reading.)**.** *Can you produce a definition of boundedness of a set* $A \subseteq \mathbb{R}^d$ *in terms of norms and then in terms of a metric? Reconsider first the definitions we made in the case* $\mathbb{R}^1$ *and then take a general norm/metric satisfying Definitions 1.8/1.9 and generalise.*

Remark 1.8. *You might think of another definition of boundedness, which makes more direct use of our knowledge of Analysis 1. Should a sequence* $(a_n) \subseteq \mathbb{R}^d$ *not be bounded if all the component sequences are bounded?*
The answer is yes. Can you work out how this is related to Definition 1.10? (Hint: Theorem 1.2.) If not, you should go ahead and read through the rest of this week's reading and come back to this question at the end.

Exercise 1.7. *Illustrating Remark 1.7, prove that a sequence which is bounded w.r.t.*[4] $\|\cdot\|_1$ *is also bounded w.r.t.* $\|\cdot\|_2$*, and* $\|\cdot\|_\infty$*.*

Definition 1.11 (Convergence of $(a_n) \subseteq \mathbb{R}^d$).
Let ρ *be a metric on* $\mathbb{R}^d$*. We say that a sequence* $(a_n) \subseteq \mathbb{R}^d$ *is* ***convergent*** *w.r.t.* ρ *if and only if there exists an* $a \in \mathbb{R}^d$ *such that*

$$\forall \varepsilon > 0 \quad \exists n_0 \in \mathbb{N} \ : \ \forall n \in \mathbb{N},\, n \geq n_0, \quad \rho(a_n, a) < \varepsilon.$$

Write the definition out in plain English:

[4] Just a reminder. The abbreviation w.r.t. stands for **with respect to**.

Remark 1.9. *In the definition above, the convergence depends on the chosen metric, which is true in general. In this module, we will (mostly) use metrics* ρ *on* $\mathbb{R}^d$ *which come from a norm, i.e. there exists a norm* $\|\cdot\|$ *such that* $\rho(x, y) = \|x - y\|$*.*
Such metrics produce all the same notion of convergence, i.e. if a sequence converges with respect to one, it converges with respect to all of them.
However, you should be aware that there exists metrics ρ*, also on* $\mathbb{R}^d$*, which do not come from a norm in the above described sense. For instance, consider the metric*

$$\rho(x, y) := \begin{cases} 1 & : \quad x \neq y \\ 0 & : \quad x = y \end{cases} . \tag{1.4.1}$$

See also Remark 1.6.

Which sequences converge in $\mathbb{R}$, if it is equipped with the above metric ρ (see (1.4.1))? (We call this metric the discrete metric.)

This discussion leads us to make the following particular definition

Definition 1.12 (Convergence of $(a_n) \subseteq \mathbb{R}^d$ w.r.t $\|\cdot\|_2$).
*We say that a sequence $(a_n) \subseteq \mathbb{R}^d$ is **convergent** (w.r.t. $\|\cdot\|_2$) if and only if there exists an $a \in \mathbb{R}^d$ such that*

$$\forall \varepsilon > 0 \quad \exists n_0 \in \mathbb{N} \ : \ \forall n \in \mathbb{N},\, n \geq n_0, \quad \|a_n - a\|_2 < \varepsilon.$$

Remark 1.10. *The convergence according to Definition 1.12 is equivalent to component-wise convergence. See the Problem Sheet.*

Remark 1.11. *By Theorem 1.2, we obtain that the use of the $\|\cdot\|_2$ norm in the above definition is for our convenience, we could use $\|\cdot\|_1$ or $\|\cdot\|_\infty$ and would have that sequences which converge with respect to the metric $\|x-y\|_i$, $i \in \{1, 2, \infty\}$ also converge with respect to the others. See also Remark 1.4.*

Exercise 1.8. *To understand the situation better, prove the assertion on the last remark, i.e. prove the following statements:*

1. *Let $(a_n) \subseteq \mathbb{R}^d$ be a sequence which converges with respect to $\|\cdot\|_2$. Prove that it then converges with respect to $\|\cdot\|_1$.*
2. *Let $(a_n) \subseteq \mathbb{R}^d$ be a sequence which converges with respect to $\|\cdot\|_1$. Prove that it then converges with respect to $\|\cdot\|_\infty$.*
3. *Let $(a_n) \subseteq \mathbb{R}^d$ be a sequence which converges with respect to $\|\cdot\|_\infty$. What can you say?*

Hint: You have to use Theorem 1.2.

Remark 1.12. *In the remainder of this module, we will use Definition 1.12 as the definition of convergence in $\mathbb{R}^d$. If necessary, we will do calculations in different norms but will always use Theorem 1.2 to get back to the ℓ^2-norm.*

We can now use Definition 1.12 to make a suitable definition of series over elements in $\mathbb{R}^d$. Before you read the definition, can you come up with a suitable one yourself?

Definition 1.13 (Series in $\mathbb{R}^d$)**.**
*Let $(a_n) \subseteq \mathbb{R}^d$. Then, the series $\sum_{k=1}^{+\infty} a_k$ is **convergent** if and only if the sequence (s_n), defined by*

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

converges in the sense of Definition 1.12.

In general, it seems difficult to check that whether a series converges or not. Try for

$$\sum_{k=1}^{+\infty} \begin{bmatrix} k^{-2} \\ k^{-3} \end{bmatrix}.$$

If you think one should be able to do that component-wise, i.e. the series should converge because we know that

$$\sum_{k=1}^{+\infty} \frac{1}{k^2} \quad \text{and} \quad \sum_{k=1}^{+\infty} \frac{1}{k^3}$$

converge, you are right. Before you turn the page, write down a theorem that formalises this intuition. Can you prove it? Consider the case $d = 2$ first to avoid unnecessary notational difficulty.

Theorem 1.3 (Componentwise convergence).
*Let $(a_n) \subseteq \mathbb{R}^d$ be a sequence. The series $\sum_{k=1}^{+\infty} a_k$ is **convergent** if and only if all the component series*

$$\sum_{k=1}^{+\infty} a_1^{(k)}, \ \ldots \ , \sum_{k=1}^{+\infty} a_d^{(k)}$$

converge, where $a_k = \left[a_1^{(k)}, \ldots, a_d^{(k)}\right]^T$.[a]

[a] The $(a_i^{(k)}) \subseteq \mathbb{R}$ are sequences with one parameter $k \in \mathbb{N}$ as we know them from Analysis 1. We call them the component sequences of $(a_k) \subseteq \mathbb{R}^d$. See also the beginning of Chapter 1.

Proof. We prove the two case separately. Read it carefully and do not forget your self-explanation training.

[⇒] Let (s_n) be the series of partial sums of $\sum_{k=1}^{+\infty} a_k$.
Since the series converges, we have that there exists an $S = [S_1, \ldots, S_d]^T \in \mathbb{R}^d$ such that for all $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that

$$\|s_n - S\|_2 = \left\| \sum_{k=1}^{n} a_k - S \right\|_2 < \frac{\varepsilon}{\sqrt{d}}.$$

Let $j \in \{1, \ldots, d\}$.
Then

$$\left| \sum_{k=1}^{+\infty} a_j^{(k)} - S_j \right| \leq \|s_n - S\|_1 \leq \sqrt{d}\|s_n - S\|_2 < \varepsilon.$$

Thus, for all $j \in \{1, \ldots, d\}$, the series $\sum_{k=1}^{+\infty} a_j^{(k)}$ converges.

[⇐] Let now $\sum_{k=1}^{+\infty} a_j^{(k)}$ be convergent for all $j \in \{1, \ldots, d\}$.
We show that $\sum_{k=1}^{+\infty} a_k$ is convergent.
Let $\varepsilon > 0$ be arbitrary.
There exist $n_1, \ldots, n_d \in \mathbb{N}$ such that for all $j \in \{1, \ldots, d\}$

$$\forall n \geq n_j \ : \quad \left| \sum_{k=1}^{n} a_j^{(k)} - S_j \right| < \frac{\varepsilon}{d} \tag{1.4.2}$$

holds.

Now, since

$$\left\| \sum_{k=1}^{n} a_k - S \right\|_2 \leq \left\| \sum_{k=1}^{n} a_k - S \right\|_1 ,$$

we get from (1.4.2) that

$$\left\| \sum_{k=1}^{n} a_k - S \right\|_2 < \frac{\varepsilon}{d} + \ldots + \frac{\varepsilon}{d} = d\frac{\varepsilon}{d} = \varepsilon.$$

Thus, $\sum_{k=1}^{+\infty} a_k$ converges.

This concludes the proof. $\square$

Let us use another definition to reduce the question to something we know.

Definition 1.14 (Absolute convergence)**.**
The series $\sum_{k=1}^{+\infty} a_k$ *is* ***absolutely convergent*** *if the series* $\sum_{k=1}^{+\infty} \|a_k\|_2$ *converges.*

Remark 1.13. *The series* $\sum_{k=1}^{+\infty} \|a_k\|_2$ *in the above definition is a series of non-negative real numbers since which can be treated with the methods of Analysis 1 and it converges if and only if the sequence* (s_n)*, defined by*

$$s_n := \sum_{k=1}^{n} \|a_k\|_2$$

converges as a sequence in $\mathbb{R}^1$*.*

We should try to prove a result, as in Analysis 1, that connects absolute convergence with convergence. Before you turn the page, try to formulate a reasonable result connecting convergence and absolute convergence and try to prove it.

Theorem 1.4 (Absolute convergence $\Rightarrow$ convergence)**.**
Let $\sum_{k=1}^{+\infty} a_k$ *be absolutely convergent. Then, the series* $\sum_{k=1}^{+\infty} a_k$ *converges.*

Proof.

Since $\sum_{k=1}^{+\infty} a_k$ converges absolutely, we have that (S_n) with

$$S_n := \sum_{k=1}^{n} \|a_k\|_2$$

converges.

To reach the result, we shall show that for all $j = 1, \ldots, d$ the series $\sum_{k=1}^{+\infty} a_k^{(j)}$ converge and then apply Theorem 1.3.

We have the estimate

$$|a_k^{(j)}| = \sqrt{|a_k^{(j)}|^2} \leq \sqrt{|a_k^{(1)}|^2 + \ldots + |a_k^{(d)}|^2} = \|a_k\|_2. \qquad (1.4.3)$$

Let now (s_n) be the sequence of partial sums of $\sum_{k=1}^{+\infty} |a_k^{(j)}|$.

By (1.4.3), we have $0 \leq s_n \leq S_n$ and we obtain that $\sum_{k=1}^{+\infty} a_k^{(j)}$ converges absolutely and, by a result of Analysis 1, therefore converges.

This concludes the proof. $\square$

1.5 The Bolzano–Weierstrass theorem

Finally, we prove a generalization of the Bolzano–Weiserstrass theorem from Analysis 1 to $\mathbb{R}^d$. Before you proceed recall the ingredients of the proof that we have done in Analysis 1 and think whether it could work in $\mathbb{R}^d$.

Theorem 1.5 (Bolzano-Weierstrass)**.**
Every bounded sequence $(x_n) \subseteq \mathbb{R}^d$ *has a convergent sub-sequence.*

Proof.

Let $(x_n) \subseteq \mathbb{R}^d$ be a bounded sequence, i.e. there exists a constant $C > 0$ such that $\|x_n\|_2 \leq C$ for all $n \in \mathbb{N}$.[5]

We denote

$$x_n = \left[x_1^{(n)} \quad , \quad \ldots \quad , \quad x_d^{(n)}\right]^T .$$

The sequence $(x_1^{(n)})$ of first components of the elements of (x_n) is a bounded real sequence.

Thus, by the Analysis 1 Bolzano–Weierstrass, there exists a convergent sub-sequence $(x_1^{(n_k)}) \subseteq (x_1^{(n)})$.

Now, we consider $(x_{n_k}) \subseteq (x_n)$.

By the same arguments[6], we can now pick a convergent subsequence of the $(x_2^{(n_k)})$ and, thus, a corresponding sub-sequence of (x_{n_k}).[7]

Continuing like this for the remaining $n-2$ components, we obtain a sub-sequence of (x_n) in which the sequences of all components converge and, therefore, the sequence itself converges in $\mathbb{R}^d$.

This concludes the proof. $\square$

[5] See Definition 1.10 and Remark 1.7.

[6] Note here that you are taking sub-sequences of sub-sequences etc. I have written the argument mostly in words as the notation would get more and more sub-indices.

[7] i.e. the sequence $(x_{n_{k_l}})$. Note that the first component remains a convergent sequence as every sub-sequence of a convergent sequence is convergent with the same limit. I leave it to you to show that as an exercise.

Chapter

2 Open, closed, and compact

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

One of the fundamental notions of modern Analysis (and a field called Topology) is the notion of an open set. To define what an open set is, we need a the simple concept of open balls. We introduce

Definition 2.1 (Open ball of radius r around x).
*We define the **open ball** of radius $r > 0$ around a point $x \in \mathbb{R}^d$ by*

$$\begin{aligned} B_r(x) &:= \left\{ y \in \mathbb{R}^d \; : \; \rho_2(x, y) < r \right\}, \\ &= \left\{ y \in \mathbb{R}^d \; : \; \|x - y\|_2 < r \right\}. \end{aligned}$$

Remark 2.1. *In the definition of the open ball, we could use different metrics than just ρ_2. However, this choice will give us the balls (and spheres) as we know them. See also Remarks 1.4 and 1.7.*

Exercise 2.1. *Can you figure out, how the unit ball looks like in $\mathbb{R}^2$ if we replace ρ_2 above by ρ_1 or ρ_∞? See Figure 1.1.*

Example 2.1. *Let us consider $d = 1$, $\rho(x, y) = |x - y|$. 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

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

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 introduced in Analysis 1, means that $y \in (x - r, x + r)$.

2.2 Open sets

Now we introduce the very important

Definition 2.2 (Open sets)**.**

*A set $\Omega \subseteq \mathbb{R}^d$ is called **open** if and only if*

Figure 2.1: Illustration of an open set.

Remark 2.2. *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 x of the set Ω we can go a, possibly very small, step in any direction without leaving the set.*

Figure 2.2: Illustration of Remark 2.2.

Example 2.2. *Let us discuss a list of examples illustrating the definition of open sets:*

- *Open interval* (a, b) *is 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* $p \in \mathbb{R}^d$ *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* $\delta \in (0, r - \rho(p, q))$.

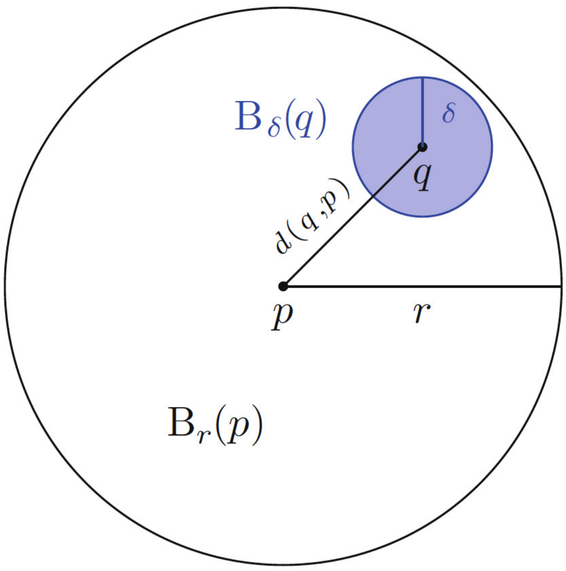


Figure 2.3: The open ball $B_r(p)$ is open.

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

2.3 Closed sets

To introduce the notion of a closed set, we need to consider points with certain properties. We call them limit points as they can be reached as a limit of a sequence whose elements are only from the considered set. We introduce

Definition 2.3 (Limit point of $\Omega \subseteq \mathbb{R}^d$).

Figure 2.4: Illustration of a limit point of Ω.

Example 2.3. *Let us discuss a list of examples illustrating the last definition:*

- *Let $\Omega = (0,1) \subseteq \mathbb{R}$. Then, every $x \in (0,1)$, $x = 0$, $x = 1$ are limit points of the set Ω.*

- *Consider the open ball $B_r(x)$. Then all points $x_0 \in B_r(x)$ are limit points of $B_r(x)$ as well as all points $x_0 \in \underbrace{\{y \in \mathbb{R}^d : \rho_2(y,x) = r\}}_{\text{Called the } (d-1)\text{-sphere of radius } r.}$.*

- *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***[1] *of the rational numbers in the real numbers.*

[1]This means that for every $x \in \mathbb{R}$, the ball $(x-r, x+r)$ contains a rational number however small one chooses $r > 0$. This was a consequence of Archimedes' theorem which we proved on a Problem Sheet in Analysis 1.

Proposition 2.1 (Characterization of limit points of $\Omega \subseteq \mathbb{R}^d$).
Let $\Omega \subseteq \mathbb{R}^d$. Then, $x \in \mathbb{R}^d$ is ***limit point*** *of Ω if and only if*

Proof.

The proof of this proposition is an exercise.

You might want to put it here.

From Proposition 2.1, we get

Corollary 2.1.

Let $\Omega \subseteq \mathbb{R}^d$ be an open set. Every $x \in \Omega$ is a limit point of Ω.

Exercise 2.2. *Prove Corollary 2.1. It might help to revisit the definition of an open set and to draw a picture.*

Remark 2.3. *Note that this corollary does not imply that all limit points are contained in Ω. It merely states that open sets can not have isolated points, see Definition 2.4. 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 $\{y \in \mathbb{R}^d : \|y\|_2 = 1\}$ are limit points and are not contained in $B_1(0)$.*

Exercise 2.3. *Prove that the set of limit points of a finite set $\{x_1, \ldots, x_n\} \subseteq \mathbb{R}^d$, $n \in \mathbb{N}$ is the set $\varnothing$. Again, it might help to draw a picture of the situation and to take into account what we have proven about limit points.*

Again from Proposition 2.1, we get immediately

Corollary 2.2.
Let $\Omega \subseteq \mathbb{R}^d$ and $x \in \mathbb{R}^d$ be a limit point of Ω. Then there exists a sequence $(x_n) \subseteq \Omega \setminus \{x\}$ such that $(x_n) \to x$.

Exercise 2.4. *Prove the converse of Corollary 2.2. That means to prove the following: Let $\Omega \subseteq \mathbb{R}^d$, $x \in \mathbb{R}^d$. If there exists a sequence $(x_n) \subseteq \Omega \setminus \{x\}$ with $(x_n) \to x$, then x is a limit point of Ω.*

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

Definition 2.4 (Isolated point of $\Omega \subseteq \mathbb{R}^d$)**.**
Let $\Omega \subseteq \mathbb{R}^d$. Then, $x \in \Omega$ is called an ***isolated point*** *of Ω if and only if*

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

Figure 2.5: Illustration of an Isolated point with neighbourhood $B_\varepsilon(x)$.

In Analysis 1, we understood that for a sequence (a_n) the notion $(a_n) \to a$ means that for all $\varepsilon > 0$ there are only finitely many a_n with $a_n \in \mathbb{R} \setminus (a - \varepsilon, a + \varepsilon)$.

Theorem 2.1.
A sequence $(a_n) \subseteq \mathbb{R}^d$ is convergent to $a \in \mathbb{R}^d$ if and only if

for all $\varepsilon > 0$ the set $\{n \in \mathbb{N} : a_n \notin B_\varepsilon(a)\}$ is finite.[a]

[a] We could alternatively state that for all $\varepsilon > 0$ the complement of the set $\{n \in \mathbb{N} : a_n \in B_\varepsilon(a)\} \subseteq \mathbb{N}$ must be finite for (a_n) to converge to a.

Proof.

[⇒]: Let $(a_n) \subseteq \mathbb{R}^d$ be convergent with $(a_n) \to a$..

Then, there exists for all $\varepsilon > 0$ an $n_0 \in \mathbb{N}$ such that

$$\forall n \in \mathbb{N}, n \geq n_0, \;\; \|a_n - a\|_2 < \varepsilon,$$

i.e $a_n \in B_\varepsilon(a)$.

Thus, for all $\varepsilon > 0$ there are at most $n_0 - 1$ elements a_n for which $a_n \notin B_\varepsilon(a)$.

[⇐]: We assume that $\{n \in \mathbb{N} : a_n \notin B_\varepsilon(a)\}$ is a finite set for all $\varepsilon > 0$.

Let $n_0 = n_0(\varepsilon)$ be the maximum of this set.

Then, for all $\varepsilon > 0$, we have $\|a_n - a\|_2 < \varepsilon$ for all $n \geq n_0 + 1$.

Thus, $a_n \in B_\varepsilon(a)$ for $n \geq n_0 + 1$.

Hence, $(a_n) \to a$.

This concludes the proof. □

Exercise 2.5. *Find an example which shows that the following statement is not true. A sequence $(a_n) \subseteq \mathbb{R}^d$ is convergent to $a \in \mathbb{R}^d$ if and only if for all $\varepsilon > 0$ the set $\{k \in \mathbb{N} : x_k \in B_\varepsilon(a)\}$ is infinite.*

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

Definition 2.5 (Closed set)**.**

A set $\Omega \subseteq \mathbb{R}^d$ is said to be ***closed*** *if and only if*

Figure 2.6: First diagrammatic thoughts about closed sets.

Example 2.4. *Let us discuss a list of examples illustrating the definition of closed sets:*

- *Closed intervals* $[a, b]$ *are closed.*

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

- *The whole space* $\mathbb{R}^d$ *is closed for any* $d \geq 1$*.*

- *The set* $\varnothing$ *is closed.*

- *All finite sets* $\{x_1, \ldots, x_n\} \subseteq \mathbb{R}^d$*,* $n \in \mathbb{N}$ *are closed. (Hint: Look back at Exercise 2.3.*

Reading 2.

The rest of this section is this week's reading. The reading ends on page 50.

Example 2.5.

Let us show that the set

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

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* $y \in \mathbb{R}^d \setminus \overline{B}_r(x)$. *Then there exists* $\varepsilon > 0$, *e.g. given by* $\varepsilon = \frac{1}{2}(\rho(x,y) - r)$, *such that* $B_\varepsilon(y) \cap \overline{B}_r(x) = \varnothing$. *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 $\overline{B}_r(x)$.

Let us state an alternative characterization of closed sets by

Proposition 2.2 (Characterization of closed sets by limits)**.**
Let $\Omega \subseteq \mathbb{R}^d$*. Then,* Ω *is closed if and only if for all Cauchy-sequences* $(x_n) \subseteq \Omega$ *we have that* $\lim_{n \to +\infty} x_n = x \in \Omega$*.*

Exercise 2.6. *Illustrate Proposition 2.2 by giving examples and counterexamples.*

Exercise 2.7. *Prove the equivalence stated in Proposition 2.2.*

Remark 2.4. ***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 $\mathbb{R}^d$ in general. Also the set $\varnothing$ is open and closed. See also Proposition 2.2.*

Theorem 2.2 (Characterization of open/closed sets)**.**
A set $\Omega \subseteq \mathbb{R}^d$ is

- *open if and only if $\Omega^c = \mathbb{R}^d \setminus \Omega$ is closed, and*
- *closed if and only if $\Omega^c = \mathbb{R}^d \setminus \Omega$ is open.*

Exercise 2.8. *Prove Theorem 2.2.*

We also have the following result regarding operations on open/closed sets.

Theorem 2.3.

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

[a]We will always assume that the open balls are defined with respect to $\|\cdot\|_2$ unless otherwise stated.

Remark 2.5. *To fully understand the following arguments, you should always try to draw a picture of the siltation.*

Especially the case $\Omega_1 \times \Omega_2$ is not trivial. Take for instance $\Omega_1 = (0,1)$, $\Omega_2 = (0,1)$ and then think about how to find a ball around a point $x = [x_1, x_2]^T \in (0,1)\times(0,1)$ by knowing that there are suitable $\varepsilon_1, \varepsilon_2 > 0$ such that $(x_1 - \varepsilon_1, x_1 + \varepsilon_1) \subseteq (0,1)$, $(x_2 - \varepsilon_2, x_2 + \varepsilon_2) \subseteq (0,1)$.

Specifically look at $(x_1 - \varepsilon_1, x_1 + \varepsilon_1) \times (x_2 - \varepsilon_2, x_2 + \varepsilon_2)$. What is the maximal radius of a ball inside this rectangle that you can find?

Proof.

Let us prove the assertions for open sets. The rest is left to you.[2]

Let Ω_1 and Ω_2 be open.

We prove the openness of $\Omega_1 \cap \Omega_2$.

If $\Omega_1 \cap \Omega_2 = \varnothing$, then the intersection is open as we have shown earlier.

If $\Omega_1 \cap \Omega_2 \neq \varnothing$, 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 $B_\varepsilon(x) \subseteq \Omega_1 \cap \Omega_2$ if $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$.

[2]Keep Theorem 2.3 and De Morgan's laws in mind. See also your notes from Mathematical Thinking.

Thus, $\Omega_1 \cap \Omega_2$ is open.

Let now $x \in \Omega_1 \cup \Omega_2$. Then, there exists $\varepsilon_1, \varepsilon_2 > 0$ such that $B_{\varepsilon_1}(x) \subseteq \Omega_1$ or $B_{\varepsilon_2}(x) \subseteq \Omega_2$.
Thus, $B_{\varepsilon_1}(x) \subseteq \Omega_1 \cup \Omega_2$ or $B_{\varepsilon_2}(x) \subseteq \Omega_1 \cup \Omega_2$.
Hence, $\Omega_1 \cup \Omega_2$ is open.

Finally, let $x \in \Omega_1 \times \Omega_2$.
Then, we have $x = (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$.
There exist $\varepsilon_1, \varepsilon_2 > 0$ such that $B_{\varepsilon_1}(\omega_1) \subseteq \Omega_1$ and $B_{\varepsilon_2}(\omega_2) \subseteq \Omega_2$.
Taking $\varepsilon := \min\{\varepsilon_1, \varepsilon_2\}$, we obtain that

$$\underbrace{B_\varepsilon(x)}_{\varepsilon-\text{ball in } \Omega_1\times\Omega_2.} \subseteq \underbrace{B_{\varepsilon_1}(\omega_1)}_{\varepsilon_1-\text{ball in } \Omega_1} \times \underbrace{B_{\varepsilon_2}(\omega_2)}_{\varepsilon_2-\text{ball in } \Omega_2} \subseteq \Omega_1 \times \Omega_2.$$

Thus, the set $\Omega_1 \times \Omega_2$ is open. $\square$

2.4 Compactness

This last section of the chapter is about compact sets. Compact sets are always a bit baffling to people when they encounter them first and it takes some time to understand their usefulness. Keep an open mind and study the following definitions carefully.

Definition 2.6 (Open cover).
Let $\Omega \subseteq \mathbb{R}^d$ be a subset and $\{\mathcal{O}_\alpha \,:\, \alpha \in I\}$ be a collection of open sets $\mathcal{O}_\alpha \subseteq \mathbb{R}^d$ indexed by some (possible uncountable) index set I. Then we say that $\{\mathcal{O}_\alpha \,:\, \alpha \in I\}$ is an ***open cover*** *of Ω if and only if*

$$\Omega \subseteq \bigcup_{\alpha \in I} \mathcal{O}_\alpha,$$

i.e. for all $x \in \Omega$ there exists an $\alpha \in I$ such that $x \in \mathcal{O}_\alpha$.

Example 2.6. *We discuss a couple of examples for open covers of subsets of $\mathbb{R}^d$.*

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.*

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 $\mathcal{O}_2 = (0,2)$.*

3. *The closed square $[0,1]^2$ is covered by the 4 open balls of radius 1 around the corner points, i.e. $\{\mathcal{O}_1, \dots, \mathcal{O}_4\} = \{\mathcal{O}_\alpha \,:\, \alpha \in \{1,2,3,4\}\}$ 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 $\mathcal{O}_4 = B_1((1,1))$.*

4. *The family $\{\mathcal{O}_\varepsilon \,:\, \varepsilon \in (0,+\infty)\}$,*

$$\mathcal{O}_\varepsilon = (-1-\varepsilon, 1+\varepsilon)$$

is an open cover of $(-1,1)$.

Definition 2.7 (Finite sub-cover)**.**
Let $\Omega \subseteq \mathbb{R}^d$ 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*** *if and only if there exists a finite subset $J \subseteq I$ such that*

$$\Omega \subseteq \bigcup_{\alpha \in J} \mathcal{O}_\alpha.$$

Example 2.7. *Consider the set $[0,1] \subseteq \mathbb{R}$. We set $\mathcal{O}_\varepsilon := (-\varepsilon, 1+\varepsilon)$ and then get*

$$\mathcal{C} := \{\mathcal{O}_\varepsilon \ : \ \varepsilon \in (0,1]\}$$

as an open cover. Clearly, every $\mathcal{O}_{\tilde{\varepsilon}}$ is a finite sub-cover of $\mathcal{C}$ if $\tilde{\varepsilon} \in (0,1]$.

Figure 2.8: A graphical interpretation of Example 2.7.

Definition 2.8 (Compact set)**.**
A set $\Omega \subseteq \mathbb{R}^d$ is called ***compact*** *if and only if every open cover $\{\mathcal{O}_\alpha \ : \alpha \in I\}$ of Ω admits a finite sub-cover.*

Remark 2.6. *One also says that $\Omega \subseteq \mathbb{R}^d$ is compact if and only if it has the Heine–Borel property. The Heine–Borel property is that any open cover of Ω admits a finite sub-cover.*

Example 2.8. *First we consider a couple of non-examples and two simple first examples.*

1. *The set $(0, 1)$ is not compact as the cover $\{(\varepsilon, 1) : \varepsilon > 0\}$ does not admit a finite sub-cover. Notice that $(\varepsilon_2, 1) \supset (\varepsilon_1, 1)$ if $0 < \varepsilon_2 < \varepsilon_1$ but never $(0, 1) \subseteq (\varepsilon, 1)$ however small one chooses $\varepsilon > 0$.*

2. *$\mathbb{R}$ is not compact as the open cover $\{(-R, R) : R \in \mathbb{R}_{>0}\}$,*
$$\mathbb{R} = \bigcup_{R>0} (-R, R),$$
does not admit a finite sub-cover.

3. *$\mathbb{R}^d$ is not compact as the open cover $\{(-R, R)^d : R \in \mathbb{R}_{>0}\}$,*
$$\mathbb{R}^d = \bigcup_{R>0} (-R, R)^d$$
does not admit a finite sub-cover.

4. *The set $\varnothing$ is compact.*

5. *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 2.2.*

Theorem 2.4.

Let $a, b \in \mathbb{R}$ with $a \leq b$. Then, the closed interval

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

is compact.

How do we prove that?

Proof.

We 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

$$[a, b] = [a, c_1] \cup [c_1, b], \quad c_1 = \frac{a+b}{2}.$$

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.

We call the non-coverable interval I_1 and repeat the bisection and the same argument.

We then obtain a sequence of intervals with

The length of I_n is given by $|b - a|2^{-n}$.

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_n) by picking an $x_n \in I_n$.
By construction, for all $m, n \geq n_0$, we have x_n, $x_m \in I_{n_0}$. Hence, we have

Thus, the sequence $(x_n) \subseteq [a, b]$ is a Cauchy sequence and, by a result of Analysis 1, there exists an $x \in [a, b]$ such that $(x_n) \to x$.

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 $\mathcal{O}_\alpha$ is open there exists an $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq \mathcal{O}_\alpha$.
Since $(x_n) \to x$, we can choose $n_0 \in \mathbb{N}$ such that

for all $n \geq n_0$.

Then, $y \in I_{n_0}$ implies

This means $I_{n_0} \subseteq B_\varepsilon(x) \subseteq \mathcal{O}_\alpha$ and, hence, I_{n_0} is covered by a single element of $\{\mathcal{O}_\alpha : \alpha \in I\}$.
This contradicts our assumption since
This concludes the proof. $\square$

Using the same technique, one can show

Theorem 2.5.

Let $n \in \mathbb{N}$ and $a_i \leq b_i$ real numbers for $i = 1, \ldots, n$. Then, the cuboid

$$[a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_d, b_d] = \bigtimes_{i=1}^{d} [a_i, b_i] \subseteq \mathbb{R}^d$$

is compact.

