Mathematical Analysis

Lars Grüne
Chair of Applied Mathematics
University of Bayreuth
95440 Bayreuth, Germany
lars.gruene@uni-bayreuth.de
num.math.uni-bayreuth.de

Lecture Notes

1st Edition

Summer Semester 2026

Preface

These lecture notes were written for a course of the same name held in the Summer Semester 2026 at the University of Bayreuth, Germany. It is a continuation of the course *Foundations of Higher Mathematics* held in the Winter Semester before. The notes are based on the notes for my German courses Analysis I and II, except for the last chapter, where I am grateful to Thomas Kriecherbauer for offering me his notes as a template. PDF versions of this notes and of the notes for the other mentioned courses are available from

`https://num.math.uni-bayreuth.de/de/team/lars-gruene/skripten/`.

For writing these notes, the textbooks mentioned in the Bibliography were used extensively. For translating the German notes into English, ChatGPT turned out to be quite useful. Of course, the responsibility for any errors in the notes lies entirely with the author. Suggestions for improvement are always welcome and appreciated.

Bayreuth, July 2026 LARS GRÜNE

Contents

Chapter 1

The Riemann Integral

Last update: July 21, 2026

Alongside differential calculus, integral calculus is the second essential concept of analysis. An intuitive motivation for integral calculus in $\mathbb{R}$ is the computation of areas, especially the areas under the graph of a function, where regions above the x-axis receive a positive value and regions below the x-axis receive a negative value.

Coming from school mathematics, it may initially seem surprising that there are several different ways to define an integral. In this lecture we will introduce the so-called *Riemann integral* from scratch and prove properties and computational rules for this integral.

1.1 Integrals for Step Functions

As already mentioned, we want to define a mathematical concept—namely the integral—that assigns to a given function $f : D \to \mathbb{R}$ the size of the area under the graph of the function. The integral is therefore a mapping from the set of functions to the real numbers; such a mapping is called a *functional*.

For the following class of functions, the computation of the area is very simple.

Definition 1.1 A function $f : [a, b] \to \mathbb{R}$ is called a *step function* if there exist points x_k, $k = 0, \ldots, n$, $n \in \mathbb{N}$, such that

$$a = x_0 < x_1 < x_2 < \ldots < x_n = b$$

and f is constant on each open interval (x_k, x_{k+1}), $k = 0, \ldots, n-1$, with value c_k for some $c_k \in \mathbb{R}$. The set of all step functions on $[a, b]$ is denoted by $T(a, b)$. The set $\{x_0, \ldots, x_n\}$ is called a *partition* of $[a, b]$. $\square$

Remark 1.2 (i) Note that the partition $\{x_0, \ldots, x_n\}$ in Definition 1.1 is not unique. For example, the required constancy condition remains satisfied if we add arbitrary additional points from $[a, b]$ to the partition.

(ii) Because of (i), for two different step functions $f, g \in T(a, b)$ we may always assume without loss of generality that they satisfy the condition from Definition 1.1 on the same partition. Indeed, if f satisfies the condition with the partition $\{x_0, \ldots, x_n\}$ and g with the partition $\{x'_0, x'_1, \ldots, x'_{n'}\}$, then both functions satisfy the condition on the partition

$$\{x_0, \ldots, x_n\} \cup \{x'_0, x'_1, \ldots, x'_{n'}\}.$$

(iii) For a constant function $f : E \to \mathbb{R}$ with $f(x) = c$ for all $x \in E$, we write $f \equiv c$. The restriction of a function $f : D \to \mathbb{R}$ to a subset $D' \subset D$ is denoted by $f|_{D'}$. The function $f|_{D'}$ therefore produces exactly the same function values as f, only the domain changes. Using these two notations, we can express the step function property concisely as $f|_{(x_k,x_{k+1})} \equiv c_k$. □

The computation of the area under the graph of a step function is so simple because the area consists only of rectangles of height c_k and width $x_{k+1} - x_k$.

If we define the area of such a rectangle as $(x_{k+1} - x_k)c_k$, we automatically obtain—just as desired—positive values for areas above the x-axis and negative values for areas below the x-axis. The integral for the entire step function on an interval $[a, b]$ can now easily be determined by adding the areas of the individual rectangles.

Definition 1.3 For a step function $f \in T(a, b)$ and a partition $a = x_0 < x_1 < \ldots < x_n = b$ with $f|_{(x_k,x_{k+1})} \equiv c_k$, we define the integral as

$$\int_a^b f(x)\,dx := \sum_{k=0}^{n-1} (x_{k+1} - x_k)c_k.$$

□

Question 1.1 *What is* $\displaystyle\int_a^b f(x)dx$ *for a step function* f *with* $c_k = 1$ *for all* k*?*

[1]Even though this definition is very intuitive, we must still verify mathematically that it yields a unique value—one then says that the defined value is *well-defined*. The reason for this is the already mentioned fact that a step function with partition $a = x_0 < x_1 < \ldots < x_n = b$ can also be a step function with respect to another partition $a = x'_0 < x'_1 < \ldots < x'_{n'} = b$. Since the definition of the integral explicitly depends on the partition, we must verify that the value of the integral does not change if we modify the partition.

To this end, consider a step function $f \in T(a, b)$ and two partitions $a = x_0 < x_1 < \ldots < x_n = b$ and $a = x'_0 < x'_1 < \ldots < x'_{n'} = b$ such that $f|_{(x_k,x_{k+1})} \equiv c_k$ for $k = 0, \ldots, n-1$ and $f|_{(x'_k,x'_{k+1})} \equiv c'_k$ for $k = 0, \ldots, n'-1$. For brevity, write

$$I := \sum_{k=0}^{n-1} (x_{k+1} - x_k)c_k \quad \text{and} \quad I' := \sum_{k=0}^{n'-1} (x'_{k+1} - x'_k)c'_k.$$

We must therefore show that $I = I'$. We consider two cases.

Case 1: $\{x_0, x_1, \ldots, x_n\} \subset \{x'_0, x'_1, \ldots, x'_{n'}\}$.

[1]Paragraphs printed in purple colour and in a smaller font were not presented at the blackboard during the lecture. They are not relevant for the exam, but may provide useful additional knowledge.

In this case, for every k there exists an index k'_k such that

$$x_k = x'_{k'_k} < x'_{k'_k+1} < \ldots < x'_{k'_{k+1}-1} = x_{k+1}.$$

It follows that

$$c_k = c'_{k'_k} = c'_{k'_k+1} = \ldots = c'_{k'_{k+1}-1}.$$

Hence,

$$I' = \sum_{k=0}^{n-1} \sum_{k'=k'_k}^{k'_{k+1}-1} (x'_{k'+1} - x'_{k'})c'_{k'} = \sum_{k=0}^{n-1} \underbrace{\sum_{k'=k'_k}^{k'_{k+1}-1} (x'_{k'+1} - x'_{k'})}_{=x_{k+1}-x_k} c_k = \sum_{k=0}^{n-1} (x_{k+1} - x_k)c_k = I.$$

Case 2: Let $\{x_0, x_1, \ldots, x_n\}$ and $\{x'_0, x'_1, \ldots, x'_{n'}\}$ be arbitrary. Then, by Remark 1.2(i), the points

$$\{x_0, x_1, \ldots, x_n\} \cup \{x'_0, x'_1, \ldots, x'_{n'}\}$$

also form a partition on which f satisfies the condition from Definition 1.1. Denote the integral corresponding to this partition by I^*. From Case 1 we obtain $I = I^*$ and $I' = I^*$. Hence $I = I'$. □

Question 1.2 *Why don't we need to make any assumptions on the values $f(x_k)$ at the partition points x_k?*

The following theorem states an important property of integrals for step functions.

Theorem 1.4 Let $f, g \in T(a, b)$ and $\lambda \in \mathbb{R}$. Then $f + g \in T(a, b)$ and $\lambda f \in T(a, b)$ and it holds that

$$\int_a^b f(x) + g(x)dx = \int_a^b f(x)dx + \int_a^b g(x)dx \quad \text{and} \quad \int_a^b \lambda f(x)dx = \lambda \int_a^b f(x)dx.$$

If we furthermore define the inequality

$$f \leq g \quad :\Leftrightarrow \quad f(x) \leq g(x) \text{ for all } x \in [a, b], \tag{1.1}$$

then the following implication holds:

$$f \leq g \quad \Rightarrow \quad \int_a^b f(x)dx \leq \int_a^b g(x)dx.$$

Proof: That λf is a step function is immediately clear. Since, according to Remark 1.2(ii), we may assume that the two step functions f and g satisfy the step function property with respect to one and the same partition, the sum of the two functions is again a step function. Moreover, we can compute the integrals of both functions based on the same partition.

All statements then follow directly from the fact that the desired (in-)equalities hold for each summand in the definition of $\int_a^b$ and therefore also for the entire integral. □

Remark 1.5 Note that in general

$$\int_a^b f(x)g(x)dx \neq \left(\int_a^b f(x)dx\right) \cdot \left(\int_a^b g(x)dx\right)$$

holds. The reason is that in general

$$\sum_{j=0}^{n} a_k b_k \neq \left(\sum_{k=0}^{n} a_k\right) \cdot \left(\sum_{k=0}^{n} b_k\right)$$

(see also the discussion at the end of [7, Section 9.1]), and this inequality transfers to the integrals via the defining sums. □

1.2 The Riemann Integral

The idea of the Riemann integral is now to approximate the area under an arbitrary function $f : [a, b] \to \mathbb{R}$ by the area under step functions. For this purpose we define — using the inequality for functions introduced in (1.1) — the following upper and lower integrals.

Question 1.3 *What objects (functions, sequences, numbers, sets, ...) are the elements of the set*

$$\left\{ \int_a^b \varphi(x)dx \,\middle|\, \varphi \in T(a, b),\, \varphi \geq f \right\} ?$$

Definition 1.6 For an arbitrary bounded function $f : [a, b] \to \mathbb{R}$ we define the *upper integral*

$$\int_a^{*b} f(x)dx := \inf \left\{ \int_a^b \varphi(x)dx \,\middle|\, \varphi \in T(a, b),\, \varphi \geq f \right\}$$

and the *lower integral*

$$\int_{*a}^{b} f(x)dx := \sup \left\{ \int_a^b \varphi(x)dx \,\middle|\, \varphi \in T(a, b),\, \varphi \leq f \right\},$$

where the integrals over φ are to be understood in the sense of Definition 1.3. □

Question 1.4 *Why do we need boundedness of f in this definition?*

From Theorem 1.4 (together with [7, Theorem 8.26]) the inequality

$$\int_a^{*b} f(x)dx \geq \int_{*a}^{b} f(x)dx$$

follows immediately.

A function is called Riemann integrable precisely when this inequality is actually an equality, or in other words, when the infimum over the approximations from above yields the same value as the supremum over the approximations from below.

Definition 1.7 We call a function $f : [a, b] \to \mathbb{R}$ *Riemann integrable* (on the interval $[a, b]$) if

$$\int_a^{*b} f(x)dx = \int_{*a}^{b} f(x)dx$$

holds. In this case we define the Riemann integral of f by

$$\int_a^b f(x)dx := \int_a^{*b} f(x)dx.$$

□

Example 1.8 (i) Every step function $f \in T(a, b)$ is Riemann integrable, because with $\varphi = f$ we have

$$\int_a^{*b} f(x)dx = \inf\left\{\int_a^b \varphi(x)dx \,\middle|\, \varphi \in T(a, b),\, \varphi \geq f\right\} \leq \int_a^b f(x)dx$$

and

$$\int_{*a}^{b} f(x)dx = \sup\left\{\int_a^b \varphi(x)dx \,\middle|\, \varphi \in T(a, b),\, \varphi \leq f\right\} \geq \int_a^b f(x)dx,$$

where the integrals over f are to be understood in the sense of Definition 1.3. Hence

$$\int_a^{*b} f(x)dx \leq \int_{*a}^{b} f(x)dx$$

and therefore equality follows (since the reverse inequality always holds). For this reason we also do not distinguish in the notation between the Riemann integral from Definition 1.7 and the integral over step functions from Definition 1.3.

(ii) Consider the function

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

on $D = [0, 1]$. Since this function takes both the value 1 and the value 0 in every interval (x_k, x_{k+1}) with $x_{k+1} > x_k$ (because every real interval contains rational numbers), the inequality $\varphi|_{(x_k,x_{k+1})} \geq 1$ holds for every step function $\varphi \geq f$. The smallest step function $\varphi \geq f$ therefore satisfies $\varphi|_{(x_k,x_{k+1})} = 1$. Analogously, the largest step function $\varphi \leq f$ satisfies $\varphi|_{(x_k,x_{k+1})} = 0$. Hence

$$\int_{*0}^{1} f(x)dx = 0 \quad \text{and} \quad \int_0^{*1} f(x)dx = 1,$$

which means that this function is not Riemann integrable.

(iii) With the help of step functions one can in principle compute the Riemann integral for many elementary functions. However, this is extremely tedious, and therefore we will not carry it out here. Instead, in the following chapter we will derive computational rules that make this much easier. □

The following theorem gives an alternative criterion for Riemann integrability that follows immediately from the definition of the Riemann integral.

Theorem 1.9 A function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if for every $\varepsilon > 0$ there exist step functions $\varphi, \psi \in T(a, b)$ with $\varphi \le f \le \psi$ such that

$$\int_a^b \psi(x)dx - \int_a^b \varphi(x)dx \le \varepsilon.$$

Proof: This follows immediately from the definition of the upper and lower integrals and from the definitions of supremum and infimum. □

With the help of this theorem we can now prove that every continuous function is Riemann integrable.

Theorem 1.10 Every continuous function $f : [a, b] \to \mathbb{R}$ is Riemann integrable.

Proof: We show that the assumption of Theorem 1.9 is satisfied. Let $\varepsilon' > 0$ be arbitrary. By [7, Theorem 14.20], f is uniformly continuous. Set $\varepsilon := \varepsilon'/(b-a)$ and choose $\delta > 0$ as in [7, Definition 14.19] for uniform continuity. Let $a = x_0 < \ldots < x_n = b$ be a partition with $x_{k+1} - x_k < \delta$ and let $c_k = \min\{f(x) \mid x \in [x_k, x_{k+1}]\}$ and $c'_k = \max\{f(x) \mid x \in [x_k, x_{k+1}]\}$. Since for any two points $x, y \in [x_k, x_{k+1}]$ the inequality $|x - y| < \delta$ holds, it follows that $|f(x) - f(y)| < \varepsilon$, and hence $c'_k - c_k < \varepsilon$.

Now consider step functions defined by

$$\varphi|_{(x_k,x_{k+1})} \equiv c_k \quad \text{and} \quad \psi|_{(x_k,x_{k+1})} \equiv c'_k$$

and $\varphi(x_k) = \psi(x_k) = f(x_k)$. Then $\varphi \le f \le \psi$. Moreover,

$$\begin{aligned} \int_a^b \psi(x)dx - \int_a^b \varphi(x)dx &= \int_a^b (\psi(x) - \varphi(x))dx \\ &= \sum_{k=0}^{n-1} (x_{k+1} - x_k)(c'_k - c_k) \\ &< \sum_{k=0}^{n-1} (x_{k+1} - x_k)\varepsilon = (b - a)\varepsilon = \varepsilon'. \end{aligned}$$

This proves the claim by Theorem 1.9. □

1.3 Riemann Sums

The construction in the proof of Theorem 1.10 provides a way to approximate the Riemann integral for continuous functions.

For this purpose, let $\varepsilon > 0$ be arbitrary and consider a partition $a = x_0 < \ldots < x_n = b$ with $x_{k+1} - x_k < \delta$, where δ is the value from the proof of Theorem 1.10 with $\varepsilon' = \varepsilon$. Now choose arbitrary points $\xi_k \in [x_k, x_{k+1}]$ and define a step function $\vartheta \in T(a, b)$ by

$$\vartheta|_{(x_k,x_{k+1})} \equiv f(\xi_k).$$

By the definition of the integral for step functions,

$$\int_a^b \vartheta(x)dx = \sum_{k=0}^{n-1}(x_{k+1}-x_k)f(\xi_k),$$

and this sum is called a *Riemann sum*.