2.4.1 Properties of compact sets

Theorem 2.6 (Compact sets are closed)**.**

Let $K \subseteq \mathbb{R}^d$ be compact. Then, K is closed, i.e. contains all its limit points.

Let us think a moment how to prove this result:

Proof.

Let $x \in \mathbb{R}^d$ be a limit point of K with $x \notin K$.

We show that K can then not be compact.

We do that by constructing an open cover of K which has no finite sub-cover.

Let

$$\begin{aligned} \mathcal{O}_\varepsilon &:= (\overline{B}_\varepsilon(x))^c = \{y \in \mathbb{R}^d : d(y,x) > \varepsilon\} \\ &= \mathbb{R}^d \setminus \{y \in \mathbb{R}^d : d(y,x) \leq \varepsilon\}. \end{aligned}$$

The sets $\mathcal{O}_\varepsilon$ are open (see Proposition 2.2) and $\{\mathcal{O}_\varepsilon : \varepsilon \in (0, +\infty)\}$ covers K since

$$\mathbb{R}^d \setminus \{x\} = \bigcup_{\varepsilon > 0} \mathcal{O}_\varepsilon.$$

Since x is a limit point of K, by Proposition 2.1, we have that for every $\varepsilon > 0$, the set $B_\varepsilon(x)$ contains infinitely many points of K.

Thus, $\{\mathcal{O}_\varepsilon : \varepsilon \in (0, +\infty)\}$ does not admit a finite sub-cover.

This concludes the proof. $\square$

Theorem 2.7 (Compact sets are bounded)**.**
Let $K \subseteq \mathbb{R}^d$ be compact. Then, K is a bounded set, i.e. there exists an $R > 0$ such that $K \subseteq B_R(0)$. [a]

[a] Clearly, if a set $A \subset \mathbb{R}^d$ satisfies $A \subseteq B_R(0)$ for some $R > 0$, then $\forall a \in A : \|a\|_2 \leq R$.

Proof.

Theorem 2.8.
If $K \subseteq \mathbb{R}^d$ is compact, then every sequence $(x_n) \subseteq K$ has a convergent subsequence with limit in K.

Exercise 2.9. *Prove Theorem 2.8. (Hint: Bolzano–Weierstrass)*

We now prove a property of closed sub-sets of compact sets.

Theorem 2.9.
Let $K \subseteq \mathbb{R}^d$ be compact and $C \subseteq K$ be closed. Then, C is compact.

Remark 2.7. *A mistake that one may make at the beginning is the following: Since K is compact, we have that every open cover $(\mathcal{O}_\alpha)_{\alpha \in I}$ contains a finite sub-cover with $K \subseteq \mathcal{O}_{\alpha_1} \cup \cdots \cup \mathcal{O}_{\alpha_k}$. Since $C \subseteq K$, we have $C \subseteq \mathcal{O}_{\alpha_1} \cup \cdots \cup \mathcal{O}_{\alpha_k}$. This does not prove the claim in Theorem 2.9 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.*

Proof of Theorem 2.9.

The main theorem of this section is the **Heine–Borel Theorem**. Before we state it, some words about its significance[3]: Students sometimes struggle with the Heine–Borel Theorem; the authors certainly did the first time it was presented to them. This theorem can be hard to motivate as the result is subtle and the applications are not obvious. Its uses may appear in different sections of the course textbook and even in different classes. **Students first seeing the theorem must accept that its value will become apparent only in time.** Indeed, the importance of the Heine–Borel Theorem cannot be overstated. It appears in every basic analysis course, and in many point-set topology, probability, and set theory courses. Borel himself wanted to call the theorem the *first fundamental theorem of measure-theory*, a title most would agree is appropriate.

Let us now state and prove the result.

Theorem 2.10 (Heine–Borel[a]).

A set $K \subseteq \mathbb{R}^d$ is compact if and only if it is closed and bounded.

[a]Named after the German mathematician Eduard Heine (1821–1881) and the French mathematician Émile Borel (1871–1956).

Proof of Theorem 2.9.

[3]This is a quote from Nicole R. Andre, Susannah M. Engdahl, and Adam E. Parker's website to be found here. Their work is titled *An Analysis of the First Proofs of the Heine-Borel Theorem* and is worth a read.

Chapter

3 Limits of functions

Let us recall some notation. We consider sets $\Omega \subseteq \mathbb{R}^d$ and functions from Ω to $\mathbb{R}^m$. In symbols, we consider

$$\begin{cases} f : \Omega \to \mathbb{R}^m \\ x \mapsto f(x) = f(x_1, \ldots, x_d) \end{cases}.$$

We recall that Ω is then called the **domain** of f, in symbols

$$\operatorname{dom}(f) = \Omega$$

and $\mathbb{R}^m$ is called the **co-domain**. Further, we define the **image** of f by

$$\operatorname{im}(f) := \{f(x) : x \in \operatorname{dom}(f)\}.$$

In general, we have $\operatorname{im}(f) \subseteq \mathbb{R}^m$.

If $f : \Omega \to \mathbb{R}^m$, that means that for every $x \in \Omega$, the value $f(x)$ belongs to $\mathbb{R}^m$. Now let us use some Linear Algebra. Let $\{e_i : i = 1, \ldots, m\}$ be the standard basis of $\mathbb{R}^m$. Then, for every $x \in \Omega$, we can write

$$f(x) = \sum_{i=1}^{m} f_i(x) e_i = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix},$$

where

$$\forall i \in \{1, \ldots, m\} : f_i(x) = \langle f(x), e_i \rangle.$$

How the properties of f and the f_i, $i = 1, \ldots, m$ are connected, will be discussed during this and further chapters of the present notes.

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

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

The co-domain and image is $\mathbb{R}^2$.

Example 3.2. *We consider the function*

$$\begin{cases} f : \mathbb{R}^3 \to \mathbb{R} \\ x \mapsto \|x\|_2 = \sqrt{x_1^2 + x_2^2 + x_3^2} \end{cases}.$$

The co-domain is $\mathbb{R}$ *and* $\mathrm{im}(f) = \{x \in \mathbb{R} : \, x \geq 0\} =: \mathbb{R}_{\geq 0}$.

3.1 Some priming

Let us suppose that we have a $f:(0,1)\to\mathbb{R}$ and would like to define

$$\lim_{x\to x_0} f(x) \tag{3.1.1}$$

for an $x_0\in(0,1)$.

Let is consider a sequence $(x_n)\subseteq(0,1)$ such that $(x_n)\to x_0$.

Then we can consider $(f(x_n))\subseteq\mathbb{R}$.

How could we define (3.1.1) with that?

Is that enough?

Can we rewrite this without sequences?

3.1.1 An abstract point of view

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

Definition 3.1 (Limit of a $\mathbb{R}^m$–valued function)**.**
Let ρ be a metric on $\mathbb{R}^d$, ρ_* be a metric on $\mathbb{R}^m$. Suppose that $\Omega \subseteq \mathbb{R}^d$ and that $x_0 \in \mathbb{R}^d$ is a limit point of Ω. Then, for $f : \Omega \to \mathbb{R}^m$, we write

$$y_0 = \lim_{x \to x_0} f(x)$$

for a $y_0 \in \mathbb{R}^m$ if and only if

$$\begin{aligned} &\forall \varepsilon > 0 \quad \exists \delta = \delta(\varepsilon) > 0 \ \text{s.t.} \\ &\qquad (\forall x \in \Omega,\ 0 < \rho(x, x_0) < \delta) \quad \Rightarrow \quad \rho_*(y_0, f(x)) < \varepsilon. \end{aligned} \tag{3.2.1}$$

Remark 3.1. *If ρ and ρ_* are given by a norm $\|\cdot\|$, i.e. $\rho(x,y) = \|x - y\|$, $\rho_*(x,y) = \|x - y\|_*$, then (9.1.2) is given by*

$$\begin{aligned} &\forall \varepsilon > 0 \quad \exists \delta = \delta(\varepsilon) > 0 \ \text{s.t.} \\ &\qquad (\forall x \in \Omega,\ 0 < \|x - x_0\| < \delta) \quad \Rightarrow \quad \|f(x) - y_0\|_* < \varepsilon. \end{aligned}$$

From now on, we will use the $\|\cdot\|_2$ norm to define both, ρ and ρ_*. Then, (3.2.1) is given by

$$\begin{aligned} &\forall \varepsilon > 0 \quad \exists \delta = \delta(\varepsilon) > 0 \ \text{s.t.} \\ &\qquad \Big(\forall x \in \Omega,\ 0 < \underbrace{\|x - x_0\|_2}_{=\sqrt{\sum_{i=1}^{d} |x_i - x_{0,i}|^2}} < \delta\Big) \quad \Rightarrow \quad \underbrace{\|f(x) - y_0\|_2}_{=\sqrt{\sum_{i=1}^{m} |f_i(x) - y_{0,i}|^2}} < \varepsilon. \end{aligned}$$

Obviously, the $\|\cdot\|_2$ are not the same even though we use the same symbols.

Remark 3.2. *To prove that a function* ***has a limit at a point*** x_0***/is continuous at*** x_0*, we have to show that*

$$\rho(x, x_0) < \delta \quad \Rightarrow \quad \rho_*(y_0, f(x)) < \varepsilon.$$

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

$$\rho_*(y_0, f(x)) \leq C\rho(x, x_0). \tag{3.2.2}$$

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

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad \forall x : \quad \rho(x, x_0) < \delta \quad \Rightarrow \quad \rho_*(y_0, f(x)) < \varepsilon.$$

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 (3.2.2) *but something of the kind*

$$\rho_*(y_0, f(x)) \leq Cg\big(\rho(x, x_0)\big)$$

for a 'well-behaved' (monotone) function g*. Then one can have* $\rho(x, x_0) < \delta$ *implies* $\rho_*(y_0, f(x)) < \varepsilon$ *if* $\delta = g^{-1}\big(\frac{\varepsilon}{C}\big)$ *. If* $f(x_0)$ *is not defined, as it happens for limits sometimes, one would ask*

$$0 < \rho(x, x_0) < \delta \quad \Rightarrow \quad \rho_*(y_0, f(x)) < \varepsilon$$

but the strategies are exactly the same.

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

3.3 Examples

Example 3.3. *Consider* $f : \mathbb{R} \to \mathbb{R}$ *given by* $f(x) = 2x + 1$. *For* $\mathbb{R}$, *we have* $\rho(x, y) = |x - y|$ *as we learned in Analysis 1.*

We show that $\lim_{x\to 1} f(x) = 3$.

We first think about what we have to prove and then do it. We keep in mind Remark 3.2.

- ***What do we have to prove?***

- ***Let us get at it then:***

Example 3.4. *Consider* $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = x_1 + x_2$.

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

Again, we first think about what we have to prove and then do it. We keep in mind Remark 3.2.

- ***What do we have to prove?***

- ***Let get at it then:***

Example 3.5. *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 $\lim_{x \to 0} f(x) = 0$.

Again, we first think about what we have to prove and then do it. We keep in mind Remark 3.2.

- ***What do we have to prove?***

- ***Let get at it then:***

Now an example that is a bit more complicated.

Example 3.6. *Consider* $f : \mathbb{R}^3 \setminus \{0\} \to \mathbb{R}^2$ *given by*

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

We show that $\lim_{x \to 0} f(x) = \left[0\,,\,1\right]$.

Again, we first think about what we have to prove and then do it. We keep in mind Remark 3.2.

- ***What do we have to prove?***

- ***Let get at it then:***

Remark 3.3. *The last example illustrates the fact mentioned in Remark 3.2 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 Characterization of limits via sequences

Let us think a moment about a function $f : (0,1) \to \mathbb{R}$. How could we define the limit of the function f at a point $x_0 \in (0,1)$ (or a limit point x_0 of $(0,1)$) if we wanted to use sequences?

Thus, we are inspired to the following

Theorem 3.1 (Sequence characterization of limits of $\mathbb{R}^m$-valued functions)**.** *Suppose $\Omega \subseteq \mathbb{R}^d$ and $x_0 \in \mathbb{R}^d$ be a limit point of Ω and let $f : \Omega \to \mathbb{R}^m$. Then*

$$\lim_{x \to x_0} f(x) = y_0$$

for an $y_0 \in \mathbb{R}^m$ if and only if

$$\lim_{n \to +\infty} f(x_n) = y_0$$

for every sequence $(x_n) \subseteq \Omega \setminus \{x_0\}$ with $(x_n) \to x_0$.

Remark 3.4. *Remember that $x_0 \in \mathbb{R}^d$, x_0 limit point of Ω means that x_0 does not necessarily belong to Ω. See also Definition 3.1. For instance, if one has $f : (0,1) \to \mathbb{R}$, one can investigate the limit of f at $x_0 = 1$ and $x_0 = 0$ even if the function is not defined there.*
Further $(x_n) \subseteq \Omega \setminus \{x_0\}$ grantees that $0 < \|x_n - x_0\|_2$ for all $n \in \mathbb{N}$. (See the proof.)

Remark 3.5. *The requirement that $f(x_n) \to y_0$ for **all** sequences $(x_n) \subseteq \Omega \setminus \{x_0\}$, $(x_n) \to x_0$ is not superfluous.*
Think about how it could happen that there are two sequences (x_n),$(\tilde{x}_n)$ as above for which $f(x_n) \to y_0$ but $f(\tilde{x}_n) \not\to y_0$.

Can you come up with an example?

Here is an example from me:

Proof of Theorem 3.1.

In the spirit of the remarks made in Remark 3.5, 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.

Example 3.7. *Let* $f : \mathbb{R}^2 \setminus \{0\} \to \mathbb{R}$ *given by*

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

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

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

does not exist.

3.5 Arithmetic rules for limits

The next theorem follows immediately from the arithmetic rules for sequences that we have proved in Analysis 1 if combined with Theorem 3.1. We define the functions fg, λf, and $f+g$ pointwise, i.e.

$$\begin{cases} f+g : \Omega \to \mathbb{R}^m \\ x \mapsto f(x)+g(x) \end{cases}, \quad \begin{cases} \lambda f : \Omega \to \mathbb{R}^m \\ x \mapsto \lambda f(x) \end{cases}, \text{ and } \begin{cases} fg : \Omega \to \mathbb{R} \\ x \mapsto f(x)g(x) \end{cases}$$

Theorem 3.2 (Arithmetic rules for limits)**.**
Let $\Omega \subseteq \mathbb{R}^d$, $x_0 \in \mathbb{R}^d$ be a limit point of Ω and

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

Suppose

$$\lim_{x\to x_0} f(x) = y_0, \quad \lim_{x\to x_0} g(x) = y_1$$

for $y_0, y_1 \in \mathbb{R}$. Then

1. $\lim_{x\to x_0} (\lambda f)(x) = \lambda y_0$, *for all* $\lambda \in \mathbb{R}$,
2. $\lim_{x\to x_0} (f+g)(x) = y_0 + y_1$,
3. $\lim_{x\to x_0} (fg)(x) = y_0 y_1$, *and*
4. $\lim_{x\to x_0} \left(\frac{f}{g}\right)(x) = \frac{y_0}{y_1}$, *provided* $y_1 \neq 0$.

Exercise 3.2. *Try to find out what results hold if f and g are $\mathbb{R}^m$-valued for $m \geq 1$. Can you prove them? (Hint: Remember Problem Sheet 1 where we have proved a result that might help. The Mock Class Test and the Class Test also contained useful results for this exercise.)*

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

Reading 3. *The entire Section 3.6 is this week's reading. If you have questions, write them down, discuss them with friends, or ask your tutor, me, or the MLSC staff.*

Definition 3.2 (Right-sided limit of f)**.**
Let $I \subseteq \mathbb{R}$, $f : I \to \mathbb{R}$, and $x_0 \in \mathbb{R}$ limit point of I. Then, we write $f(x_0+) = y_0$ for a $y_0 \in \mathbb{R}$ if and only if

$$\lim_{n\to+\infty} f(x_n) = y_0$$

for any sequence $(x_n) \subseteq \mathbb{R}$ with $x_n \in I$ and $x_n > x_0$ for all $n \in \mathbb{N}_0$ and $(x_n) \to x_0$.

In the same way, we define

Definition 3.3 (Left-sided limit of f)**.**
Let $I \subseteq \mathbb{R}$, $f : I \to \mathbb{R}$, and $x_0 \in \mathbb{R}$ limit point of I. Then, we write $f(x_0-) = y_0$ for a $y_0 \in \mathbb{R}$ if and only if

$$\lim_{n\to+\infty} f(x_n) = y_0$$

for any sequence $(x_n) \subseteq \mathbb{R}$ with $x_n \in I$ and $x_n < x_0$ for all $n \in \mathbb{N}_0$ and $(x_n) \to x_0$.

We may also write

$$\begin{aligned} f(x_0+) &= \lim_{x\to x_0+} f(x) = \lim_{x\to x_0+0} f(x), \\ f(x_0-) &= \lim_{x\to x_0-} f(x) = \lim_{x\to x_0-0} f(x). \end{aligned}$$

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 $x_0 \in [a, b)$ iff*

$$\exists y_0 \in \mathbb{R} \text{ s.t. } \forall \varepsilon > 0 \ \exists \delta > 0 : \quad 0 < h < \delta \quad \Rightarrow \quad |f(x_0 + h) - y_0| < \varepsilon.$$

Similarly, one rephrases Definition 3.3.

We state the following result without proof.

Theorem 3.3 (Limits by one-sided limits)**.**
Let $I \subseteq \mathbb{R}$, $f : I \to \mathbb{R}$, and $x_0 \in \mathbb{R}$ be a limit point of I. Then,

$$\lim_{x \to x_0} f(x) = y_0$$

for a $y_0 \in \mathbb{R}$ if and only if

$$\lim_{x \to x_0 -} f(x) = y_0 \quad \text{and} \quad \lim_{x \to x_0 +} f(x) = y_0.$$

Example 3.8. *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 3.9. *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

Often, students have heard about continuity in school and in other modules. Sometimes that means that students think they know what continuity is and fail to pay their utmost attention while I am explaining. The result is often that students still think they know what continuity is but that I am bad at explaining it because they do not understand what I am saying. This is the wrong approach. Continuity is, once properly understood, not the most difficult concept but it is also far from being trivial and far from the 'school concept' of not taking pens from paper while graphing. Mathematicians grappled centuries with its precise formulation.

Be open to challenge your beliefs and pay your utmost attention. One reason for the guided notes is to give you the time to engage in thinking about the contents of the lecture while the lecture unfolds.

4.1 We start by thinking about non-continuous functions

Let us think about the following situation. Consider $f : [-1, 1] \to \mathbb{R}$ given by

$$f(x) = \begin{cases} c_1 & : \quad x < 0 \\ c_2 & : \quad x \geq 0 \end{cases}$$

where c_1 and c_2 are arbitrary real numbers with $c_1 \neq c_2$. To understand that function, let us make a sketch.

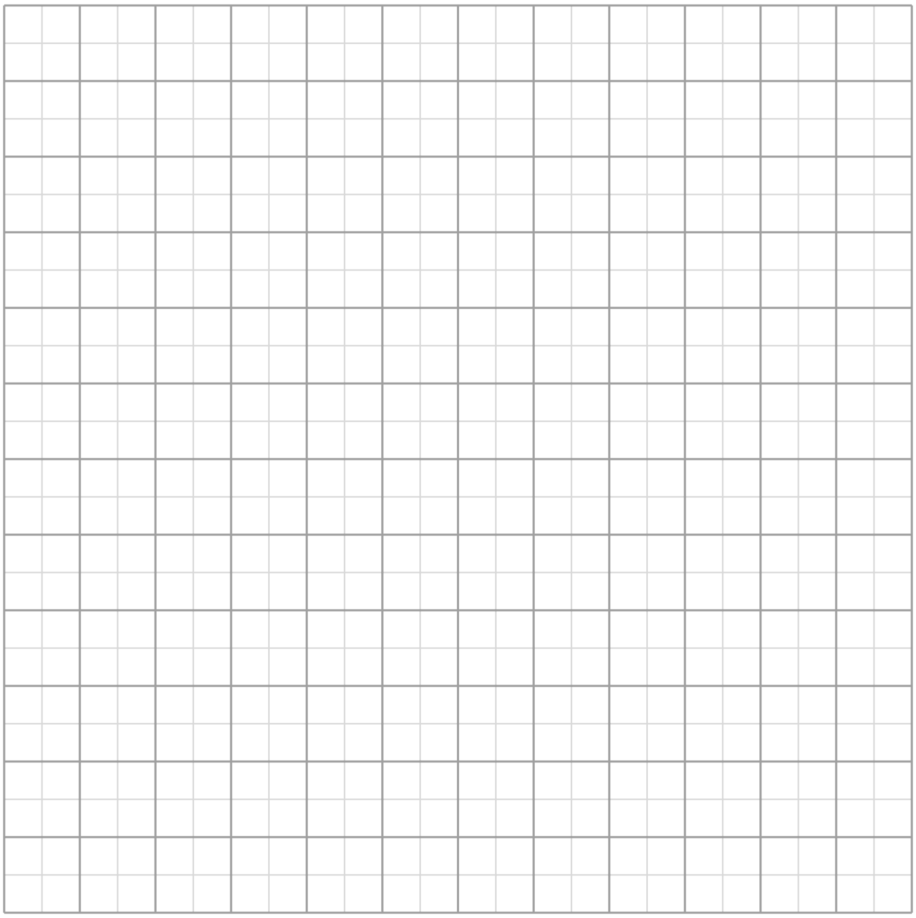


Figure 4.1: Sketch of the function f.

We say that the function has a jump at $x = 0$. Be careful with your intuition though, the function

$$\begin{cases} f : [-1,1] \setminus \{0\} \to \mathbb{R} \\ x \mapsto \begin{cases} c_1 & : \; x < 0 \\ c_2 & : \; x > 0 \end{cases} \end{cases}$$

does not have a jump. Once we stated the definition of continuity, you should be able to see that the function is actually continuous on its domain.

How could we formalise what it means to have a jump?

What else could is described by this definition?

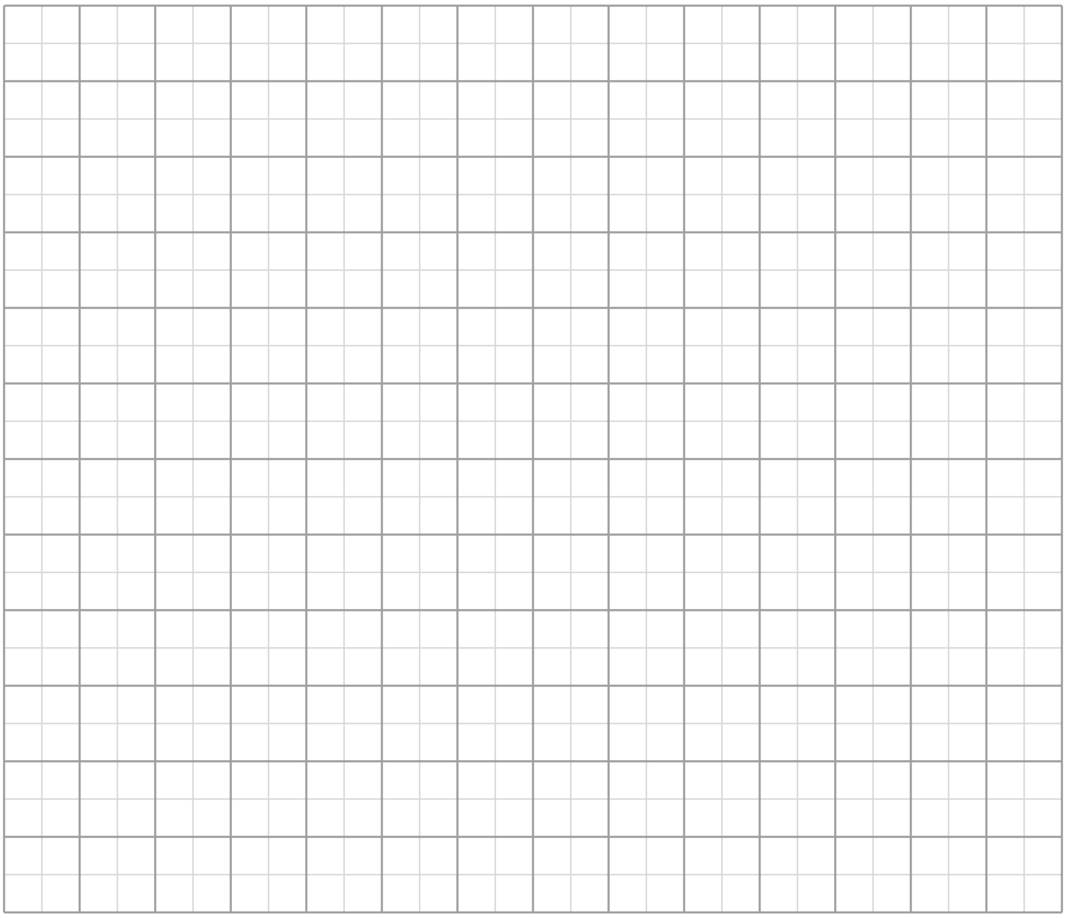


Figure 4.2: Another kind of discontinuity.

4.2 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[a]). *Let ρ and ρ_* be metrics on [b] on $\mathbb{R}^d$ and $\mathbb{R}^m$ respectively[c] and $\Omega \subseteq \mathbb{R}^d$ with $x_0 \in \Omega$. Then, $f : \Omega \to \mathbb{R}^m$ is called **continuous at** x_0 if and only if*

$$\begin{aligned} &\forall \varepsilon > 0 \quad \exists \delta = \delta(\varepsilon) > 0 \text{ s.t.} \\ &\qquad \big(x \in \Omega : \ \rho(x, x_0) < \delta\big) \quad \Rightarrow \quad \rho_*(f(x), f(x_0)) < \varepsilon. \end{aligned} \tag{4.2.1}$$

[a] This definition of continuity is called the (ε, δ)-definition of continuity.
[b] See Definition 1.9.
[c] Also note that the metrics ρ and ρ_* are generic and not necessary the ones on the basis of $\|\cdot\|_p$ as in Example 1.5.

Let us rewrite the (ε, δ)-definition in plain English:

Remark 4.1. *If ρ and ρ_* are given by norms $\|\cdot\|$ and $\|\cdot\|_*$, i.e. $\rho(x, y) = \|x - y\|$, $\rho_*(x, y) = \|x - y\|_*$. Then (9.1.2) is given by*

$$\begin{aligned} &\forall \varepsilon > 0 \quad \exists \delta = \delta(\varepsilon) > 0 \text{ s.t.} \\ &\qquad (\forall x \in \Omega, \ \|x - x_0\| < \delta) \quad \Rightarrow \quad \|f(x) - f(x_0)\|_* < \varepsilon. \end{aligned}$$

From now on, we will use the $\|\cdot\|_2$ norm to define both, ρ and ρ_*. Thus, (9.1.2) is given by

$$\begin{aligned} &\forall \varepsilon > 0 \quad \exists \delta = \delta(\varepsilon) > 0 \text{ s.t.} \\ &\qquad (\forall x \in \Omega, \ \|x - x_0\|_2 < \delta) \quad \Rightarrow \quad \|f(x) - f(x_0)\|_2 < \varepsilon. \end{aligned}$$

Obviously, the $\|\cdot\|_2$ *are not the same even though we use the same symbols:*

$$\begin{aligned} \|x - x_0\|_2 &= \sqrt{\sum_{i=1}^{d} |x_i - x_{0,i}|^2}, \\ \|f(x) - f(x_0)\|_2 &= \sqrt{\sum_{i=1}^{m} |f(x)_i - f(x_0)_i|^2}. \end{aligned}$$

Here, we denoted

$$x_0 = \left[x_{0,1}, \ldots, x_{0,d}\right]^T$$

and the symbols $f(x)_i$, $f(x_0)_i$ *denote the* i^{th} *component of the vectors* $f(x)$ *and* $f(x_0)$ *respectively. With the notation introduced in Chapter 3, we can write* $f(x)_i = f_i(x)$ *and* $f(x_0)_i = f_i(x_0)$.

Remark 4.2. *In the definition of convergence, see Definition 1.11, the* n_0 *is a function of* ε. *Here, the* δ *is, in general, a function of* ε *(and* x_0*).*

Remark 4.3. *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 x_0 where $f(x_0)$ does not exist. For example, the function*

$$f(x) = \begin{cases} -1 & : \quad x < 0 \\ 1 & : \quad x > 0 \end{cases} \tag{4.2.2}$$

is continuous on $\mathbb{R} \setminus \{0\}$, therefore, it makes no sense to ask for continuity at $x_0 = 0$. However, one can ask whether the limits of $f(x)$ for $x \to x_0$ exists, i.e. whether

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

exists.

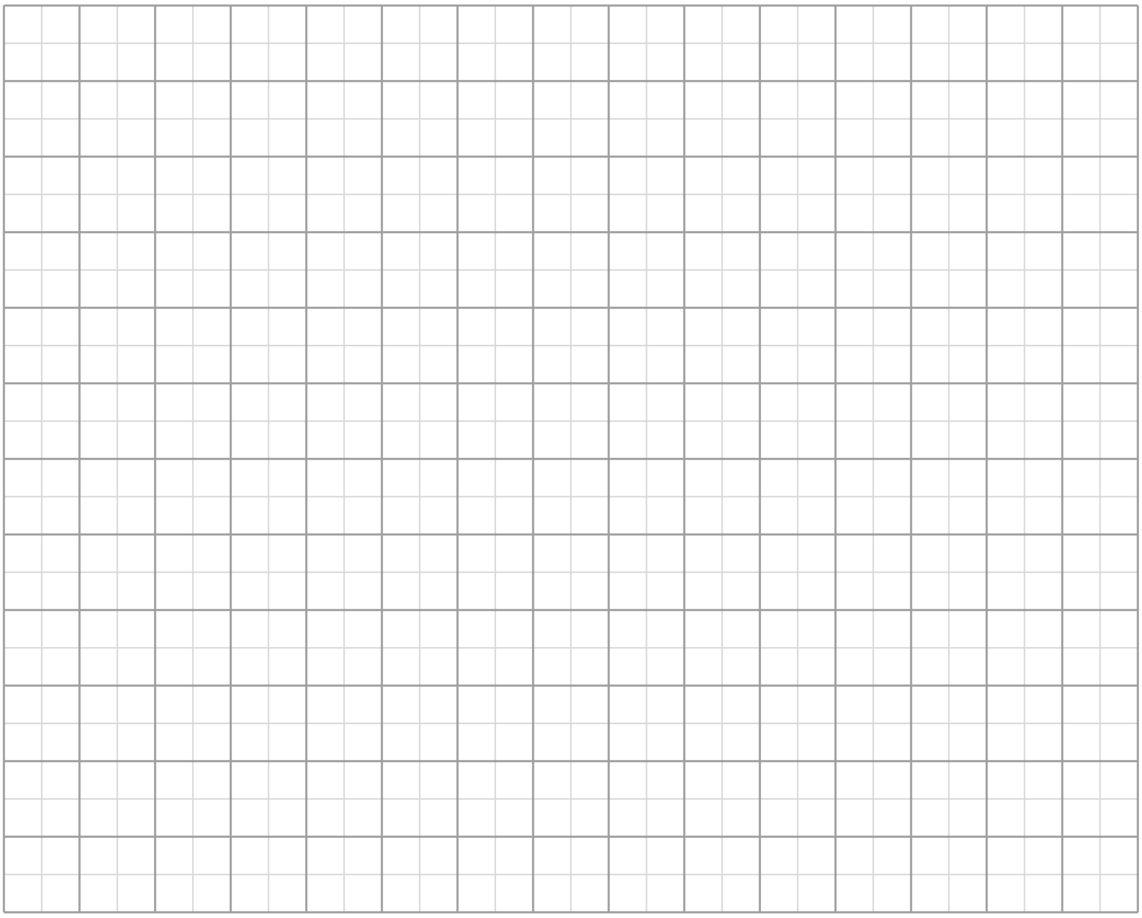


Figure 4.3: Drawing of (4.2.2).

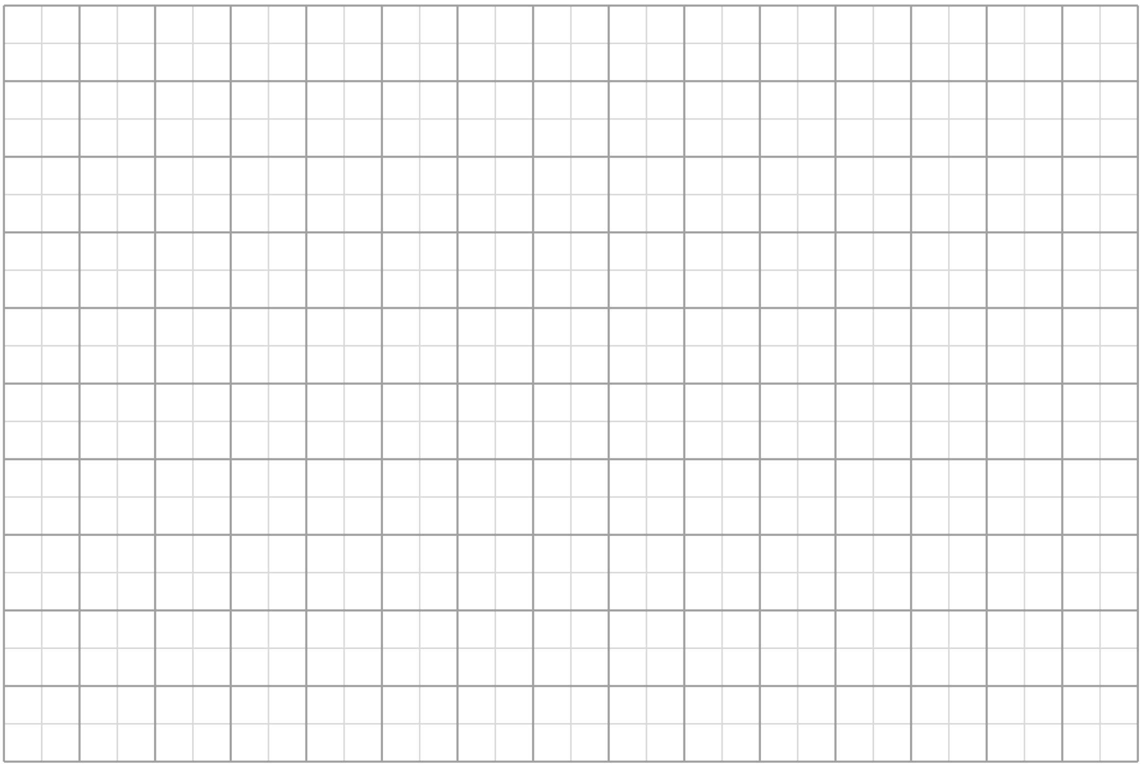


Figure 4.4: Illustration of the definition of continuity for $m = d = 1$.

With the notion introduced in Section 2.1, we can rewrite

$$\rho(x, x_0) < \delta \quad \Rightarrow \quad \rho_*(f(x), f(x_0)) < \varepsilon$$

as

$$x \in B_\delta(x_0) \quad \Rightarrow \quad f(x) \in B_\varepsilon(f(x_0))$$

or

$$f(B_\delta(x_0)) \subseteq B_\varepsilon(f(x_0)). \tag{4.2.3}$$

Here, we use additionally the following notation. Let $f : A \to B$ and $E \subseteq A$. Then, the set $f(E)$ is defined by

$$f(E) := \{f(x) \ : \ x \in E\} \subseteq B.$$

Definition 4.2 (Rephrasing Definition 4.1 for Ω open)**.**
Let $\Omega \subseteq \mathbb{R}$ open, $x_0 \in \Omega$. A function $f : \Omega \to \mathbb{R}^m$ is continuous at x_0 if and only if

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \qquad f\left(B_\delta(x_0)\right) \subseteq B_\varepsilon(f(x_0)).$$

Remark 4.4. *This definition is, as the title says just as rephrasing. It does not 'replace' Definition 4.1. When asked for the definition of continuity, you give Definition 4.1 adapted to the proper context, i.e. with the right metrics.*

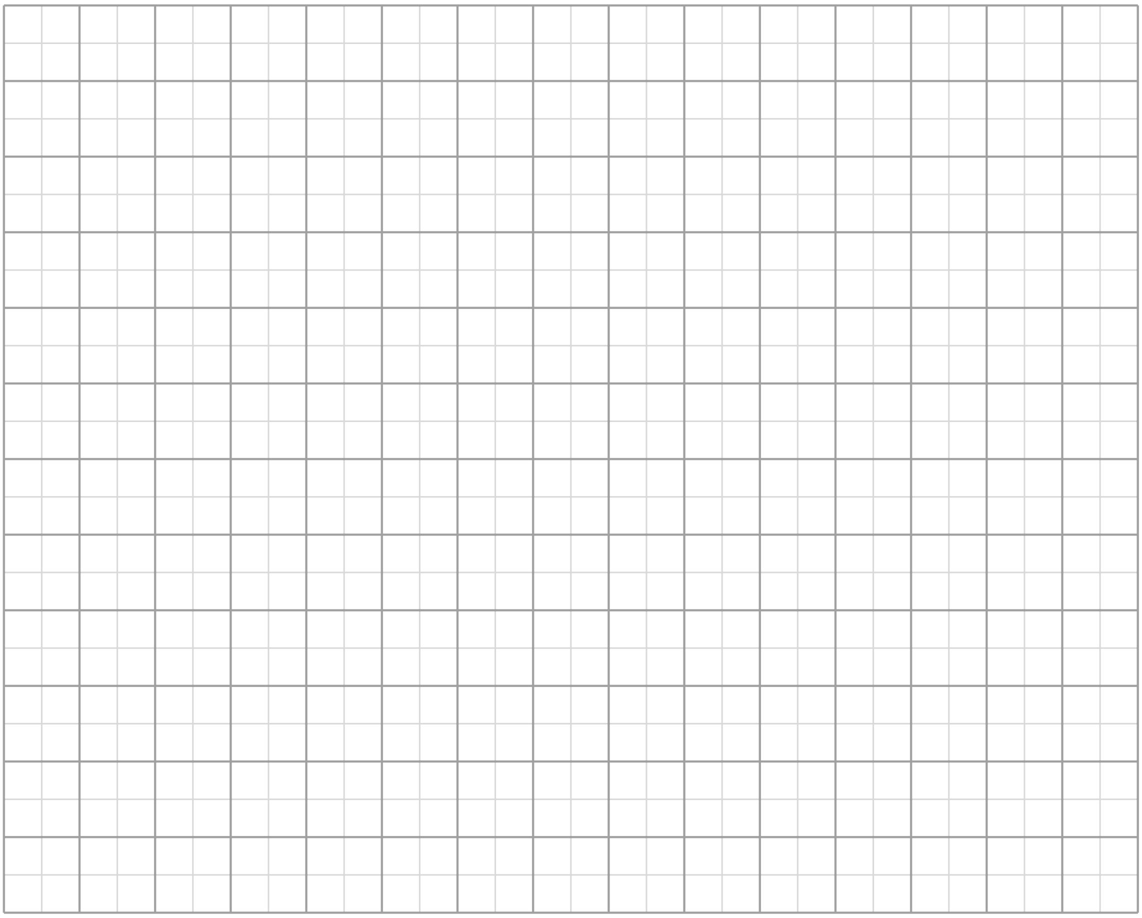


Figure 4.5: Illustration of continuity. Especially (4.2.3).

Definition 4.3 (Continuity of $\mathbb{R}^m$ valued functions)**.**
Let $\Omega \subseteq \mathbb{R}^d$. Then, a function $f : \Omega \to \mathbb{R}^m$ is said to be ***continuous (on Ω)*** *if and only if it is continuous at all $x_0 \in \Omega$.*[a]

[a] See Definition 4.1.

Definition 4.4 (Definition 4.3 as (ε, δ)-criterion w.r.t. $\|\cdot\|_2$)**.**
Let $\Omega \subseteq \mathbb{R}^d$. Then, a function $f : \Omega \to \mathbb{R}^m$ is said to be ***continuous (on Ω)*** *if and only if*

$$\forall x_0 \in \Omega \quad \forall \varepsilon > 0 \quad \exists \delta = \delta(x_0, \varepsilon) > 0 \text{ s.t.}$$
$$\big(x \in \Omega : \ \|x - x_0\|_2 < \delta\big) \quad \Rightarrow \quad \|f(x) - f(x_0)\|_2 < \varepsilon.$$

Using Definition 3.1 and Theorem 3.1, it holds immediately

Theorem 4.1 (Limit characterization of continuity)**.**
Let $\Omega \subseteq \mathbb{R}^d$, $f : \Omega \to \mathbb{R}^m$, *and* $x_0 \in \Omega$ *limit point of* Ω. *Then,* f *is continuous at* x_0 *if and only if*

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

and if and only if for all $(x_n) \subseteq \Omega \setminus \{x_0\}$ *with* $(x_n) \to x_0$ *holds*

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

Exercise 4.1. *In the case* $d = 1$, *we have that* $f : (-1,1) \to \mathbb{R}$ *is continuous at* $x_0 \in (-1,1)$ *if and only if*

$$\begin{aligned} &\forall \varepsilon > 0 \quad \exists \delta > 0 \text{ s.t.} \\ &\qquad \Big(\forall x \in (-1,1),\, |x - x_0| < \delta\Big) \quad \Rightarrow \quad |f(x) - f(x_0)| < \varepsilon. \end{aligned} \tag{4.2.4}$$

Prove, as stated in Theorem 4.1, that (4.2.4) *is equivalent to*

$$\forall (x_n) \subseteq (-1,1) \setminus \{x_0\},\ (x_n) \to x_0 \quad \Rightarrow \quad \big(f(x_n)\big) \to f(x_0).$$

4.3 Examples

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

- *What do we have to show?*

- *Let us show it then. We take Remark 3.2 into consideration.*

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

Example 4.2. *Consider $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$. We show that this function is continuous at all $x_0 \in \mathbb{R}$.*
We still have to prove that

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \, : \; |x - x_0| < \delta \quad \Rightarrow \quad |f(x) - f(x_0)| < \varepsilon.$$

We reformulate $|x - x_0| < \delta$ by setting $x = x_0 + \delta'$. Then

Thus, we get

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

If we restrict $\delta \leq 1$, we finally have to choose

Example 4.3. *Consider $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 2x^2 - 8x + 6$. Again, let us prove continuity everywhere.*

- *What do we have to prove?*

- *What is the plan?*

- *Let us work it out then.*

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

- *What do we have to show?*

- *Working it out.*

  Let $x_0 = 0$ *and* $\varepsilon > 0$*. We get*

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

  if we choose $\delta = \varepsilon^2$*.*

  Now let $x_0 \in (0, +\infty)$ *and* $\varepsilon > 0$*.*

  We estimate

  Thus, we get

  $$|x - x_0| < \delta \quad \Rightarrow \quad |\sqrt{x} - \sqrt{x_0}| < \varepsilon$$

  for

  $$\delta =$$

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

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

We want to show, that this function is continuous at $x_0 = \begin{bmatrix} 0 & , & 0 & , & 0 \end{bmatrix}^T$.

- *What do we have to show?*

- *Let us work it out then.*

4.4 Discontinuity

Reading 4. *This section is this week's reading. Make sure you read it carefully. Make sketches of the functions discussed. Record your questions and discuss them with friends and/or the MLSC staff. If they remain, you can also discuss them with your tutors and me.*
Prove the claimed properties in the examples below where the details have been left open.

Definition 4.5 (Discontinuous at a point for $\mathbb{R}^m$ valued functions)**.**
Let $\Omega \subseteq \mathbb{R}^d$ and $x_0 \in \Omega$. Then, $f : \Omega \to \mathbb{R}^m$ is called ***discontinuous*** *at x_0 if and only if f is not continuous at x_0.*

Exercise 4.3. *Write the definition of discontinuity explicitly as the negation of the definition of continuity at a point x_0. Write also a 'sequence-version'.*

Definition 4.6 (Discontinuity (explicit version))**.**

Definition 4.7 (Discontinuity (sequence version)).

Example 4.5. *The function* $f : \mathbb{R} \to \mathbb{R}$*, 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 3.6.

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

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

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

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

Example 4.7. *The Dirichlet-function*[1] *(which is the characteristic function*[2] *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 4.4. *Prove that the function* $\chi_{\mathbb{Q}}$ *from the example above is nowhere continuous.*

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.

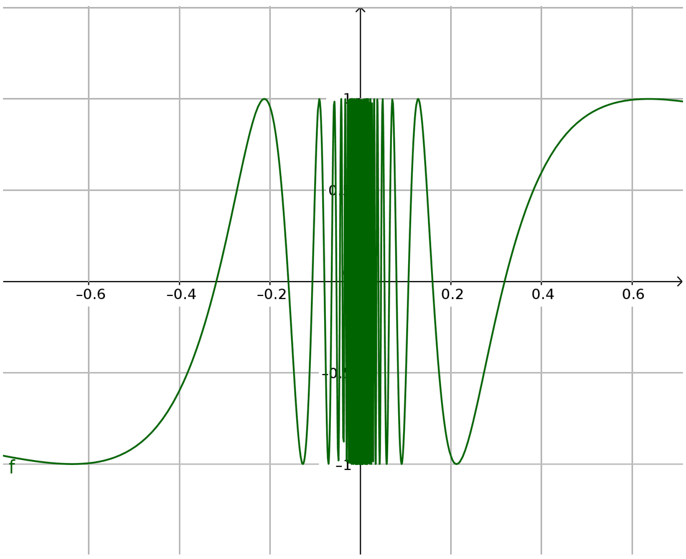


Figure 4.6: The function $f(x) = \sin\left(\frac{1}{x}\right)$.

[1] Johann Peter Gustav Lejeune Dirichlet (1805–1859) is a German mathematician.
[2] Also called indicator function.

4.4.1 Classification of discontinuities

We give some names to different types of discontinuities.

Definition 4.8 (Classification of discontinuities)**.**
Let $I \subseteq \mathbb{R}$, $x_0 \in I$, and $f : I \to \mathbb{R}$. Suppose f is discontinuous at x_0. Then we have the following classification:

*1) If $\lim\limits_{x \to x_0} f(x)$ exists, then f has a **removable discontinuity** at x_0.*

*2) If $\lim\limits_{x \to x_0} f(x)$ does not exist but both $\lim\limits_{x \to x_0+} f(x)$ and $\lim\limits_{x \to x_0-} f(x)$ exist, then f has a **discontinuity of the first kind**, or **jump discontinuity**.*

*3) If at least one of the limits $\lim\limits_{x \to x_0+} f(x)$ or $\lim\limits_{x \to x_0-} f(x)$ do not exist, then f has a **discontinuity of the second kind**.*

Exercise 4.5. *Give examples for all types of discontinuities defined in Definition 4.8. Investigate the examples of this section with respect to this definition.*

4.4.2 Further counterexamples in continuity

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

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

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 $x_1, \ldots, x_n$, then $f(x)\chi_{\mathbb{Q}}(x)$ is continuous at the points $x_1, \ldots, x_n$.

Exercise 4.6. *Find a function that is defined on $\mathbb{R}$ and only continuous at $x_0 = 3$.*

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

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

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*

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

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

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

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

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

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*

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

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 $x_0 \in \mathbb{R}$.

4.5 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

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

We have

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

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

Then, f is continuous at $x_0 \in \Omega$ if and only if the functions $f_k(x)$ are continuous at x_0 for all $k \in \{1, \ldots, m\}$.

Proof.

The proof is an exercise.

You might want to put the main points of reasoning here.

Remark 4.5. *Theorem 4.2 is useful to decide whether a function 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 as one has only to check the the component functions

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

at the components of $f(x_0)$ which is easier.

4.6 Operations with continuous functions

Given two functions $f, g : \Omega \to \mathbb{R}^m$, one can define addition $f + g$, multiplication fg, and multiplication by scalars ($\lambda \in \mathbb{R}$) pointwise by

$$\begin{cases} f + g : \Omega \to \mathbb{R}^m \\ x \mapsto f(x) + g(x) \end{cases}, \begin{cases} \lambda f : \Omega \to \mathbb{R}^m \\ x \mapsto \lambda f(x) \end{cases}, \text{ and } \begin{cases} fg : \Omega \to \mathbb{R} \\ x \mapsto f(x)g(x) \end{cases}$$

In other words, the set of functions from Ω to $\mathbb{R}^m$ is a real vector space.

Theorem 4.3 (Arithmetic rules of continuous functions)**.**
Let $\Omega \subseteq \mathbb{R}^d$, $x_0 \in \Omega$. Consider $f, g : \Omega \to \mathbb{R}$ which are continuous (at x_0). Then

- *$\lambda f + \mu g$ is continuous (at x_0) for all $\mu, \lambda \in \mathbb{R}$,*
- *fg is continuous (at x_0), and*
- *$\dfrac{f}{g}$ is continuous (at x_0) if $g(x_0) \neq 0$.*

Proof of Theorem 4.3.

Using Theorem 3.2, the proof follows whenever $x_0 \in \Omega$ is a limit point of the set Ω.

If $x_0 \in \Omega$ is an isolated points, the result is easy to prove. (Exercise!)

This concludes the proof. □

Exercise 4.7. *Similar to Exercise 3.2, find a suitable statement in the spirit of Theorem 4.3 for functions*

$$f, g : \Omega \to \mathbb{R}^m$$

and prove it.

With that we are able to make the following definition:

Definition 4.9 (The space $C(\Omega, \mathbb{R}^m)$)**.**
Let $\Omega \subseteq \mathbb{R}^d$. Then, the set of all continuous[a] *functions from Ω to $\mathbb{R}^m$ is denoted by $C(\Omega, \mathbb{R}^m)$. If $m = 1$, we write $C(\Omega, \mathbb{R}) = C(\Omega)$.*

[a]See Definition 4.3.

Remark 4.6. *Theorem 3 shows that $C(\Omega, \mathbb{R}^m)$ is a vector space.*
Since we can multiply $\mathbb{R}$-valued continuous functions, $C(\Omega)$ is even an algebra. For further details see your Linear Algebra lecture notes.

Exercise 4.8. *Prove that the functions $f(x) = c$, $c \in \mathbb{R}$ (constant function) and $g(x) = x$ are continuous as functions from $\mathbb{R}$ to $\mathbb{R}$. Then, conclude that polynomials of degree N are continuous. In other words, show that functions*

$$p_N(x) = \sum_{i=1}^{N} a_i x^i,$$

where the coefficients $a_i \in \mathbb{R}$ for all $i = 1, \ldots, N$ are constants, are continuous.

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

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

Further suppose that f *is continuous at* $x_0 \in \Omega_1$ *and* g *is continuous at* $f(x_0) \in \Omega_2$*. Then,* $g \circ f$ *is continuous at* x_0*.*

Figure 4.7: A graphical analysis of the composition theorem Theorem 4.4 and its proof.

Proof of Theorem 4.4.

4.7 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 **Extremal Value Theorem** (EVT) which is also commonly called *Theorem of Weierstrass*[3].

The proofs will reuse some strategies that we have seen in Analysis 1 and will heavily depend on the completeness axiom. It is worthwhile to think about the fact that continuous functions on $[a, b] \cap \mathbb{Q}$ would not satisfy these theorems.

Connected to what you are learning in the module Numbers, you might want to have a look at *Continuity and Irrational Numbers* of Richard Dedekind[4].

[3] Karl Weierstraß (1815–1897), German mathematician.
[4] Richard Dedekind (1831–1916), German mathematician .

4.7.1 Two important properties of continuous functions

To motivate the proof of our first result, we shall first prove first the following result on sequences with strictly positive limits.

Lemma 4.1.
Let $(a_n) \subseteq \mathbb{R}$, $a \in \mathbb{R}$, $a > 0$. *Suppose that* $(a_n) \to a$. *Then there exists* $n_0 \in \mathbb{N}$ *such that* $a_n > 0$ *for all* $n \geq n_0$.

Proof.

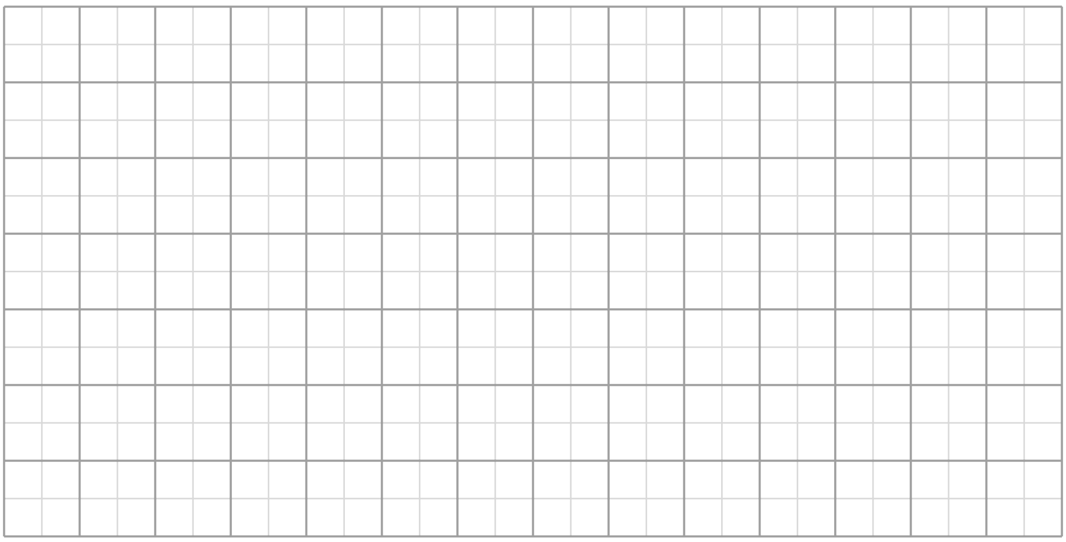


Figure 4.8: Continuous function which is positive at a point x_0.

Now, let us state this quite intuitive insight about continuous functions as a theorem.

Theorem 4.5 (Preservation of sign)**.**
Let $I \subseteq \mathbb{R}$ $f : I \to \mathbb{R}$ *be continuous at* $x_0 \in I$*. If* $f(x_0) > 0$*, then there exists a* $\delta > 0$ *such that*

$$f(x) > 0 \quad \textit{for all} \quad x \in [x_0 - \delta, x_0 + \delta] \cap I.$$

Proof.

Now, we also introduce the notion of a bounded function. This is essentially a recap of results of Chapter 1 of Analysis I (Bounded sets).

Definition 4.10 (Bounded from above).
Let $I \subseteq \mathbb{R}$. Then, a function $f : I \to \mathbb{R}$ is called ***bounded from above*** *if and only if there exists a constant $C \in \mathbb{R}$ such that for all $f(x) \leq C$ for all $x \in I$.*

Definition 4.11 (Bounded from below).
Let $I \subseteq \mathbb{R}$. Then, a function $f : I \to \mathbb{R}$ is called ***bounded from below*** *if and only if $-f$ is bounded from above.*

Definition 4.12 (Bounded).
Let $I \subseteq \mathbb{R}$. Then, a function $f : I \to \mathbb{R}$ is called ***bounded*** *if and only if f is bounded from above and bounded from below.*

Remark 4.7. *Let us set $f(I) := \{f(x) : x \in I\}$. 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.

Now we prove that continuous functions $f : [a, b] \to \mathbb{R}$ are locally bounded.

Theorem 4.6 (Local boundedness).
Let $I \subseteq \mathbb{R}$ and $f : I \to \mathbb{R}$ be continuous at $x_0 \in I$. Then, there exist a constant $C > 0$ and a $\delta > 0$ such that

$$|f(x)| \leq C \quad \text{for all} \quad x \in [x_0 - \delta, x_0 + \delta] \cap I.$$

Proof.

4.7.2 The Intermediate Value Theorem (IVT)

Theorem 4.7 (Intermediate value theorem (IVT))**.**
Let $f : [a, b] \to \mathbb{R}$ be continuous and suppose that $f(a) < 0$ and $f(b) > 0$. Then, there exists an $x_0 \in [a, b]$ such that $f(x_0) = 0$.

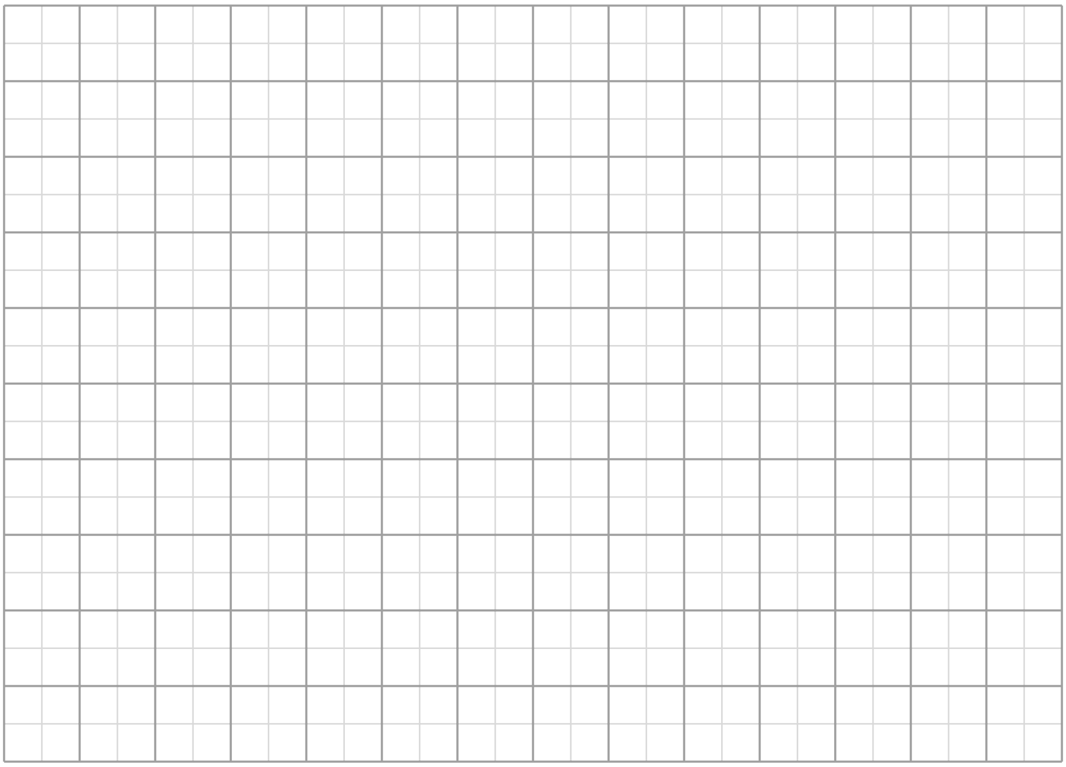


Figure 4.9: Illustration of Theorem 4.7.

Exercise 4.9. *Think about whether the converse of Theorem 4.7 is true. Then study whether the continuity is really needed in general. See also the Review/Problem Sheet of this week and have a look in the book* Counterexamples in Analysis *from the Reading List.*

Remark 4.8. *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*[5]*. If we look at continuous functions on $\mathbb{Q}$, i.e. $f : [a, b] \cap \mathbb{Q} \to \mathbb{R}$, which are continuous, then the IVT is false.*
For example, the function $f(x) = x^2 - 2$ on $[1, 2] \cap \mathbb{Q}$ does not have a zero in $[1, 2] \cap \mathbb{Q}$, i.e. there is no $x \in [1, 2] \cap \mathbb{Q}$ such that $f(x) = 0$ since $\sqrt{2}$ is irrational.

Proof of Theorem 4.7.

[5] Remember that means that all bounded above, non-empty sub-sets of $\mathbb{R}$ have a supremum.

From the IVT, we get immediately

Corollary 4.1 (Reformulated 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 an* $x_0 \in [a, b]$ *such that* $f(x_0) = y_0$.

4.7.3 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.*

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

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

*is **surjective**.*[6]

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

To state another important conclusion, we need the notion of a fixed point.

Definition 4.13 (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 *if and only if* $f(x_0) = x_0$.

Theorem 4.8 (Brouwer's Fixed Point Theorem[a])**.**

Let $f : [a, b] \to [a, b]$ *be continuous. Then there exists a* $x_0 \in [a, b]$ *such that*

$$f(x_0) = x_0.$$

[a]Named after the Dutch mathematician Luitzen Egbertus Jan Brouwer (1881–1966).

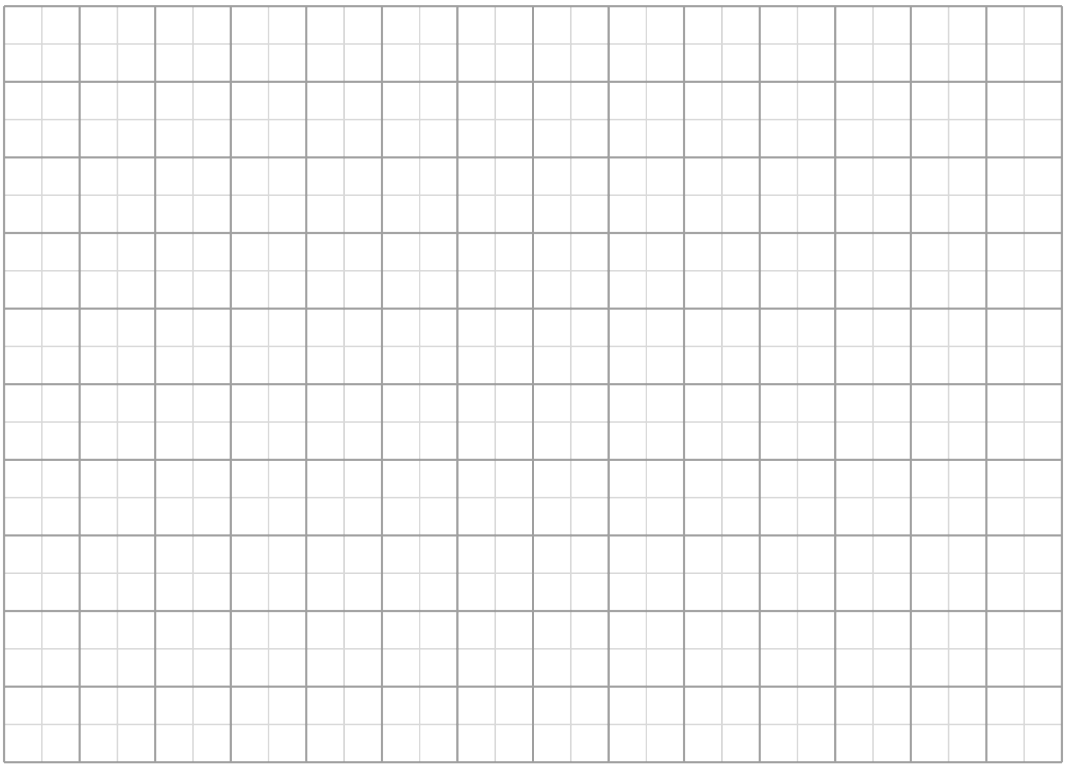


Figure 4.10: Illustration of the Brouwer's Fixed Point Theorem.

Proof of Theorem 4.8.

Decide which of the following statements is true.

- All continuous functions $f : [-1, 1] \to [-1, 1]$ have a fixed point.
- All continuous functions $f : [-2, 3] \to [0, 3]$ have a fixed point.
- All continuous functions $f : [0, 1] \to [0, 3]$ have a fixed point.

4.7.4 Weierstrass' Extremal Value Theorem

We introduce the notions of local and global maximum.