From the construction of the functions φ and ψ in the proof of Theorem 1.10 it follows that

$$\varphi(x) \leq \vartheta(x) \leq \psi(x)$$

for all $x \in [a,b] \setminus \{x_0,\ldots,x_n\}$.

From the definition of the integral for step functions we obtain

$$\int_a^b \varphi(x)dx \leq \int_a^b \vartheta(x)dx \leq \int_a^b \psi(x)dx \leq \int_a^b \varphi(x)dx + \varepsilon.$$

Since also

$$\int_a^b \varphi(x)dx \leq \int_a^b f(x)dx \leq \int_a^b \psi(x)dx$$

holds, it follows that

$$\left|\int_a^b f(x)dx - \sum_{k=0}^{n-1}(x_{k+1}-x_k)f(\xi_k)\right| \leq \varepsilon.$$

Thus we can approximate[2] the integral by the Riemann sum

$$\sum_{k=0}^{n-1}(x_{k+1}-x_k)f(\xi_k).$$

Since $\varepsilon > 0$ in (1.3) is arbitrary, it follows in particular that the value of the Riemann sum converges to the value of the integral as $\delta \to 0$.

Example 1.11 We consider the function $f(x) = x$ on the interval $[0,y]$ for $y > 0$. Let $\delta = y/n$ for some $n \in \mathbb{N}$ and choose the partition $x_k = k\delta$. Then $x_{k+1} - x_k = \delta$. If we choose $\xi_k = x_k$, we obtain

$$\begin{aligned}
\sum_{k=0}^{n-1}(x_{k+1}-x_k)f(\xi_k) &= \sum_{k=0}^{n-1}(x_{k+1}-x_k)x_k = \sum_{k=0}^{n-1}(\delta)\delta k \\
&= \delta^2\sum_{k=0}^{n-1}k = \delta^2\frac{n(n-1)}{2} = \delta^2\frac{n^2}{2} - \delta^2\frac{n}{2} \\
&= \frac{y^2}{2} - \delta\frac{y}{2}.
\end{aligned}$$

As $\delta \to 0$, this expression converges to $y^2/2$, and therefore we obtain

$$\int_0^y x\,dx = \frac{y^2}{2}.$$

□

[2]With a more involved proof (for which we refer, for example, to Forster [6, §18]) one can show that this approximation works for all Riemann-integrable functions, not only for continuous ones.

1.4 Mean Value Theorem for Integrals

Similar to differential calculus, there is also a mean value theorem in integral calculus. With this theorem we conclude this chapter.

Theorem 1.12 Let $f, g : [a, b] \to \mathbb{R}$ be continuous functions with $g \geq 0$. Then there exists $\xi \in [a, b]$ such that

$$\int_a^b f(x)g(x)dx = f(\xi) \int_a^b g(x)dx.$$

In the special case $g \equiv 1$ we obtain

$$\int_a^b f(x)dx = f(\xi)(b - a).$$

Proof: Let

$$m := \min\{f(x) \,|\, x \in [a, b]\}, \qquad M := \max\{f(x) \,|\, x \in [a, b]\}.$$

Then $mg \leq fg \leq Mg$, and therefore

$$m \int_a^b g(x)dx \leq \int_a^b f(x)g(x)dx \leq M \int_a^b g(x)dx.$$

Hence there exists $\mu \in [m, M]$ such that

$$\mu \int_a^b g(x)dx = \int_a^b f(x)g(x)dx.$$

By the intermediate value theorem for continuous functions, there exists $\xi \in [a, b]$ with $f(\xi) = \mu$, which proves the claim. $\square$

Chapter 2

Differential and Integral Calculus

Last update: July 21, 2026

In the previous chapter we computed very few examples. This is because the area-based definition of the Riemann integral via lower and upper sums is very cumbersome for computational purposes. In this chapter we will first show that integration is precisely the inverse operation of differentiation. This will allow us to determine the integral for many elementary functions and to derive computational rules for evaluating more complicated integrals.

In order to establish the connection between differential and integral calculus, we must first clarify the relationship between integrals over different intervals. This is done in the following theorem.

Theorem 2.1 A function $f : [a,b] \to \mathbb{R}$ is Riemann integrable if and only if for any $c \in (a,b)$ the functions $f|_{[a,c]}$ and $f|_{[c,b]}$ are Riemann integrable. In this case we have

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx. \tag{2.1}$$

Proof: Clearly, by adding the point c to the partition if necessary, every step function in $T(a,b)$ can be split into two step functions in $T(a,c)$ and $T(c,b)$. Conversely, any two step functions in $T(a,c)$ and $T(c,b)$ can be combined to form a step function in $T(a,b)$. Hence the statement holds for the integrals of step functions, and therefore also for the upper and lower integrals and thus for the Riemann integral. $\square$

Question 2.1 *Draw a sketch that illustrates the statement of Theorem 2.1 with the area under the graph of f.*

The restriction $a < c < b$ can be avoided if we define the value of an integral whose upper limit is less than or equal to the lower limit.

Definition 2.2 For a Riemann integrable function $f : [a, b] \to \mathbb{R}$ and arbitrary points $c, d \in [a, b]$ with $c > d$ we define

$$\int_c^d f(x)dx := -\int_d^c f(x)dx.$$

Moreover, for all $c \in [a, b]$ we define

$$\int_c^c f(x)dx := 0.$$

□

With these definitions the formula

$$\int_{x_1}^{x_3} f(x)dx = \int_{x_1}^{x_2} f(x)dx + \int_{x_2}^{x_3} f(x)dx \tag{2.2}$$

holds for all Riemann integrable functions $f : [a, b] \to \mathbb{R}$ and for arbitrary points $x_1, x_2, x_3 \in [a, b]$. This can be proved by case distinction for the possible $<$ and $\leq$ relations between x_1, x_2, x_3. As an example we only consider the case $x_1 < x_3 < x_2$. In this case, by (2.1) with $a = x_1$, $b = x_2$ and $c = x_3$, we have

$$\int_{x_1}^{x_2} f(x)dx = \int_{x_1}^{x_3} f(x)dx + \int_{x_3}^{x_2} f(x)dx$$

and therefore

$$\int_{x_1}^{x_3} f(x)dx = \int_{x_1}^{x_2} f(x)dx - \int_{x_3}^{x_2} f(x)dx = \int_{x_1}^{x_2} f(x)dx + \int_{x_2}^{x_3} f(x)dx.$$

An immediate consequence of these formulas is the continuity of the mappings

$$c \mapsto \int_c^d f(x)dx \quad \text{and} \quad d \mapsto \int_c^d f(x)dx,$$

for $c, d \in [a, b]$, because, for example, if $c_n \to c$ then

$$\int_{c_n}^d f(x)dx = \int_c^d f(x)dx + \int_{c_n}^c f(x)dx.$$

Now consider step functions $\varphi \leq f \leq \psi$ on $[a, b]$. As step functions attain only finitely many values, there exists $M > 0$ with $\varphi(x) \geq -M$ and $\psi(x) \leq M$ for all $x \in [a, b]$ and from the definition of the upper and the lower integral it follows that

$$\left|\int_{c_n}^c f(x)dx\right| \leq \max\left\{\left|\int_{c_n}^c \varphi(x)dx\right|, \left|\int_{c_n}^c \psi(x)dx\right|\right\} \leq |c_n - c|M \to 0$$

for $c_n \to c$.

2.1 The Fundamental Theorem of Calculus

In the following let $I \subset \mathbb{R}$ always denote an arbitrary interval containing more than one point.

Theorem 2.3 Let $f : I \to \mathbb{R}$ be a continuous function and $a \in I$. For $x \in I$ define

$$F(x) := \int_a^x f(y)dy.$$

Then $F : I \to \mathbb{R}$ is differentiable and $F' = f$.

Proof: For $h \neq 0$ with $x + h \in I$, we have

$$\begin{aligned} \frac{F(x+h) - F(x)}{h} &= \frac{1}{h}\left(\int_a^{x+h} f(y)dy - \int_a^x f(y)dy\right) \\ &= \frac{1}{h}\left(\int_a^x f(y)dy + \int_x^{x+h} f(y)dy - \int_a^x f(y)dy\right) \\ &= \frac{1}{h}\int_x^{x+h} f(y)dy. \end{aligned}$$

By the mean value theorem for integrals (with $g \equiv 1$) there exists $\xi \in [x, x+h]$ if $h > 0$ or $\xi \in [x+h, x]$ if $h < 0$ such that

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{h}\int_x^{x+h} f(y)dy = \frac{1}{h}(hf(\xi)) = f(\xi).$$

As $h \to 0$ we have $\xi \to x$, and by continuity of f it follows that $f(\xi) \to f(x)$. This proves the claim. □

Example 2.4 We already saw in Example 1.11 that for the function $f(x) = x$ we have

$$F(x) = \int_0^x f(y)dy = \frac{x^2}{2}.$$

Since

$$F'(x) = \frac{2x}{2} = x = f(x),$$

this confirms the statement of Theorem 2.3. □

Definition 2.5 A differentiable function $F : I \to \mathbb{R}$ is called an *antiderivative* of an integrable function $f : I \to \mathbb{R}$ if $F'(x) = f(x)$ for all $x \in I$. □

Question 2.2 *Suppose $f : I \to \mathbb{R}$ is n-times differentiable. How often is its antiderivative $F : I \to \mathbb{R}$ differentiable?*

Example 2.6 (i) For $f(x) = \sin x$, the function $F(x) = -\cos x$ is an antiderivative since

$$F'(x) = -(-\sin x) = \sin x.$$

Similarly, $F(x) = \sin x$ is an antiderivative of $f(x) = \cos x$.

(ii) For every polynomial $f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0$ the function

$$F(x) = \frac{a_n}{n+1}x^{n+1} + \frac{a_{n-1}}{n}x^n + \ldots + a_0 x$$

is an antiderivative, as can easily be verified by differentiation.

Some further antiderivatives are given in Example 2.9. □

Remark 2.7 Theorem 2.3 shows that an antiderivative of a continuous function f always exists. Moreover, the antiderivative of a function f is unique up to an additive constant: if F_1 and F_2 are two antiderivatives, then $F_1'(x) = f(x) = F_2'(x)$ for all $x \in I$. Hence $(F_1 - F_2)' \equiv 0$, and therefore $F_1 - F_2$ is a constant function. Thus there exists $C \in \mathbb{R}$ with $F_2(x) = F_1(x) + C$ for all $x \in I$. □

Theorem 2.8 (Fundamental Theorem of Calculus) Let $f : I \to \mathbb{R}$ be continuous and let F be an antiderivative of f. Then for all $a, b \in I$ we have

$$\int_a^b f(x)dx = F(b) - F(a),$$

which we write briefly as

$$F(b) - F(a) =: F(x)\Big|_a^b.$$

Proof: For any $x \in I$ the function

$$F_1(x) := \int_a^x f(y)dy$$

is, by Theorem 2.3, an antiderivative of f with

$$F_1(a) = 0 \quad \text{and} \quad F_1(b) = \int_a^b f(y)dy.$$

Hence the statement holds for $F = F_1$.

For any other antiderivative F there exists $C \in \mathbb{R}$ with $F(x) = F_1(x) + C$. Thus

$$F(b) - F(a) = F_1(b) + C - F_1(a) - C = F_1(b) - F_1(a) = \int_a^b f(y)dy.$$

□

Example 2.9 (i) For $f(x) = \sin x$ it holds that (cf. Example 2.6(i))

$$\int_0^\pi \sin x = -\cos x\Big|_0^\pi = -\cos\pi - (-\cos 0) = -(-1) - (-1) = 2.$$

(ii) For the function $f(x) = x^3$ it holds that (cf. Example 2.6(i))

$$\int_a^b f(x) = \frac{x^4}{4}\Big|_a^b = \frac{1}{4}(b^4 - a^4).$$

(iii) It holds that

$$\int_a^b e^x dx = e^x\Big|_a^b.$$

For, e.g., $a = 0$ and $b = 1$ this implies

$$\int_0^1 e^x dx = e^1 - 1 \approx 1,71828182846$$

(iv) It holds that

$$\int_a^b \frac{1}{x} dx = \ln x\Big|_a^b \qquad \text{for } a, b > 0,$$

since $\ln' x = 1/x$ for $x > 0$. Analogously,

$$\int_a^b \frac{1}{x} dx = \ln(-x)\Big|_a^b \qquad \text{for } a, b < 0.$$

since $(\ln(-x))' = 1/x$ for $x < 0$. In summary, we obtain

$$\int_a^b \frac{1}{x} dx = \ln|x|\Big|_a^b \qquad \text{for } a < b \text{ and } 0 \notin [a, b].$$

□

Question 2.3 *What goes wrong in (iv) when $0 \in [a, b]$?*

2.2 Computation Rules

From the rules of differential calculus one can now derive corresponding computation rules for integrals by exploiting the relationship between differential and integral calculus. We begin with the chain rule of differential calculus, whose counterpart in integral calculus is the following substitution rule.

Theorem 2.10 (Substitution Rule) Let $f : I \to \mathbb{R}$ be continuous and let $\varphi : [a, b] \to \mathbb{R}$ be continuously differentiable with $\varphi([a, b]) \subset I$. Then we have[1]

$$\int_a^b f(\varphi(x))\varphi'(x)dx = \int_{\varphi(a)}^{\varphi(b)} f(y)dy.$$

[1] Using the symbolic shorthand $d\varphi(x) := \varphi'(x)dx$ this can be written more briefly and is easier to remember as

$$\int_a^b f(\varphi(x))d\varphi(x) = \int_{\varphi(a)}^{\varphi(b)} f(y)dy.$$

Proof: Let F be an antiderivative of f. Then by the chain rule

$$(F \circ \varphi)'(x) = F'(\varphi(x))\varphi'(x) = f(\varphi(x))\varphi'(x).$$

Thus $F \circ \varphi$ is an antiderivative of $x \mapsto f(\varphi(x))\varphi'(x)$, and therefore

$$\int_a^b f(\varphi(x))\varphi'(x)dx = F \circ \varphi(x)\Big|_a^b = F(\varphi(b)) - F(\varphi(a)) = F(y)\Big|_{\varphi(a)}^{\varphi(b)} = \int_{\varphi(a)}^{\varphi(b)} f(y)dy.$$

□

Note that renaming the integration variable from x to y in the second integral is not necessary for mathematical correctness. However, it helps to avoid confusion when applying substitution. As a reminder one may use the equations $y = \varphi(x)$ and $dy = d\varphi(x) = \varphi'(x)dx$.

Example 2.11 (i) Applying Theorem 2.10 with $\varphi(x) = x + c$ ($\Rightarrow \varphi'(x) = 1$) yields

$$\int_a^b f(x+c)dx = \int_{a+c}^{b+c} f(y)dy.$$

(ii) Applying Theorem 2.10 with $\varphi(x) = cx$, $c \neq 0$ ($\Rightarrow \varphi'(x) = c$) gives

$$c\int_a^b f(cx)dx = \int_{ca}^{cb} f(y)dy.$$

(iii) With $\varphi(x) = x^2$ and $\varphi'(x) = 2x$ we obtain

$$2\int_a^b xf(x^2)dx = \int_{a^2}^{b^2} f(y)dy.$$

(iv) With $f(x) = 1/x$ and an arbitrary continuously differentiable function $\varphi : [a,b] \to \mathbb{R}$ with $\varphi(x) \neq 0$ for $x \in [a,b]$ we obtain

$$\int_a^b \frac{\varphi'(x)}{\varphi(x)}dx = \int_a^b f(\varphi(x))\varphi'(x)dx = \int_{\varphi(a)}^{\varphi(b)} f(y)dy = \ln|y|\Big|_{\varphi(a)}^{\varphi(b)} = \ln|\varphi(x)|\Big|_a^b.$$

With $\varphi(x) = \cos x$ and $[a,b] \subset (-\pi/2, \pi/2)$ one can, for example, compute

$$\int_a^b \tan x dx = \int_a^b \frac{\sin x}{\cos x}dx = -\ln\cos x\Big|_a^b.$$

(v) **(Computation of the area of a circle)** The boundary of the circle with radius 1 centered at $(0,0)$ consists of the points $x, y \in \mathbb{R}$ satisfying $x^2 + y^2 = 1$. If we consider only the points with $y \geq 0$ (the "upper" semicircle), they can be written as pairs $(x, \sqrt{1-x^2})$, i.e. $(x, f(x))$ with the function $f(x) = \sqrt{1-x^2}$.