Definition 4.14 (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*

$$\forall x \in (x_0 - \delta, x_0 + \delta) \cap I \;:\; f(x) \leq f(x_0).$$

Exercise 4.10. *Why is the $\cap I$ necessary in the above definition. When could we leave it away?*

Definition 4.15 (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 maximum** if*

$$\forall x \in I \;:\; f(x) \leq f(x_0).$$

Exercise 4.11. *Write down the according definition for a local minimum, Definition 4.16 and a global minimum, Definition 4.17, for a function $f : I \to \mathbb{R}$, where $I \subseteq \mathbb{R}$.*

Definition 4.16 (Local minimum)**.**

Definition 4.17 (Global minimum)**.**

Exercise 4.12. *Draw pictures and illustrate the above defined notions of local and global maximum/minimum. Convince yourself that neither local nor global maxima/minima must be unique.*

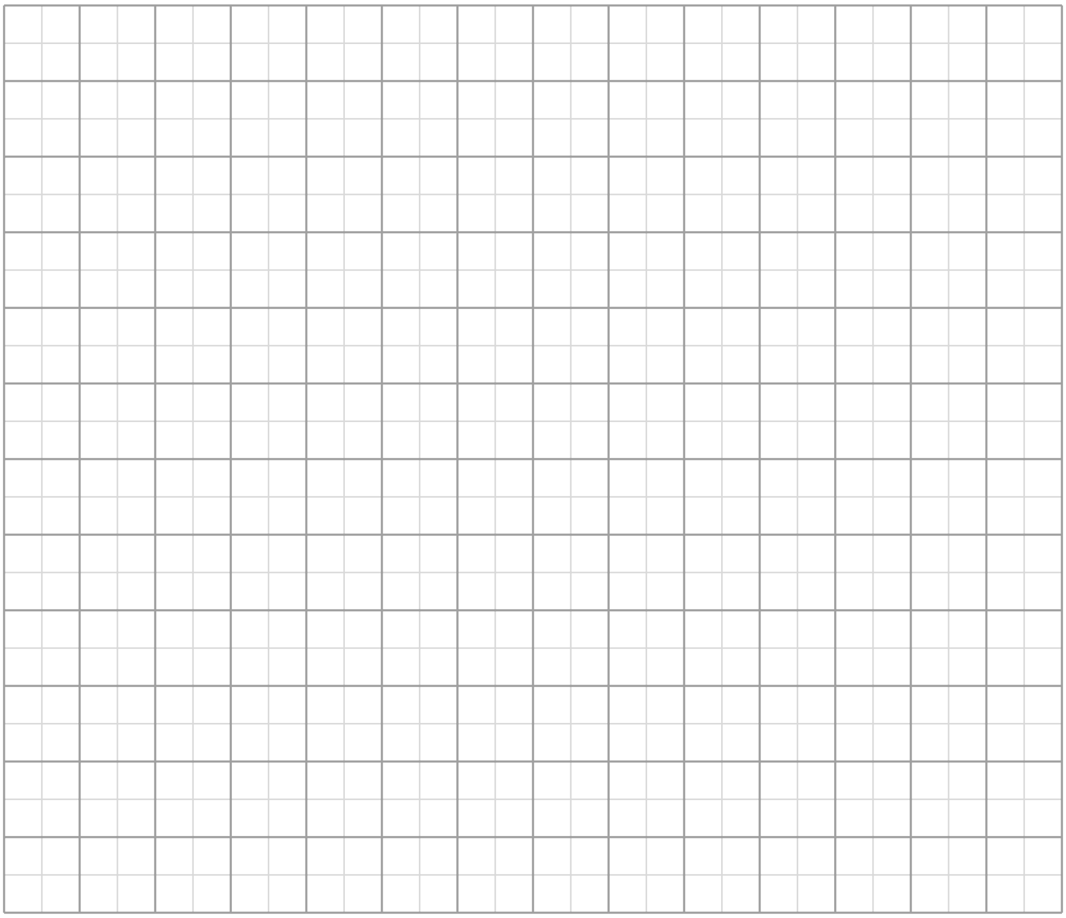


Figure 4.11: Illustration of local/global maxima/minima.

The next important theorem is Weierstrass' Extremal Value Theorem.

Theorem 4.9 ((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 exist* x_1, $x_2 \in [a, b]$ *such that*

$$f(x_1) = \sup_{x \in [a,b]} f(x), \quad f(x_2) = \inf_{x \in [a,b]} f(x).$$

Exercise 4.13. *As for the IVT, think about whether the converse of the EVT is true. Further, explore the necessity of the continuity. See also the Review/Problem Sheet of this week and have a look in the book* Counterexamples in Analysis *from the Reading List.*

Proof of Theorem 4.9.

First, we prove (i).

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

$$\forall x \in [a, b] \ : \ |g(x)| \leq M.$$

This means, by the definition of g that

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

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

Remark 4.9. *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 4.10. *The continuity of the function f in the statement of the EVT is essential. For instance, consider $f:[0,1]\to\mathbb{R}$ with*

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

where $\frac{m}{n}$ in the definition of the function is regarded to be in lowest terms. In every neighbourhood 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 4.11. *Consider $f:[0,1]\to\mathbb{R}$ given by*

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

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 4.9. 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 $I\subseteq[0,1]$.

4.8 Continuity of linear maps

Reading 5. *This section is this week's reading. Please take a stack of paper and a pen and go through all calculations by hand. They are not very difficult but are best understood if performed by you and not just read through.*

We convince ourselves that a linear map $L : \mathbb{R}^d \to \mathbb{R}^m$ is continuous. For the definition of Linear Map see Definition A.15. Since $Lx - Lx_0 = L(x - x_0)$, it is sufficient to prove that L is continuous at $x_0 = 0$. If the matrix L is the 0 matrix, then the continuity is obvious.

Let us denote

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

With that, we obtain

$$Lh = \begin{bmatrix} l_{11} & \dots & l_{1d} \\ \vdots & & \vdots \\ l_{m1} & \dots & l_{md} \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_d \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{d} l_{1j} h_j \\ \vdots \\ \sum_{j=1}^{d} l_{mj} h_j \end{bmatrix}.$$

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

$$h \in \mathbb{R}^d,\ \|h\|_2 < \delta \quad \Rightarrow \quad \|Lh\|_2 < \varepsilon. \tag{4.8.1}$$

We have

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

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

$$\left| \sum_{j=1}^{d} l_{ij} h_j \right| \ \leq \ \sum_{j=1}^{d} |l_{ij}| |h_j|$$

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

$$|l_{ij}| \leq \max_{\substack{i=1,\dots,m \\ j=1,\dots,d}} |l_{ij}| =: \mathcal{L}.$$

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

$$\left|\sum_{j=1}^{d} l_{ij}h_j\right| \leq \mathcal{L}\sum_{j=1}^{d}|h_j|.$$

Hence, we obtain

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

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

$$\|Lh\|_2 \leq C_L\|h\|_2$$

and then (4.8.1) by choosing $\delta = \frac{\varepsilon}{C_L}$. The constant C_L is not zero as the matrix is not the zero matrix. The argument is complete. $\square$

Chapter

5 Differentiability and derivative on $\mathbb{R}^1$

5.1 Historical comments on Differential Calculus

The history of differential (and integral) calculus is quite long and convoluted. Everyone knows that Leibniz (1646–1716) and Newton (1643–1727) are the main inventors of calculus in the 17. century. However, elements of it can be traced back as far as Archimedes (ca. 287–212 BC) and Brahmagupta (ca. 598–665).

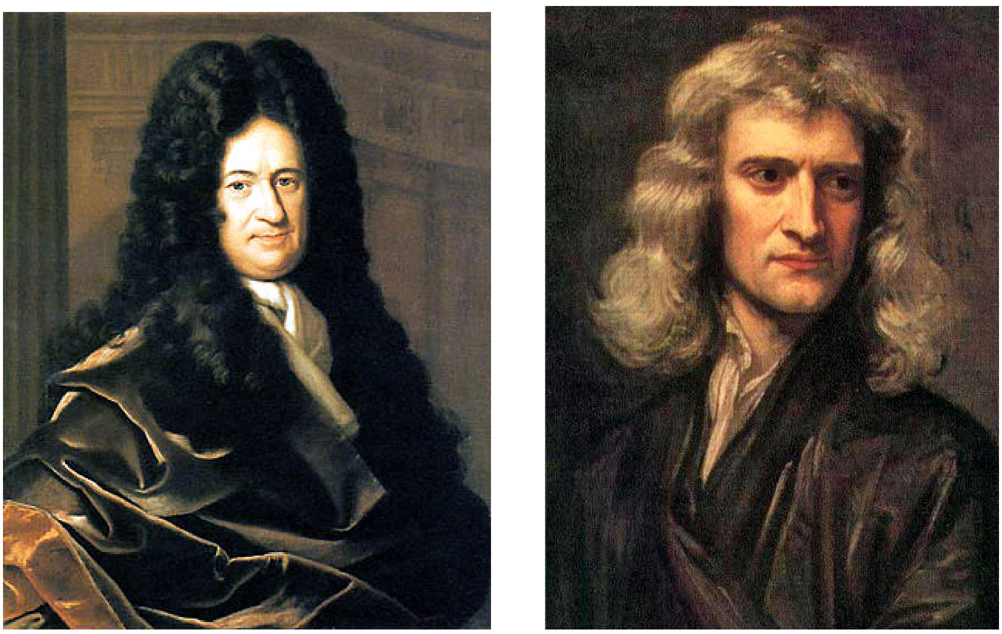


Table 5.1: Gottfried Wilhelm Leibniz and Isaac Newton

The ideas of calculus did not come out of the blue, contributions were made by Descartes (1596–1650), Fermat (1607–1665), Pascal (1623–1662) and others. The foundations of calculus, the infinitesimals, where mysterious at the time and had many critics. Success in answering otherwise much harder to answer questions gave the mathematicians following Leibniz and Newton, foremost the Bernoullis (Jacob 1654–1705, Johann 1667–1748), Euler (1707–1783) and Gauss (1777–1855), the confidence in the methods even with the shaky grounds; there was also some success to make the methods more rigorous step by step. A complete solid foundation was not laid until the work of Cauchy (1789–1857) and Weierstraß (1815–1897) who defined limits arithmetically and by that laid the discussions about the ghosts of departed quantities at rest. Their approach is what we will develop in this chapter.

Table 5.2: Augustin-Louis Cauchy and Karl Weierstraß

See also [14, 3, 6] and the references therein.

An interesting lecture, Titled *Ghosts of Departed Quantities: Calculus and its Limits* from Prof. Raymond Flood from Gresham College at the topic can be found here.

A modern treatment following the ideas closer to Leibniz original intuition than the modern one build on limits is the so-called Non-standard Analysis.

5.2 Some thinking

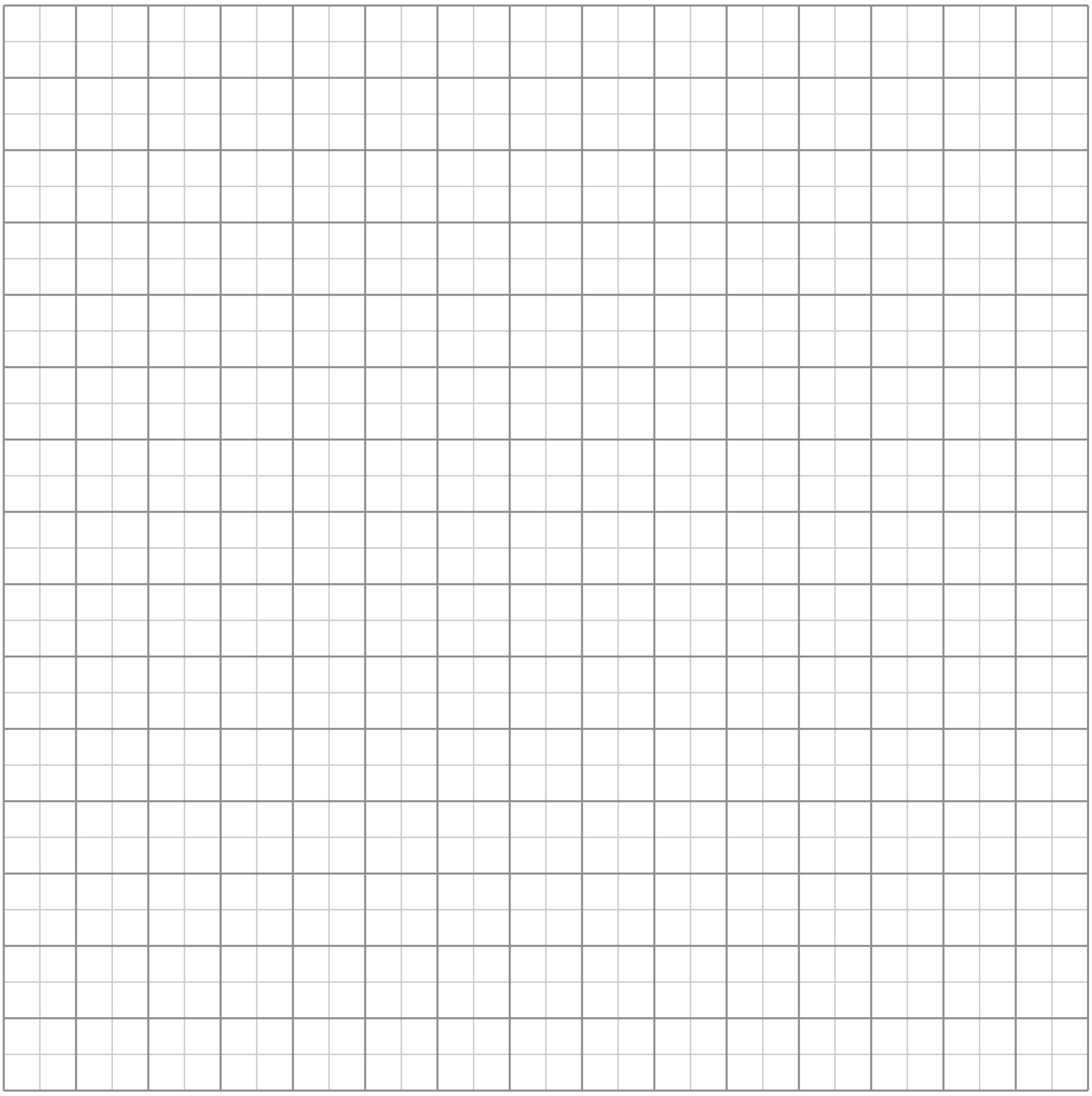


Figure 5.1: Differentiability and derivative.

5.3 Definition

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*** $x_0 \in (a, b)$ *if and only if*

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} \tag{5.3.1}$$

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

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}. \tag{5.3.2}$$

Graphically, as discussed in the last section, 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 3.6. 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* (5.3.1) *can be written as*

$$f'(x_0) = \lim_{x\to x_0} \frac{f(x) - f(x_0)}{x - x_0}. \tag{5.3.3}$$

2. *Definition 5.1 can be restated as: $f : (a, b) \to \mathbb{R}$ is differentiable at $x_0 \in (a, b)$ if and only if 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 (5.3.2). *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 (5.3.3), *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) neighbourhood of x_0, 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 Chapter 3, 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) if and only if f is differentiable at x_0 for all $x_0 \in (a,b)$.*
The derivative $f' : (a,b) \to \mathbb{R}$ is given by the map

$$x \mapsto f'(x)$$

where $f'(x)$ is defined in (5.3.2).

Remark 5.4. *In the sequel, 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.*

5.3.1 Examples of derivatives

Exercise 5.2. *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.*

- *What do we have to investigate?*

- *Let us work it out then.*

Remark 5.6. *The converse of Example 5.2 is also true. For that see Lemma 5.1.*

Example 5.1. *Let us compute the derivative of $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x$ with the differential quotient. We have*

Example 5.2. *Let us compute the derivative of $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$ with the differential quotient. We have*

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

Example 5.4. *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*

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

See also Example 3.8.

Example 5.5. *There exist also function which are differentiable at one point but are not continuous anywhere else. Let*

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

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

$$\lim_{h\to 0} \frac{f(0+h) - f(0)}{h}. \tag{5.3.4}$$

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 (5.3.4) exists and is equal to 0.

Remark 5.7. *As we have seen, e.g. with Example 5.4, there are more continuous than differentiable 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.

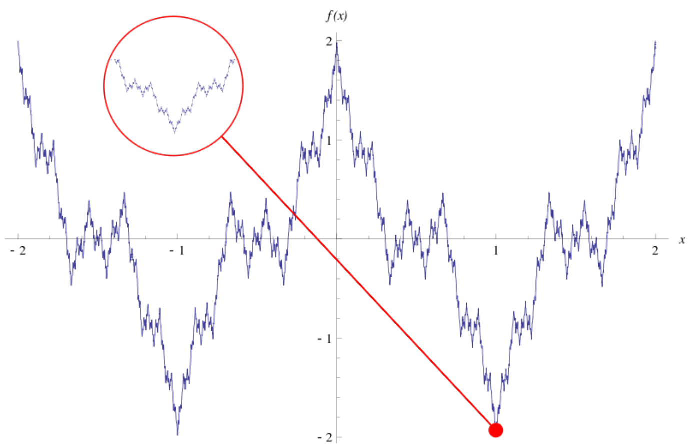


Figure 5.2: A glimpse into the Weierstrass Monster Function.

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

5.3.2 Differentiability $\Rightarrow$ 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.

Remark 5.9. *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.*

5.3.3 Are derivatives continuous?

Reading 6. *This section constitutes this week's reading. Read the section carefully and work out all the details, i.e. redo calculations, build your own examples, etc. pp. If you have questions, ask your tutors, talk to the staff in the MLSC or come to me.*

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

$$f(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & : \ x \in [-1, 1] \setminus \{0\} \\ 0 & : \ x = 0 \end{cases} . \tag{5.3.5}$$

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

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

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

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

does not exist.

Exercise 5.3. *Prove that the function* (5.3.5) *is differentiable at* $x = 0$*.*

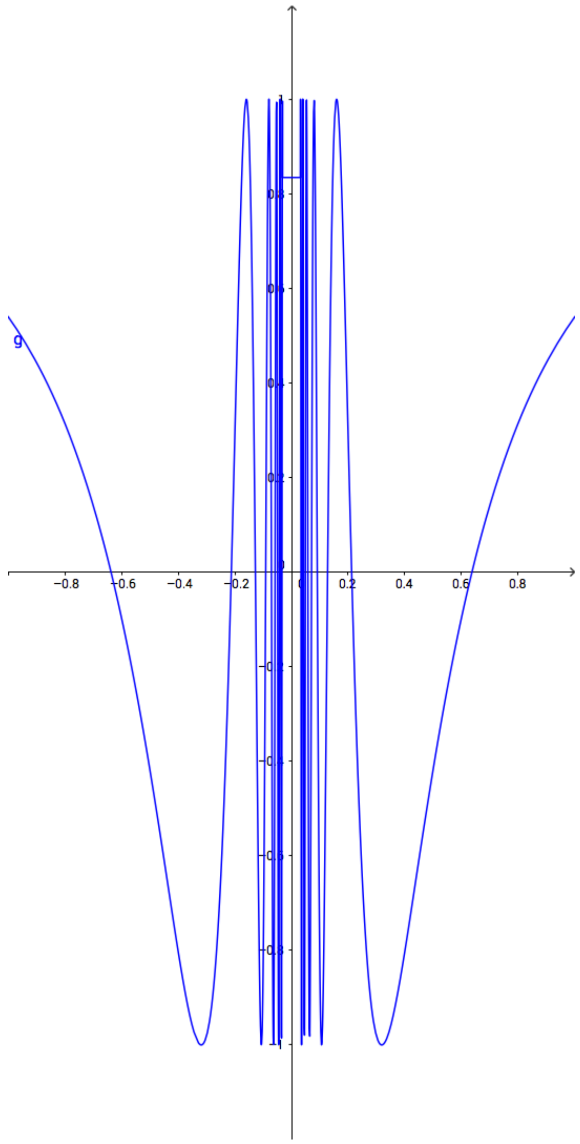


Figure 5.3: The function $\cos(\frac{1}{x})$. As x approaches zero, the functions tries to take all values between -1 and 1 and can therefore not have a limit.

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.

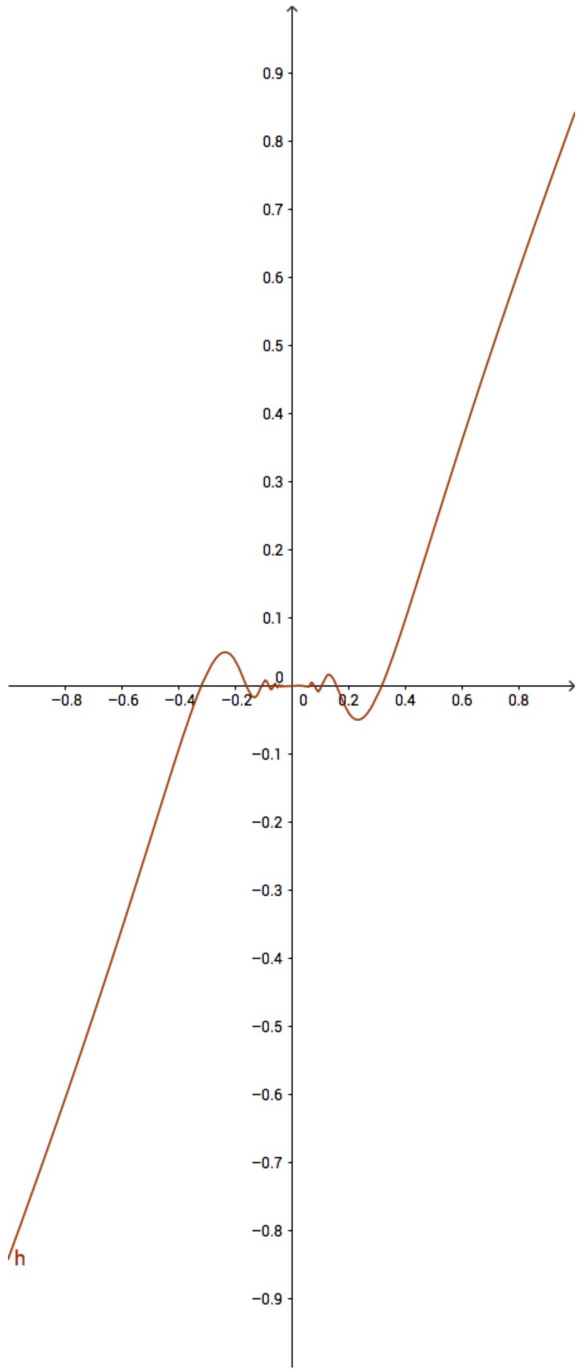


Figure 5.4: A plot of $f(x) = x^2 \sin\left(\frac{1}{x}\right)$.

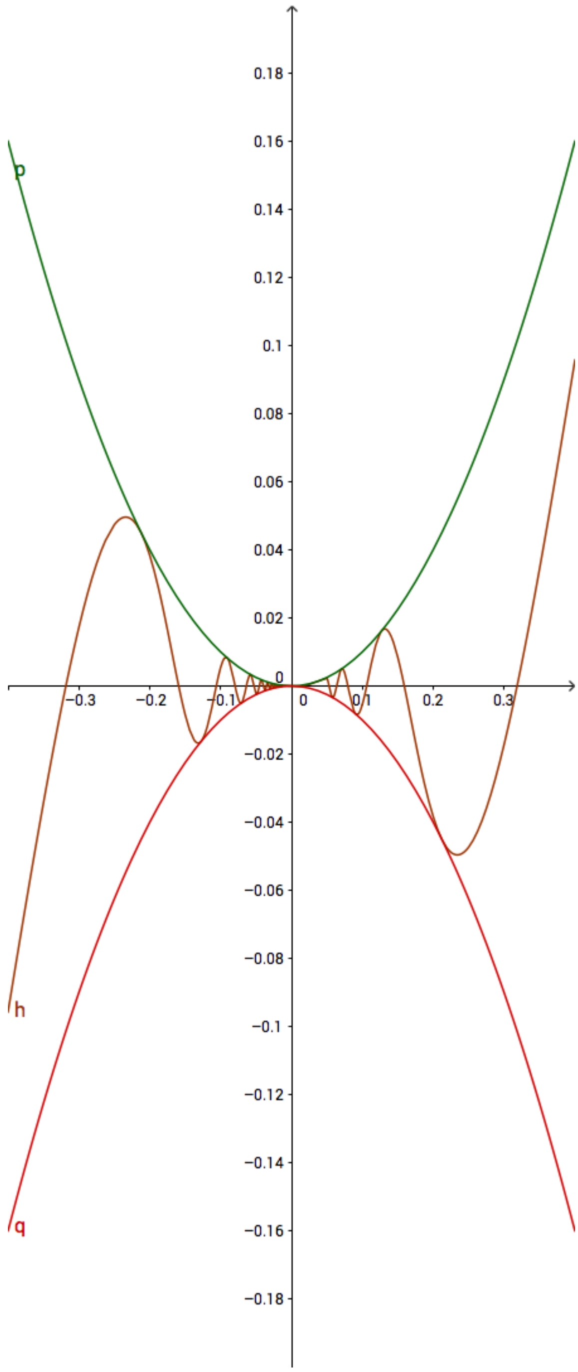


Figure 5.5: A more detailed plot of $f(x) = x^2 \sin\left(\frac{1}{x}\right)$ around $x = 0$ with the envelope $\pm x^2$.

5.4 Operations with differentiable functions

As we have done in Section 4.6 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 (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) $(f \cdot g)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$,

4) $\left(\frac{f}{g}\right)'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g^2(x_0)}$, *provided* $g(x_0) \neq 0$.

Proof.

Exercise 5.4. *Use Theorem 5.2 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*

$$p(x) = \sum_{k=0}^{n} a_k x^k,$$

where the a_k, $k = 0, \ldots, n$ 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.)

Theorem 5.3 (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 5.10. *In the composition $f \circ g(x) = f(g(x))$ 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 5.11. *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:*

$$\exists\varepsilon_0>0\ \forall\varepsilon\in(0,\varepsilon_0]$$
$$\exists x\in(x_0-\varepsilon,x_0+\varepsilon)\setminus\{x_0\}\text{ s.t. } g(x)=g(x_0).$$

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

$$\frac{f(g(x)) - f(g(x_0))}{x - x_0} = \frac{f(g(x)) - f(g(x_0))}{g(x) - g(x_0)} \frac{g(x) - g(x_0)}{x - x_0}. \tag{5.4.1}$$

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

$$\lim_{x\to x_0} \frac{g(x) - g(x_0)}{x - x_0} = 0. \tag{5.4.2}$$

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 sub-sequences $(x_{n_l})_{l\in\mathbb{N}} \subseteq (x_n)_{n\in\mathbb{N}}$ 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 $g(x_{n_k}) \neq g(x_0)$.[1] 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 (5.4.1) to get

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

[1] One of them could be empty, the argument proceeds nevertheless. Do you see that?

by (5.4.2). 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 (5.4.1) and basic limit calculus. □

Example 5.6. *Let us compute a couple of examples with the chain rule:*

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

5.5 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.

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 5.7. *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 5.6: Illustration of Rolle's Theorem.

Proof of Rolle's Theorem:

Remark 5.12. *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. 4.9), 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 $\mathbb{Q}$.*

An immediate consequence of Rolle's Theorem is the so-called **Mean Value Theorem** which also has many important applications.

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 $(b, f(b))$.

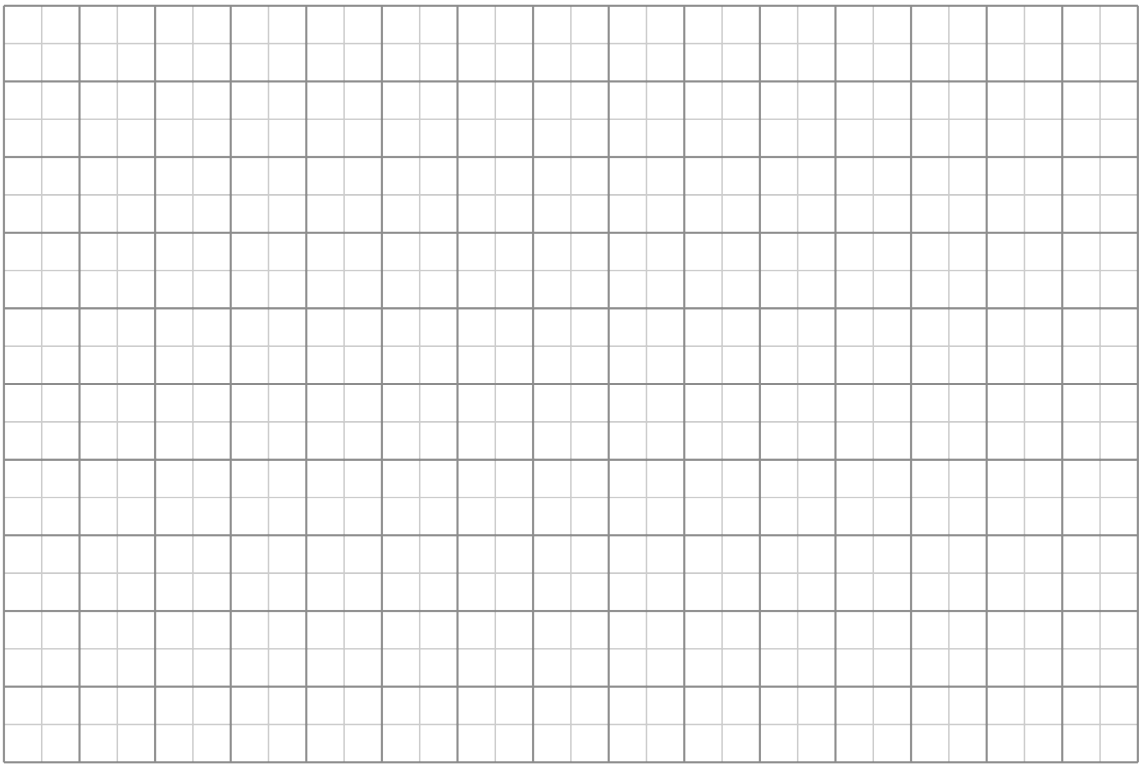


Figure 5.7: Illustration of the Mean Value Theorem.

Proof of Theorem 5.6.

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 $x \in [a, b]$. $\square$

5.6 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 neighbourhood[2] $(x_0 - \delta, x_0 + \delta)$ of x_0, and $f^{(k-1)}$ must be differentiable at x. Since $f^{(k-1)}$ must exist in neighbourhood of x_0, $f^{(n-2)}$ must be differentiable in that neighbourhood.

5.7 Function spaces

Reading 7. *This section constitutes this week's reading. It is mostly the introduction of notation but you need to read it carefully nonetheless. If you are confused by anything, please ask your tutors, see the staff in the MLSC or come by see me.*

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 $\mathbb{R}^d$.

Definition 5.4 (The space $C(I)$)**.**
Let $I \subseteq \mathbb{R}$ be an interval. Then, by $C(I) = C^0(I)$, we denote the set of all functions $f : I \to \mathbb{R}$ which are continuous on I.

[2] or in a one-sided neighbourhood $(x_0 - \delta, x_0]$, $[x_0, x_0 + \delta)$ if x_0 is a boundary point of an interval on which f is defined.

Remark 5.13. *If* $I = [a, b]$*, we also write* $C[a, b]$ *instead of* $C([a, b])$*.*

Exercise 5.5. *Convince yourself that* $C[a, b]$ *is a real vector space, where* $+$ *is the usual pointwise definition*

$$\begin{aligned} + : C[a, b] \times C[a, b] &\to C[a, b] \\ (f, g) &\mapsto f + g, \end{aligned}$$

where $(f + g)(x) := f(x) + g(x)$ *for all* $x \in [a, b]$*, and*

$$\begin{aligned} \cdot : \mathbb{R} \times C[a, b] &\to C[a, b] \\ (\lambda, f) &\mapsto \lambda f, \end{aligned}$$

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 1.8 from $\mathbb{R}^d$ 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 $C[a,b]$ with the properties

(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 the derivatives* $f^{(0)}, f^{(1)}, f^{(2)} \ldots, f^{(k)}$ *exist and are all continuous on* $[a, b]$*.*

Exercise 5.6. *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* $C^k[a, b]$ *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

$$C^\infty[a,b] = \bigcap_{k\in\mathbb{N}_0} C^k[a,b].$$

Remark 5.14. *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 5.15. *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 Function sequences

This chapter contains more information than I will examine. For the development of integrals, we need the notions of pointwise- and uniform convergence of sequences of functions. This will be discussed in Section 6.1 of this Chapter. The remainder of the chapter is for your general education and I recommend reading it to get a comprehensible overview but the contents will not be examined.

6.1 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, $(f_n(x))_{n\in\mathbb{N}_0}$ 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.

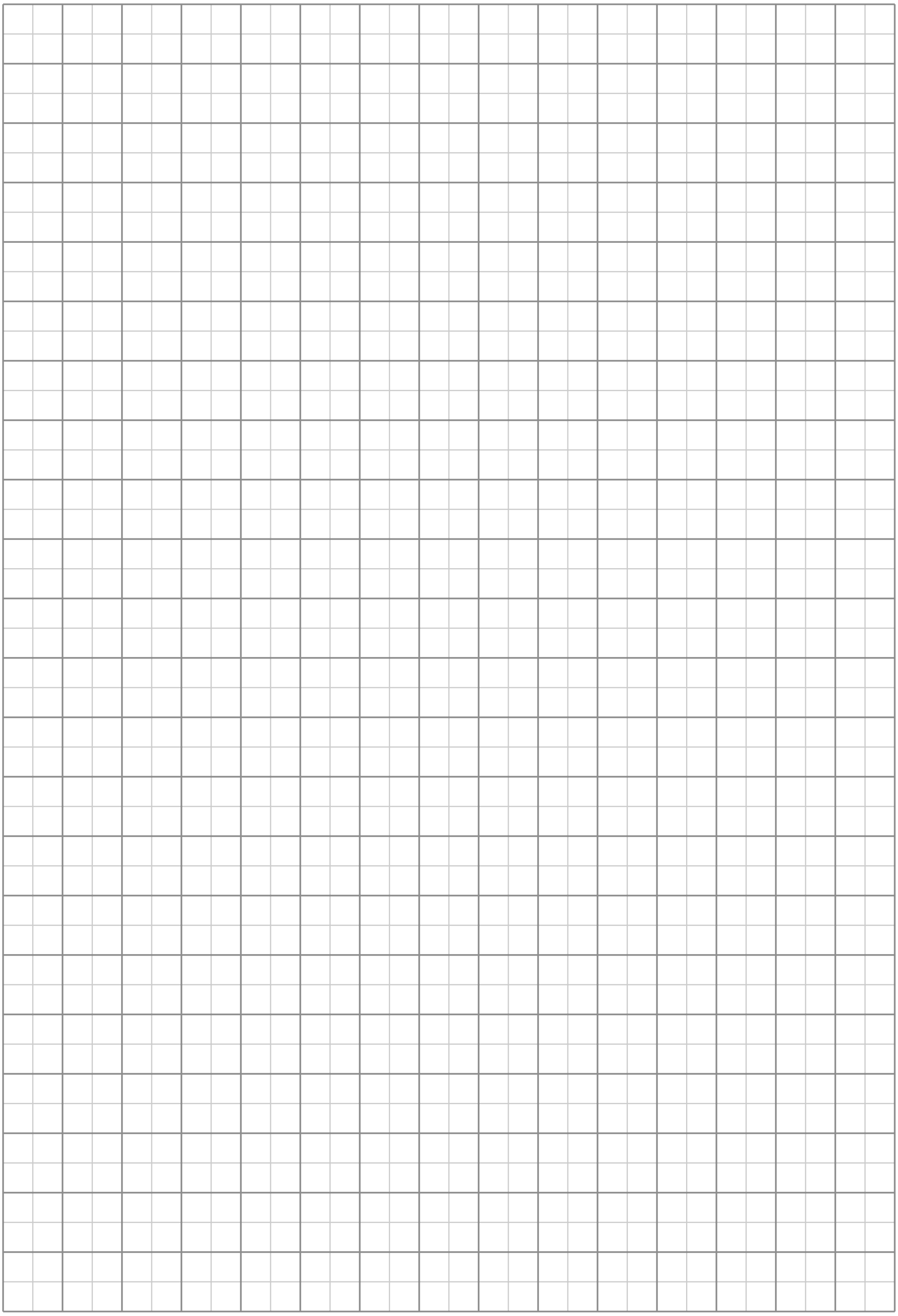


Figure 6.1: Illustration of function sequences.

Let us make this notion of convergence more precise with

Definition 6.1 (Pointwise convergence)**.**

Let (f_n) *be a sequence of functions* $f_n : [a, b] \to \mathbb{R}$ *and* $f : [a, b] \to \mathbb{R}$. *We say* $f_n \to f$ ***pointwise*** *as* $n \to +\infty$ *if and only if for all* $x \in [a, b]$, *we have*

$$\lim_{n \to +\infty} f_n(x) = f(x). \tag{6.1.1}$$

The line (6.1.1) can be rewritten as

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 behaviour if we ask $|f_n(x) - f(x)| < \varepsilon$ not for one point but over all $x \in [a, b]$, 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. In formula

$$n \geq n_0 \quad \Rightarrow \quad \sup_{x \in [a,b]} |f_n(x) - f(x)| < \varepsilon.$$

To be more precise, let us introduce

Definition 6.2 (Uniform convergence)**.**
Let (f_n) *be a sequence of functions* $f_n : [a,b] \to \mathbb{R}$ *and* $f : [a,b] \to \mathbb{R}$. *We then say* $f_n \to f$ ***uniformly*** *as* $n \to +\infty$ *if and only if*

$$\begin{aligned} &\forall \varepsilon > 0 \quad \exists n_0 = n_0(\varepsilon) \in \mathbb{N} \ : \\ &\qquad \forall n \in \mathbb{N}, \ n \geq n_0 \quad \Rightarrow \quad \sup_{x \in [a,b]} |f(x) - f_n(x)| < \varepsilon. \end{aligned} \tag{6.1.2}$$

Using the notation of the norm introduced in 5.7, we can restate (6.1.2) 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 6.1 (Uniform $\Rightarrow$ pointwise)**.**
Let (f_n) be a sequence of functions $f_n : [a, b] \to \mathbb{R}$ such that there exists a function $f : [a, b] \to \mathbb{R}$ with $f_n \to f$ uniformly. Then $f_n \to f$ pointwise.

Exercise 6.1. *Prove Proposition 6.1.*

We can of course formulate a Cauchy-criterion

Definition 6.3 (Cauchy criterion)**.**

Let (f_n) be a sequence of functions $f_n : [a,b] \to \mathbb{R}$. Then, we say that (f_n) is ***uniformly Cauchy*** *on $[a,b]$ of for every $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that*

$$\forall m, n \geq n_0 \ : \ \sup_{x\in[a,b]} |f_n(x) - f_m(x)| < \varepsilon.$$

One can prove the following

Theorem 6.1.

Suppose (f_n) is a sequence of functions $f_n : [a,b] \to \mathbb{R}$. Then, (f_n) is uniformly convergent if and only if it is uniformly Cauchy.

Exercise 6.2. *Prove Theorem 6.1. First assume that (f_n) is uniformly convergent and deduce that it is Cauchy. Review the case for real sequences. For the converse prove first that $(f_n(x)) \subseteq \mathbb{R}$ is Cauchy in $\mathbb{R}$ and then set $f(x) = \lim_{n\to+\infty} f_n(x)$ for all $x \in [a,b]$. Then show the uniform convergence $f_n \to f$.*

The central result of this section is

Theorem 6.2.

Suppose (f_n) is a sequence of continuous functions $f_n : [a,b] \to \mathbb{R}$. Assume that $f_n \to f$ uniformly for a function $f : [a,b] \to \mathbb{R}$. Then, f is continuous.

Proof. We assume that there exists a function f such that (f_n) is uniformly convergent to f. We show that f is continuous, i.e. that for all $x_0 \in [a,b]$ and all $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$x \in [a,b],\ |x - x_0| < \delta \quad \Rightarrow \quad |f(x) - f(x_0)| < \varepsilon.$$

We compute

$$\begin{aligned} |f(x) - f(x_0)| &= |f(x) - f_n(x) + f_n(x) - f(x_0)| \\ &\leq |f(x) - f_n(x)| + |f_n(x) - f(x_0)| \\ &\leq \sup_{x\in[a,b]} |f(x) - f_n(x)| + |f_n(x) - f(x_0)|. \end{aligned}$$

Since $f_n \to f$ uniformly, there exists $n_1 \in \mathbb{N}$ such that

$$\forall n \geq n_1 \ :\ \sup_{x\in[a,b]} |f(x) - f_n(x)| < \frac{\varepsilon}{3}.$$

Thus, we have for all $n \geq n_1$ that

$$\begin{aligned} |f(x) - f(x_0)| &< \frac{\varepsilon}{3} + |f_n(x) - f(x_0)| \\ &= \frac{\varepsilon}{3} + |f_n(x) - f_n(x_0) + f_n(x_0) - f(x_0)| \\ &\leq \frac{\varepsilon}{3} + |f_n(x) - f_n(x_0)| + |f_n(x_0) - f(x_0)|. \end{aligned}$$

By Proposition 6.1, there exists $n_2 \in \mathbb{N}$ such that

$$\forall n \geq n_2 \ :\ |f_n(x_0) - f(x_0)| < \frac{\varepsilon}{3}$$

for all $n \geq n_2$. Further, the f_n are continuous on $[a,b]$, hence, there exists for all $n \geq n_0 = \max\{n_1, n_2\}$ a $\delta = \delta(n,\varepsilon) > 0$ such that $|f_n(x) - f_n(x_0)| < \frac{\varepsilon}{3}$. Thus,

$$|f(x) - f(x_0)| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon$$

whenever $|x - x_0| < \delta$. This concludes the proof. ☐

6.2 Examples & counterexamples

Example 6.1 (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 6.2. *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) *with* $f_n : [0,1] \to \mathbb{R}$ *defined by*

$$f_n(x) = \begin{cases} 2n^2x & : \ 0 \le x \le \frac{1}{2n} \\ 2n^2\left(\frac{1}{n} - x\right) & : \ \frac{1}{2n} < x < \frac{1}{n} \\ 0 & : \ \frac{1}{n} \le x \le 1 \end{cases}.$$

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 6.3. *Consider* (f_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* $n \ge \left\lfloor \frac{1}{\varepsilon} \right\rfloor + 1$*, where* $\left\lfloor \frac{1}{\varepsilon} \right\rfloor$ *means the largest integer smaller or equal to* $\frac{1}{\varepsilon}$*.*

6.3 (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 6.3 ((Real) Power series)**.**
Let $(a_n) \subseteq \mathbb{R}$ *be a sequence of numbers and suppose that*

$$\sum_{k=0}^{+\infty} |a_n| R^n < +\infty. \tag{6.3.1}$$

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 (6.3.1). We have

$$\begin{aligned} \sup_{x\in[-R,R]} |f_n(x) - f(x)| &= \sup_{x\in[-R,R]} \left| \sum_{k=n+1}^{+\infty} a_k x^k \right| \\ &\leq \sum_{k=n+1}^{+\infty} |a_k| \|x^k\|_\infty \\ &= \sum_{k=n+1}^{+\infty} |a_k| R^k. \end{aligned}$$

Since (6.3.1), 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 6.3. *The series*

$$\sum_{k=0}^{+\infty} \frac{x^k}{k!} \tag{6.3.2}$$

converges uniformly. We define the exponential function $e^x = \exp(x)$ by

$$e^x := \sum_{k=0}^{+\infty} \frac{x^k}{k!} \tag{6.3.3}$$

.

Exercise 6.4. *Check that (6.3.2) converges uniformly. Use the Definition (6.3.3) to prove $e^{x+y} = e^x e^y$. To see that the product is well-defined, you should look up the definition of Cauchy product.*

Example 6.4. *We define (see Taylor's Theorem and the Mathematical Methods II module) the following functions*

$$\sin(x) := \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1}, \tag{6.3.4}$$

$$\cos(x) := \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k)!} x^{2k}. \tag{6.3.5}$$

Exercise 6.5. *Check that the series in the definitions (6.3.4) and (6.3.5) 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

7 Integration on $\mathbb{R}^1$

7.1 Step functions and their integrals

First, we remind ourselves about a definition from set theory and introduce the notion of an indicator function.

Definition 7.1 (Indicator function)**.**
Let $A \subseteq \mathbb{R}^d$. Then, the function

$$\chi_A(x) = \begin{cases} 1 & : \quad x \in A \\ 0 & : \quad x \notin A \end{cases}.$$

is called indicator function of A.

Remark 7.1. *The letter χ is a the Greek letter* chi.

The following properties of characteristic functions are useful.

Proposition 7.1 (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 7.1. *Prove Proposition 7.1 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 7.2 (Step-function)**.**
A function $f : [a,b] \to \mathbb{R}$ is called a ***step-function*** *if and only if f can be written as*

$$f(x) = \sum_{i=1}^{N} c_i \, \chi_{A_i}(x) \tag{7.1.1}$$

for $c_i \in \mathbb{R}$ and $A_i \subseteq [a,b]$ intervals for $i = 1, \ldots, n$.[a]

[a] We assume that $A_1 \cup \ldots \cup A_N = [a,b]$.

Example 7.1. *The function $f : [0,10] \to \mathbb{R}$, given by*

$$f(x) = \begin{cases} 1 & : \quad 0 \le x \le 3 \\ -3 & : \quad 3 < x < 4 \\ 8 & : \quad 4 \le x \le 9 \\ 2 & : \quad 9 < x \le 10 \end{cases}$$

is a step function as we can write

$$f(x) = \chi_{[0,3]}(x) - 3\chi_{(3,4)}(x) + 8\chi_{[4,9]}(x) + 2\chi_{(9,10]}(x).$$

Exercise 7.2. *The intervals* A_i *in Definition 7.2 can always assumed to be disjoint. We have*

$$f : [0,3] \to \mathbb{R}, \quad f(x) = \chi_{[0,2]}(x) + 3\chi_{[1,3]}(x)$$

which can be written as

$$f(x) = \chi_{[0,1)}(x) + 4\chi_{[1,2]}(x) + 3\chi_{(2,3]}(x).$$

Remark 7.2. *Suppose* f *is a step function as in Definition 7.2. Then, the representation in (7.1.1) 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*

$$f(x) = \sum_{i=1}^{N} c_i \, \chi_{A_i}(x) = \sum_{i=1}^{M} \tilde{c}_i \, \chi_{\tilde{A}_i}(x).$$

Definition 7.3 (The collective set $\mathcal{S}[a,b]$ of all step functions on $[a,b]$.).
Denote by $\mathcal{S}[a,b]$ *the set of all step functions* $f : [a,b] \to \mathbb{R}$ *as given in Definition 7.2.*

Theorem 7.1 ($\mathcal{S}[a,b]$ is a vector space and an algebra).
The set $\mathcal{S}[a,b]$ *of step-functions on* $[a,b]$ *is a real vector space, i.e. for all* f, $g \in \mathcal{S}[a,b]$, *we have*

$$\forall \lambda, \mu \in \mathbb{R} \,:\, \lambda f + \mu g \in \mathcal{S}[a,b].$$

Moreover, $\mathcal{S}[a,b]$ *is an algebra, i.e. for all* $f, g \in \mathcal{S}[a,b]$, *we have*

$$f \cdot g \in \mathcal{S}[a,b].$$

Equipped with

$$\|f\|_\infty = \sup_{x\in[a,b]} |f(x)|,$$

the space $\mathcal{S}[a,b]$ *is a normed vector space.*

Remark 7.3. *If* $f \in \mathcal{S}[a,b]$ *with*

$$f(x) = \sum_{i=1}^{N} c_i \chi_{A_i}(x),$$

where $A_i \cap A_j = \varnothing$, $i \neq j$, *then*

$$\|f\|_\infty = \max\{|c_1|, \ldots, |c_N|\}.$$

Proof of Theorem 7.1.

This follows immediately from Proposition 7.1. You are advised to carry out some details.

We state

Proposition 7.2.

Let $f \in \mathcal{S}[a,b]$. There exists a finite family of intervals $\{B_i : i = 0, \ldots, M\}$ such that $B_i \cap B_j = \varnothing$ for $i \neq j$, $B_1 \cup \ldots \cup B_M = [a,b]$, and

$$f(x) = \sum_{i=1}^{M} d_i \, \chi_{B_i}(x).$$

Proof (Sketch).

Let $f \in \mathcal{S}[a,b]$ and

$$f(x) = \sum_{i=1}^{N} c_i \chi_{A_i}(x).$$

The A_i are if the form $[a_i, b_i]$, (a_i, b_i), $(a_i, b_i]$, or $[a_i, b_i)$. We consider $\tilde{A}_i = (a_i, b_i)$ if $a_i < b_i$. To treat the 'lost' boundary points $\{x_1, \ldots, x_L\}$, we set $d_i \chi_{[x_i, x_i]}(x)$ with

$$d_i = \sum_{j \colon x_i \in A_j} c_j.$$

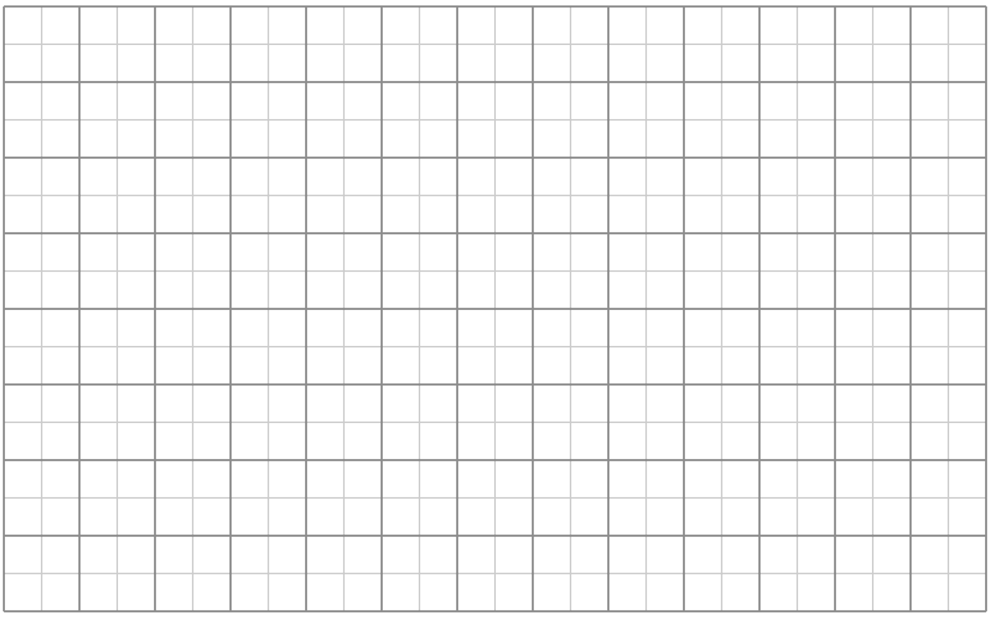


Figure 7.1: Illustration of the argument so far.

To divide the $\tilde{A}_i$, we relabel them such that

$$a = a_1 < a_2 < \ldots < a_K = b. \tag{7.1.2}$$

We start with b_1 and but in (7.1.2) where it belongs, then take b_2, b_3 and so forth. Finally, we get a partition of $[a, b]$ of the type

$$a = x_1 < x_2 < \ldots < x_L = b.$$

The intervals (x_i, x_{i+1}), together with the one-point sets from the start, are the sought after B_i in the proposition.

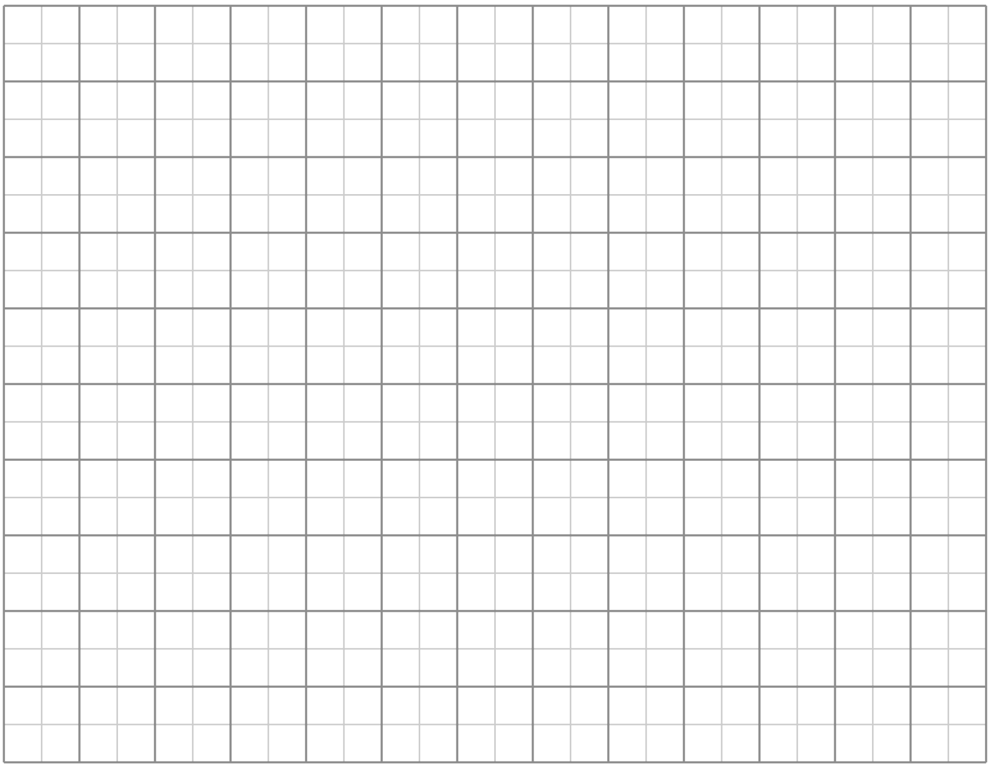


Figure 7.2: Illustration of the rest of the argument.

To define the integral of step functions, we have to introduce the diameter/length of an integral.

Definition 7.4 (Diameter $\mathrm{diam}(I)$ of an interval)**.**
Let $I \subseteq \mathbb{R}$ be an interval. Then, the ***length/diameter*** *of the interval I, in symbols $\mathrm{diam}(I)$, is defined to be*

$$\mathrm{diam}(I) := \sup\{|x-y| : \; x, y \in I\} = \sup_{x,y \in I} |x-y|.$$

Remark 7.4. *To avoid confusion, we will use $\mathrm{diam}(I)$. However, many text use the notation $|I|$ for the length/diameter of an interval which is not to be confused with the absolute value.*
It is easy to check that

$$\begin{aligned} \mathrm{diam}([a,b]) &= \mathrm{diam}([a,b)) = \mathrm{diam}((a,b)) \\ &= \mathrm{diam}((a,b]) = |b-a|. \end{aligned}$$

In general that means that

$$\mathrm{diam}(I) = |\sup(I) - \inf(I)|.$$

Next, we define the integral of a step function. We start with some thinking:

Taking

$$\mathcal{S}[a,b] \ni f(x) = \sum_{i=1}^{N} c_i \chi_{A_i}(x)$$

and thinking intuitively, we would like to set

$$\mathcal{I}(f) := \sum_{i=1}^{N} c_i \operatorname{diam}(A_i). \tag{7.1.3}$$

However, as said in Remark 7.2, we need to know that $\mathcal{I}(f)$ does not change if we write f differently. If we have two arbitrary representations

$$f(x) = \sum_{i=1}^{N} c_i \chi_{A_i}(x) = \sum_{i=1}^{M} \tilde{c}_i \chi_{\tilde{A}_i}(x). \tag{7.1.4}$$

we can subtract them and are left to prove that the 0-function has integral 0 no matter how we write it as a step function. To do that, we state

Lemma 7.1.

Let $f \in \mathcal{S}[a,b]$. *Then,*

$$|\mathcal{I}(f)| \leq |b-a| \|f\|_\infty$$

independently of the representation of f.

With that, we immediately get

Corollary 7.1.

If $f \in \mathcal{S}[a,b]$ *with* $f \equiv 0$, *we have* $\mathcal{I}(f) = 0$.

which shows that (7.1.3) is independent of the representation of f, i.e.

$$\sum_{i=1}^{N} c_i \operatorname{diam}(A_i) = \sum_{i=1}^{M} \tilde{c}_i \operatorname{diam}(\tilde{A}_i)$$

for (7.1.4).

Thus we can state

Definition and Theorem 7.2 (Step function integral)**.**

Let $I \subseteq \mathbb{R}$ *be an interval and* $f : I \to \mathbb{R}$ *be a step function. Then, the integral of* f*, denoted by* $\mathcal{I}(f)$*, is defined as*

$$\mathcal{I}(f) = \sum_{i=1}^{N} c_i \, \mathrm{diam}(A_i),$$

where

$$f(x) = \sum_{i=1}^{N} c_i \, \chi_{A_i}(x). \tag{7.1.5}$$

The value of $\mathcal{I}(f)$ *does not depend on the particular representation* (7.1.5)*.*

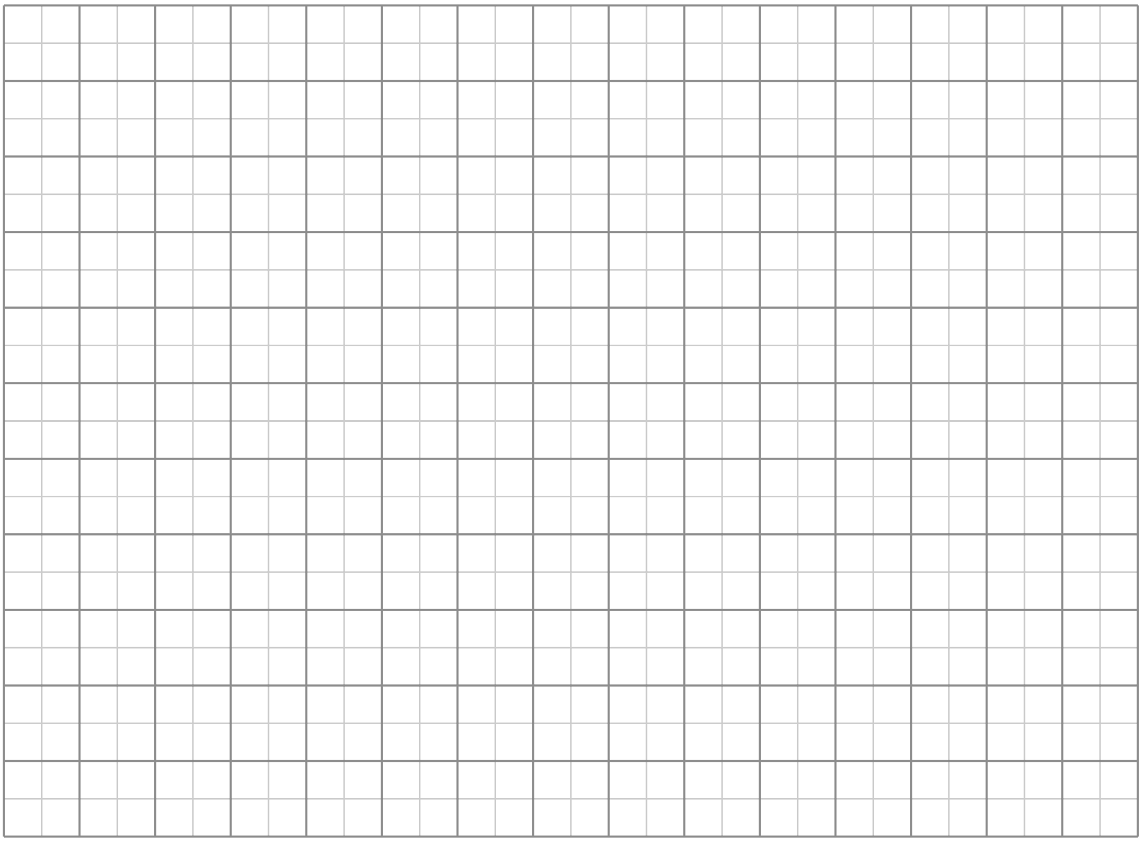


Figure 7.3: Illustration of the step-function integral.

7.1.1 Properties of the step function integral $\mathcal{I}(f)$

Let us collect and prove a couple of properties of the step-function integral.

Theorem 7.3 (Properties of $\mathcal{I}(f)$)**.**
The following properties hold for

$$\mathcal{I} : \mathcal{S}[a,b] \to \mathbb{R}.$$

(i) If $f \in \mathcal{S}[a,b]$, then $|f| \in \mathcal{S}[a,b]$. Further, it holds

$$\mathcal{I}(f) \leq \mathcal{I}(|f|).$$

(ii) The operator $\mathcal{I}$ is linear, i.e. for all $f, g \in \mathcal{S}[a,b]$ and all $\alpha, \beta \in \mathbb{R}$, we have

$$\mathcal{I}(\alpha f + \beta g) = \alpha \mathcal{I}(f) + \beta \mathcal{I}(g).$$

(iii) For all $f \in \mathcal{S}[a,b]$, we have

$$|\mathcal{I}(f)| \leq |b-a| \|f\|_\infty.$$

(iv) If $f \in \mathcal{S}[a,b]$ and $f(x) \geq 0$ for all $x \in [a,b]$, then

$$\mathcal{I}(f) \geq 0.$$

(v) If $f, g \in \mathcal{S}[a,b]$ and $f(x) \geq g(x)$ for all $x \in [a,b]$, then

$$\mathcal{I}(f) \geq \mathcal{I}(g).$$

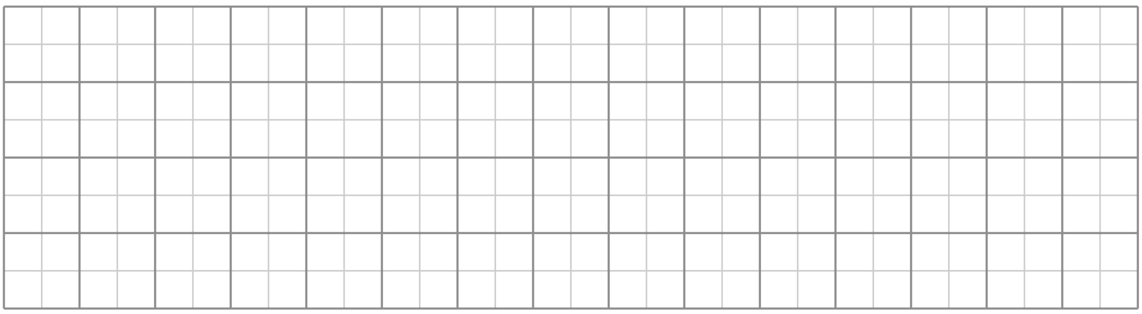


Figure 7.4: Illustration of (iii).

Proof of Theorem 7.3.

(i) If f is written as

$$f(x) = \sum_{i=1}^{N} c_i \chi_{A_i}(x)$$

with disjoint A_i (see Proposition 7.2), then $|f|$ can be written as

$$|f|(x) = \sum_{i=1}^{N} |c_i| \chi_{A_i}(x).$$

Using Theorem 7.2 and

$$\forall x \in [a, b] \; : \; \sum_{i=1}^{N} c_i \chi_{A_i}(x) \leq \sum_{i=1}^{N} |c_i| \chi_{A_i}(x),$$

we get the result.

(ii) This is an immediate consequence of Theorem 7.1.

(iii)

(iv) We rewrite f with non-intersecting intervals

$$f(x) = \sum_{i=1}^{N} c_i \chi_{A_i}(x).$$

Thus, $c_i \geq 0$ for all $i = 1, \ldots, N$. Thus, by Definition 7.1.3, we get $\mathcal{I}(f) \geq 0$ as the sum of non-negative numbers is non-negative.

(v) Follows from (iii) realising that $f - g \in \mathcal{S}[a, b]$ (see Theorem 7.1) and that $f(x) - g(x) \geq 0$ for all $x \in [a, b]$.

7.2 Regulated functions and their integrals

Using step functions and uniform convergence, we introduce a new class of functions for which we can easily construct an integral.

Definition 7.5 (Regulated function)**.**
*A function $f : [a, b] \to \mathbb{R}$ is said to be **regulated** if and only if there exists a sequence $(f_n) \subseteq \mathcal{S}[a, b]$ such that $f_n \to f$ uniformly as $n \to +\infty$.*
The set of all regulated functions on $[a, b]$ is denoted by $\mathcal{R}[a, b]$.