By computing the integral

$$\int_{-1}^1 \sqrt{1-x^2}dx$$

we can therefore determine the area of the semicircle (and by doubling the result the area of the entire circle).

We apply the substitution formula with $f(x) = \sqrt{1-x^2}$, $\varphi(x) = \sin x$, and initially arbitrary limits $a < b$. This gives

$$\int_a^b \sqrt{1-\sin^2 x}\,\sin' x\,dx = \int_a^b f(\varphi(x))\varphi'(x)dx = \int_{\varphi(a)}^{\varphi(b)} f(y)dy = \int_{\sin a}^{\sin b} \sqrt{1-y^2}dy.$$

Choosing $a := \arcsin(-1) = -\pi/2$ and $b := \arcsin(1) = \pi/2$ gives $\sin a = -1$ and $\sin b = 1$, and hence

$$\int_{-1}^{1} \sqrt{1-x^2}dx = \int_a^b \sqrt{1-\sin^2 x}\,\sin' x\,dx = \int_a^b \cos^2 x\,dx.$$

Since

$$\cos^2 x = \left(\frac{e^{ix}+e^{-ix}}{2}\right)^2 = \frac{1}{4}(e^{2ix}+e^{-2ix}) + \frac{1}{2} = \frac{1}{2}(\cos 2x + 1),$$

we obtain

$$\begin{aligned} \int_{-1}^{1} \sqrt{1-x^2}dx &= \int_a^b \frac{1}{2}(\cos 2x + 1)dx \\ &= \frac{1}{4}\sin 2x\Big|_a^b + \frac{1}{2}x\Big|_a^b \\ &= \frac{1}{4}(\sin\pi - \sin(-\pi)) + \frac{1}{2}\left(\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right) = \frac{\pi}{2}, \end{aligned}$$

since $\sin\pi = \sin(-\pi) = 0$.

Thus the area of a semicircle of radius 1 is—not entirely unexpectedly—exactly $\pi/2$. □

Question 2.4 *How would the computation in Part (v) of this example look for the lower semicircle? Could you also do the computation for the left or right semicircle?*

We now turn to the second important rule for derivatives, the product rule. This rule is the basis for the method of integration by parts.

Theorem 2.12 (Integration by Parts) Let $f, g : [a,b] \to \mathbb{R}$ be two continuously differentiable functions. Then[2]

$$\int_a^b f(x)g'(x)dx + \int_a^b g(x)f'(x)dx = f(x)g(x)\Big|_a^b.$$

[2]Using the shorthand notation introduced in Theorem 2.10, this can also be written as

$$\int_a^b f(x)dg(x) + \int_a^b g(x)df(x) = f(x)g(x)\Big|_a^b.$$

Proof: Let $F := fg$. By the product rule we have

$$F'(x) = f'(x)g(x) + f(x)g'(x).$$

By Theorem 2.8 it follows that

$$\int_a^b f(x)g'(x)dx + \int_a^b g(x)f'(x)dx = \int_a^b F'(x)dx = F(x)\Big|_a^b = f(x)g(x)\Big|_a^b,$$

which proves the claim. □

Question 2.5 *Which integration rule for continuous functions can be derived from the sum rule for the derivative [7, First rule in Theorem 16.8]?*

Example 2.13 (i) With $f(x) = \ln x$, $g(x) = x$ and $b > a > 0$ we obtain

$$\begin{aligned}
\int_a^b \ln x\,dx &= \int_a^b f(x)g'(x)dx \\
&= f(x)g(x)\Big|_a^b - \int_a^b g(x)f'(x)dx \\
&= x\ln x\Big|_a^b - \int_a^b 1\,dx \\
&= x\ln x\Big|_a^b - x\Big|_a^b = x(\ln x - 1)\Big|_a^b.
\end{aligned}$$

(ii) With $f(x) = \arctan x$, $g(x) = x$ we obtain

$$\int_a^b \arctan x\,dx = x\arctan x\Big|_a^b - \int_a^b x\,\arctan'(x)dx.$$

Since $\arctan' x = 1/(1+x^2)$, we obtain

$$\int_a^b x\,\arctan'(x)dx = \int_a^b \frac{x}{1+x^2}dx.$$

With $\varphi(x) = x^2 + 1$ ($\Rightarrow \varphi'(x) = 2x$) and Example 2.11(iv) we obtain

$$\int_a^b \frac{x}{1+x^2}dx = \frac{1}{2}\int_a^b \frac{\varphi'(x)}{\varphi(x)}dx = \frac{1}{2}\ln|\varphi(x)|\Big|_a^b = \frac{1}{2}\ln(x^2+1)\Big|_a^b.$$

Altogether we obtain

$$\int_a^b \arctan x\,dx = x\arctan x - \frac{1}{2}\ln(x^2+1)\Big|_a^b.$$

□

An application of integration by parts is given by the following theorem.

Theorem 2.14 (Trapezoidal Rule) Let $f : [a,b] \to \mathbb{R}$ be twice continuously differentiable. Then there exists $\xi \in [a,b]$ such that

$$\int_a^b f(x)dx = \frac{b-a}{2}(f(a)+f(b)) - \frac{(b-a)^3}{12}f''(\xi).$$

Proof: We first consider the case $a = 0$, $b = 1$. Let

$$g(x) = \frac{x(1-x)}{2}.$$

Then $g'(x) = \frac{1}{2} - x$ and $g''(x) = -1$, and therefore

$$\int_0^1 f(x)dx = -\int_0^1 g''(x)f(x)dx.$$

Applying integration by parts gives

$$-\int_0^1 g''(x)f(x)dx = -g'(x)f(x)\Big|_0^1 + \int_0^1 g'(x)f'(x)dx.$$

The first term on the right-hand side becomes

$$-g'(x)f(x)\Big|_0^1 = \frac{1}{2}(f(0)+f(1)).$$

For the second term we apply integration by parts once more:

$$\int_0^1 g'(x)f'(x)dx = g(x)f'(x)\Big|_0^1 - \int_0^1 g(x)f''(x)dx.$$

Here the first term on the right-hand side is zero. For the second term we obtain, by the mean value theorem,

$$\int_0^1 g(x)f''(x)dx = f''(\xi)\int_0^1 g(x)dx = \frac{f''(\xi)}{12}.$$

Combining these results yields

$$\int_0^1 f(x)dx = \frac{1}{2}(f(0)+f(1)) - \frac{f''(\xi)}{12},$$

which proves the claim for $[0,1]$.

For arbitrary a, b we perform the substitution $\varphi(x) = a + x(b-a)$ with $\varphi'(x) = b-a$. Then by the first part

$$\begin{aligned}\int_a^b f(y)dy &= \int_0^1 f(\varphi(x))\varphi'(x)dx \\ &= \frac{1}{2}\big(f(\varphi(0))\varphi'(0) + f(\varphi(1))\varphi'(1)\big) - ((f\circ\varphi)\varphi')''(\xi)\frac{1}{12}.\end{aligned}$$

From the usual differentiation rules we obtain

$$((f \circ \varphi)\varphi')''(\xi) = (b-a)^3 f''(\varphi(\xi)),$$

which proves the claim if we replace ξ by $\varphi(\xi)$. $\square$

The trapezoidal rule can be used to approximate integrals by splitting the integration interval $[a, b]$ into $(b-a)/h$ subintervals of length h, applying the formula on each interval $[a + kh, a + (k+1)h]$, and adding up the values.

Let $C \geq 0$ be a bound for $|f''|$ on $[a, b]$. On each subinterval the error is then $Ch^3/12$. On the entire interval with $(b-a)/h$ subintervals the total error is therefore

$$C\frac{b-a}{h}\frac{h^3}{12} = C\frac{b-a}{12}h^2.$$

Thus, the smaller we choose the subinterval length h, the more accurate the approximation becomes. The factor h^2 shows that when h is halved, the error decreases by a factor of 1/4.

Question 2.6 *We want to approximate the integral*

$$\int_0^1 x^2 dx = \left.\frac{x^3}{3}\right|_0^1 = \frac{1}{3}.$$

Consider the partition $x_0 = 0$, $x_1 = 1/4$, $x_2 = 1/2$, $x_3 = 3/4$, and $x_4 = 1$ of the interval $[0, 1]$. Approximate the integral by the Riemann sum with $\xi_k = x_k$ and with $\xi = x_{k+1}$, and by the trapezoidal rule on subintervals with length $h = 1/4$. Which approximation is the most accurate?

Chapter 3

Properties and Extensions of the Riemann Integral

3.1 Properties of the Riemann Integral

Last update: July 21, 2026

In the following we derive several properties of the Riemann integral from corresponding properties of the upper and lower integrals. The following lemma ensures that, when investigating properties of the upper and lower integrals, it suffices to consider only one of them.

Lemma 3.1 We have

$$\int_{*a}^{b} f(x)dx = -\int_{a}^{*b} (-f(x))dx.$$

Proof: This follows from the equivalence

$$\varphi \leq f \quad \Leftrightarrow \quad -\varphi \geq -f$$

and the identities

$$\int_a^b \varphi(x)dx = -\int_a^b (-\varphi(x))dx$$

as well as — for arbitrary sets $A \subset \mathbb{R}$ —

$$\inf\{a \mid a \in A\} = -\sup\{-a \mid a \in A\}.$$

□

Theorem 3.2 For bounded functions $f, g : [a,b] \to \mathbb{R}$ and $\lambda > 0$ the following holds for the upper integral:

$$\int_a^{*b} (f+g)(x)dx \leq \int_a^{*b} f(x)dx + \int_a^{*b} g(x)dx$$

as well as

$$\int_a^{*b} (\lambda f)(x)dx = \lambda \int_a^{*b} f(x)dx.$$

Proof: For the first statement it suffices to show that for every $\varepsilon > 0$ the inequality

$$\int_a^{*b} (f+g)(x)dx \leq \int_a^{*b} f(x)dx + \int_a^{*b} g(x)dx + \varepsilon$$

holds.

By the definition of the upper integral and the definition of the infimum, there exist step functions $\varphi, \psi \in T(a,b)$ with $\varphi \geq f$, $\psi \geq g$ and

$$\int_a^b \varphi(x)dx \leq \int_a^{*b} f(x)dx + \frac{\varepsilon}{2} \quad \text{and} \quad \int_a^b \psi(x)dx \leq \int_a^{*b} g(x)dx + \frac{\varepsilon}{2}.$$

Since $\varphi \geq f$ and $\psi \geq g$ imply $\varphi + \psi \geq f + g$, we obtain

$$\begin{aligned}
\int_a^{*b} (f+g)(x)dx &\leq \int_a^b (\varphi + \psi)(x)dx \\
&= \int_a^b \varphi(x)dx + \int_a^b \psi(x)dx \\
&\leq \int_a^{*b} f(x)dx + \int_a^{*b} g(x)dx + \varepsilon.
\end{aligned}$$

For the second statement it suffices to show that for all $\varepsilon > 0$ the inequalities

$$\lambda \int_a^{*b} f(x)dx - \varepsilon \leq \int_a^{*b} (\lambda f)(x)dx \leq \lambda \int_a^{*b} f(x)dx + \varepsilon$$

hold.

Let $\varphi \in T(a,b)$ with $\varphi \geq f$ and

$$\int_a^b \varphi(x)dx \leq \int_a^{*b} f(x)dx + \frac{\varepsilon}{\lambda}.$$

Then $\lambda\varphi \geq \lambda f$ and therefore

$$\int_a^{*b} (\lambda f)(x)dx \leq \int_a^b (\lambda\varphi)(x)dx = \lambda \int_a^b \varphi(x)dx \leq \lambda \int_a^{*b} f(x)dx + \varepsilon.$$

Analogously, let $\psi \in T(a,b)$ with $\psi \geq \lambda f$ and

$$\int_a^b \psi(x)dx \leq \int_a^{*b} (\lambda f)(x)dx + \varepsilon.$$

Then $\psi/\lambda \geq f$ and it follows that

$$\int_a^{*b} (\lambda f)(x)dx \geq \int_a^b \psi(x)dx - \varepsilon = \lambda \int_a^b (\psi/\lambda)(x)dx - \varepsilon \geq \lambda \int_a^{*b} f(x)dx - \varepsilon.$$

$\square$

Question 3.1 *Can you find bounded functions $f, g : [a, b] \to \mathbb{R}$ with*

$$\int_a^{*b} (f+g)(x)dx < \int_a^{*b} f(x)dx + \int_a^{*b} g(x)dx?$$

Hint: Use f from Example 1.8(ii).

Theorem 3.3 For Riemann-integrable functions $f, g : [a, b] \to \mathbb{R}$ and $\lambda \in \mathbb{R}$, the functions $f + g$ and λf are also Riemann-integrable, and

$$\int_a^b (f+g)(x)dx = \int_a^b f(x)dx + \int_a^b g(x)dx$$

as well as

$$\int_a^b (\lambda f)(x)dx = \lambda \int_a^b f(x)dx.$$

If in addition $f \leq g$, then

$$\int_a^b f(x)dx \leq \int_a^b g(x)dx.$$

Proof: For $f + g$ we obtain from Theorem 3.2

$$\int_a^{*b} (f+g)(x)dx \leq \int_a^{*b} f(x)dx + \int_a^{*b} g(x)dx = \int_a^b f(x)dx + \int_a^b g(x)dx.$$

Using Lemma 3.1 and Theorem 3.2 we also obtain

$$\begin{aligned}
\int_{*a}^{b} (f+g)(x)dx &= -\int_a^{*b} (-f-g)(x)dx \\
&\geq -\int_a^{*b} (-f)(x)dx - \int_a^{*b} (-g)(x)dx \\
&= \int_{*a}^{b} f(x)dx + \int_{*a}^{b} g(x)dx \\
&= \int_a^b f(x)dx + \int_a^b g(x)dx.
\end{aligned}$$

Thus

$$\int_{*a}^{b} (f+g)(x)dx \leq \int_a^{*b} (f+g)(x) \leq \int_a^b f(x)dx + \int_a^b g(x)dx \leq \int_{*a}^{b} (f+g)(x)dx,$$

which implies equality and hence Riemann integrability and the claimed identity.

For λf, the statement for $\lambda > 0$ follows immediately from Theorem 3.2, since

$$\int_a^{*b} (\lambda f)(x)dx = \lambda \int_a^{*b} f(x)dx = \lambda \int_a^b f(x)dx$$

and

$$\int_{*a}^{b}(\lambda f)(x)dx = -\int_{a}^{*b}(\lambda(-f))(x)dx = -\lambda\int_{a}^{*b}(-f)(x)dx = \lambda\int_{*a}^{b}f(x)dx = \lambda\int_{a}^{b}f(x)dx.$$

For $\lambda = 0$ the statement is clear (since $0 \cdot f$ is a step function), and for $\lambda < 0$ the statement follows from

$$\int_{*a}^{b}(\lambda f)(x)dx = -\int_{a}^{*b}((-\lambda)f)(x)dx = \lambda\int_{a}^{*b}f(x)dx = \lambda\int_{a}^{b}f(x)dx$$

and

$$\int_{a}^{*b}(\lambda f)(x)dx = -\int_{*a}^{b}((-\lambda)f)(x)dx = \lambda\int_{*a}^{b}f(x)dx = \lambda\int_{a}^{b}f(x)dx.$$

The stated inequality follows since, because $f-g \leq 0$, the function $\varphi \equiv 0$ is a step function with $\varphi \geq f-g$. Hence

$$\int_{a}^{b}f(x)dx - \int_{a}^{b}g(x)dx = \int_{a}^{b}(f-g)(x)dx = \int_{a}^{*b}(f-g)(x)dx \leq \int_{a}^{b}\varphi(x)dx = 0,$$

which proves the claim. □

The first two properties established in Theorem 3.3 state that the integral is a linear mapping from the set of Riemann-integrable functions to $\mathbb{R}$: sums and real multiples of functions are mapped to the sums and real multiples of their integral values. For this reason the Riemann integral is called a *linear functional*. Because of the third property, the integral is also a *monotone functional*.