Exercise 7.3. *Prove that regulated functions are bounded.*

Remark 7.5. *One can characterize regulated functions as follows: a function $f : [a,b] \to \mathbb{R}$ is regulated if and only if both the limit*

$$\lim_{x\to c-} f(x) \quad \text{and} \quad \lim_{x\to c+} f(x)$$

exist for all $c \in (a,b)$ as well as $f(a+)$ and $f(b-)$. This is due to Dieudonné.[1]

Theorem 7.4 ($\mathcal{R}[a,b]$ is a vector space and an algebra)**.**
The set $\mathcal{R}[a,b]$ is a real vector space, i.e. for all f, $g \in \mathcal{R}[a,b]$, we have

$$\forall \lambda, \mu \in \mathbb{R} \;:\; \lambda f + \mu g \in \mathcal{R}[a,b].$$

Moreover, $\mathcal{R}[a,b]$ is an algebra, i.e. for all $f, g \in \mathcal{R}[a,b]$, we have

$$f \cdot g \in \mathcal{R}[a,b].$$

Equipped with

$$\|f\|_\infty = \sup_{x\in[a,b]} |f(x)|,$$

the space $\mathcal{R}[a,b]$ is a normed vector space.

Exercise 7.4. *Prove Theorem 7.4.*

[1] Jean Dieudonné (1906–1992) French mathematician.

We slightly extend the notion of a continuous function to piecewise continuous functions.

Definition 7.6 (Piecewise continuous function)**.**
A function $f : [a, b] \to \mathbb{R}$ *is said to be* ***piecewise continuous*** *if and only if there exist finitely many points* $x_1, \ldots, x_n \in [a, b]$ *such that*

(i) f *is continuous on* $[a, b] \setminus \{x_1, \ldots, x_n\}$ *and*

(ii) the limits

$$\lim_{x \to x_k+} f(x) \quad \text{and} \quad \lim_{x \to x_k-} f(x)$$

exist for all $k = 1, \ldots, n$.

We denote the set of all piecewise continuous functions $f : [a, b] \to \mathbb{R}$ *by* $PC[a, b]$.

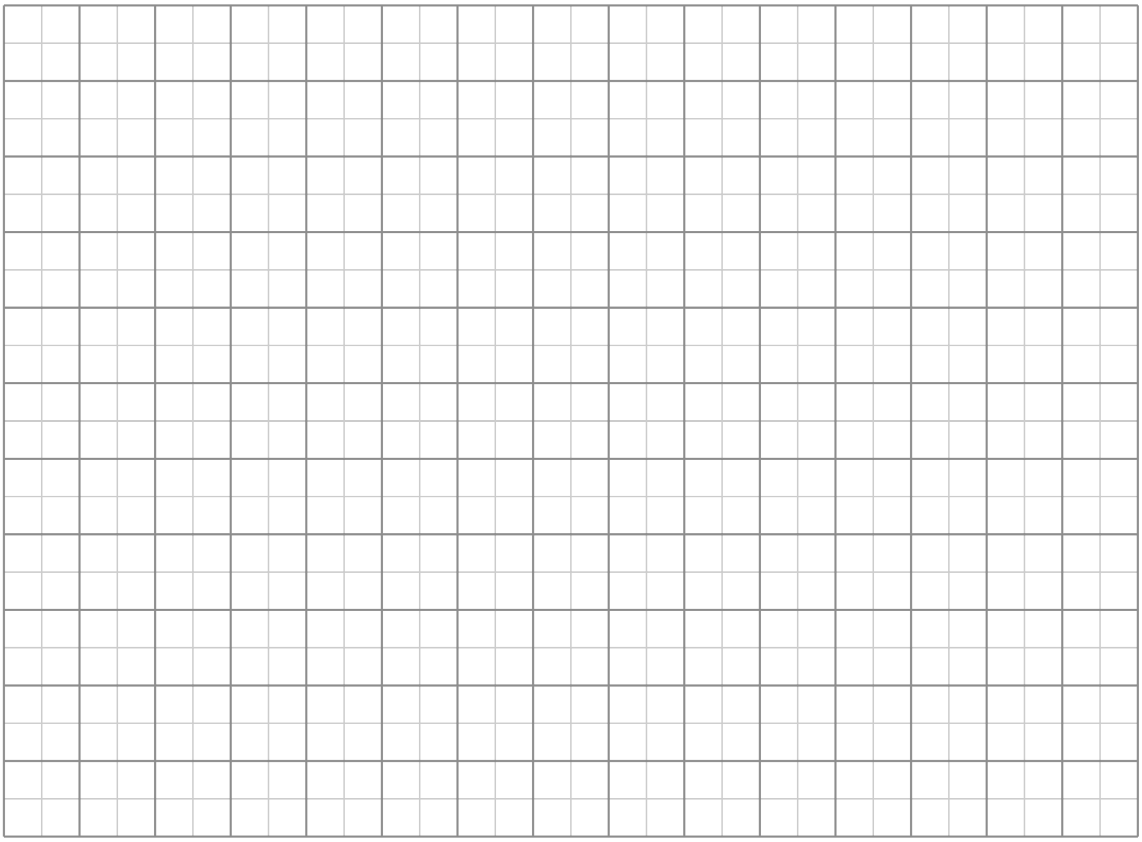


Figure 7.5: Illustration of piecewise continuous functions.

Theorem 7.5 ($PC[a,b] \subseteq \mathcal{R}([a,b])$).
Let $f \in PC[a,b]$. Then $f \in \mathcal{R}[a,b]$.

To prove Theorem 7.5, we introduce the notion of uniform continuity and then Heine's theorem which says that all functions which are continuous on a compact set are uniformly continuous.

Definition 7.7 (Uniform continuity).
*Let I be an interval. A function $f : I \to \mathbb{R}$ said to be **uniformly continuous** if and only if for all $\varepsilon > 0$ there exists $\delta > 0$ such that*

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

Exercise 7.5. *Let $I \subseteq \mathbb{R}$ be an interval and f be uniformly continuous on I. Show that f is continuous on I.*

Remark 7.6. *Let us contrast the definition of continuity again against uniform continuity: for the first, we have*

$$\forall x \in [a,b] \quad \forall \varepsilon > 0 \quad \exists \delta > 0 \,:\, |x - y| < \delta \quad \Rightarrow \quad |f(x) - f(y)| < \varepsilon$$

while uniform continuity means

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad \forall x, y \in [a,b] \,:\, |x - y| < \delta \quad \Rightarrow \quad |f(x) - f(y)| < \varepsilon.$$

Theorem 7.6 (Heine's theorem).
Let $f : [a,b] \to \mathbb{R}$ be continuous. Then, f is uniformly continuous on $[a,b]$.

Proof of Theorem 7.6.

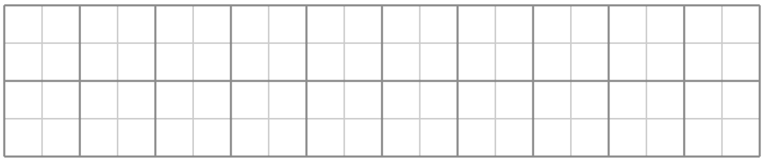
Since f is continuous at every $x \in [a,b]$, there exists $\delta_x > 0$ such that

$$y \in [a,b],\ |x-y| < \delta_x \quad \Rightarrow \quad |f(x) - f(y)| < \frac{\varepsilon}{2}.$$

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 $B_{\frac{\delta_{x_1}}{2}}(x_1)$, ..., $B_{\frac{\delta_{x_k}}{2}}(x_k)$ still covering $[a,b]$. Let now



If now $x, y \in [a,b]$ with $|x-y| < \delta$, we claim that we have

$$|f(x) - f(y)| < \varepsilon.$$

Since the $B_{\frac{\delta_{x_1}}{2}}(x_1)$, ..., $B_{\frac{\delta_{x_k}}{2}}(x_k)$ cover $[a,b]$, we have that there exists a $j \in \{1, \ldots, k\}$ such that $x \in B_{\frac{\delta_{x_j}}{2}}(x_j)$, i.e. $|x - x_j| < \frac{\delta_{x_j}}{2} < \delta_{x_j}$ and, therefore, $|f(x) - f(x_j)| < \frac{\varepsilon}{2}$.
Moreover, we have

$$\begin{aligned} |y - x_j| &\leq |y-x| + |x - x_j| \\ &< \frac{\delta_{x_j}}{2} + \frac{\delta_{x_j}}{2} = \delta_{x_j} \end{aligned}$$

which then gives $|f(y) - f(x_j)| < \frac{\varepsilon}{2}$.
Thus,

$$\begin{aligned} |f(y) - f(x)| &\leq |f(x) - f(x_j)| + |f(y) - f(x_j)| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Proof of Theorem 7.5.

1. We reduce the prove to show that a continuous function on a compact interval is regulated.
   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 $\lim\limits_{x \to x_i+} f(x)$, $\lim\limits_{x \to x_i-} f(x)$ exist for $i = 1, \ldots, n-1$.
   We prove that the restriction of f to the interval (x_i, x_{i+1}), denoted by $f|_{(x_i,x_{i+1})}$, can be uniformly approximated by step functions on $[x_i, x_{i+1}]$. We can extend $f|_{(x_i,x_{i+1})}$ to a function $f|_{[x_i,x_{i+1}]}$ since we 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 $[a,b]$.

2. So, for every $\varepsilon > 0$, we have to construct a step-function $s \in \mathcal{S}[a,b]$ such that $\|f - s\|_\infty < \varepsilon$. By Heine's Theorem, we have that there exists a $\delta > 0$ such that
   $$x, y \in [a,b],\ |x-y| < \delta \quad \Rightarrow \quad |f(x) - f(y)| < \varepsilon.$$
   Thus, we divide $[a,b]$ into sub-intervals: Pick $N \in \mathbb{N}$ so large that
   $$\frac{|b-a|}{N-1} < \frac{\delta}{2}$$
   and set
   $$x_k = a + (k-1)\frac{b-a}{N-1}, \quad k = 1, \ldots, N.$$
   This gives us the partition
   $$\begin{aligned}[a,b] = [x_1, x_2) \cup [x_2, x_3) \cup \ldots \\ \ldots \cup [x_{N-2}, x_{N-1}) \cup [x_{N-1}, x_N].\end{aligned}$$

Now, we can define a step function $s : [a, b] \to \mathbb{R}$ by

$$s(x) = \sum_{k=1}^{N-1} c_k \, \chi_{A_k}(x),$$

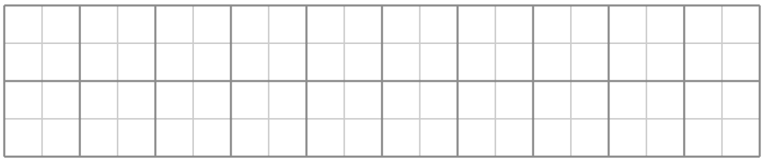
where

$$\forall k \in \{1, \ldots, N-1\} \ : \ c_k := f(x_k)$$

and

$$\forall k \in \{1, \ldots, N-1\} \ : \ A_k = [x_k, x_{k+1})$$

and $A_{N-1} := [a_{N-1}, a_N]$. Then, by construction,



for a y with $|x - y| \leq \frac{\delta}{2} < \delta$. Thus, $|f(x) - s(x)| < \varepsilon$ and, hence,

$$\|f - s\|_\infty = \sup_{x \in [a,b]} |f(x) - s(x)| < \varepsilon.$$

This concludes the proof.

Corollary 7.2 ($C[a,b] \subseteq \mathcal{R}[a,b]$).
Continuous functions $f : [a,b] \to \mathbb{R}$ are regulated on $[a,b]$.

Remark 7.7. *We say that a set $A \subseteq \mathbb{R}^d$ is dense in $\mathbb{R}^d$ if for every $x \in \mathbb{R}^d$ there exists a sequence $(x_n) \subseteq A$ such that $(x_n) \to x$ as $n \to +\infty$. An example is $\mathbb{Q}$ in $\mathbb{R}$ or $\mathbb{Q}^d$ in $\mathbb{R}^d$.*
Here, we have a similar situation. The space $\mathcal{S}[a,b]$ is dense in $PC[a,b]$ since we clearly have $\mathcal{S}[a,b] \subseteq PC[a,b]$ and for every $f \in PC[a,b]$ there exists a sequence $(f_n) \subseteq \mathcal{S}[a,b]$ with $f_n \to f$ uniformly, i.e. using $\operatorname{dist}(f,g) = \|f-g\|_\infty$ as a distance. This is the content of Theorem 7.5.

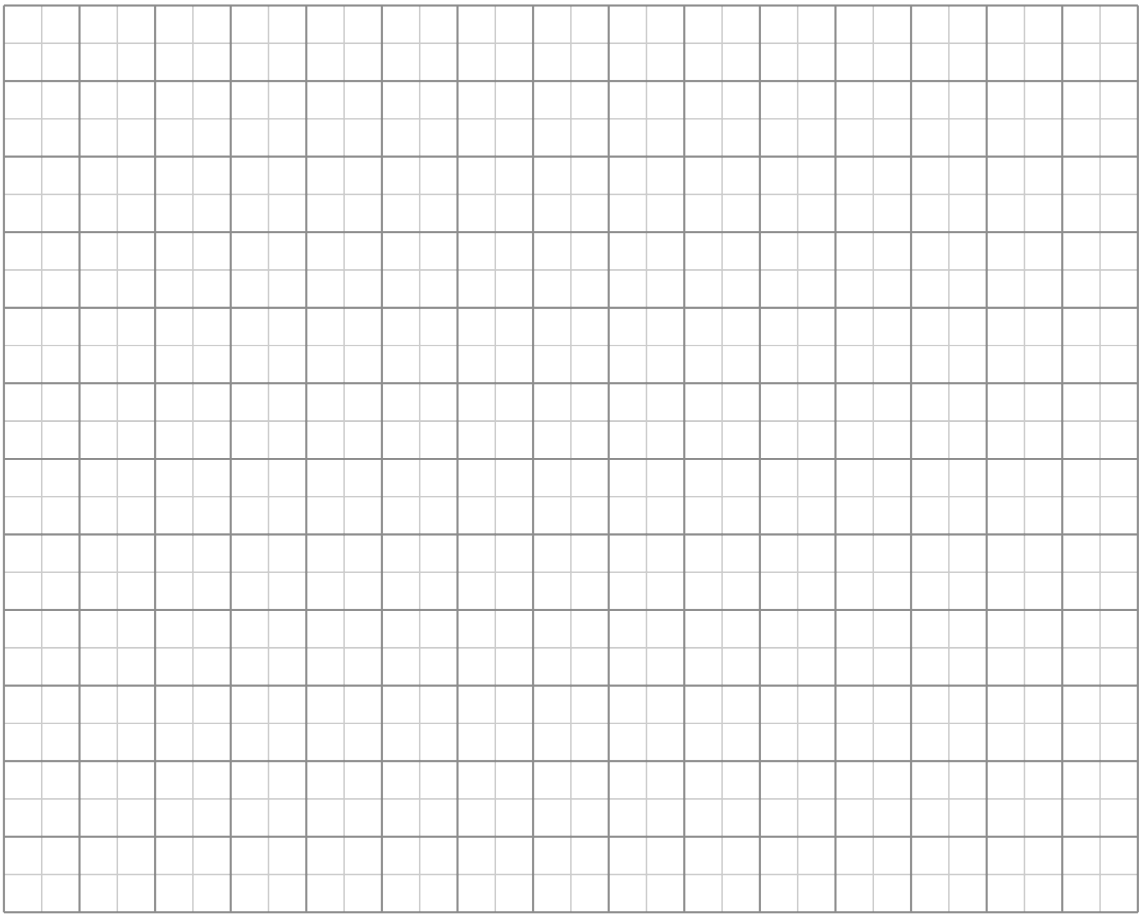


Figure 7.6: Illustration of the relation in Remark 7.7.

Theorem 7.7 (Construction of the Regulated Integral)**.**

Suppose $f \in \mathcal{R}[a, b]$ *and let* $(f_n) \subseteq \mathcal{S}[a, b]$ *be a sequence of step-functions such that* $f_n \to f$ *uniformly as* $n \to +\infty$.

Then the sequence $(\mathcal{I}(f_n)) \subseteq \mathbb{R}$ *of step-function integrals converges and the limit is independent of the particular sequence* (f_n).[a,b]

[a]The Regulated Integral is also sometimes called the Cauchy-Riemann Integral.

[b]The integral $\mathcal{I}(f_n)$ is constructed in Definition/Theorem 7.2.

Proof.

Definition 7.8 (Regulated Integral)**.**

Let $f \in \mathcal{R}[a,b]$ and $(f_n) \subseteq \mathcal{S}[a,b]$ be an arbitrary sequence such that $f_n \to f$ uniformly as $n \to +\infty$. Then, the ***regulated integral***[a,b] *is defined by*

$$\int_a^b f(x)\mathrm{d}x := \lim_{n\to+\infty} \mathcal{I}(f_n).$$

[a]The Regulated Integral is also sometimes called the Cauchy-Riemann Integral.

[b]The regulated integral was introduced by Dieudonne in *Foundations of Modem Analysis* (1960) and N. Bourbaki *Fonctions d'une Variable Réelle* (1976).

We agree on the following conventions

- We may write
$$\int_{[a,b]} f(x)\mathrm{d}x = \int_a^b f(x)\mathrm{d}x.$$
- If $a > b$, we define
$$\int_a^b f(x)\mathrm{d}x := -\int_b^a f(x)\mathrm{d}x.$$
- We also set
$$\int_a^a f(x)\mathrm{d}x = 0.$$

Theorem 7.8 (Properties of the Regulated Integral).

Let $f, g \in \mathcal{R}[a,b]$. Then

(i) for all $\lambda, \mu \in \mathbb{R}$, we have

$$\int_a^b \big(\lambda f(x) + \mu g(x)\big) dx = \lambda \int_a^b f(x)dx + \mu \int_a^b g(x)dx.$$

(ii) for any $c \in [a,b]$, we have

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx.$$

(iii) if $f(x) \geq 0$ for all $x \in [a,b]$ then

$$\int_a^b f(x)dx \geq 0.$$

(iv) if $f(x) \geq g(x)$ for all $x \in [a,b]$, then

$$\int_a^b f(x)dx \geq \int_a^b g(x)\mathrm{d}x.$$

(v) $\left|\int_a^b f(x)\mathrm{d}x\right| \leq \int_a^b |f(x)|\mathrm{d}x$*,*

(vi) $\left|\int_a^b f(x)\mathrm{d}x\right| \leq |b-a|\|f\|_\infty$*, and*

(vii) $\left|\int_a^b f(x)g(x)\mathrm{d}x\right| \leq \int_a^b |g(x)||f(x)|\mathrm{d}x \leq \|g\|_\infty \int_a^b |f(x)|\mathrm{d}x.$

Remark 7.8. *The last Theorem implies in particular that the Regulated Integral (Cauchy–Riemann Integral) is a linear map from* $PC[a,b]$ *to* $\mathbb{R}$ *in the sense of Definition A.15.*

Exercise 7.6. *Prove Theorem 7.8. (You should attempt this if you feel well prepared and have scored* 80 *or higher in Analysis 1.)*

7.2.1 Higher dimensions

It is clear that one can develop a similar theory on cuboids $Q := [a_1, b_1] \times \ldots \times [a_d, b_d]$ by using the characteristic functions

$$\chi_Q(x_1, \ldots, x_d) = \chi_{[a_1,b_1]}(x_1) \cdot \ldots \cdot \chi_{[a_d,b_d]}(x_d).$$

One needs to say what regulated functions are, define their integral and prove that some well-known classes as $C(Q)$ are regulated.

To treat more general domains, one needs to prove that certain classes of sub-sets of $\mathbb{R}^d$ can be well approximated by cuboids and then uses a second limit process to define the integral.

7.3 The Riemann-Integral

To get to one of the most important integral definitions, we need to introduce some notation.

Definition 7.9 (Upper- and lower integral)**.**
Let $f : [a, b] \to \mathbb{R}$*. Then, we say that*

$$\overline{\int_a^b} f(x)\mathrm{d}x = \inf\left\{ \int_a^b h(x)\mathrm{d}x \;:\; h \in \mathcal{S}[a, b],\; h \geq f \right\}$$

is the ***upper integral*** *of* f *and*

$$\underline{\int_a^b} f(x)\mathrm{d}x = \sup\left\{ \int_a^b h(x)\mathrm{d}x \;:\; h \in \mathcal{S}[a, b],\; h \leq f \right\}$$

the ***lower integral*** *of* f*.*

Remark 7.9. *We can reformulate the above definition in a more compact form*[2]*: Consider* $f : [a, b] \to \mathbb{R}$ *and define the* ***upper integral***

$$\overline{\mathcal{I}(f)} := \overline{\int_a^b} f(x)\mathrm{d}x = \inf_{\substack{h \in \mathcal{S}[a,b] \\ h \geq f}} \mathcal{I}(h)$$

and the ***lower integral***

$$\underline{\mathcal{I}(f)} := \underline{\int_a^b} f(x)\mathrm{d}x = \sup_{\substack{h \in \mathcal{S}[a,b] \\ h \leq f}} \mathcal{I}(h).$$

Example 7.2. *For any* $f \in \mathcal{S}[a, b]$*, we have that*

$$\overline{\int_a^b} f(x)\mathrm{d}x = \underline{\int_a^b} f(x)\mathrm{d}x = \int_a^b f(x)\mathrm{d}x.$$

[2] I am presenting this notation here since some of the solutions to older exam questions use it. You may entirely use the notation in Definition 7.9. However, for proofs it is sometimes suitable to use the shorter notation to safe some space.

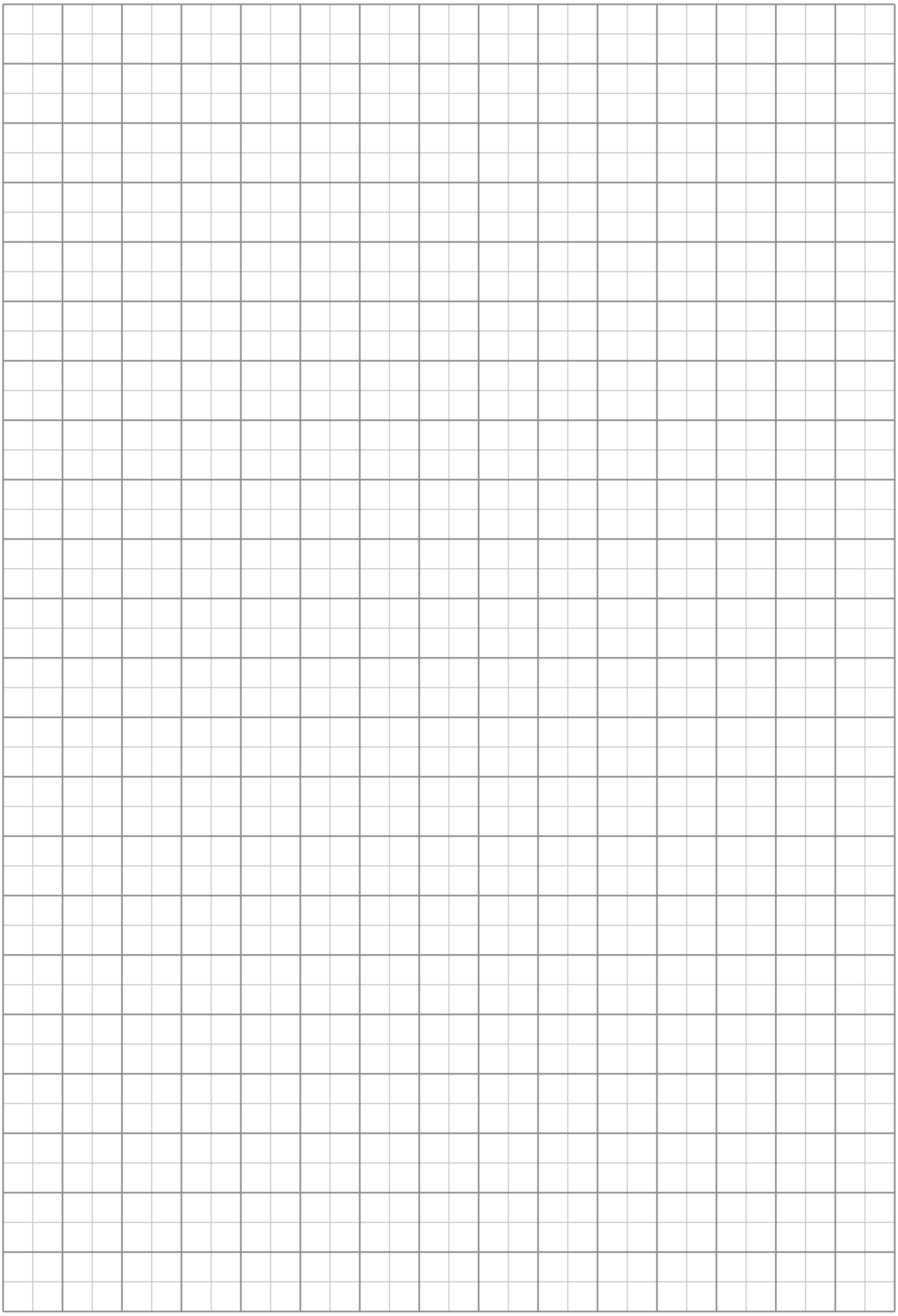


Figure 7.7: A bit of thinking on upper- and lower integrals

Remark 7.10. *We clearly have that*

$$\underline{\int_a^b} f(x)\mathrm{d}x \leq \overline{\int_a^b} f(x)\mathrm{d}x \tag{7.3.1}$$

for any function $f : [a, b] \to \mathbb{R}$*.*

Definition 7.10 (Riemann[a] integrability)**.**

A bounded function $f : [a, b] \to \mathbb{R}$ *is said to be* ***Riemann-integrable*** *if and only if*

$$\overline{\int_a^b} f(x)\mathrm{d}x = \underline{\int_a^b} f(x)\mathrm{d}x.$$

In that case, we set

$$\int_a^b f(x)\mathrm{d}x := \overline{\int_a^b} f(x)\mathrm{d}x.$$

[a]Bernhard Riemann (1826–1866), German mathematician.

As on p. 196, we agree on the following conventions

- We may write
$$\int\limits_{[a,b]} f(x)\mathrm{d}x = \int\limits_a^b f(x)\mathrm{d}x.$$
- If $a > b$, we define
$$\int\limits_a^b f(x)\mathrm{d}x := -\int\limits_b^a f(x)\mathrm{d}x.$$
- We also set
$$\int\limits_a^a f(x)\mathrm{d}x = 0.$$

Example 7.3. *Not all functions are Riemann integrable. The Dirichlet function* $f(x) := \chi_{\mathbb{Q}}(x)$ *is not Riemann-integrable on any compact interval* $[a,b] \subseteq \mathbb{R}$*,* $a < b$*.*
We have that for any step-function $h \in \mathcal{S}[a,b]$ *with* $h \geq f$ *must hold that* $h(x) \geq 1$ *for all but finitely many* $x \in [a,b]$*. Thus,*

Further, for all $h \in \mathcal{S}[a,b]$ *with* $h \leq f$ *must hold that* $h(x) \leq 0$ *for all but finitely many* $x \in [a,b]$*. Thus,*

Example 7.4. *The most basic example of a Riemann-integrable function is a step-function. See Example 7.2.*

To find more classes of functions which are Riemann-integrable, we state that a close investigation of the proof of Theorem 7.5 yields, with some small changes

Theorem 7.9.
Let $f \in C[a,b]$. Then, for every $\varepsilon > 0$, there exist $s_1, s_2 \in \mathcal{S}[a,b]$ such that

$$\forall x \in [a,b] \; : \; s_1(x) \leq f(x) \leq s_2$$

(shorter $s_1 \leq f \leq s_2$) and

$$\|s_1 - s_2\|_\infty < \varepsilon.$$

Now we inspect the definition of Riemann integrability and obtain the following characterisation of bounded Riemann-integrable functions.

Theorem 7.10 (Characterisation of Riemann integrability)**.**
A function $f : [a,b] \to \mathbb{R}$ is Riemann integrable if and only if it is bounded and for every $\varepsilon > 0$ there exist step-functions $s_1, s_2 \in \mathcal{S}[a,b]$ such that $s_1 \leq f \leq s_2$ and

$$\int_a^b s_2(x)\mathrm{d}x - \int_a^b s_1(x)\mathrm{d}x < \varepsilon.$$

Form that, we get the immediate corollary

Corollary 7.3 (Riemann integrability of regulated functions)**.**
Let $f \in \mathcal{R}[a, b]$, then f is Riemann integrable and

$$\underbrace{\int_a^b f(x)\mathrm{d}x}_{\text{Riemann}} = \underbrace{\lim_{n\to+\infty} \mathcal{I}(f_n)}_{\text{Regulated}}$$

where $(f_n) \subseteq \mathcal{S}[a, b]$ is an arbitrary sequence with $f_n \to f$ uniformly.

From that and Theorem 7.5 and Corollary 7.2 (or combining Theorem 7.9 with Theorem 7.10) we obtain

Corollary 7.4.
If $f \in PC[a, b]$, then f is Riemann integrable and

$$\underbrace{\int_a^b f(x)\mathrm{d}x}_{\text{Riemann}} = \underbrace{\lim_{n\to+\infty} \mathcal{I}(f_n)}_{\text{Regulated}}$$

where $(f_n) \subseteq \mathcal{S}[a, b]$ is an arbitrary sequence with $f_n \to f$ uniformly.

and

Corollary 7.5.
If $f \in C[a, b]$, then f is Riemann integrable and

$$\underbrace{\int_a^b f(x)\mathrm{d}x}_{\text{Riemann}} = \underbrace{\lim_{n\to+\infty} \mathcal{I}(f_n)}_{\text{Regulated}}$$

where $(f_n) \subseteq \mathcal{S}[a, b]$ is an arbitrary sequence with $f_n \to f$ uniformly.

Now one could think that Riemann Integral and Regulated Integral are the same notions. That is is, however, not the case. In the next example, we show that there exist Riemann-integrable functions which are not regulated. Thus, the Riemann integral is more general than the integral for regulated functions.

Reading 8. *Example 7.5 forms the reading of this week. Please study it carefully with some extra paper and a pencil. Work out all the details you need and try to draw some illuminating pictures.*

Example 7.5. *We give an example of a function which is Riemann integrable but not regulated. Consider* $f : [0,1] \to \mathbb{R}$*, defined by*

$$f(x) = \begin{cases} 1 & : \ x = \frac{1}{n}, \ n \in \mathbb{N} \\ 0 & : \ x \neq \frac{1}{n}, \ n \in \mathbb{N} \end{cases} .$$

Though f *is not regulated, the Riemann integral exists*

$$\int_0^1 f(x)\mathrm{d}x = 0.$$

First we discuss that the Riemann integral indeed exists and is equal to 0*. To this end, we define*

$$h_n(x) = \begin{cases} 1 & : \ x \in \left[0, \frac{1}{n}\right] \\ f(x) & : \ x \in \left(\frac{1}{n}, 1\right] \end{cases} .$$

By construction, we have $h_n \geq f$*. Further, since* h_n *is constant equal to* 1 *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

$$\overline{\int_a^b} f(x)\mathrm{d}x = \inf_{\substack{h \in \mathcal{S}[0,1] \\ h \geq f}} \int_0^1 h(x)dx \leq \inf_{n \in \mathbb{N}} \int_0^1 h_n(x)dx = \inf_{n \in \mathbb{N}} \frac{1}{n} = 0.$$

Since $0 \leq f(x)$, we also have

$$0 \leq \sup_{\substack{h \in \mathcal{S}([a,b]) \\ h \leq f}} \int\limits_0^1 h(x)dx = \underline{\int_a^b} f(x)\mathrm{d}x.$$

By (7.3.1), we get

$$0 \leq \underline{\int_a^b} f(x)\mathrm{d}x \leq \overline{\int_a^b} f(x)\mathrm{d}x \leq 0.$$

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

$$g(x) = \sum_{i=1}^{N} c_i \chi_{I_i}(x)$$

with $I_1 \cup \ldots \cup I_N = [0,1]$ and $I_i \cap I_j = \varnothing$ for $i \neq j$ (see Proposition 7.2) such that $\|f - g\|_\infty < \varepsilon$. Then, the I_i, $i = 1, \ldots, N$ define a partition

$$0 < x_1 < x_2 < \ldots < x_n = b$$

of $[0,1]$. 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

$$|f(x_1) - g(x_1)| = |c_1| \quad \textit{and} \quad |f(x_2) - g(x_2)| = |1 - c_1|. \tag{7.3.2}$$

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 (7.3.2) that implies $\varepsilon > \frac{1}{2}$. Thus, f can not be uniformly approximated by a step function.

7.3.1 Properties of the Riemann Integral

Using Theorem 7.10, we can show

Theorem 7.11 (Properties of the Riemann Integral)**.**
Let $f, g : [a, b] \to \mathbb{R}$ be Riemann integrable Then

(i) for $\lambda, \mu \in \mathbb{R}$, the function $\lambda f + \mu g$ is Riemann-integrable and

$$\int_a^b \lambda f(x) + \mu g(x)\mathrm{d}x = \lambda \int_a^b f(x)\mathrm{d}x + \mu \int_a^b g(x)\mathrm{d}x.$$

(ii) the functions $f\big|_{[a,c]}, f\big|_{[c,b]}$ [a] are Riemann-integrable for any $c \in [a, b]$ and

$$\int_a^b f(x)\mathrm{d}x = \int_a^c f(x)\mathrm{d}x + \int_c^b f(x)\mathrm{d}x.$$

(iii) $\int_a^b f(x)\mathrm{d}x \geq 0$ if $f(x) \geq 0$ for all $x \in [a, b]$.

(iv) $\int_a^b f(x)\mathrm{d}x \geq \int_a^b g(x)\mathrm{d}x$ if $f(x) \geq g(x)$ for all $x \in [a, b]$.

(v) the function $|f|$ is Riemann-integrable and

$$\left|\int_a^b f(x)\mathrm{d}x\right| \leq \int_a^b |f(x)|\mathrm{d}x.$$

(vi) the function f is bounded and

$$\left|\int_a^b f(x)\mathrm{d}x\right| \leq |b - a| \|f\|_\infty.$$

(vii) the function fg is Riemann-integrable and

$$\left|\int_a^b f(x)g(x)\mathrm{d}x\right| \leq \int_a^b |g(x)||f(x)|\mathrm{d}x \leq \|g\|_\infty \int_a^b |f(x)|\mathrm{d}x.$$

[a] Let $f : A \to B$ and $C \subseteq A$. Then, $f\big|_C$ denotes the restriction of f to the set C, i.e the function $g : C \to B$ with $g(x) = f(x)$ for $x \in C$.

7.4 The fundamental theorem of calculus and primitives

We start with the introduction of the notion of a

Definition 7.11 (Primitive of a function).
Let $f : [a, b] \to \mathbb{R}$ *be a function. A function* $F : [a, b] \to \mathbb{R}$ *is called a* ***primitive*** *or* ***indefinite integral***[a] *of* f *if* F *is differentiable and* $F'(x) = f(x)$ *for all* $x \in [a, b]$.

[a]Some people say antiderivative, though that is not a good name.

By our previous considerations, the following theorem holds true for the notion of Regulated Integral and Riemann-Integral and both give the same results.

Theorem 7.12 (Fundamental Theorem of Calculus).
Let $f \in C[a, b]$ *and define* $F : [a, b] \to \mathbb{R}$ *by*

$$F(x) := \int_a^x f(y)\mathrm{d}y. \tag{7.4.1}$$

Then, F *is (uniformly) continuous on* $[a, b]$ *and differentiable on* $[a, b]$ *and* $F'(x) = f(x)$ *for all* $x \in [a, b]$.

Before we prove this result, let us state some other formulation and corollaries to clear up the notion of primitive and its relation to definite integrals.

Remark 7.11. *In (7.4.1), the choice of the lower limit does not matter. In fact, one can prove the theorem for*

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

where $x_0 \in [a,b]$ *is arbitrary but fixed. All these* F *are primitives of* f*.*

An equivalent formulation of Theorem 7.12 is

Theorem 7.13.
Let $F \in C^1[a,b]$ *with* $F'(x) = f(x)$ *for all* $x \in [a,b]$*. Then*

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

Remark 7.12. *We also sometimes write*

$$\int_a^b f(x)\mathrm{d}x = F(x)\Big|_a^b = \Big[F(x)\Big]_a^b = F(b) - F(a).$$

Proof of Theorem 7.13.

We define an auxiliary function $G : [a,b] \to \mathbb{R}$ by

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

Then,

$$G'(x) = f(x) - f(x) = 0.$$

by Theorem 7.12. Thus, by Lemma 5.1, G is constant and since $G(a) = 0$ is identically zero. This concludes the proof.

An immediate corollary (actually just a rewrite) of Theorem 7.13 is the following reconstruction formula

Corollary 7.6.
Let $f \in C^1[a, b]$*, then*

$$f(x) = f(a) + \int_a^x f'(y)\mathrm{d}y.$$

Remark 7.13. *The usual notation for a primitive* F *of* f *is*

$$F(x) = \int f(x)\mathrm{d}x + C. \tag{7.4.2}$$

The C *is due to the fact that two primitives may differ by an additive constant as one can see from Definition 7.11 and Remark 7.11.*
Be aware that the $=$ *sign in (7.4.2) equates a function with a set and is therefore in principal without meaning. The meaning we give it here is that for* F*, you may choose any*

$$\int_a^x f(y)\mathrm{d}y + C$$

for any C*. Similarly, as discussed in Remark 7.11, one can choose different lower limits as discussed in .*

Remark 7.14. *I would like to draw attention to the fact that not all Riemann integrable functions have primitives. The signum function*

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

is integrable on $[-1, 1]$ *but there exists no function* F *such that* $F'(x) = \mathrm{sgn}(x)$ *for all* $x \in [-1, 1]$*. 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 7.12 guarantees the existence of a primitive.*

Proof of Theorem 7.12.

7.5 Some Rules for Integration

Differentiation is mechanics, integration is art.

I assume that you are familiar with these rules from school and mathematical Methods I & II. However, for your convenience I state them here and give some proofs for the sake of completeness.

The next theorem is in some sense the inverse version of the product-rule for differentiation. See Theorem 5.2. Given f and g are differentiable, then we get

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

and, slightly handwavingly, we obtain

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

which leads to

$$\int f \cdot g' \mathrm{d}x = f \cdot g - \int f' \cdot g \mathrm{d}x.$$

Using Theorem 7.12, we can prove a version for definite integrals.

Theorem 7.14 (Integration by parts)**.**
Suppose $f, g \in C^1[a, b]$. *Then*

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

Proof.

The proof follows from the product rule in Theorem 5.2 and integration over $[a, b]$ using the Fundamental Theorem of Calculus, Theorem 7.12.

The correct use of Theorem 7.14 takes some exercise. With experience, the reader will learn what kinds of integrals are to be attacked by integration by parts.

Example 7.6. *Let us suppose that we want to compute the integral*

$$\int_0^\pi \cos^2(x)\mathrm{d}x.$$

We obtain

$$\begin{aligned} \int_0^\pi \cos^2(x)\mathrm{d}x &= \Big[\sin(x)\cos(x)\Big]_0^\pi + \int_0^\pi \sin^2(x)\mathrm{d}x \\ &= \int_0^\pi \sin^2(x)\mathrm{d}x. \end{aligned} \tag{7.5.1}$$

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*

$$\begin{aligned} \int_0^\pi \sin^2(x)\mathrm{d}x &= \int_0^\pi \big(1 - \cos^2(x)\big)\mathrm{d}x = \int_0^\pi 1\mathrm{d}x - \int_0^\pi \cos^2(x)\mathrm{d}x \\ &= \pi - \int_0^\pi \cos^2(x)\mathrm{d}x. \end{aligned}$$

With that, from (7.5.1)*, we finally obtain*

$$\int_0^\pi \cos^2(x)\mathrm{d}x = \frac{\pi}{2}$$

and in fact we also showed that

$$\int_0^\pi \sin^2(x)\mathrm{d}x = \frac{\pi}{2}.$$

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 7.15 (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

$$\int_{g(a)}^{g(b)} f(x)\mathrm{d}x = \int_a^b f(g(x))g'(x)\mathrm{d}x. \tag{7.5.2}$$

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 7.11), we have that the right hand side of (7.5.2) is well defined as is the left hand side. Let F be a primitive of f. Since F is differentiable (Theorem 7.12), we obtain that $F\circ g : [a,b] \to \mathbb{R}$ is differentiable and by the chain rule (Theorem 5.3), we get

$$\frac{d}{\mathrm{d}x}F(g(x)) = f(g(x))g'(x)\mathrm{d}x.$$

Thus, we have that

$$\int_a^b f(g(x))g'(x)\,\mathrm{d}x = \int_a^b \frac{d}{dx}F(g(x))\,\mathrm{d}x.$$

By Corollary 7.6 (or Theorem 7.13), we obtain

$$\int_a^b \frac{d}{dx}F(g(x))\,\mathrm{d}x = F(g(b)) - F(g(a)) = \int_{g(a)}^{g(b)} f(x)\mathrm{d}x.$$

This concludes the proof. □

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 sight of the slightest problem as integration is really an art. A fantastic book on the matters of this section is [11].

Example 7.7. *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*

$$\begin{aligned}\int\limits_a^b e^{Ax+B}\mathrm{d}x &= \frac{1}{A}\int\limits_a^b Ae^{Ax+B}\mathrm{d}x = \frac{1}{A}\int\limits_a^b g'(x)e^{g(x)}\mathrm{d}x \\ &= \frac{1}{A}\int\limits_{Aa+B}^{Ab+B} e^y \mathrm{d}y = \frac{1}{A}\left(e^{Ab+B} - e^{Aa+B}\right).\end{aligned}$$

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

$$\int x^2 e^{-x^3}\mathrm{d}x, \quad \text{or} \quad \int xe^{x^2}\mathrm{d}x.$$

Do you see how to solve it? The "trick" is that one realises that these integrals are essentially of the form $\int f'(x)e^{f(x)}\mathrm{d}x$ which makes the solution obvious:

$$\int f'(x)e^{f(x)}\mathrm{d}x = e^{f(x)} + C.$$

Compare also Theorem 7.15 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:

$$\int x^2 e^{-x^3}\mathrm{d}x = -\frac{1}{3}\int(-3x^2)e^{-x^3}\mathrm{d}x = -\frac{1}{3}e^{-x^3} + C.$$

The same "tricks" work obviously for

$$\int f'(x)\sin(f(x))\mathrm{d}x, \quad \text{and} \quad \int f'(x)\cos(f(x))\mathrm{d}x.$$

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.

7.6 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 on the related Problem Sheet.

In this section we consider the questions under which conditions we can exchange integrals with limits, i.e. given a sequence (f_n) of functions we ask: When is

$$\lim_{n\to+\infty}\int_a^b f_n(x)dx = \int_a^b \lim_{n\to+\infty} f_n(x)dx \tag{7.6.1}$$

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.

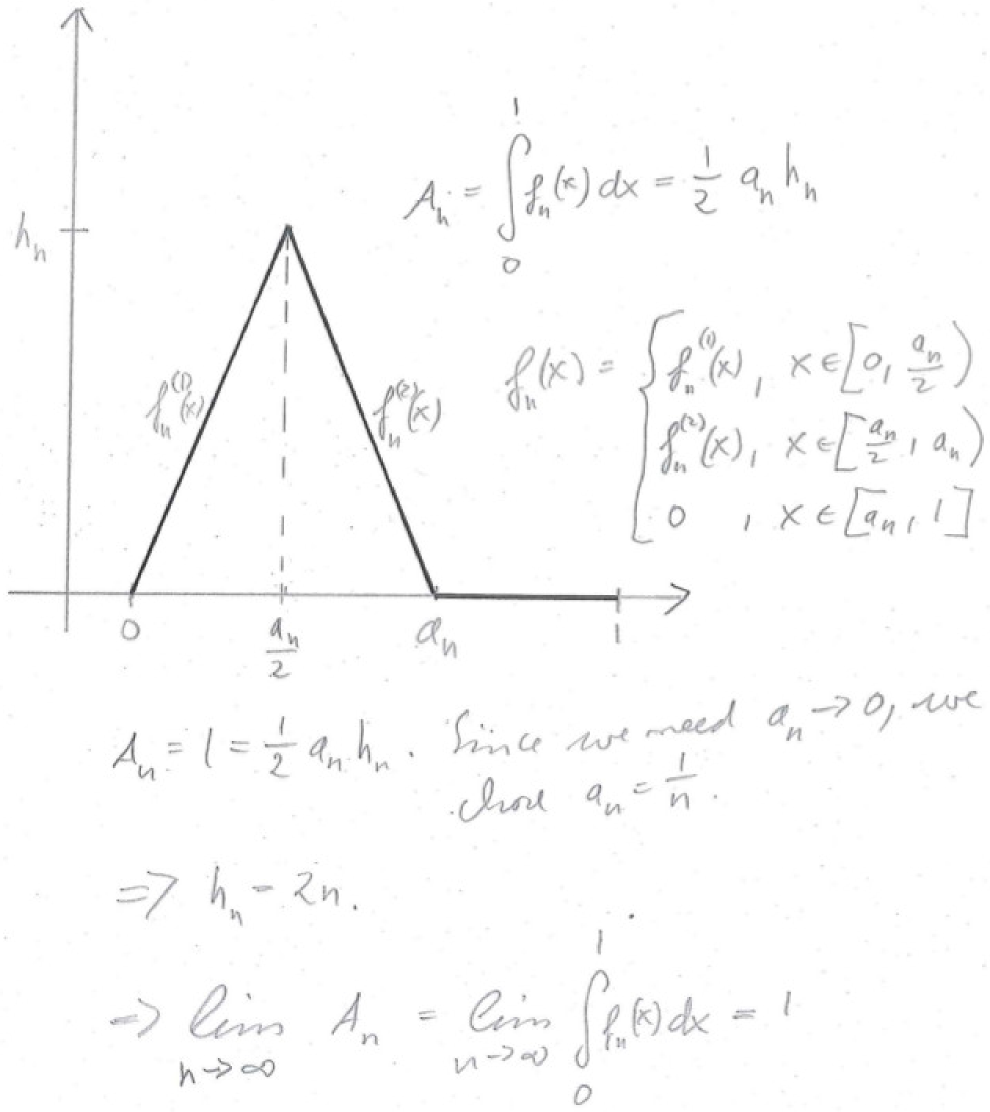


Figure 7.8: A sequence of continuous f_n with $f_n \to 0$ pointwise but $\int\limits_0^1 f_n(x)dx = 1.$

In fact, one can even construct a sequence of continuous functions (f_n) which converge pointwise to $f \equiv 0$ but the limit of the integrals grows unbounded.

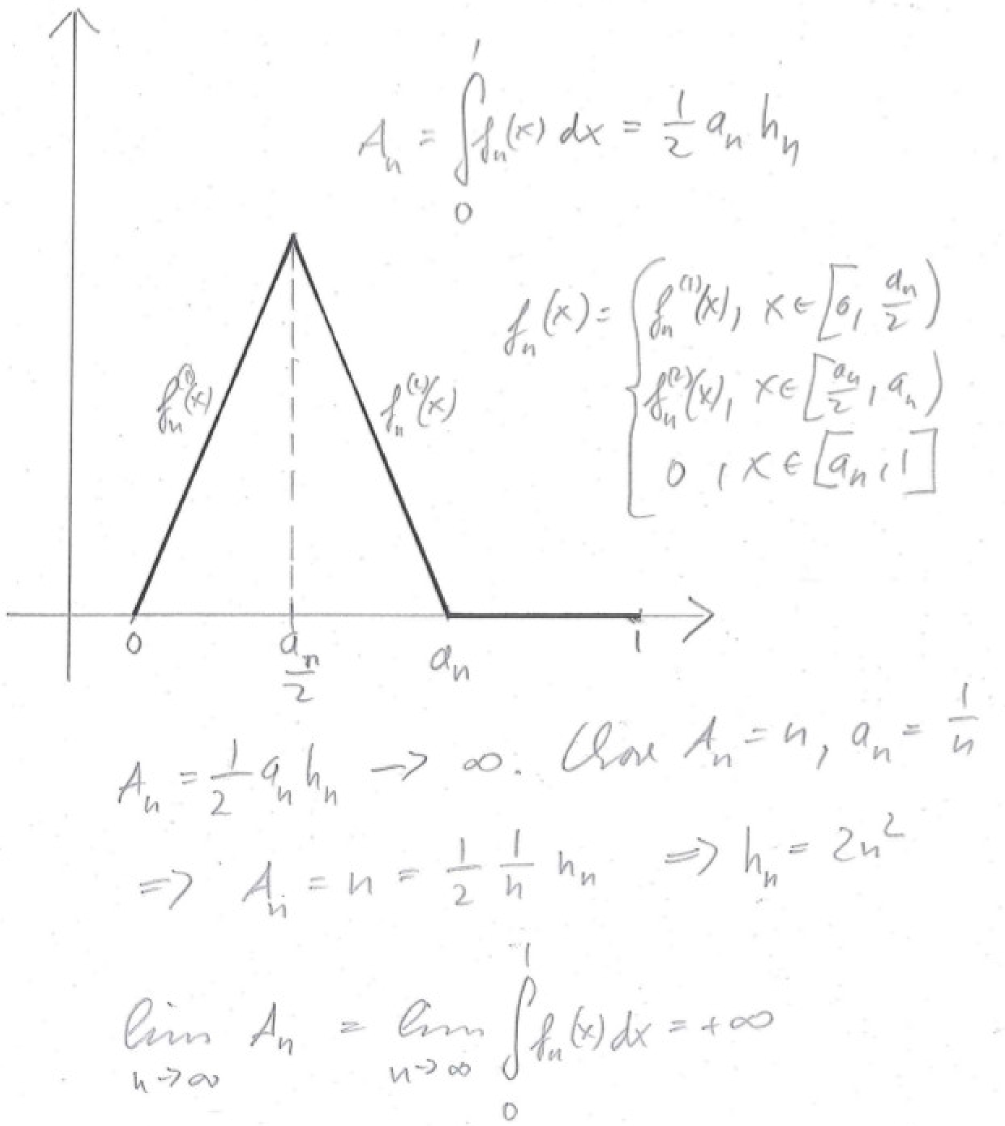


Figure 7.9: A sequence of continuous f_n with $f_n \to 0$ pointwise but $\int_0^1 f_n(x)dx = +\infty$.

Thus, in the subsequent theorems, we will be concerned with uniform convergence.

Theorem 7.16.

Let $[a, b]$ *be a compact interval and let* $(f_n) \subseteq \mathcal{R}[a, b]$*. Suppose that there exists* $f : [a, b] \to \mathbb{R}$ *such that* $f_n \to f$ *uniformly. Then*

$$\lim_{n\to+\infty} \int_a^b f_n(x)dx = \int_a^b \lim_{n\to+\infty} f_n(x)dx. \tag{7.6.2}$$

Remark 7.15. *Since regulated functions are Riemann integrable, we get, under the hypothesis of Theorem 7.16, that*

$$\lim_{n\to+\infty} \int_a^b f_n(x)dx = \int_a^b \lim_{n\to+\infty} f_n(x)dx.$$

for Riemann integrals.

To prove Theorem 7.16, we first need another theorem.

Theorem 7.17.

Let $(f_n) \subseteq \mathcal{R}[a, b]$ *and 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 $\varepsilon > 0$ be arbitrary. 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 $n_0 \in \mathbb{N}$ 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. $\square$

Remark 7.16. *Theorem 7.17 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 $(f_n) \subseteq \mathcal{R}[a,b]$ 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 $n_0 \in \mathbb{N}$ such that*

$$\forall m,n \geq n_0 \;:\; \sup_{x\in[a,b]} |f_n(x) - f_m(x)| < \varepsilon.$$

Thus, for all $x \in [a,b]$, we have that $(f_n(x))$ 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 7.17, we have that $f \in \mathcal{R}[a,b]$. □

Proof of Theorem 7.16.

By Theorem 7.17, 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

$$\begin{aligned} \left|\int_a^b (f_n(x) - f(x))dx\right| &\leq \int_a^b |f_n(x) - f(x)|dx \\ &\leq |b-a| \sup_{x\in[a,b]} |f_n(x) - f(x)|. \end{aligned}$$

Since $f_n \to f$ uniformly, we have that for all $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that

$$\forall n \geq n_0 \;:\; \|f_n - f\|_\infty < \frac{\varepsilon}{|b-a|}.$$

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, (7.6.3) holds. (See Definition A.16.) □

Since continuous functions are regulated, we also get

Corollary 7.7.

Let $[a, b]$ *be a compact interval and let* $(f_n) \subseteq C[a, b]$*. Suppose that there exists* $f : [a, b] \to \mathbb{R}$ *such that* $f_n \to f$ *uniformly. Then*

$$\lim_{n\to+\infty} \int_a^b f_n(x)dx = \int_a^b \lim_{n\to+\infty} f_n(x)dx. \tag{7.6.3}$$

The integral can be interpreted as Regulated (Cauchy-Riemann) or Riemann integral.

Exercise 7.7. *Corollary 7.7 follows from Theorem 7.16 by Theorem 7.5. However, use the proof of Theorem 7.16 above as a blue print to prove Corollary 7.7 again in detail by bare hands.*

Chapter

8 Improper Integrals on on $\mathbb{R}^1$

Improper integrals are integrals over non-compact sets that are written as limits of integrals over compact sets.

Example 8.1. *Consider*

$$\int_{(0,1]} \frac{1}{\sqrt{x}} dx$$

is an improper integral since $\lim_{x \to 0+} \frac{1}{\sqrt{x}} = +\infty$*. The integral should be understood as*

$$\int_0^1 \frac{1}{\sqrt{x}} dx = \lim_{\varepsilon \to 0+} \int_\varepsilon^1 \frac{1}{\sqrt{x}} dx. \tag{8.0.1}$$

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* (8.0.1)*, 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 Regulated Integral (or its Riemann integral) which would of course depend on* ε*.*
So, (8.0.1) *says that we say that the integral over* $(0, 1]$ *exists if the limit* $\lim_{\varepsilon \to 0+} \mathcal{I}_\varepsilon(f)$ *exists.*

Example 8.2. *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. Note that this is not to say that

$$\lim_{R\to+\infty}\int_{-R}^R f(x)\mathrm{d}x$$

must exist as the latter could exist without the two limits

$$\lim_{R\to+\infty}\int_{-R}^0 f(x)dx \quad \text{and} \quad \lim_{R\to+\infty}\int_0^R f(x)dx$$

existing. Consider for example

$$f(x) = \begin{cases} \frac{1}{x} & : \; x \in \mathbb{R}\setminus[-1,1] \\ x & : \; x\in[-1,1] \end{cases}.$$

To make the above a bit more precise, we say that $\int_I f(x)\mathrm{d}x$ is an improper integral of $f : I \to \mathbb{R}$ if $\mathrm{diam}(I) = +\infty$ or there exists an $x_0 \in \mathrm{cl}(I)$ such that $\lim_{x\to x_0}|f(x)| = +\infty$. Examples are

$$\int_0^1 \frac{1}{x}\mathrm{d}x, \quad \int_1^{+\infty}\frac{1}{x}\mathrm{d}x.$$

Definition 8.1 (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) be a sequence of compact intervals $I_n \subseteq I$ with $I_{n+1} \supseteq I_n$ such that $\mathrm{cl}(I) = \bigcup_{n=1}^{\infty} I_n$. If $f\big|_{I_n}$ is regulated (or Riemann integrable) for every $n \geq 1$, then $\int_I f(x)dx$ is said to be convergent (or f is integrable over I) if the sequence $(A_n) \subseteq \mathbb{R}$, defined by

$$A_n := \int\limits_{I_n} f(x)dx,$$

converges.[a]

[a]By $\mathrm{cl}(I)$ we denote the closure of I, i.e. the smallest closed set that contains I.

A specialisation of the above integral is given by

Definition 8.2 (Special case of Definition 8.1)**.**
Let $f : [0, +\infty) \to \mathbb{R}$ be continuous. Then, we say that

$$\int\limits_0^{+\infty} f(x)\mathrm{d}x$$

converges *if and only if*

$$\lim_{R\to+\infty} \int\limits_0^R f(x)\mathrm{d}x$$

exists.

Example 8.3. *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

$$\int_{I_n} f(x) dx = \sum_{k=1}^{n} (-1)^k \int_{k\pi}^{(k+1)\pi} \left| \frac{\sin(x)}{x} \right| dx.$$

In analogy to our treatment of sequences, we further introduce the notion of absolute integrability.

Definition 8.3 (Absolute Integrability on $\mathbb{R}$).
Let $I \subseteq \mathbb{R}$ be an interval and let (I_n), be a sequence of compact intervals $I_n \subseteq I$ such that $\mathrm{cl}(I) = \bigcup_{n=1}^{\infty} I_n$, $I_{n+1} \supseteq I_n$. *Then, a function $f : I \to \mathbb{R}$ with the property that $f\big|_{I_n}$ is regulated (or Riemann-integrable) for any $n \in \mathbb{N}$ is said to be* ***absolutely integrable***[a] *if the sequence $(A_n)_{n\in\mathbb{N}_0}$, defined by*

$$A_n := \int_{I_n} |f(x)| dx,$$

converges.[b]

[a] We also say absolutely convergent.
[b] By $\mathrm{cl}(I)$ we denote the closure of I, i.e. the smallest closed set that contains I.

In the spirit of Definition 8.2, we state a specialisation of the above integral by

Definition 8.4 (Special case of Definition 8.3).
Let $f : [0, +\infty) \to \mathbb{R}$ be continuous. Then, we say that

$$\int_0^{+\infty} f(x)\mathrm{d}x$$

converges absolutely *if and only if*

$$\lim_{R\to+\infty} \int_0^R |f(x)|\mathrm{d}x$$

exists.

Example 8.4. *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 8.5. *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

$$\int_1^{+\infty} \left| \frac{\sin(x)}{x^2} \right| dx \leq \int_1^{+\infty} \frac{dx}{x^2} < +\infty$$

as seen above.

Example 8.6. *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.

8.1 Appendix : The closure of sets

Definition 8.5 (Closure).
Let $\Omega \subseteq \mathbb{R}^d$ 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 8.1. *By definition we have that $\mathrm{cl}(\Omega)$ is closed and that $\Omega \subseteq \mathrm{cl}(\Omega)$.*

Remark 8.2. *The closure $\mathrm{cl}(\Omega)$ of a set Ω is the smallest closed set that contains Ω, i.e.*

$$\mathrm{cl}(\Omega) = \bigcap_{\substack{C \supseteq \Omega \\ C \text{ closed}}} C.$$

Can you prove that?

Example 8.7. *Simple examples are $\mathrm{cl}((a,b)) = [a,b]$, $\mathrm{cl}(B_r(x_0)) = \{x \in \mathbb{R}^d : \|x - x_0\|_2 \leq r\}$. Can you find more yourself?*

CHAPTER

9 Differentiability and derivative on $\mathbb{R}^d$

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 $m \geq 1$.

9.1 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 9.1 (Total Derivative).
*Let $\Omega \subseteq \mathbb{R}^d$ be an open set and let $f : \Omega \to \mathbb{R}^m$ be a function. Then, f is said to be **differentiable at** $x_0 \in \Omega$ if there exists a linear map $L : \mathbb{R}^d \to \mathbb{R}^m$ and a function $r : \mathbb{R}^d \to \mathbb{R}^m$ such that*

$$f(x_0 + h) = f(x_0) + Lh + r(h) \tag{9.1.1}$$

and

$$\lim_{h\to 0} \frac{\|r(h)\|_2}{\|h\|_2} = 0. \tag{9.1.2}$$

The linear map L is called the total derivative of f and is denoted by $Df(x_0)$.

Remark 9.1. *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 9.1. *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

By that we have that f *is differentiable for all* x_0 *if we can show*

$$\lim_{h\to 0} \frac{\|r(h)\|_2}{\|h\|_2} = 0 \tag{9.1.3}$$

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 9.1. *Prove* (9.1.3) *in the previous example.*

One can give an alternative definition of differentiability by

Definition 9.2 (Total Derivative II)**.**
Let $\Omega \subset \mathbb{R}^d$ *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* $L : \mathbb{R}^d \to \mathbb{R}^m$ *and a function* $r : \mathbb{R}^d \to \mathbb{R}^m$ *such that*

$$\lim_{h \to 0} \frac{\|f(x_0 + h) - f(x_0) - Lh\|_2}{\|h\|_2} = 0. \tag{9.1.4}$$

Lemma 9.1.
Definitions 9.1 and 9.2 are equivalent.

Proof. [$\Rightarrow$] Let $\Omega \subseteq \mathbb{R}^d$ be an open set and let $f : \Omega \to \mathbb{R}^m$ be a function. Let f be differentiable by Definition 9.1. From (9.1.1), 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 (9.1.4) by (9.1.2). Thus, f is differentiable at x_0 by Definition 9.2.
[$\Leftarrow$]: Let $\Omega \subseteq \mathbb{R}^d$ be an open set and let $f : \Omega \to \mathbb{R}^m$ be a function. Let f be differentiable by Definition 9.2. Take the L from (9.1.1); 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 (9.1.4) by

$$\lim_{h \to 0} \frac{\|r(h)\|_2}{\|h\|_2} = \lim_{h \to 0} \frac{\|f(x_0 + h) - f(x_0) - Lh\|_2}{\|h\|_2} = 0.$$

$\square$

Lemma 9.2 (Uniqueness of the total derivative).

Let $\Omega \subseteq \mathbb{R}^d$ be open and $f : \Omega \to \mathbb{R}^m$ be differentiable at x_0. Then, L in Definition 9.2 (and by the last Lemma in Definition 9.1) is uniquely determined.

Proof. Let L, $M : \mathbb{R}^d \to \mathbb{R}^m$ be linear maps satisfying at x_0 (9.1.1). Then

$$\begin{aligned}
&\lim_{h\to 0} \frac{\|Lh - Mh\|_2}{\|h\|_2} \\
&\quad = \lim_{h\to 0} \frac{\| - f(x_0 + h) + f(x_0) + Lh - Mh + f(x_0 + h) - f(x_0)\|_2}{\|h\|_2} \\
&\quad \leq \lim_{h\to 0} \frac{\| - f(x_0 + h) + f(x_0) + Lh\|_2}{\|h\|_2} \\
&\qquad + \lim_{h\to 0} \frac{\| - Mh + f(x_0 + h) - f(x_0)\|_2}{\|h\|_2} = 0.
\end{aligned}$$

Now, let $v \in \mathbb{R}^d \setminus \{0\}$ 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. $\square$

Remark 9.2. *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 9.3. *As in the 1D case, we can write (9.1.1) 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:

$$f(x_0 + h) = f(x_0) + Lh + o(h).$$

9.1.1 Directional Derivative

Restricting the directions of h in Definition 9.1, we can give

Definition 9.3 (Directional Derivative)**.**
Let $v \in \mathbb{R}^d \setminus \{0\}$, $\Omega \subseteq \mathbb{R}^d$ open, $x_0 \in \Omega$, and $f : \Omega \to \mathbb{R}^m$. Then, if

$$\lim_{t \to 0} \frac{1}{t}\big(f(x_0 + tv) - f(x_0)\big) \tag{9.1.5}$$

exists, it is called the directional derivative of f in direction v at x_0 and denoted by $D_v f(x_0)$.

Theorem 9.1.
Let $\Omega \subseteq \mathbb{R}^d$ 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 $v \in \mathbb{R}^d \setminus \{0\}$ and we have $D_v f(x_0) = Df(x_0)v$.