Question 3.2 *Theorem 3.3 is easier to prove if f and g are continuous. Can you explain why?*

3.2 Further Riemann-Integrable Functions

In Chapter 1 we have already seen that step functions and continuous functions are Riemann-integrable, whereas the function defined by $f(x) = 1$ for $x \in \mathbb{Q}$ and $f(x) = 0$ for $x \in \mathbb{R} \setminus \mathbb{Q}$ is not. In this section we consider further classes of functions that are Riemann-integrable.

Theorem 3.4 Every monotone function $f : [a, b] \to \mathbb{R}$ is Riemann-integrable.

Proof: We prove the statement for monotonically increasing functions; the proof for monotonically decreasing functions is analogous.

For arbitrary $n \in \mathbb{N}$ we define the partition

$$x_k = a + k\frac{b-a}{n}, \quad k = 0, \ldots, n.$$

Define $c_k := f(x_k)$, $c'_k := f(x_{k+1})$ and the associated step functions φ and ψ as in the proof of Theorem 1.10. From the monotonicity of f we obtain the inequalities $\varphi \leq f \leq \psi$. Moreover,

$$\begin{aligned}\int_a^b \psi(x)dx - \int_a^b \varphi(x)dx &= \int_a^b \psi(x) - \varphi(x)dx \\ &= \sum_{k=0}^{n-1}(x_{k+1} - x_k)(c'_k - c_k) \\ &= \sum_{k=0}^{n-1} \frac{b-a}{n}\big(f(x_{k+1}) - f(x_k)\big) \\ &= \frac{b-a}{n}\big(f(b) - f(a)\big).\end{aligned}$$

Given arbitrary $\varepsilon > 0$, choose

$$n > \frac{(b-a)(f(b) - f(a))}{\varepsilon}.$$

Then

$$\int_a^b \psi(x)dx - \int_a^b \varphi(x)dx < \varepsilon,$$

and the claim follows from Theorem 1.9. □

Definition 3.5 For a function $f : D \to \mathbb{R}$ we define

$$f^+(x) := \begin{cases} f(x), & \text{if } f(x) > 0, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$f^-(x) := \begin{cases} -f(x), & \text{if } f(x) < 0, \\ 0, & \text{otherwise.} \end{cases}$$

□

From this definition the identities

$$f = f^+ - f^- \qquad \text{and} \qquad |f| = f^+ + f^-$$

follow immediately.

Theorem 3.6 Let $f, g : [a, b] \to \mathbb{R}$ be Riemann-integrable functions. Then the following statements hold:

(i) The functions f^+ and f^- are Riemann-integrable.

(ii) For every $p \in [1,\infty)$ the function $|f|^p$ is Riemann-integrable. Moreover, the inequality (triangle inequality for integrals)

$$\left|\int_a^b f(x)dx\right| \leq \int_a^b |f(x)|dx$$

holds.

(iii) The function $fg : [a,b] \to \mathbb{R}$ is Riemann-integrable. In general,

$$\int_a^b f(x)g(x)dx \neq \left(\int_a^b f(x)dx\right)\left(\int_a^b g(x)dx\right).$$

Question 3.3 *Find two continuous functions $f, g : \mathbb{R} \to \mathbb{R}$ with*

$$\int_0^1 f(x)g(x)dx \neq \left(\int_0^1 f(x)dx\right)\left(\int_0^1 g(x)dx\right).$$

Hint: This is possible with $f = g$.

Proof:

(i) By Theorem 1.9, for every $\varepsilon > 0$ there exist step functions $\varphi, \psi \in T(a,b)$ with $\varphi \leq f \leq \psi$ and

$$\int_a^b \psi(x) - \varphi(x)dx \leq \varepsilon.$$

Clearly φ^+ and ψ^+ are also step functions with $\varphi^+ \leq f^+ \leq \psi^+$ and

$$\int_a^b \psi^+(x) - \varphi^+(x)dx \leq \int_a^b \psi(x) - \varphi(x)dx \leq \varepsilon.$$

Hence f^+ is integrable. The integrability of f^- is shown analogously.

(ii) By (i) and Theorem 3.3, the function $|f| = f^+ + f^-$ is integrable, so the statement holds for $p = 1$.

For $p > 1$ it suffices to consider functions f with $0 \leq |f| \leq 1$. Since $|f|$ is bounded, this can always be achieved by multiplying $|f|$ by a sufficiently small $\lambda > 0$; by Theorem 3.3, the scaled function is integrable if and only if the original one is integrable.

Since $|f|$ is integrable, by Theorem 1.9 there exist, for every $\varepsilon > 0$, step functions $\varphi, \psi \in T(a,b)$ with $\varphi \leq |f| \leq \psi$ and

$$\int_a^b \psi(x) - \varphi(x)dx \leq \frac{\varepsilon}{p}.$$

Because $0 \leq |f| \leq 1$, we may choose $\varphi \geq 0$ and $\psi \leq 1$.

Then φ^p and ψ^p are step functions satisfying $\varphi^p \leq |f|^p \leq \psi^p$. For the function $g(x) = x^p$, the mean value theorem implies that for $0 \leq x < y \leq 1$

$$y^p - x^p = g(y) - g(x) = g'(\xi)(y - x) = p\xi^{p-1}(y - x) \leq p(y - x),$$

since $\xi \in (x, y) \subset [0, 1]$ implies $\xi^{p-1} \in [0, 1]$.

Thus

$$\psi^p - \varphi^p \leq p(\psi - \varphi),$$

and therefore

$$\int_a^b \psi^p(x) - \varphi^p(x) dx \leq p \int_a^b \psi(x) - \varphi(x) dx \leq \varepsilon.$$

Hence $|f|^p$ is Riemann-integrable by Theorem 1.9.

The stated inequality for $p = 1$ follows directly from the monotonicity and linearity of the integral together with $f \leq |f|$ and $-f \leq |f|$.

(iii) The claim follows from (ii) since

$$fg = \frac{1}{4}\big((f+g)^2 - (f-g)^2\big) = \frac{1}{4}\big(|f+g|^2 - |f-g|^2\big).$$

$\square$

Question 3.4 *Can you show that* $h : [a, b] \to \mathbb{R}$ *given by* $h(x) = \max\{f(x), g(x)\}$ *is Riemann integrable if* $f, g : [a, b] \to \mathbb{R}$ *are Riemann integrable?*
Hint: The term $(x + y + |x - y|)/2$ *can be useful to answer this question.*

3.3 Improper Integrals

Improper integrals are integrals in which either the interval of integration is unbounded or the integrand is not defined at one (or both) endpoints of the interval of integration. Using the concept of limits, both situations can be handled easily.

Definition 3.7 (i) Let $f : [a, \infty) \to \mathbb{R}$ be a function that is Riemann-integrable on every interval $[a, b]$, $b > a$. We then call the function *Riemann-integrable on* $[a, \infty)$ or say that *the Riemann integral on* $[a, \infty)$ *exists* if the limit

$$\int_a^\infty f(x)\, dx := \lim_{b\to\infty} \int_a^b f(x)\, dx$$

exists (in the proper sense, i.e., it is finite).

Analogously we define Riemann integrability on $(-\infty, b]$ by

$$\int_{-\infty}^b f(x)\, dx := \lim_{a\to-\infty} \int_a^b f(x)\, dx$$

and on $(-\infty, \infty)$ by

$$\int_{-\infty}^\infty f(x)\, dx := \lim_{a\to-\infty} \lim_{b\to\infty} \int_a^b f(x)\, dx = \lim_{a\to-\infty} \int_a^c f(x)\, dx + \lim_{b\to\infty} \int_c^b f(x)\, dx,$$

where $c \in \mathbb{R}$ is arbitrary.

(ii) Let $f : (a, b] \to \mathbb{R}$ be a function that is Riemann-integrable on every interval $[c, b]$, $c \in (a, b)$. We call the function *Riemann-integrable on* $(a, b]$ if the limit

$$\int_a^b f(x)\,dx := \lim_{\substack{c\to a\\ c>a}} \int_c^b f(x)\,dx$$

exists.

Analogously to (i) we define Riemann integrability on $[a, b)$, (a, b), (a, ∞) and $(-\infty, b)$. □

Example 3.8 (i) The Riemann integral $\int_1^\infty 1/x^s\,dx$ exists for all $s > 1$. Indeed,

$$\int_1^b \frac{1}{x^s}\,dx = \frac{1}{1-s}\frac{1}{x^{s-1}}\Big|_1^b = \frac{1}{s-1}\left(1 - \frac{1}{b^{s-1}}\right) \to \frac{1}{s-1}$$

as $b \to \infty$.

(ii) The Riemann integral $\int_1^\infty 1/x^s\,dx$ does not exist for any $s \le 1$. For $s < 1$ the same calculation as above yields

$$\int_1^b \frac{1}{x^s}\,dx = \frac{1}{s-1}\left(1 - \frac{1}{b^{s-1}}\right) = \frac{1}{s-1}\left(1 - b^{1-s}\right) \to \infty$$

as $b \to \infty$, and for $s = 1$ we have

$$\int_1^b \frac{1}{x^s}\,dx = \ln x\Big|_1^b = \ln b \to \infty$$

as $b \to \infty$.

(iii) The Riemann integral $\int_0^1 1/x^s\,dx$ exists for all $s < 1$. Indeed, for $c \in (0, 1)$ we have

$$\int_c^1 \frac{1}{x^s}\,dx = \frac{1}{1-s}\frac{1}{x^{s-1}}\Big|_c^1 = \frac{1}{1-s}\left(1 - c^{1-s}\right) \to \frac{1}{1-s}$$

as $c \to 0$.

(iv) With calculations similar to those in (i)–(iii) one shows that the Riemann integral $\int_0^1 1/x^s\,dx$ does not exist for any $s \ge 1$.

(v) The integral

$$\int_{-1}^1 \frac{1}{\sqrt{1-x^2}}\,dx$$

exists, since[1]

$$\begin{aligned}
\int_{-1}^1 \frac{1}{\sqrt{1-x^2}}\,dx &= \lim_{\substack{a\to -1\\ a>-1}} \int_a^0 \frac{1}{\sqrt{1-x^2}}\,dx + \lim_{\substack{b\to 1\\ b<1}} \int_0^b \frac{1}{\sqrt{1-x^2}}\,dx \\
&= -\lim_{\substack{a\to -1\\ a>-1}} \arcsin(a) + \lim_{\substack{b\to 1\\ b<1}} \arcsin(b) \\
&= -\left(-\frac{\pi}{2}\right) + \frac{\pi}{2} = \pi.
\end{aligned}$$

[1] According to [7, Theorem 16.10],

$$\arcsin'(x) = \frac{1}{\sin'(\arcsin(x))} = \frac{1}{\cos(\arcsin(x))} = \frac{1}{\sqrt{1-\sin^2(\arcsin(x))}} = \frac{1}{\sqrt{1-x^2}}.$$

□

Question 3.5 *What can you say about the existence of* $\displaystyle\int_0^\infty \frac{1}{x^s} dx$ *for different* $s > 0$*?*

The definition of the integral $\int_0^\infty f(x)\,dx$ via a limit is reminiscent of the definition of the limit of an infinite series. Indeed, a series is nothing other than an integral over a step function. But even for some functions that are not step functions one can establish a relationship between the convergence of series and integrals.

Theorem 3.9 Let $f : [1, \infty) \to \mathbb{R}$ be a monotonically decreasing function with $f \geq 0$. Then the limit $\sum_{k=1}^\infty f(k)$ exists if and only if the integral $\int_1^\infty f(x)\,dx$ exists.

Proof: From the monotonicity we obtain

$$f(k) \leq \int_{k-1}^k f(x)\,dx \leq f(k-1).$$

Summing over k gives

$$\sum_{k=2}^n f(k) \leq \int_1^n f(x)\,dx \leq \sum_{k=1}^{n-1} f(k).$$

If $\int_1^\infty f(x)\,dx$ exists, then the sequence $\sum_{k=1}^n f(k) = f(1) + \sum_{k=2}^n f(k)$ is bounded. Since the series is also monotone because $f(k) \geq 0$, it is therefore convergent.

Conversely, if the limit $\sum_{k=1}^\infty f(k)$ exists, then $\int_1^a f(x)\,dx$ is bounded and monotone in a. We claim that this implies that it is convergent for $a \to \infty$: Monotonicity and boundedness imply that the sequence $\left(\int_1^n f(x)\,dx\right)_{n\geq 1}$ converges monotonically as $n \to \infty$. For any monotone increasing integer sequence n_k converging to ∞ as $k \to \infty$, after removing duplicate entries, which does not change the limit, the subsequence $\int_1^{n_k} f(x)\,dx$ converges to the same limit as $\int_1^n f(x)\,dx$. For an arbitrary sequence $a_k \to \infty$, $a_k \geq 1$, we can find monotone increasing integer sequences $n_k \leq a_k \leq N_k$ that converge to ∞. This implies

$$\int_1^{n_k} f(x)\,dx \leq \int_1^{a_k} f(x)\,dx \leq \int_1^{N_k} f(x)\,dx$$

and since the first and the last sequence in this inequality chain converge to the same limit as $k \to \infty$, the middle one must converge to this limit, too. This shows the claim. □

Question 3.6 *Does the theorem remain true when "monotonically decreasing" is changed to "monotonically increasing" (and everything else remains the same)?*

Chapter 4

Sequences of Functions

Last update: July 21, 2026

Sequences can be defined not only for real and complex numbers, but also for functions. In this chapter, we examine the essential properties of such sequences of functions.

4.1 Definitions

Definition 4.1 A sequence $(f_n)_{n\in\mathbb{N}}$ of functions $f_n : D \to \mathbb{R}$ is called a *sequence of functions.* $\square$

We give some simple examples of sequences of functions.

Example 4.2 (i) $f_n(x) = x^n$, $D = \mathbb{R}$ or $D = [0, 1]$.

(ii) $f_n(x) = |x|^{1+1/n}$, $D = \mathbb{R}$, $n \geq 1$.

(iii) $f_n(x) = (1 + 1/n)x^2$, $D = \mathbb{R}$, $n \geq 1$.

(iv) Sequences of functions arising from series:

$$f_n(x) = \sum_{k=0}^{n} g_k(x), \quad g_k : D \to \mathbb{R}.$$

(v) Power series:

$$f_n(x) = \sum_{k=0}^{n} c_k (x - x_0)^k$$

with $c_k \in \mathbb{R}$, $x_0 \in \mathbb{R}$, and $D \subset \mathbb{R}$. Note that this is a special case of (iv). $\square$

The following definition introduces a notion of convergence for sequences of functions.

Definition 4.3 We say that a sequence of functions $(f_n)_{n\in\mathbb{N}}$ *converges pointwise* to a function $f : D \to \mathbb{R}$, if

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

for all $x \in D$. The function f is then called the *limit function.* $\square$

Example 4.4 We consider the pointwise convergence of the sequences from Example 4.2 as $n \to \infty$.

(i) $f_n(x) = x^n$: For $|x| > 1$, this sequence clearly does not converge because $|x^n| = |x|^n \to \infty$. On $[0, 1]$, however, it converges pointwise to

$$f_n(x) \to \begin{cases} 0, & x \in [0,1), \\ 1, & x = 1. \end{cases}$$

(ii) $f_n(x) = |x|^{1+1/n}$: For all $x \in \mathbb{R}$,

$$f_n(x) = |x|^{1+1/n} \to |x| =: f(x).$$

(iii) $f_n(x) = (1 + 1/n)x^2$: For each $x \in \mathbb{R}$,

$$f_n(x) = (1 + 1/n)x^2 = x^2 + x^2/n \to x^2 =: f(x).$$

(iv) and (v) The functions defined by the series converge pointwise to the corresponding series limits, if these exist. For a power series, this holds at least on the restricted domain $D = \{x \in \mathbb{R} \mid |x - x_0| < r\}$, where r is the radius of convergence [7, Theorem 9.23]. □

Question 4.1 *If every f_n in a sequence of functions is bounded and a pointwise limit function f exists, is f also bounded?*

4.2 Continuity and Differentiability

In this section, we investigate two important questions: If all functions f_n in a sequence are continuous or differentiable, and the sequence converges to a limit function f, is f necessarily continuous or differentiable?

We begin with continuity and consider Example 4.2(i), i.e., $f_n(x) = x^n$ on $D = [0, 1]$. Since each f_n is a polynomial, it is continuous on $\mathbb{R}$ and, in particular, on D. However, by Example 4.4(i), the sequence converges pointwise to the limit function

$$f(x) = \begin{cases} 0, & x \in [0,1), \\ 1, & x = 1, \end{cases}$$

which is clearly discontinuous at $x = 1$.

Hence, the limit of a pointwise convergent sequence of continuous functions is not automatically continuous. To ensure continuity, we need the stronger concept of convergence given below.

Definition 4.5 A sequence of functions $f_n : D \to \mathbb{R}$ is called *uniformly convergent* to a limit function $f : D \to \mathbb{R}$, if

$$\lim_{n\to\infty} \sup_{y\in D} |f_n(y) - f(y)| = 0.$$

□

Note that uniform convergence implies pointwise convergence because

$$|f_n(x) - f(x)| \leq \sup_{y \in D} |f_n(y) - f(y)| \to 0 \quad \text{for all } x \in D.$$

The converse is not true, as Example 4.2(i) shows:

For $n \geq 1$, define $x_n := \sqrt[n]{1/2} \in (0, 1)$. Then $f_n(x_n) = (\sqrt[n]{1/2})^n = 1/2$ and for the limit function f from Example 4.4(i) we get

$$\sup_{y \in D} |f_n(y) - f(y)| \geq |f_n(x_n) - f(x_n)| = |1/2 - 0| = 1/2$$

for all $n \in \mathbb{N}$, so

$$\sup_{y \in D} |f_n(y) - f(y)| \not\to 0.$$

In Example 4.2(ii), uniform convergence does hold if we restrict the domain to $D = [-d, d]$ for any $d > 0$. For all $y \in [-d, d]$,

$$|f_n(y) - f(y)| = \left| |y|^{1+1/n} - |y| \right| = |y| \left| |y|^{1/n} - 1 \right|.$$

To show that the supremum over $y \in [-d, d]$ tends to 0, we consider three cases for an arbitrary $\varepsilon \in (0, 1)$:

Case 1: $|y| < \varepsilon$. Then $||y|^{1/n} - 1| \leq 1$, so

$$|y| \left| |y|^{1/n} - 1 \right| \leq |y| < \varepsilon.$$

Case 2: $|y| \in [\varepsilon, 1]$. Then $|y| \leq 1$, $|y|^{1/n} \leq 1$, and $|y|^{1/n} \geq \varepsilon^{1/n}$, hence

$$|y| \left| |y|^{1/n} - 1 \right| = |y|(1 - |y|^{1/n}) \leq 1 - \varepsilon^{1/n} < \varepsilon$$

for all sufficiently large n, since $\varepsilon^{1/n} \to 1$ as $n \to \infty$.

Case 3: $|y| \in [1, d]$. Then $|y|^{1/n} \geq 1$, so

$$|y| \left| |y|^{1/n} - 1 \right| = |y|(|y|^{1/n} - 1) \leq d(d^{1/n} - 1) < \varepsilon$$

for all sufficiently large n, since $d^{1/n} \to 1$.

Combining the three cases shows that

$$\sup_{y \in [-d,d]} |y| \left| |y|^{1/n} - 1 \right| \to 0,$$

hence the convergence is uniform.

Theorem 4.6 Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of continuous functions $f_n : D \to \mathbb{R}$ that converges uniformly to a limit function $f : D \to \mathbb{R}$. Then f is also continuous.

Proof: Let $x \in D$. We use the ε-δ criterion of continuity. Given $\varepsilon > 0$, we must find $\delta > 0$ such that

$$|x - y| < \delta \implies |f(x) - f(y)| < \varepsilon \quad \forall y \in D.$$

Since f_n converges uniformly to f, there exists $n \in \mathbb{N}$ with

$$\sup_{y \in D} |f_n(y) - f(y)| \leq \varepsilon/3,$$

so $|f_n(y) - f(y)| \leq \varepsilon/3$ for all $y \in D$.

Since f_n is continuous, there exists $\delta > 0$ with

$$|x - y| < \delta \implies |f_n(x) - f_n(y)| < \varepsilon/3.$$

Then for $|x - y| < \delta$,

$$|f(x) - f(y)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| < \varepsilon.$$

$\square$

Question 4.2 *If every f_n in a sequence of functions is bounded and a uniform limit function f exists, is f also bounded?*

Since uniform convergence is so useful, we introduce a notation to write it compactly.

Definition 4.7 For $f : D \to \mathbb{R}$, define the *supremum norm* as

$$\|f\|_\infty := \sup_{y \in D} |f(y)|.$$

$\square$

Then the uniform convergence of f_n to f can be written concisely as

$$\lim_{n \to \infty} \|f_n - f\|_\infty = 0.$$

For sequences defined by series, as in Example 4.2(iv), the following criterion ensures uniform convergence.

Theorem 4.8 [Weierstrass M-test] Let $g_k : D \to \mathbb{R}$, $k \in \mathbb{N}$, be functions such that

$$\sum_{k=0}^{\infty} \|g_k\|_\infty$$

converges. Then the sequence

$$f_n(x) := \sum_{k=0}^{n} g_k(x)$$

converges absolutely[1] and uniformly to a limit function $f : D \to \mathbb{R}$.

[1] A series of functions converges absolutely if $\sum_{k=0}^{\infty} |g_k(x)|$ converges pointwise.

Proof: For each $x \in D$, $|g_k(x)| \leq \|g_k\|_\infty$, so the series $\sum_{k=0}^\infty g_k(x)$ converges absolutely. Define the limit function

$$f(x) := \sum_{k=0}^{\infty} g_k(x) = \lim_{n\to\infty} f_n(x).$$

To show uniform convergence, let $\varepsilon > 0$. There exists $N(\varepsilon)$ with

$$\sum_{k=n+1}^{\infty} \|g_k\|_\infty \leq \varepsilon/2 \quad \forall n \geq N(\varepsilon),$$

hence for all $y \in D$,

$$|f_n(y) - f(y)| = \left| \sum_{k=n+1}^{\infty} g_k(y) \right| \leq \sum_{k=n+1}^{n} \|g_k\|_\infty \leq \varepsilon/2,$$

so $\|f_n - f\|_\infty \leq \varepsilon/2 < \varepsilon$. ∎

Finally, consider differentiability. Uniform convergence of f_n does not in general imply differentiability of the limit function f, as Example 4.2(ii) shows.

Indeed, for

$$f_n(x) = |x|^{1+1/n} = \exp((1+1/n)\ln|x|),$$

we compute for $x > 0$,

$$f_n'(x) = (1+1/n)x^{1/n} = (1+1/n)|x|^{1/n},$$

and similarly for $x < 0$:

$$f_n'(x) = (1+1/n)|x|^{1/n}.$$

For $x = 0$, the left and right derivatives are zero. Hence f_n is differentiable for all $n \in \mathbb{N}$. The limit function $f(x) = |x|$ is not differentiable at $x = 0$, even though f_n converges uniformly to f on $D = [-d, d]$.

Question 4.3 *Consider the functions $s_n(x) = \frac{\ln(1+e^{nx})}{n}$. Are they differentiable? To which function do they converge pointwise?*

Uniform convergence of the functions alone is not sufficient to guarantee differentiability of the limit function. The following theorem provides an additional condition under which differentiability holds.

Theorem 4.9 Let $f_n : D \to \mathbb{R}$ be a sequence of continuously differentiable functions such that f_n and f_n' converge uniformly to functions $f : D \to \mathbb{R}$ and $f^* : D \to \mathbb{R}$, respectively. Then f is continuously differentiable and

$$f' = f^*.$$

Proof: Let $x \in D$ and $h \in \mathbb{R}$ with $h \neq 0$ such that $x + h \in D$. Choose $n(|h|)$ such that

$$\|f_n - f\|_\infty < h^2 \quad \text{and} \quad \|f_n' - f^*\|_\infty < |h|$$

for $n = n(|h|)$. Then

$$\frac{f(x+h) - f(x)}{h} = \frac{f_n(x+h) - f_n(x)}{h} + R_1(h)$$

with

$$R_1(h) = \frac{f(x+h) - f_n(x+h)}{h} + \frac{f_n(x) - f(x)}{h}.$$

We have

$$|R_1(h)| \leq \frac{\|f_n - f\|_\infty}{|h|} + \frac{\|f_n - f\|_\infty}{|h|} < 2|h|,$$

so $R_1(h) \to 0$ as $h \to 0$.

By the Mean Value Theorem,

$$\frac{f(x+h) - f(x)}{h} = f_n'(\xi) + R_1(h) = f^*(\xi) + R_1(h) + R_2(h),$$

for some $\xi \in [x - |h|, x + |h|] \subset D$, where

$$|R_2(h)| = |f_n'(\xi) - f^*(\xi)| < |h|.$$

Hence $R_2(h) \to 0$ as $h \to 0$.

Since f_n' converges uniformly, f^* is continuous by Theorem 4.6, and $\xi \to x$ as $h \to 0$, we obtain

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \big(f^*(\xi) + R_1(h) + R_2(h) \big) = f^*(x).$$

This proves the claim. $\square$

4.3 Power Series

We now consider power series, as in Example 4.2(v). These have already appeared in [7, Section 9.E]. Because of the special structure of the summands, we can give a particularly simple criterion for uniform convergence. The following theorem strengthens [7, Theorem 9.23] in that it makes a statement about uniform convergence and also considers the derivative of the power series. For simplicity of exposition, we limit ourselves to real power series here.

Theorem 4.10 Let

$$f_n(x) = \sum_{k=0}^{n} c_k (x - x_0)^k$$

be a power series with $c_k, x_0 \in \mathbb{R}$, which converges at some $x_1 \neq x_0$.

Then the series converges absolutely and uniformly on any interval $D = [x_0 - \rho, x_0 + \rho]$ with $0 < \rho < |x_1 - x_0|$. In particular, the radius of convergence r satisfies $r \geq |x_1 - x_0|$. Moreover, the series

$$\sum_{k=1}^{n} kc_k(x - x_0)^{k-1}$$

also converges absolutely and uniformly on D.

Proof: The proof of [7, Theorem 9.23] shows that there are $M > 0$ and $\theta < 1$ such that $g_k(x) := c_k(x - x_0)^k$ satisfies for all $x \in D$ the inequality

$$|g_k(x)| = |c_k(x - x_0)^k| = |c_k(x_1 - x_0)^k| \cdot \left|\frac{x - x_0}{x_1 - x_0}\right|^k \leq M\theta^k$$

Hence $\|g_k\|_\infty \leq M\theta^k$. Since $\sum M\theta^k$ converges, the series $\sum \|g_k\|_\infty$ converges, and Theorem 4.8 gives absolute and uniform convergence of f_n.

For the derivative series $\sum_{k=1}^{n} kc_k(x - x_0)^{k-1}$, set $g_k(x) = kc_k(x - x_0)^{k-1}$. Analogously to the proof of [7, Theorem 9.23] we obtain $\|g_k\|_\infty \leq kM\theta^{k-1}$, and since

$$\frac{(k+1)M\theta^k}{kM\theta^{k-1}} = \frac{k+1}{k}\theta \to \theta < 1,$$

the series $\sum kM\theta^{k-1}$ converges by the ratio test, giving absolute and uniform convergence of the derivative series as well. □

Question 4.4 *Consider a power series where $c_k \neq 0$ only for finitely many k. How large is the convergence radius? And how do we call its limit function?*

Corollary 4.11 Under the assumptions of Theorem 4.10, define

$$f(x) := \sum_{k=0}^{\infty} c_k(x - x_0)^k.$$

Then f is continuously differentiable (and hence continuous), with

$$f'(x) = \sum_{k=1}^{\infty} kc_k(x - x_0)^{k-1}.$$

Proof: By Theorem 4.10, the functions

$$f_n(x) := \sum_{k=0}^{n} c_k(x - x_0)^k \quad \text{and} \quad \tilde{f}_n(x) := \sum_{k=1}^{n} kc_k(x - x_0)^{k-1}$$

converge uniformly to f and some limit function f^*, respectively. As polynomials, f_n and $\tilde{f}_n$ are continuously differentiable, and by [7, Example 16.9(a)] and the chain rule, $f_n' = \tilde{f}_n$. The result follows from Theorem 4.9. □

In other words, this corollary allows us to differentiate power series term by term to compute their derivatives.

Example 4.12 For the cosine function,

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

This is the power series

$$\cos x = \sum_{k=0}^{\infty} c_k x^k$$

with $c_k = (-1)^{k/2}/k!$ for even k and $c_k = 0$ for odd k.

Since this series converges for all $x \in \mathbb{R}$, Corollary 4.11 applies to any interval $D = [-\rho, \rho]$ with $\rho > 0$, hence on all of $\mathbb{R}$. The derivative of $\cos x$ is

$$\cos' x = \sum_{k=1}^{\infty} k c_k x^{k-1} = 0 + 2c_2 x + 0 + 4c_4 x^3 + 0 + 6c_6 x^5 + \cdots = -\sin x,$$

as expected. □

4.4 The Riemann Integral for Sequences of Functions

We have seen that continuity and differentiability of a sequence of functions f_n carry over to the limit function if the convergence of f_n—and in the case of differentiability also of f_n'—is uniform. We now examine the situation for the Riemann integral. The following theorem holds.

Theorem 4.13 Let $f_n : [a, b] \to \mathbb{R}$ be a uniformly convergent sequence of continuous functions with limit function f. Then

$$\lim_{n\to\infty} \int_a^b f_n(x)\, dx = \int_a^b f(x)\, dx = \int_a^b \lim_{n\to\infty} f_n(x)\, dx.$$

For uniformly convergent continuous functions, integration and taking the limit may thus be interchanged.

Proof: By Theorem 4.6, f is continuous and hence integrable. Moreover, for all $x \in [a, b]$, we have $|f(x) - f_n(x)| \le \|f - f_n\|_\infty$. Hence

$$\left| \int_a^b f(x)\, dx - \int_a^b f_n(x)\, dx \right| \le \int_a^b |f(x) - f_n(x)|\, dx \le \int_a^b \|f - f_n\|_\infty\, dx = (b-a)\|f - f_n\|_\infty,$$

and since $\lim_{n\to\infty} \|f - f_n\|_\infty = 0$, the claim follows. □

Example 4.14 The exponential series $f_n(x) = \sum_{k=0}^{n} x^k/k!$ converges uniformly on every bounded interval $[a, b]$ to e^x (this follows from Theorem 4.10 and the fact that the exponential series converges for all $x \in \mathbb{R}$). Therefore,

$$\begin{aligned}\int_a^b e^x\,dx &= \int_a^b \sum_{k=0}^{\infty} \frac{x^k}{k!}\,dx = \int_a^b \lim_{n\to\infty} \sum_{k=0}^{n} \frac{x^k}{k!}\,dx \\ &= \lim_{n\to\infty} \int_a^b \sum_{k=0}^{n} \frac{x^k}{k!}\,dx = \lim_{n\to\infty} \sum_{k=0}^{n} \int_a^b \frac{x^k}{k!}\,dx = \sum_{k=0}^{\infty} \int_a^b \frac{x^k}{k!}\,dx,\end{aligned}$$

where in the third step we applied Theorem 4.13 and in the fourth step linearity of the integral (Theorem 3.3).

Hence we can integrate the exponential function by integrating each term of the infinite series individually and then summing the results. Similarly, any power series may be integrated term by term on the interval given in Theorem 4.10. □

There are many sequences of functions that converge only pointwise but not uniformly, for which the statement of Theorem 4.13 still holds.

An example is the sequence from Example 4.2(i). Another example is the sequence $f_n : [0,2] \to \mathbb{R}$, $n \geq 1$, given by

$$f_n(x) = \begin{cases} nx, & x \in [0, 1/n] \\ 2 - nx, & x \in (1/n, 2/n] \\ 0, & x > 2/n \end{cases}.$$

This sequence converges pointwise to the zero function $f \equiv 0$, since for $x = 0$ we always have $f_n(x) = 0$, and for $x > 0$, $f(x) = 0$ for all $n > 2/x$. It does not converge uniformly, however, because

$$\|f_n - f\|_\infty \geq |f_n(1/n) - f(1/n)| = |1 - 0| = 1 \not\to 0 \quad \text{as } n \to \infty.$$

The graph of each f_n forms a triangle with the x-axis and vertices at $(0,0)$, $(1/n, 1)$, and $(2/n, 0)$. The area under this function, and hence the integral, is $1/n$. Therefore,

$$\int_0^2 f_n(x)\,dx \to 0 = \int_0^2 f(x)\,dx \quad \text{as } n \to \infty.$$

Thus, the statement of Theorem 4.13 holds even though only pointwise convergence occurs.