9.1.2 Partial Derivatives

An important special case of directional derivatives, according to Definition 9.3, 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 $\mathbb{R}^d$, 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

$$\delta_{kj} = \begin{cases} 0 & : \ j \neq k \\ 1 & : \ j = k \end{cases}.$$

Then, letting $\tilde{e}_j$ be the canonical basis of $\mathbb{R}^m$, we can write

If f is differentiable (see Def. 9.1), we obtain from (9.1.5) that

$$\begin{aligned}
&\lim_{t\to 0} \frac{1}{t}\big(f(x_0 + te_k) - f(x_0)\big) \\
&\qquad = \sum_{j=1}^{m} \lim_{t\to 0} t^{-1}\big(f_j(x_0 + te_k) - f_j(x_0)\big)\tilde{e}_j \\
&\qquad = \begin{bmatrix} \lim_{t\to 0} \frac{1}{t}\big(f_1(x_0 + te_k) - f_1(x_0)\big) \\ \vdots \\ \lim_{t\to 0} \frac{1}{t}\big(f_m(x_0 + te_k) - f_m(x_0)\big) \end{bmatrix}.
\end{aligned}$$

We set

We give these special directional derivatives a name by

Definition 9.4 (Partial Derivatives)**.**
Let $\Omega \subseteq \mathbb{R}^d$ open, $p \in \Omega$, and $f : \Omega \to \mathbb{R}^m$. If the limit

$$\lim_{h\to 0} \frac{f_j(p_1, \ldots, p_{k-1}, p_k + h, p_{k+1}, \ldots, p_n) - f_j(p_1, \ldots, p_n)}{h} \tag{9.1.6}$$

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 9.4. *As you can see in (9.1.6), that means that the partial derivative $\frac{\partial f_j}{\partial x_k}(p)$ 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*

$$\frac{\partial f}{\partial x_k}(p) = \lim_{h\to 0} \frac{f(p_1, \ldots, p_{k-1}, p_k + h, p_{k+1}, \ldots, p_n) - f(p_1, \ldots, p_n)}{h},$$

where $p = [p_1, \ldots p_n]^T \in \Omega$. Compare Definition 5.1.

Example 9.2. *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:*

Exercise 9.2. *Calculate, using the definition, the partial derivative with respect to the second variable of $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = 2x^2 + 3y$, $(x_0, y_0) \in \mathbb{R}^2$.*

Theorem 9.2.
Let $\Omega \subseteq \mathbb{R}^d$ 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 9.5. *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*

$$f(x,y) = \begin{cases} \frac{x^2y}{x^4+y^2} & : \ x^2+y^2 \neq 0 \\ 0 & : \ x = y = 0 \end{cases}.$$

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 $(0, 0)$.

9.2 Meaning of Differentiability in $\mathbb{R}^d$

Reading 9. *This section constitutes this week's reading. Read the section carefully and work out all the details, i.e. redo calculations, build your own examples, etc. pp. Use GeoGebra or other programs to look at the examples yourself.*
If you have questions, ask your tutors, talk to the staff in the MLSC or come to me.

If $f : \Omega \to \mathbb{R}$, then the function graph

$$\left\{(x, f(x)) \in \mathbb{R}^{n+1} : x \in \mathbb{R}^d\right\}$$

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 $\mathbb{R}^d$, we distinguish the different planes. in $\mathbb{R}^4$ we have the hyperplanes which are $4 - 1 = 3$-dimensional, we have 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 $V = \{(x, Lx) : x \in \mathbb{R}^d\}$ 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 $\mathbb{R}^3$.

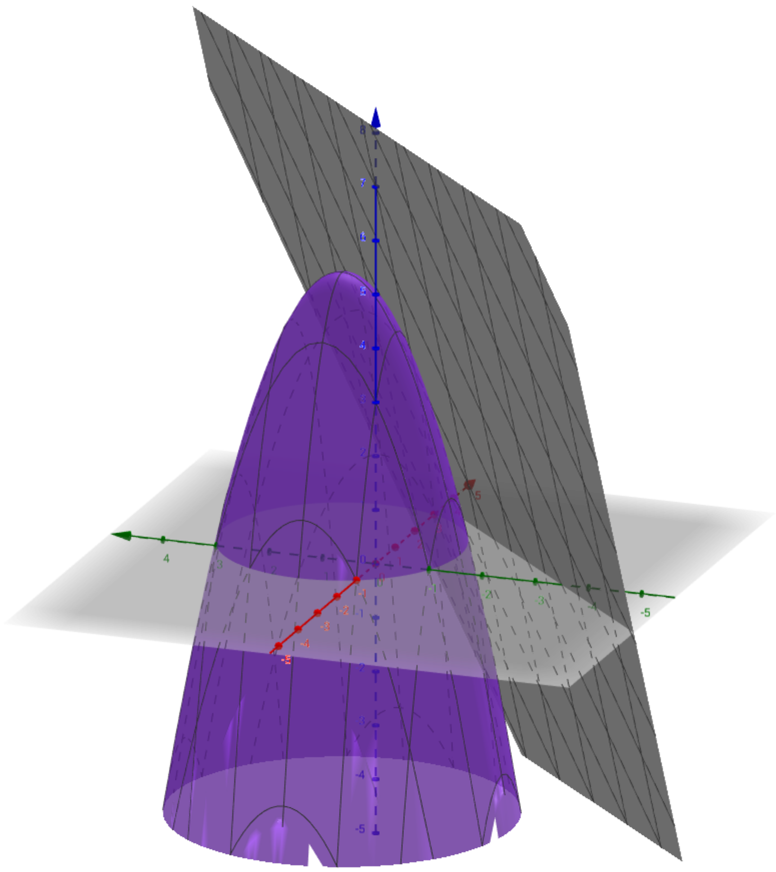


Figure 9.1: The function f with its tangent plane $E(x, y) = 7 - 2x + 2y$ at $(x_0, y_0) = (2, 0)$.

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 $f : \Omega \to \mathbb{R}$, $\Omega \subseteq \mathbb{R}^d$ open, one may say

$$\begin{aligned} & f(x_1 + \Delta x_1, \dots x_n + \Delta x_n) \\ & \quad \approx f(x_1, \dots, x_n) + \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) & \frac{\partial f}{\partial x_2}(x) & \dots & \frac{\partial f}{\partial x_n}(x) \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \\ \vdots \\ \Delta x_n \end{bmatrix} \\ & \quad = f(x_1, \dots, x_n) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_1, \dots, x_n) \Delta x_i. \end{aligned}$$

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

Appendix

A Prerequisites

A.1 Some notation used in this notes

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 A.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
$\mathbb{C}$	complex numbers

Table A.2: Notation of certain sets.

We have the following inclusions

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

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

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

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.

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

One can also consider the half-open cases

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

and

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

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

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

which is the same as

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

A.1.1 Operations on sets

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

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

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

Definition A.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 A.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 A.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.

Exercise A.1. *Prove that for any set A, one has $\varnothing \subseteq A$.*

Definition A.3 (The set A^B)**.**
Let A and B be two sets. Then, we denote the set of all functions $f : B \to A$ as A^B.

Example A.1. *The set $\mathbb{R}^{\mathbb{R}}$ are all functions from $\mathbb{R}$ to $\mathbb{R}$. The set $\mathbb{R}^{\mathbb{N}}$ is the set of all real sequences seen as functions from $\mathbb{N}$ to $\mathbb{R}$. With the latter we can write $(a_n) \in \mathbb{R}^{\mathbb{N}}$ and vice versa.*

The last notion on operations on sets, we introduce

Definition A.4 (Cartesian product)**.**
*Let A and B be sets. Then, the **Cartesian product** of A and B, $A \times B$, is the set of all ordered pairs (a, b) with $a \in A$ and $b \in B$, i.e.*

$$A \times B = \{(a, b) \; : \; a \in A, \, b \in B\} \, .$$

Example A.2. *Let $A = \{1, 2, 3\}$, $B = \{3, 4\}$. Then*

$$A \times B = \big\{(1, 3), (1, 4), (2, 3), (2, 4), (3, 3), (3, 4)\big\}.$$

Exercise A.2. *Draw a picture of $[0, 1] \times [0, 1]$.*

A.1.2 Some properties of sets

Definition A.5 (Bounded below/bounded above/bounded).
Let $A \subseteq \mathbb{R}$.

1. *If there exists $l \in \mathbb{R}$ such that $l \leq a$ for all $a \in A$, then we say that A is* ***bounded below*** *and l is said to be a* ***lower bound****.*
2. *If there exists $u \in \mathbb{R}$ such that $a \leq u$ for all $a \in A$, then we say that A is* ***bounded above*** *and u is said to be an* ***upper bound****.*
3. *We say that A is* ***bounded*** *iff A is bounded below and bounded above.*

Definition A.6 (Supremum/Infimum).
Let $A \subseteq \mathbb{R}$, $A \neq \varnothing$.

1. *Let $L \in \mathbb{R}$ be such that*
   a) $L \leq a$ *for all* $a \in A$*, and*
   b) *for all lower bounds l of A, we have $l \leq L$.*

   Then we call L the ***infimum*** *of A.*
2. *Let $L \in \mathbb{R}$ be such that*
   a) $a \leq U$ *for all* $a \in A$*, and*
   b) *for all upper bounds u of A, we have $U \leq u$.*

   Then we call U the ***supremum*** *of A.*

Remark A.2. *As discussed in Analysis 1, we should write* an infimum/a supremum *instead of* the*. Why are we justified in using the language above and in using $\sup(A)$ and $\inf(A)$ to denote* the *supremum and infimum of A respectively.*

An important property of suprema/infima is

Lemma A.1.
Let $A \subseteq \mathbb{R}$.

1. *If* A *has a supremum, then for all* $\varepsilon > 0$ *there exists an* $a \in A$ *such that*
$$\sup(A) - \varepsilon < a \leq \sup(A).$$
2. *If* A *has an infimum, then for all* $\varepsilon > 0$ *there exists an* $a \in A$ *such that*
$$\inf(A) \leq a < \inf(A) + \varepsilon.$$

Exercise A.3. *Prove Lemma A.1.*

From the **completeness axiom**, we have that every non-empty, bounded above set has a supremum. As discussed in Analysis 1, this implies the existence of an infimum for non-empty bounded below sets. For further information consult your Analysis 1 notes.

This year, you will attend the module Numbers (MAA245). In the last chapter, you will introduce the real numbers by completing the rational numbers by Dedekind cuts. The Dedekind cuts are exactly the suprema and infima that we, in out axiomatic system, presuppose to exist.

You may also have a look into the original literature. Richard Dedekind wrote a very readable account of the construction of the real numbers in the book *Essays on the Theory of Numbers*. Especially *Continuity of irrational numbers* is worth a read.

A.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 A.7 (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 α, $\beta \in \mathbb{R}$ 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 α, $\beta \in \mathbb{R}$ 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 A.4. *Convince yourself that* $(\mathbb{R}^n, \mathbb{R})$ *with the addition of vectors*

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

and multiplication with a scalar

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

is a real vector space in the sense of Definition A.7. Be reminded that

$$
\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T .
$$

Exercise A.5. *Convince yourself that* $(\mathbb{R}^{\mathbb{N}}, \mathbb{R})$ *(see Definition A.3) with*

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

and

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

is a real vector space in the sense of Definition A.7. Thus, sequences can be seen as points *in a vector space and we can try to apply intuitive geometric reasoning to this setting. See the theorem about arithmetical rules for convergent sequences in Analysis 1 and the subsequent remarks.*

Remark A.3. *The space* $(\mathbb{R}^{\mathbb{N}}, \mathbb{R})$ *in the example above is a neat example of an infinite dimensional vector space. For further details on dimension compare your Linear Algebra and Analysis 1 notes.*

We introduce the important notion of the scalar product. In school you probably heard about it as dot product of vectors. This might also be the name that is use in Mathematical Methods.

Definition A.8 (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:*

(i) 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)

(ii) For all u, $v \in V$, we have $\langle u, v \rangle = \langle v, u \rangle$. (Symmetry)

(iii) 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 A.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.*

Exercise A.6. *Prove the assertion made in Remark A.4.*

Example A.3. *The most important example in our context is*

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i. \tag{A.2.1}$$

This scalar product induces the Euclid length of a vector in $\mathbb{R}^n$*:*

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

This scalar product is, especially in applied mathematics, often denoted by a thick dot and called dot-product: $\langle x, y\rangle = x \cdot y = x^T y$*, where the last multiplication is a matrix multiplication. Since* (A.2.1) *is not the only possibility, we introduced* $\langle \cdot, \cdot \rangle$ *to symbolise a scalar product.*

Remark A.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*

$$x = \Big[1, 2\Big]^T, \quad y = \Big[-3, 1\Big]^T.$$

However, in this module, only the standard scalar product given by (A.2.1) *is of importance on* $\mathbb{R}^n$*.*

Proposition A.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 on* V *by*

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

Exercise A.7. *Use the definition of a scalar product to prove that* $\|x\| = \sqrt{\langle x, x\rangle}$ *defines a norm as stated in the last proposition. For the definition of a norm, see 1.8.*

Remark A.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 1.3.*

Remark A.7. *Scalar products satisfy a very important property for Analysis, the so-called Cauchy-Schwarz inequality. See Theorem A.10. This inequality says that*

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

The Cauchy-Schwarz inequality implies

$$-1 \leq \frac{\langle v, w\rangle}{\|v\|_2\|w\|_2} \leq 1$$

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

$$\cos(\angle(v, w)) = \frac{\langle v, w\rangle}{\|v\|_2\|w\|_2}.$$

A.2.1 Basis and dimension

Definition A.9 (Linear combination).

Let V be a real vector space and $\{v_1, \ldots, v_k\} \subseteq V$ 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 $\alpha_i \in \mathbb{R}$, $i = 1, \ldots, k$ a ***linear combination*** *of $v_1, \ldots, v_k$.*

Definition A.10 (Linear dependence).

A set $\{v_1, \ldots, v_k\} \subseteq V$ is called ***linearly dependent*** *if it is possible to write*

$$0 = \sum_{i=1}^{k} \alpha_i v_i, \tag{A.2.2}$$

where not all α_i are equal to zero. If (A.2.2) is only possible $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$, the set is called ***linearly independent***.

Definition A.11 (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 A.12 (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 $\sharp(B)$, in symbols: $\dim(V) = \sharp(B)$.*[a]

[a] The function $\sharp(B)$ returns the number of elements in B.

Definition A.13 (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 A.14 (Span/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 A.2.
Let V be a real vector space and $\mathcal{B} \subseteq V$ be a basis. Then $\operatorname{span}(\mathcal{B}) = V$.

Example A.4 (Canonical/standard basis)**.** *We define*

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

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

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

Then, the set $\{e_i : ni \in \{1, \ldots, n\}\}$ 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*

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

A.2.2 Linear maps

Definition A.15 (Linear map)**.**
Let U and V be two real vector spaces. Then, a map $L : U \to V$ is called a ***linear map*** *if and only if*

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

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

Remark A.8. *Following convention, we will usually write Lu instead of $L(u)$ to denote the image of u under L if L is a linear map. If we apply the map to a linear combination, we will use brackets to avoid ambiguities.*

A.3 Sequences

A sequence $(a_n)_{n\in\mathbb{N}}$ is an element of $\mathbb{R}^{\mathbb{N}}$, i.e. a map from $\mathbb{N}$ to $\mathbb{R}$. We can also think about a sequence as an infinite list of real numbers (indexed by the natural numbers):

$$(a_n)_{n\in\mathbb{N}} = (a_1, a_2, \ldots, a_n, \ldots).$$

We denote sequences $(a_n)_{n\in\mathbb{N}}$ also by (a_n) in short. For further information, see your Analysis 1 Lecture Notes.

Definition A.16 (Convergence of sequences)**.**
*A sequence $(a_n) \subseteq \mathbb{R}$ **converges** iff there exists an $a \in \mathbb{R}$ such that any $\varepsilon > 0$ there exists an index $n_0 \in \mathbb{N}$ such that $|a_n - a| < \varepsilon$ for all $n \geq n_0$. The value a is called the limit of (a_n). In symbols, we write*

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

or

$$a_n \to a \quad \text{as} \quad n \to +\infty,$$

respectively

$$(a_n) \to a.$$

Proposition A.3 (Uniqueness of limits)**.**
Let $(a_n) \subseteq \mathbb{R}$ be convergent. Then the limit is unique.

Exercise A.8. *Prove Proposition A.3. (Hint: Assume there are two different limits and show that they must be equal.)*

Exercise A.9. *Find a sequence that has three sub-sequences that converge all to different limits. (Hint: it is not always necessary to think in formulas.)*

Definition A.17 (Sub-sequence)**.**
Let $(a_n) \subseteq \mathbb{R}$ and $(n_k) \subseteq \mathbb{N}$ be a strictly increasing sequence. Then, (a_{n_k}) is a ***subsequence*** *of (a_n). We denote that by $(a_{n_k}) \subseteq (a_n)$.*

Proposition A.4 (Convergence of sub-sequences)**.**
Let $(a_n) \subseteq \mathbb{R}$ be a convergent sequence. Then, all sub-sequences converge every sub-sequence converges to the same limit.

Proof.

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

$$\forall \varepsilon > 0 \quad \exists n_0 \in \mathbb{N} \, : \, \forall n \geq n_0, \; |a_n - a| < \varepsilon \quad . \tag{A.3.1}$$

Let now $(a_{n_k}) \subseteq (a_n)$ be a subsequence of $(a_n)_{n \in \mathbb{N}}$.
Since $n_k \geq k$, we have if $k \geq n_0$ that $n_k \geq n_0$.
Thus, from (A.3.1), we get $|a_{n_k} - a| < \varepsilon$ for $k \geq n_0$.
Thus, $a_{n_k} \to a$ as $k \to +\infty$.
This concludes the proof. □

The converse of the last proposition is also true and yields

Theorem A.1.
A sequence $(a_n) \subseteq \mathbb{R}$ is convergent iff every subsequence converges to the same limit.

We further introduce

Definition A.18 (Monotonically decreasing/increasing)**.**
Let $(a_n) \subseteq \mathbb{R}$.

1. *If* $a_{n+1} \geq a_n$ *for all* $n \in \mathbb{N}$, *we say that* (a_n) *is (monotonically)* ***increasing****.*
   If $a_{n+1} > a_n$ *for all* $n \in \mathbb{N}$, *we say that* (a_n) *is* ***strictly*** *(monotonically)* ***increasing****.*

2. *If* $a_{n+1} \leq a_n$ *for all* $n \in \mathbb{N}$, *we say that* (a_n) *is (monotonically)* ***decreasing****.*
   If $a_{n+1} < a_n$ *for all* $n \in \mathbb{N}$, *we say that* (a_n) *is* ***strictly*** *(monotonically)* ***decreasing****.*

Definition A.19 (Monotone sequence)**.**
We say a sequence is ***monotone*** *if it is increasing or decreasing.*

Theorem A.2.
Every sequence $(a_n) \subseteq \mathbb{R}$ *has a monotone sub-sequence.*

Exercise A.10. *Prove Theorem A.2. If you do not remember how to do this, revise the proof in your Analysis 1 notes.*

We give

Definition A.20 (Boundedness of sequences)**.**
A sequence $(a_n) \subseteq \mathbb{R}$ *is* ***bounded*** *if and only if there exits a* $C > 0$ *such that*

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

Proposition A.5 (Boundedness of convergent sequences)**.**
Let $(a_n) \subseteq \mathbb{R}$ *be convergent. Then,* (a_n) *is bounded.*

Proof.

Let $a \in \mathbb{R}$ be the limit of (a_n) and fix $\varepsilon = 1$.

Then there exists $n_0 \in \mathbb{N}$ such that $|a_m - a| < 1$ for all $n \geq n_0$.

By the reverse triangle inequality, we get $|a_n| \leq 1 + |a|$ for all $n \geq n_0$.

Then, a bound for the sequence (a_n) is given by

$$M := \max \big\{ |a_1|, \ldots, |a_{n_0-1}|, 1 + |a| \big\},$$

i.e. $|a_n| \leq M$ for all $n \in \mathbb{N}$. □

Exercise A.11. *Give examples and counterexamples for the converse of Proposition A.5.*

Theorem A.3 (Arithmetic properties of sequences)**.**
Let $(a_n) \subseteq \mathbb{R}$ *and* $(b_n) \subseteq \mathbb{R}$ *be sequences with* $a_n \to a$ *and* $b_n \to b$ *as* $n \to +\infty$*. Then, one has*

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

(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* $n \in \mathbb{N}$ *and* $b \neq 0$*.*

Proof. See your notes from Analysis 1 or regard it as an exercise. □

Theorem A.4 (Monotone convergence)**.**

Every bounded and monotone sequence is convergent.

Exercise A.12. *Prove Theorem A.4, give examples, and give counterexamples for the converse.*

Theorem A.5 (Bolzano–Weierstrass)**.**

Every bounded sequence $(a_n) \subseteq \mathbb{R}$ has a convergent subsequence.

Proof. See your notes from Analysis 1. We will prove a more general version later this semester. Can you recall the ingredients of the proof before you look the details up? □

Definition A.21 (Cauchy-sequence/Fundamental sequence)**.**

*Let $(a_n) \subseteq \mathbb{R}$ be a sequence. We say that (a_n) is a **Cauchy-sequence** (fundamental sequence) if and only if 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*

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

Exercise A.13. *Prove that every sub-sequence of a Cauchy sequence is a Cauchy sequence.*

Proposition A.6 (Boundedness of Cauchy sequences)**.**

Let $(a_n) \subseteq \mathbb{R}$ be a Cauchy sequence. Then, $(a_n)_{n\in\mathbb{N}}$ is bounded.

Exercise A.14. *Prove Proposition A.6, give examples and counterexamples for the converse.*

Proposition A.7.
Let $(a_n) \subseteq \mathbb{R}$ be a convergent sequence with limit a. Then $(a_n) \subseteq \mathbb{R}$ is a Cauchy sequence.

Exercise A.15. *Before reading the proof, draw a picture of the situation and see whether you can reproduce the proof yourself.*

Proof of Proposition A.7.

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

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

This concludes the proof. $\square$

We also have a converse to Proposition A.7.

Proposition A.8.
Let $(a_n) \subseteq \mathbb{R}$ be a Cauchy sequence. Then there exists an $a \in \mathbb{R}$ such that $a_n \to a$ as $n \to +\infty$.

Proof.

From Theorem A.6, we know that (a_n) is bounded.
From Bolzano–Weierstrass we know that there exist a convergent subsequence $(a_{n_k}) \subseteq (a_n)$.

Now, we show that (a_n) has the same limit. Let a be the limit of the convergent subsequence (a_{n_k}).
Let $\varepsilon > 0$ and choose $k_1 \in \mathbb{N}$ such that $|a_{n_k} - a| < \frac{\varepsilon}{2}$ for all $k \geq k_1$.
Choose $k_2 \in \mathbb{N}$ such that $|a_{n_k} - a_m| < \frac{\varepsilon}{2}$ for all m, $k \geq k_2$.
Let $k_0 = \max\{k_1, k_2\}$. Then

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

this concludes the proof. □

Thus, taking the last two propositions together, we have

Theorem A.6 (Cauchy criterion for convergence)**.**
Let $(a_n) \in \mathbb{R}$ be s sequence. There exists an $a \in \mathbb{R}$ such that $a_n \to a$ as $n \to +\infty$ if and only if $(a_n)_{n \in \mathbb{N}} \in \mathbb{R}$ is a Cauchy sequence.

Remark A.9. *Theorem A.6 is only valid because we are in $\mathbb{R}$. 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, $x_n \to \sqrt{2}$, $n \to +\infty$. Thus, the limit exists only if the sequence is considered to be in $\mathbb{R}$. This property is called completeness.

Theorem A.7 (Limits respect inequalities)**.**
Let $(a_n) \subseteq \mathbb{R}$ be a sequence and suppose that $a_n \leq C$ for all $n \in N$. Then, if $\lim_{n\to+\infty} a_n$ exists, it holds

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

Remark A.10. *The $\leq$ in Theorem A.7 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 A.16. *Prove Theorem A.7.*

A.4 Series

Let $(a_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$ 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 $\sum\limits_{n=0}^{+\infty} a_n$.

Definition A.22 (Convergence of series)**.**
*We say that $\sum\limits_{n=0}^{+\infty} a_n$ **converges** iff its sequence of partial sums $(S_N)_{N\in\mathbb{N}_0}$ converges. We say that $\sum\limits_{n=0}^{+\infty} a_n$ is **absolutely convergent** if $\sum\limits_{n=0}^{+\infty} |a_n|$ is convergent.*

Definition A.23 (Divergent series)**.**
A series $\sum_{n=0}^{+\infty} a_n$ *is said divergent iff it is not convergent.*

Since we have Theorem A.6, we can conclude

Corollary A.1.
The series $\sum_{n=0}^{+\infty} a_n$ *converges iff the sequence of its partial sums* (S_N) *is a Cauchy sequence.*

Remark A.11. *Writing Corollary A.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*

$$\left| \sum_{i=m+1}^{n} a_i \right| < \varepsilon \tag{A.4.1}$$

for all $n \geq m \geq N_0$*. Equivalently (prove that), we can ask*

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

in place of (A.4.1).

An important though quite obvious property is stated in

Proposition A.9.
Let $\sum_{n=0}^{+\infty} a_n$ *be a convergent series. Then* $(a_n) \to 0$.

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 A.3, we obtain

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

This concludes the proof. $\square$

Remark A.12. *One can not say much more than $(a_n) \to 0$ for a convergent series. If the summands satisfy $a_n \geq a_{n+1} > 0$, then we can show that $(na_n) \to 0$.*

We have

Theorem A.8 (Absolute convergence $\Rightarrow$ convergence)**.**
If $\sum\limits_{n=0}^{+\infty} a_n$ converges absolutely, then it converges.

A final result we will need from Analysis 1 is

Theorem A.9 (Comparison theorem)**.**
Let $\sum\limits_{n=1}^{+\infty} a_n$, $\sum\limits_{n=1}^{+\infty} b_n$ two series such that $0 \leq a_n \leq b_n$ for all $n \in \mathbb{N}_0$. Then,

1. *if $\sum\limits_{n=1}^{+\infty} b_n$ converges, then $\sum\limits_{n=1}^{+\infty} a_n$ converges.*

2. *if $\sum\limits_{n=1}^{+\infty} a_n$ diverges, then $\sum\limits_{n=1}^{+\infty} b_n$ diverges.*

In case 1, we then have

$$\sum_{n=1}^{+\infty} a_n \leq \sum_{n=1}^{+\infty} b_n.$$

Exercise A.17. *Prove Theorem A.9.*

A.5 Elementary properties of functions of one variable

Let us recall what a function is. First, we have to define what a relation is and then we will use that to give a precise definition of functions.

Definition A.24 (Relation)**.**
A relation on a set A is a sub-set $\rho \subseteq A \times A$. We abbreviate the statement $(x, y) \in \rho$ by $x\rho y$.

Example A.5 (The relation $>$ on $\mathbb{Z}$.)**.**
Consider the set $\rho = \{(x, y) \in \mathbb{Z} \times \mathbb{Z} : x - y \in \mathbb{N}\} \subseteq \mathbb{Z} \times \mathbb{Z}$. This is the $>$ relation on the set $A = \mathbb{Z}$. It is infinite because there are infinitely many ways to have $x > y$ where x and y are integers.

With relations, we can define what a function is.

Definition A.25 (Function)**.**
Let A and B be sets. Then, a function f from A to B, in symbols $f : A \to B$ is a relation $f \subseteq A \times B$ from A to B satisfying the property that for each $a \in A$ the relation f contains exactly one ordered pair of the form (a, b). The statement $(a, b) \in f$ is abbreviated by $f(a) = b$.

Exercise A.18. *Interpret the definition of function graphically. Draw functions and non-functions to understand what* for each $a \in A$ the relation f contains exactly one ordered pair of the form (a, b) *means.*

We can add, subtract, multiply, and divide functions by pointwise definition. Another important operation is the composition: let $f : A \to B$ and $g : B \to C$ be functions. Then, $g \circ f(x) = g(f(x))$ is a function from A to C. Unless C is contained in A, the function $f \circ g$ is not defined.

Example A.6. *Let $f(x) = e^x$ and $g(x) = x^2 + 1$. Both functions are defined on $\mathbb{R}$. We can consider*

$$f \circ g(x) = f(g(x)) = e^{x^2+1},$$
$$g \circ f(x) = g(f(x)) = e^{2x} + 1.$$

That also shows that, in general, $f \circ g$ is not equal to $g \circ f$ should both be definable.

A.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.

A.5.2 Monotonic functions

Definition A.26 (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 A.13. *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* $x \geq y$*. Compare that to our use of language in sequences in Analysis 1.*

Remark A.14. *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.*

A.5.3 Odd and even functions

Definition A.27 (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 A.15. *Odd/even functions can be defined on intervals too but one needs to take some care with respect to the symmetry property.*

Remark A.16. *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.

A.5.4 Convex and concave functions

Definition A.28 (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.*

Remark A.17. *In school you might have heard about convex up/down of concave up/down. There names are not used in the mathematical literature above school level and should be dropped.*

Exercise A.19. *Interpret the definition of convex/concave graphically.*
Consider $f(x) = x^2$ *and* $g(x) = \sqrt{x}$ *and say on which intervals they are concave/convex if on any.*

Proposition A.10 (Jensen's inequality[a]).

Let $f : \mathbb{R} \to \mathbb{R}$ *be convex. Suppose that* $x_1, \dots, x_n \in \mathbb{R}$ *and* $\lambda_1, \dots, \lambda_n \in [0, +\infty])$ *with* $\sum_{i=1}^{n} \lambda_i = 1$. *Then,*

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

holds.

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

A.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 we have the trivial estimates $2\min\{a, b\} \leq a + b \leq 2\max\{a, b\}$.
- 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

  $$|a + b| \leq |a| + |b| \tag{A.6.1}$$

  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 A.7. *We need these rather obvious inequalities often in the analysis of sequences. For example, there exists an* $n_0 \in \mathbb{N}$ *such that*

$$\forall n \geq n_0, \quad \frac{1}{n^2+n} \leq \frac{1}{n^2}.$$

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*

$$\forall n \geq n_0, \quad \frac{1}{n^2-n} = \frac{1}{\frac{n^2}{2} + \frac{n^2}{2} - n} \leq \frac{1}{n^2}.$$

Exercise A.20. *Analise 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 $\displaystyle\sum_{n=1}^{+\infty} \frac{1}{n^3}$.

Exercise A.21. *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 1.3.[1]

[1] See Theorem 1.2.

A.7 Cauchy–Schwarz, Minkowski, and Hölder

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

Theorem A.10 (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 A.8 in Section A.2.

Proof. The proof of this important inequality is on the problem sheet for you to find. The solutions of the sheet contain two different proofs an here I will present another one that one student of the year 2016/17 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 $\mathbb{R}^n \setminus \{0\}$.

Then define $v = |a|b - |b|a$.

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

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

Thus, we obtain $2|a|^2|b|^2 \geq 2|a||b|\langle a, b\rangle$.

Dividing by $2|a||b|$, we get the Cauchy–Schwarz inequality.

This concludes the proof. □

Exercise A.22. *Work out what properties of the scalar product have been used where in the proof above. See Definition A.8).*

Exercise A.23. *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 A.24. *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 $v = x - y$.

The next theorem proves that $\|\cdot\|_p$, $p \in [1, +\infty]$ satisfies the triangle inequality. See also Definition 1.8 and Example 1.3.

Theorem A.11 (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 1.3. With the appropriate changes, the inequality remains true for $p = \infty$, i.e.

$$\max_{i=1,\dots,n} |a_i + b_i| \leq \max_{i=1,\dots,n} |a_i| + \max_{i=1,\dots,n} |b_i|,$$
$$\|a + b\|_\infty \leq \|a\|_\infty + \|b\|_\infty.$$

Remark A.18. *The case $p = \infty$ is very easy to prove. It only relies on the triangle inequality for real numbers (see (A.6.1)): $|a_i + b_i| \leq |a_i| + |b_i|$ for all $i = 1, \dots, n$ implies*

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

The final inequality stated in this section is

Theorem A.12.

Suppose that $1 \le p, q \le \infty$ *with* $p^{-1} + q^{-1} = 1$. *Then, for* $a, b \in \mathbb{R}^n$, *we have*

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

Shorter, we can write

$$\sum_{i=1}^{n} |a_i b_i| \le \|a\|_p \|b\|_q.$$

Exercise A.25. *Show that the Cauchy–Schwarz inequality is a special case of the Hölder inequality.*

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