Question 4.5 *What happens if we scale all functions f_n by n?*

This raises the question of whether the assumptions of Theorem 4.13 can be weakened. In fact, they can, if the integral is defined in a different (but even more complicated) way. This leads to the so-called Lebesgue integral, which is, however, not a topic of this lecture.

4.5 Appendix: A Taylor Series with Convergence Radius 0

Using the techniques introduced in this chapter, we can construct a function whose Taylor series has radius of convergence 0 (thanks to Prof. Kriecherbauer for this example).

We use a function $\chi : \mathbb{R} \to \mathbb{R}$ with the following properties:

- $0 \leq \chi(x) \leq 1$ for all $x \in \mathbb{R}$
- $\chi(x) = 1$ for $x \in [-1/2, 1/2]$
- $\chi(x) = 0$ for $x \notin [-1, 1]$
- χ is infinitely differentiable on $\mathbb{R}$

Such a function can be explicitly constructed by composing constant and exponential functions, proving that it exists. Since its precise form is irrelevant for our purposes, no explicit construction is needed.

Because χ is constant zero outside $[-1, 1]$, all higher derivatives $\chi^{(m)}$ also vanish outside this interval, so

$$\|\chi^{(m)}\|_\infty = \sup\{|\chi^{(m)}(x)| : x \in [-1, 1]\} < \infty.$$

Define

$$C_m := \max\{\|\chi^{(n)}\|_\infty : n = 0, \ldots, m\}.$$

Now, for $k \in \mathbb{N}$, define

$$g_k(x) := (kx)^k \chi(k^2 x).$$

Using induction, the product and chain rules, and the binomial theorem, the m-th derivatives are

$$g_k^{(m)}(x) = \sum_{j=0}^{m} \binom{m}{j} k^k \frac{k!}{(k-j)!} x^{k-j} \chi^{(m-j)}(k^2 x) k^{2(m-j)}.$$

For $x \in [-1/k^2, 1/k^2]$, $\chi^{(m-j)}(k^2 x)$ may be nonzero. Using C_m, we obtain

$$\sup_{x \in \mathbb{R}} |x^{k-j} \chi^{(m-j)}(k^2 x)| \leq C_m k^{2(j-k)}.$$

Hence

$$\|g_k^{(m)}\|_\infty \leq \sum_{j=0}^{m} \binom{m}{j} k^k \frac{k!}{(k-j)!} C_m k^{2(m-k)} \leq k^{k+m} k^{2(m-k)} \sum_{j=0}^{m} \binom{m}{j} = C k^{3m-k} 2^m.$$

By the ratio test, $\sum_{k=0}^{\infty} C k^{3m-k} 2^m$ converges, so the Weierstrass M-test (Theorem 4.8) implies that

$$f_n^{(m)}(x) := \sum_{k=0}^{n} g_k^{(m)}(x)$$

converges uniformly to a limit $f^{(m)}$, and Theorem 4.9 gives $(f^{(m)})' = f^{(m+1)}$. Hence $f = f^{(0)}$ is infinitely differentiable on $\mathbb{R}$, and we may form the Taylor series at $x_0 = 0$.

Since χ is constant 1 on $[-1/2, 1/2]$, we have $\chi(0) = 1$ and $\chi^{(m)}(0) = 0$ for $m \geq 1$. Evaluating $g_k^{(m)}$ at 0 gives $g_k^{(m)}(0) = 0$ for $m \neq k$ and $g_k^{(k)}(0) = k! k^k$. Therefore,

$$f^{(m)}(0) = \sum_{k=0}^{\infty} g_k^{(m)}(0) = m! m^m.$$

Thus the Taylor series of f at 0 is

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

For any $x \neq 0$, $|k^k x^k| = |kx|^k \geq 2^k$ for $k \geq |2/x|$, which does not converge to zero. Hence the series diverges for all $x \neq 0$, and the radius of convergence is 0.

Chapter 5

Functions in $\mathbb{R}^n$ and Topological Foundations

Last update: July 21, 2026

In this and the following chapters we want to generalize the concept of functions, which we have previously considered in $\mathbb{R}$, to $\mathbb{R}^n$.

To define higher-dimensional sets in general, we use the Cartesian product.

Definition 5.1 For two sets A_1, A_2 we define the *Cartesian product* as the set of all pairs

$$A_1 \times A_2 := \{(a_1, a_2) \mid a_1 \in A_1,\, a_2 \in A_2\}.$$

For $n \geq 3$ sets $A_1, \ldots, A_n$ we analogously define the *Cartesian product* as the set of all n-tuples

$$A_1 \times A_2 \times \ldots \times A_n := \{(a_1, a_2, \ldots, a_n) \mid a_i \in A_i\}.$$

□

With this definition we define

$$\mathbb{R}^n := \underbrace{\mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R}}_{n \text{ times}}.$$

We often write the elements of $\mathbb{R}^n$ not as tuples but as column vectors

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \tag{5.1}$$

with real entries $x_1, x_2, \ldots, x_n \in \mathbb{R}$—that is, we identify the points in $\mathbb{R}^n$ with the corresponding vectors. For this identification to be unique, we must choose a basis. Unless explicitly stated otherwise, we always use the standard basis consisting of the unit vectors

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \ldots, \quad e_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \tag{5.2}$$

so that $x = x_1 e_1 + x_2 e_2 + \ldots + x_n e_n$ for x from (5.1).

We denote transposed vectors with a superscript "T", i.e.

$$(x_1, x_2, \ldots, x_n)^T = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}^T = (x_1, x_2, \ldots, x_n).$$

Just as a real function assigns to every real number x from a domain $D \subset \mathbb{R}$ a number $f(x) \in \mathbb{R}$, a "higher-dimensional" function assigns to every vector x from a given domain $D \subset \mathbb{R}^n$ a vector $f(x) \in \mathbb{R}^m$. Here n and m need not coincide, i.e. the vectors in the image set

$$f(D) := \{f(x) \,|\, x \in D\}$$

may have a different dimension than the vectors from the domain. Note that such an f can be interpreted as a function with one vector argument x or with n scalar arguments $x_1, \ldots, x_n$.

Since we can also write the real numbers as $\mathbb{R}^1$ (because every one-dimensional vector $x = (x_1)$ corresponds exactly to a real number x_1), our previous real-valued functions form a special case of this more general class of functions.

Formally, a vector-valued function $f : D \to \mathbb{R}^m$ is written as an m-dimensional vector

$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{pmatrix}$$

of functions $f_i : D \to \mathbb{R}$, $i = 1, \ldots, m$.

Example 5.2 (i) $n = 2$, $m = 1$. Such a function assigns a real number $f(x) \in \mathbb{R}$ to each vector $x \in D \subset \mathbb{R}^2$. An example with $D = \mathbb{R}^2$ is the function

$$f(x) = x_1^2 + 2x_2^2.$$

For every vector (for example $x = (x_1, x_2)^T = (2, 1)^T$) we obtain a real number by substituting the components into the function (in the numerical example $f((2, 1)^T) = 2^2 + 2 \cdot 1^2 = 4 + 2 = 6$).

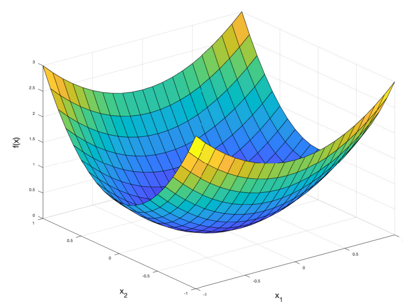

Figure 5.1: Graph of $f(x) = x_1^2 + 2x_2^2$

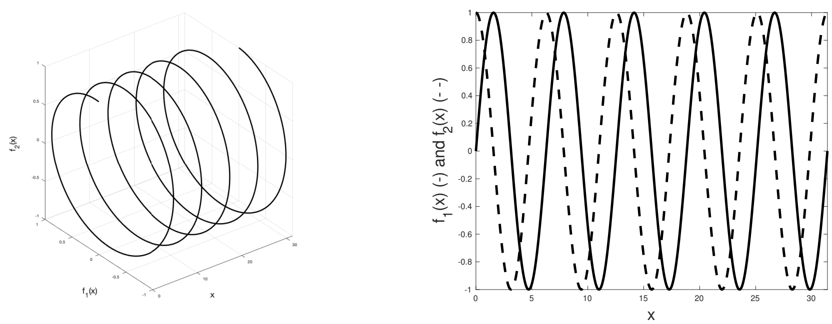

Figure 5.2: Graph of $f(x) = (\sin x, \cos x)^T$

The graph of a function with $n = 2$ and $m = 1$ can be represented graphically as a surface over the (x_1, x_2)-plane, see Figure 5.1.

(ii) $n = 1$, $m = 2$. In this case each $x \in D \subset \mathbb{R}$ is assigned a vector $f(x) \in \mathbb{R}^2$. An example with $D = \mathbb{R}$ is the function

$$f(x) = \begin{pmatrix} \sin x \\ \cos x \end{pmatrix},$$

whose component functions are therefore $f_1(x) = \sin x$ and $f_2(x) = \cos x$. Such a function can also be represented graphically in 3d, see Figure 5.2(left).

Question 5.1 *How does the graph change when you change f to $f(x) = \begin{pmatrix} \cos x \\ \sin x \end{pmatrix}$?*

The representation in Figure 5.2(right), in which each of the real component functions $f_i : \mathbb{R} \to \mathbb{R}$ is shown separately as a "normal" real function, can in general be used for $n = 1$ and arbitrary $m \geq 2$. However, in that case one obtains m graphs in a single figure, which may become rather confusing when m is large.

(iii) For $n \geq 2$ and $m \geq 2$ it is generally no longer possible to produce a meaningful graphical representation of the entire function. Nevertheless, one can still perform useful

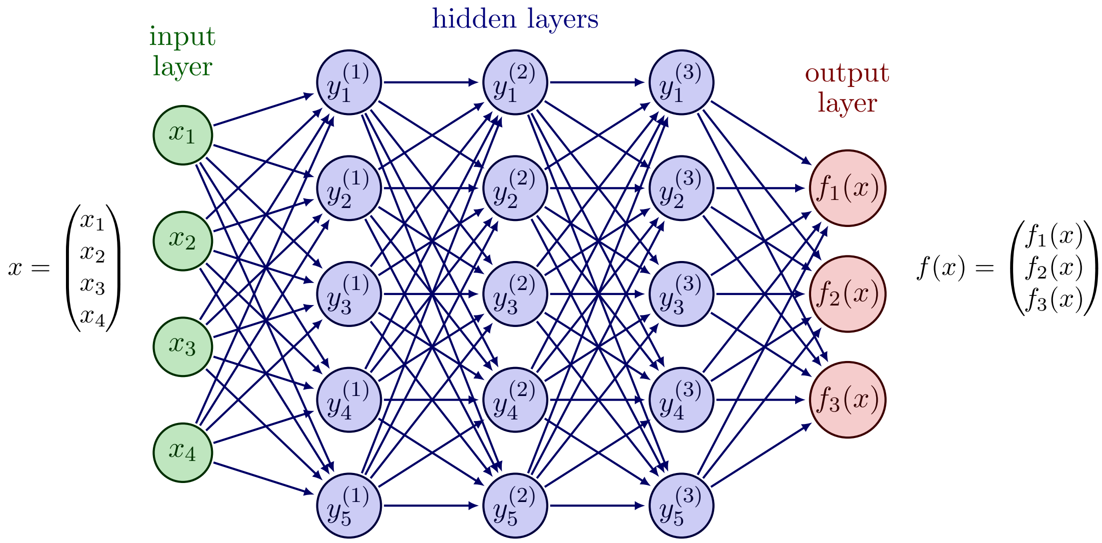

Figure 5.3: Schematic representation of a feedforward neural network

computations with such functions. This will be the main topic of the following chapters, although for illustration we will mostly refer to one of the two special cases from (i) and (ii).

(iv) There are many applications for functions in $\mathbb{R}^n$ even when only very elementary mathematics is used. For example, the surface area of a rectangular box with edge lengths x_1, x_2, and x_3 is given by

$$f_1(x) = 2(x_1x_2 + x_1x_3 + x_2x_3)$$

(where, as usual, we use the vector notation $x = (x_1, x_2, x_3)^T$). This is already a function from $\mathbb{R}^3$ to $\mathbb{R}$.

The volume of the same box is given by

$$f_2(x) = x_1x_2x_3,$$

again a function from $\mathbb{R}^3$ to $\mathbb{R}$. If we want to represent both quantities in a single function, we can define

$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \begin{pmatrix} 2(x_1x_2 + x_1x_3 + x_2x_3) \\ x_1x_2x_3 \end{pmatrix}.$$

This is then a function from $\mathbb{R}^3$ to $\mathbb{R}^2$ that assigns to each edge-length vector from $\mathbb{R}^3$ a vector in $\mathbb{R}^2$ whose entries represent the surface area and the volume of the rectangular box.

A more advanced application is a feedforward neural network. This is a function from $D \subset \mathbb{R}^n$ to $\mathbb{R}^m$, whose structure can be represented graphically as in Figure 5.3 (there with $n = 4$ and $m = 3$).

In this neural network every "neuron" (represented by the circles) stands for a scalar value in $\mathbb{R}$. These neurons are arranged in layers, from left to right. In order to define the

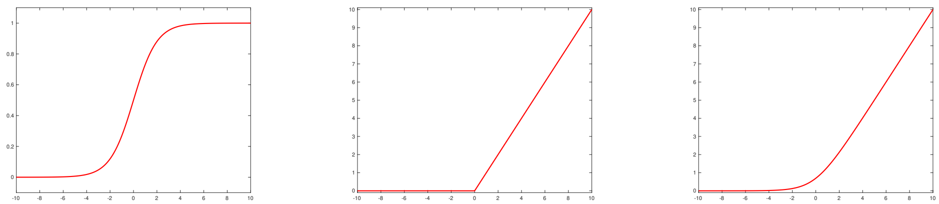

Figure 5.4: Graph of the sigmoid, ReLU, and softplus function (left to right)

function $f : D \to \mathbb{R}^m$ represented by this neural network, we need to specify how these scalar values are computed from the values in the layer on the left. This is done using the formulas

$$y_i^{(1)} = \sigma\left(w_{i,1}^{(0)} x_1 + w_{i,2}^{(0)} x_2 + \ldots + w_{i,n}^{(0)} x_n + b_i^{(0)}\right) = \sigma\left(\sum_{j=1}^{n} w_{i,j}^{(0)} x_j + b_i^{(0)}\right) \tag{5.3}$$

$$y_i^{(l+1)} = \sigma\left(w_{i,1}^{(l)} y_1^{(l)} + w_{i,2}^{(l)} y_2^{(l)} + \ldots + w_{i,n_l}^{(l)} y_{n_l}^{(l)} + b_i^{(l)}\right) = \sigma\left(\sum_{j=1}^{n_l} w_{i,j}^{(l)} y_j^{(l)} + b_i^{(l)}\right) \tag{5.4}$$

$$f_i(x) = w_{i,1}^{(L)} y_1^{(L)} + w_{i,2}^{(L)} y_2^{(L)} + \ldots + w_{i,n_L}^{(l)} y_{n_L}^{(L)} + b_i^{(L)} = \sum_{j=1}^{n_L} w_{i,j}^{(L)} y_j^{(L)} + b_i^{(L)} \tag{5.5}$$

In Figure 5.3 we have $n = 4$ (dimension of x), $n_1 = n_2 = n_3 = 5$ (dimensions of the hidden layers), $L = 3$ (number of hidden layers). The so-called *activation functions* $\sigma : \mathbb{R} \to \mathbb{R}$ are scalar functions, which may be chosen identical for each l (as in the formulas (5.3) and (5.4)) or different in each layer l. Popular choices of σ are

$$\text{sigmoid: } \sigma(y) = \frac{1}{1+e^{-y}}, \quad \text{ReLU: } \sigma(y) = \max\{0, y\}, \quad \text{or softplus: } \sigma(y) = \log(1+e^y),$$

whose graphs are depicted in Figure 5.4.

Question 5.2 *Are all of these activation functions differentiable?*

□

5.1 Metric and Norm

If we want to work with functions in $\mathbb{R}^n$ in the following, we must first define some fundamental objects in $\mathbb{R}^n$ itself. We need to determine how distances and lengths are measured in $\mathbb{R}^n$ (which we need, for example, to define convergence) or how the concept of open and closed intervals can be generalized to $\mathbb{R}^n$ (which is important, for example, for differentiability, since one would like to avoid the occurrence of one-sided derivatives whenever possible). The branch of mathematics that deals with these topics is called topology. In fact, topology can be studied not only in $\mathbb{R}^n$ but also in general sets or vector spaces, and in the following we will treat it in this general form. Nevertheless, we will always consider $\mathbb{R}$ and $\mathbb{R}^n$ separately as the most important special cases.

The first concept we define for this purpose is the distance between two points in an arbitrary set X. Mathematically, this is formalized by the concept of a metric.

Definition 5.3 Let X be a set. A *metric* on X is a map

$$d : X \times X \to \mathbb{R}, \qquad (x, y) \mapsto d(x, y),$$

with the following properties.

(i) $d(x, y) = 0$ holds if and only if $x = y$.

(ii) For all $x, y \in X$ we have $\quad d(x, y) = d(y, x) \quad$ ("symmetry").

(iii) For all $x, y, z \in X$ we have $\quad d(x, z) \leq d(x, y) + d(y, z) \quad$ ("triangle inequality").

The pair (X, d) is then called a *metric space*. □

The interpretation of the mapping d is that for any two points $x, y \in X$ it measures the distance $d(x, y)$ between x and y. Note that for a metric space (X, d) and every subset $A \subset X$, the pair (A, d_A) is again a metric space, where d_A denotes the restriction of d to $A \times A$.

From the triangle inequality and symmetry it follows for two arbitrary points x, y that

$$d(x, y) = \frac{1}{2}(d(x, y) + d(y, x)) \geq \frac{1}{2} d(x, x) = 0.$$

Thus the metric always yields a value ≥ 0, which is also sensible in view of the interpretation as a distance. Likewise, from the triangle inequality and symmetry one obtains the reverse triangle inequality

$$d(x, y) \geq |d(x, z) - d(y, z)|. \tag{5.6}$$

As we already know, a metric on the real numbers $X = \mathbb{R}$ is given for example by

$$d(x, y) = |x - y|.$$

We say "for example" because although this is the natural choice, it is by no means the only possibility. On any set X one can also define a metric by

$$d(x, y) = \begin{cases} 0, & x = y, \\ 1, & x \neq y. \end{cases} \tag{5.7}$$

Although this metric is rather useless for a meaningful notion of distance, it satisfies all the conditions of the above definition.

Question 5.3 *Check the triangle inequality for the metric from (5.7).*

Thus one can construct rather abstract metrics. In practice, however, one usually prefers metrics that have an intuitive geometric interpretation as a distance. Such metrics are typically derived from so-called norms, which are defined as follows.

Definition 5.4 Let V be a real vector space[1]. A *norm* on V is a map

$$\|\cdot\| : V \to \mathbb{R}, \qquad x \mapsto \|x\|,$$

with the following properties for all $x, y \in V$.

(i) $\|\lambda x\| = |\lambda|\,\|x\|$ for all $\lambda \in \mathbb{R}$.

(ii) $\|x\| = 0 \Leftrightarrow x = 0$.

(iii) $\|x + y\| \le \|x\| + \|y\|$ ("triangle inequality").

The pair $(V, \|\cdot\|)$ is then called a *normed vector space.* □

From the triangle inequality one can easily derive the reverse triangle inequality

$$\big|\,\|x\| - \|y\|\,\big| \le \|x - y\|. \tag{5.8}$$

The proof of the following theorem follows immediately from the properties of norms.

Theorem 5.5 Let $(V, \|\cdot\|)$ be a normed vector space. Then

$$d(x, y) := \|x - y\|$$

defines a metric on V. Thus every normed vector space is also a metric space.

Intuitively, one can think of a norm $\|x\|$ as a measure of the length of the vector x. The metric just defined therefore measures the distance between two points as the length of the vector $x - y$, which—after an appropriate translation—can be interpreted as the vector with "tip" at x and "base point" at y.

How can a norm in $\mathbb{R}^n$ look in practice? The following definition provides an entire family of such norms.

Definition 5.6 Let $x \in \mathbb{R}^n$.

(i) For every $p \in [1, \infty)$ we define

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

(ii) For $p = \infty$ we define

$$\|x\|_\infty := \max\{\,|x_i| : i = 1, \ldots, n\}.$$

□

[1] That is, we can add elements of V and multiply them by real numbers, and we always obtain elements of V again.

Question 5.4 *Consider* $x = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ *and compute* $\|x\|_p$ *for larger and larger* p*. What do you observe? Can you explain the notation for the norm* $\|x\|_\infty$ *with your observation?*

For the case $p = 2$, the norm in (i) is called the *Euclidean norm*. For a geometric interpretation of some of these norms see Example 5.13.

That Definition 5.6 actually defines norms satisfying all properties from Definition 5.4 is not obvious. Properties (i) and (ii) are easy to verify (which we will not do here in detail). For the triangle inequality (iii), however, we need to do some additional work. We first consider the simpler case $p = \infty$.

Theorem 5.7 For all $x, y \in \mathbb{R}^n$ we have

$$\|x + y\|_\infty \leq \|x\|_\infty + \|y\|_\infty.$$

Proof: Let $z = x + y$ and let x_{i^*}, y_{j^*} and z_{k^*} denote the entries of x, y and z with the largest absolute values. Then

$$\|x + y\|_\infty = \|z\|_\infty = |z_{k^*}| = |x_{k^*} + y_{k^*}| \leq |x_{k^*}| + |y_{k^*}| \leq |x_{i^*}| + |y_{j^*}| = \|x\|_\infty + \|y\|_\infty.$$

□

For $p < \infty$ we first need a lemma and a theorem based on it in order to prove the triangle inequality.

Lemma 5.8 Let $p, q \in (1, \infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$. Then for all $a, b \geq 0$ the inequality

$$a^{1/p} b^{1/q} \leq \frac{a}{p} + \frac{b}{q}$$

holds.

Proof: If $a = 0$ or $b = 0$, the statement is clear. For $a, b > 0$ we use that the logarithm is concave (this follows from $\ln''(x) = -1/x^2 < 0$, since this implies a negative curvature, cf. [7, End of Section 17.A]). Applying the definition of concavity with $\lambda = 1/p$ and $1 - \lambda = 1/q$ yields

$$\ln\left(\frac{1}{p}a + \frac{1}{q}b\right) \geq \frac{1}{p}\ln a + \frac{1}{q}\ln b.$$

The claim then follows by applying the exponential function to both sides. □

Theorem 5.9 (Hölder's inequality) Let $p, q \in (1, \infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$. Then for all $x, y \in \mathbb{R}^n$ we have

$$\sum_{i=1}^{n} |x_i y_i| \leq \|x\|_p \, \|y\|_q.$$

Proof: If $x = 0$ or $y = 0$, the statement is clear. Otherwise we set

$$v_i := \frac{|x_i|^p}{\|x\|_p^p}, \qquad w_i := \frac{|y_i|^q}{\|y\|_q^q},$$

for $i = 1, \ldots, n$. Then

$$\sum_{i=1}^n v_i = 1, \qquad \sum_{i=1}^n w_i = 1.$$

Applying Lemma 5.8 with $a = v_i$ and $b = w_i$ gives

$$\frac{|x_i y_i|}{\|x\|_p \|y\|_q} = v_i^{1/p} w_i^{1/q} \leq \frac{v_i}{p} + \frac{w_i}{q}$$

for $i = 1, \ldots, n$. Summing over i yields

$$\frac{1}{\|x\|_p \|y\|_q} \sum_{i=1}^n |x_i y_i| \leq \frac{1}{p} + \frac{1}{q} = 1,$$

which proves the claimed inequality. □

Remark 5.10 In the special case $p = q = 2$, we obtain for the

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

defined *Euclidean inner product* the inequality

$$|\langle x, y \rangle| \leq \|x\|_2 \, \|y\|_2,$$

which is also known as the *Cauchy–Schwarz inequality.* □

Question 5.5 *Consider the vector* $x = e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. *Find vectors* $y \in \mathbb{R}^2$ *such that* $\langle x, y \rangle = 0$. *Can you find other pairs of vectors* $x, y \in \mathbb{R}^2$ *(not equal to* 0*) with* $\langle x, y \rangle = 0$*? What do you observe when you visualize these vectors in a coordinate system?*

An application of the Cauchy–Schwarz inequality is the following lemma. If we define the integral of a function $f : [a, b] \to \mathbb{R}^n$ with integrable component functions $f_i : [a, b] \to \mathbb{R}$ componentwise by

$$\int_a^b f(x)\, dx := \begin{pmatrix} \int_a^b f_1(x)\, dx \\ \vdots \\ \int_a^b f_n(x)\, dx \end{pmatrix}, \tag{5.9}$$

then the question arises whether the triangle inequality for integrals also holds for norms. In general this is not the case, but for the Euclidean norm $\| \cdot \|_2$ it holds at least for continuous functions, as shown in the following lemma.

Lemma 5.11 Let $f : [a, b] \to \mathbb{R}^n$ be a continuous function. Then

$$\left\| \int_a^b f(x)\, dx \right\|_2 \leq \int_a^b \|f(x)\|_2\, dx.$$

Proof: Let $u := \int_a^b f(x)\, dx$ and $K := \|u\|_2$. If $K = 0$, the claim follows immediately. Otherwise, using the linearity of the integral and the Cauchy–Schwarz inequality we obtain

$$\begin{aligned} K^2 = \langle u, u \rangle &= \left\langle \int_a^b f(x)\, dx,\, u \right\rangle \\ &= \int_a^b \langle f(x), u \rangle\, dx \leq \int_a^b \|f(x)\|_2\, \|u\|_2\, dx = K \int_a^b \|f(x)\|_2\, dx. \end{aligned}$$

Dividing by K yields the desired result. □

Using Hölder's inequality we can now prove the triangle inequality for the $\|\cdot\|_p$ norms for $p \in [1, \infty)$, which is also known as the *Minkowski inequality*.

Theorem 5.12 (Minkowski inequality) For all $p \in [1, \infty)$ and all $x, y \in \mathbb{R}^n$ we have

$$\|x + y\|_p \leq \|x\|_p + \|y\|_p.$$

Proof: For $p = 1$ the statement follows immediately from the triangle inequality for the absolute value. For $p > 1$ the statement is clear if $x = 0$, $y = 0$, or $x + y = 0$. Otherwise let

$$q := \frac{p}{p-1},$$

so that $\frac{1}{p} + \frac{1}{q} = 1$. We define the vector $z = (z_1, \ldots, z_n)^T$ by

$$z_i := |x_i + y_i|^{p-1}, \qquad i = 1, \ldots, n.$$

Then

$$z_i^q = |x_i + y_i|^{q(p-1)} = |x_i + y_i|^p$$

and therefore

$$\|z\|_q = \|x + y\|_p^{p/q}.$$

Using the triangle inequality for the absolute value and Hölder's inequality we obtain

$$\sum_{i=1}^n |x_i + y_i||z_i| \leq \sum_{i=1}^n |x_i||z_i| + \sum_{i=1}^n |y_i||z_i| \leq (\|x\|_p + \|y\|_p)\|z\|_q.$$

From the definition of z it follows that

$$\|x + y\|_p^p \leq (\|x\|_p + \|y\|_p)\|x + y\|_p^{p/q}.$$

Since $p - p/q = 1$, the claim follows after dividing both sides by $\|x + y\|_p^{p/q}$. □

Thus we have shown that all norms from Definition 5.6 are indeed norms in the sense of Definition 5.4. In the following example we consider three special cases and the corresponding metrics.

Example 5.13 (i) For $p = 2$ we obtain the Euclidean norm

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

This is precisely the Euclidean length of the vector with tip x and base point y and therefore corresponds to the usual notion of distance in $\mathbb{R}^n$ known from school (there typically for $n = 2$ and $n = 3$). The basis of this formula is the theorem of Pythagoras, which is easily seen for $n = 2$: Drawing a right-angled triangle with hypotenuse x and shorter sides of length x_1 and x_2, Pythagoras' theorem implies $\|x\|_2^2 = x_1^2 + x_2^2$ and thus the above formula. In higher dimensions, the theorem must be applied iteratively.

The corresponding metric is given by

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2},$$

which is the length of straight line connecting the points x and y.

(ii) For $p = 1$ we obtain the norm

$$\|x\|_1 := \sum_{i=1}^{n} |x_i|$$

with the associated metric

$$d(x, y) = |x_1 - y_1| + |x_2 - y_2| + \ldots + |x_n - y_n|.$$

Here the distance between two points is the sum of the coordinate distances. This distance can also be useful in practice. For example, consider a city with a rectangular street grid (as is common in the USA and also in some German cities at least in certain districts). Then this norm corresponds exactly to the distance one must travel when driving from point x to point y along the streets.

(iii) For $p = \infty$ we obtain the norm

$$\|x\|_\infty := \max\{|x_i| : i = 1, \ldots, n\}$$

with the metric

$$d(x, y) = \max\{|x_i - y_i| : i = 1, \ldots, n\}.$$

□

Question 5.6 *Find two vectors $x, y \in \mathbb{R}^2$ for which $\|x+y\|_1 = \|x\|_1 + \|y\|_1$ but $\|x+y\|_\infty < \|x\|_\infty + \|y\|_\infty$ holds.*

The following theorem clarifies the relationships between these norms and thus also between the corresponding metrics.

Theorem 5.14 For the norms from Definition 5.6 the following inequalities hold for all $p \in (1, \infty)$:

(i) $\|x\|_1 \geq \|x\|_p \geq \|x\|_\infty$

(ii) $\|x\|_p \leq n^{\frac{1}{p}}\|x\|_\infty$

(iii) $\|x\|_1 \leq n^{\frac{p-1}{p}}\|x\|_p$

(iv) $\|x\|_1 \leq n\|x\|_\infty$

Proof:

(i) The first inequality follows by writing the vector x as the sum of the vectors

$$y^{(1)} := x_1 e_1, \quad y^{(2)} := x_2 e_2, \quad \ldots, \quad y^{(n)} := x_n e_n$$

with the standard basis vectors e_i from (5.2). From the definitions we obtain

$$x = \sum_{i=1}^n y^{(i)}, \qquad \|x\|_1 = \sum_{i=1}^n \|y^{(i)}\|_1,$$

and

$$\|y^{(i)}\|_1 = |x_i| = (|x_i|^p)^{1/p} = \|y^{(i)}\|_p.$$

The triangle inequality for the p-norm therefore gives

$$\|x\|_p \leq \|y^{(1)}\|_p + \ldots + \|y^{(n)}\|_p = \|y^{(1)}\|_1 + \ldots + \|y^{(n)}\|_1 = \|x\|_1.$$

For the second inequality let x_{i^*} denote the component of x with maximal absolute value. Then

$$\|x\|_\infty = |x_{i^*}| = (|x_{i^*}|^p)^{1/p} \leq \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} = \|x\|_p.$$

(ii) With x_{i^*} as above we obtain

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \leq \left(\sum_{i=1}^n |x_{i^*}|^p\right)^{1/p} = (n|x_{i^*}|^p)^{1/p} = n^{1/p}\|x\|_\infty.$$

(iii) Let $y = (1, 1, \ldots, 1)^T$. Then Hölder's inequality with $q = p/(p-1)$ gives

$$\|x\|_1 = \sum_{i=1}^n |x_i y_i| \leq \|x\|_p \|y\|_q = \|x\|_p \, n^{1/q}.$$

(iv) This follows by combining (ii) and (iii). $\square$

Question 5.7 *Find a vector $x \in \mathbb{R}^2$ for which equality holds in (ii) with $p = 2$, i.e., $\|x\|_2 = \sqrt{2}\|x\|_\infty$.*

We conclude this section with an example of a norm in a vector space that is not $\mathbb{R}^n$.

Example 5.15 Let $C([a,b],\mathbb{R})$ denote the set of continuous real-valued functions on an interval $[a,b]$. This set is a vector space over $\mathbb{R}$, since for any two functions $f, g \in C([a,b],\mathbb{R})$ and any $\lambda \in \mathbb{R}$ the functions $f+g$ and λf are again continuous and therefore belong to $C([a,b],\mathbb{R})$.

Recall Definition 4.7:

$$\|f\|_\infty := \max_{y\in[a,b]} |f(y)|$$

(we can write "max" here since f is continuous and $[a,b]$ is compact, cf. [7, Theorem 14.14]).

We claim that this defines a norm. Properties (i) and (ii) of a norm are immediate. The triangle inequality is proved exactly as for the $\|\cdot\|_\infty$ norm for vectors (see the proof of Theorem 5.7), where instead of indices such as i^* one uses the points at which the respective functions attain their maxima. $\square$

5.2 Open and Closed Sets

After clarifying how lengths and distances can be measured in $\mathbb{R}^n$ (and also in more general spaces), we now turn to the question of how the concept of open and closed intervals can be transferred to $\mathbb{R}^n$. For this purpose we first need the auxiliary notions of a *ball* and a *neighborhood*, which can be defined for general metric spaces.

Definition 5.16 Let (X,d) be a metric space and let $x \in X$.

(i) For $\varepsilon > 0$ we call the set

$$B_\varepsilon(x) := \{\, y \in X \mid d(y,x) < \varepsilon \,\}$$

the *(open) ball around x with radius ε* (or simply the *ε-ball around x*).

(ii) A subset $U \subset X$ is called a *neighborhood* of x if there exists $\varepsilon > 0$ such that

$$B_\varepsilon(x) \subset U.$$

$\square$

The notion of a "ball" becomes intuitive if we take $X = \mathbb{R}^3$ and choose d as the Euclidean metric. Then $B_\varepsilon(x)$ is indeed a ball with center x and radius $\varepsilon > 0$ in the everyday sense (where by definition the boundary of the ball is not included).

In $\mathbb{R}^2$ with the Euclidean metric, $B_\varepsilon(x)$ is a disk (without boundary), and in $\mathbb{R}$ with the metric defined by the absolute value, $B_\varepsilon(x)$ is nothing other than the open interval

$$(x-\varepsilon,\, x+\varepsilon).$$

The exact geometric shape of a ball depends on the metric used. For example, it is interesting to draw the balls in $\mathbb{R}^2$ (for which visualization on a board or a sheet of paper is easier than in $\mathbb{R}^3$) for the metrics corresponding to the norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$.

Question 5.8 *Draw the balls with radius* $\varepsilon = 1$ *around* $x = 0$ *in* $\mathbb{R}^2$ *for the metrics generated by the norms* $\|\cdot\|_1$ *and* $\|\cdot\|_\infty$.

The following theorem shows a property of neighborhoods of two points $x, y \in X$.

Theorem 5.17 (Hausdorff separation axiom) Let (X, d) be a metric space and let $x, y \in X$ with $x \neq y$. Then there exist neighborhoods U and V of x and y such that

$$U \cap V = \emptyset.$$

Proof: From the definition of a neighborhood it follows that every ball $B_\varepsilon(x)$ with arbitrary $\varepsilon > 0$ is a neighborhood of x. Now choose $\varepsilon = d(x, y)/2$ and set

$$U := B_\varepsilon(x), \qquad V := B_\varepsilon(y).$$

Then these are neighborhoods of x and y, respectively. We prove that their intersection is empty by contradiction.

Assume there exists a point $z \in U \cap V$. Then by definition

$$d(x, z) < \varepsilon, \qquad d(y, z) < \varepsilon.$$

Using the triangle inequality and the symmetry of the metric we obtain

$$d(x, y) \leq d(x, z) + d(z, y) < \varepsilon + \varepsilon = d(x, y),$$

that is, $d(x, y) < d(x, y)$, which is clearly impossible and therefore yields the desired contradiction. $\square$

The following definition generalizes the notion of an open interval.

Definition 5.18 A subset $A \subset X$ of a metric space (X, d) is called *open* if for every $x \in A$ there exists $\varepsilon > 0$ such that

$$B_\varepsilon(x) \subset A.$$

$\square$

In other words, a set A is open if and only if it is a neighborhood of each of its points.

Example 5.19 (i) In $\mathbb{R}$ with the metric $d(x, y) = |x - y|$, every open interval of the form (a, b), $(-\infty, b)$ or (a, ∞) with $a, b \in \mathbb{R}$ is an open set in the sense of Definition 5.18.

Proof: Let $x \in (a, b)$ be arbitrary. Then $x > a$ and $x < b$, and therefore

$$\varepsilon := \min\{x - a,\ b - x\} > 0$$

(if one of the interval boundaries is infinite, the corresponding term is omitted in the minimum). By construction of ε, every

$$y \in B_\varepsilon(x) = (x - \varepsilon, x + \varepsilon)$$

is greater than a and smaller than b, and therefore contained in (a,b). Hence $B_\varepsilon(x) \subset (a,b)$.

(ii) In $\mathbb{R}$ with the metric $d(x,y) = |x-y|$, unions and intersections of finitely many open intervals are again open sets. This follows because each such set consists of finitely many disjoint open intervals to which the argument from (i) can be applied.

(iii) Intervals of the form $(a,b]$, $[a,b)$ or $[a,b]$ (with finite interval endpoints) are not open. For example, in the case of $(a,b]$ and $x=b$, every interval of the form $(x-\varepsilon, x+\varepsilon)$ extends beyond the right endpoint of $(a,b]$ and therefore is not contained in $(a,b]$.

(iv) In a metric space, for every $r>0$ and every $y \in X$, the open ball $B_r(y)$ is an open set.

Proof: Let $x \in B_r(y)$. Then by definition of the ball we have $d(x,y) < r$. Define

$$\varepsilon := r - d(x,y) > 0.$$

For every $z \in B_\varepsilon(x)$ we have $d(z,x) < r - d(x,y)$, and by the triangle inequality

$$d(z,y) \leq d(z,x) + d(x,y) < r - d(x,y) + d(x,y) = r.$$

Hence every $z \in B_\varepsilon(x)$ lies in $B_r(y)$ and thus $B_\varepsilon(x) \subset B_r(y)$.

(v) Using a proof analogous to (iv), but employing the reverse triangle inequality (5.6), one can show that for every $r > 0$ and $y \in X$ the set

$$\{\, x \in X \mid d(x,y) > r \,\}$$

is also open. □

Remark 5.20 For $X = \mathbb{R}^n$ the following holds: Although the shape of the ball $B_\varepsilon(x)$ depends on the chosen metric or the underlying norm (for example it looks different for $\|\cdot\|_1$ than for $\|\cdot\|_\infty$), the definition of open sets leads to the same collections of sets for every norm $\|\cdot\|_p$ from Definition 5.6.

To **prove** this, we denote the balls corresponding to the p-norms by $B_{p,\varepsilon}(x)$. By Theorem 5.14, for any two norms $\|\cdot\|_p$ and $\|\cdot\|_q$ from Definition 5.6 and every ball $B_{p,\varepsilon}(x)$ one can find an $\varepsilon' > 0$ such that

$$B_{q,\varepsilon'}(x) \subset B_{p,\varepsilon}(x)$$

(for example, for $p = 1$ and $q = \infty$ this works with $\varepsilon' = \varepsilon/n$). Thus, if for a set A and every $x \in A$ there exists $\varepsilon > 0$ such that $B_{p,\varepsilon}(x) \subset A$, then there also exists $\varepsilon' > 0$ with $B_{q,\varepsilon'}(x) \subset A$.

Consequently, in $\mathbb{R}^n$ we can speak of open sets without having to specify each time which norm from Definition 5.6 we are using. Therefore, in the case of $\mathbb{R}^n$ we will simply denote the norm by $\|\cdot\|$ in the following and specify it only when necessary (for example for a concrete calculation).

In metric spaces other than $\mathbb{R}^n$, or for metrics that are not defined via norms, this is not necessarily the case (for example, consider what the balls and open sets in $\mathbb{R}$ look like for the "strange" metric (5.7)). □

Example 5.19(iii) shows that for intervals of the form $(a,b]$ in $\mathbb{R}$ the boundary point $b \in (a,b]$ prevents the interval from being open. This phenomenon actually holds in general once we formally define what a boundary point of a set is.

Definition 5.21 Let (X, d) be a metric space and $A \subset X$. A point $x \in X$ is called a *boundary point* of A if every neighborhood U of x contains both points from A and points from $X \setminus A$, i.e., if for every neighborhood U of x both

$$U \cap A \neq \emptyset \quad \text{and} \quad U \cap (X \setminus A) \neq \emptyset$$

hold.

The set of all boundary points of A is called the *boundary* of A and is denoted by ∂A. □

Question 5.9 *What are the boundary points of* $[0, 1] \cup (2, \infty)$*?*

Remark 5.22 (i) Note that according to this definition the boundary points do not necessarily have to lie in A.

(ii) From the definition it follows that the boundary points of the set A coincide with the boundary points of the set $X \setminus A$. □

Theorem 5.23 Let (X, d) be a metric space. A set $A \subset X$ is open if and only if it contains no boundary point of A.

Proof:

"⇒": Let A be open and let $x \in A$ be arbitrary. Then there exists a ball $B_\varepsilon(x) \subset A$. Hence $U = B_\varepsilon(x)$ is a neighborhood that contains no points $y \in X \setminus A$. Therefore x is not a boundary point.

"⇐": Let A be a set that contains no boundary points of itself. Let $x \in A$ be arbitrary. Since A contains no boundary points, x is not a boundary point. Hence there exists a neighborhood U of x such that either $U \subset A$ or $U \subset X \setminus A$. Because $x \in U$ and $x \in A$, we have $U \cap A \neq \emptyset$, so the case $U \subset X \setminus A$ is impossible. Since U by definition contains a ball $B_\varepsilon(x)$ for some $\varepsilon > 0$, we obtain

$$B_\varepsilon(x) \subset U \subset A.$$

Hence A is open. □

The following theorem lists several properties of open sets.

Theorem 5.24 For a metric space (X, d) the following holds:

(i) $\emptyset$ and X are open sets.

(ii) For finitely many open sets $A_1, \ldots, A_m$, the intersection

$$\bigcap_{i=1,\ldots,m} A_i$$

is also open.

(iii) For arbitrarily many[2] open sets A_α, $\alpha \in B$, the union

$$\bigcup_{\alpha \in B} A_\alpha$$

is open.

[2] even uncountably many

Proof:

(i) For the empty set there is nothing to prove because it contains no points, and the entire space X contains every ε-ball $B_\varepsilon(x)$ around every point $x \in X$.

(ii) Let $x \in \bigcap_{i=1,\dots,m} A_i$. Then x lies in each A_i, and since the sets are open there exist $\varepsilon_1, \dots, \varepsilon_m$ with $B_{\varepsilon_i}(x) \subset A_i$. For

$$\varepsilon := \min\{\varepsilon_i \mid i = 1, \dots, m\}$$

we obtain

$$B_\varepsilon(x) \subset B_{\varepsilon_i}(x) \subset A_i$$

for all $i = 1, \dots, m$, and therefore

$$B_\varepsilon(x) \subset \bigcap_{i=1,\dots,m} A_i.$$

Thus the intersection is open.

(iii) Let $x \in \bigcup_{\alpha \in B} A_\alpha$. Then x lies in one of the sets A_α, and since this set is open there exists ε with $B_\varepsilon(x) \subset A_\alpha$. Hence also

$$B_\varepsilon(x) \subset \bigcup_{\alpha \in B} A_\alpha,$$

so the union is open. □

Example 5.25 Statement (ii) does not hold in general for infinitely many open sets $A_1, A_2, A_3, \dots$. An example in $X = \mathbb{R}$ is given by the intervals $A_k = (-1/k, 1/k)$. Each of these intervals is open, but their intersection is

$$\bigcap_{k=1}^{\infty} A_k = \{0\},$$

which is not open since it contains no ball $B_\varepsilon(0)$. □

After obtaining a generalization of open intervals via open sets, we now define the corresponding generalization of closed sets.

Definition 5.26 A set A is called *closed* if the set $X \setminus A$ is open. □

Example 5.27 (i) In every metric space (X, d) the set

$$\overline{B}_\varepsilon(x) := \{\, y \in X \mid d(x, y) \le \varepsilon \,\}$$

is closed, since the complement

$$X \setminus \overline{B}_\varepsilon(x) = \{\, y \in X \mid d(x, y) > \varepsilon \,\}$$

is open by Example 5.19(v). The set $\overline{B}_\varepsilon(x)$ is called a *closed ball.*

(ii) In every metric space (X, d) the sets $\emptyset$ and X are closed, since their complements X and $\emptyset$ are open. □

From Remark 5.22(ii) and Theorem 5.23 the following theorem immediately follows.

Theorem 5.28 Let (X, d) be a metric space. A set $A \subset X$ is closed if and only if it contains every boundary point of A.

Proof: From Remark 5.22(ii) it follows that A contains every boundary point of A if and only if $X \setminus A$ contains no boundary point of $X \setminus A$. By Theorem 5.23 this is equivalent to $X \setminus A$ being open, i.e., to A being closed. ∎

Example 5.29 (i) Every closed interval of the form $[a, b]$, $[a, \infty)$, $(-\infty, b]$ or $(-\infty, \infty)$ is closed in $\mathbb{R}$ since the boundary points a and/or b are contained (note that $\pm\infty$ are not boundary points, since boundary points in $\mathbb{R}$ must lie in $\mathbb{R}$ by definition, whereas $\pm\infty$ are not real numbers).

(ii) The interval $(a, b] \subset \mathbb{R}$ with $a, b \in \mathbb{R}$ is neither open nor closed: since the boundary point a is not contained in $(a, b]$, the interval is not closed, and since the boundary point b lies in $(a, b]$, the interval is not open. □

From Theorems 5.23 and 5.28 the following theorem immediately follows.

Theorem 5.30 Let (X, d) be a metric space and $A \subset X$. Then

(i) the set $A \setminus \partial A$ is open,

(ii) the set $A \cup \partial A$ is closed,

(iii) the boundary ∂A is closed.

Proof:

(i) Let $x \in A \setminus \partial A$. Then there exists a neighborhood U of x with $U \subset A$, hence also a ball $B_\varepsilon(x) \subset U \subset A$. Let $y \in B_\varepsilon(x)$. Since $B_\varepsilon(x)$ is open, there exists a ball $B_{\varepsilon'}(y) \subset B_\varepsilon(x) \subset A$. Thus $y \notin \partial A$ and therefore $B_\varepsilon(x) \subset A \setminus \partial A$. Hence $A \setminus \partial A$ is open.

(ii) This follows from (i) since

$$X \setminus (A \cup \partial A) = (X \setminus A) \setminus \partial A = (X \setminus A) \setminus \partial(X \setminus A)$$

is open.

(iii) For any set $B \subset X$ the identity

$$B = (A \cup B) \setminus (A \setminus B)$$

holds. Applying this to ∂A gives

$$\partial A = (A \cup \partial A) \setminus (A \setminus \partial A).$$

Moreover, for arbitrary sets $C, D \subset X$ we have

$$X \setminus (C \setminus D) = (X \setminus C) \cup D.$$

Hence

$$X \setminus \partial A = (X \setminus (A \cup \partial A)) \cup (A \setminus \partial A).$$

By (i) and (ii) both sets in this union are open, so $X \setminus \partial A$ is open and therefore ∂A is closed. □

This theorem motivates the following definition.

Definition 5.31 For a set $A \subset X$ we define the *interior* of A by

$$\operatorname{int} A := A \setminus \partial A$$

and the *closure* of A by

$$\overline{A} := A \cup \partial A.$$

□

By Theorem 5.30, $\operatorname{int} A$ is always an open set and $\overline{A}$ is always a closed set.

Question 5.10 *What are* $\operatorname{int} A$ *and* $\overline{A}$ *for* $A = [0, 1] \cup (2, \infty)$ *(cf. Question 5.9).*

Finally, we note that open sets can also be defined without using a metric. For an arbitrary set X one chooses a collection T of subsets of X (so the elements of T are themselves sets) and declares each set $A \in T$ to be "open". The collection T must be chosen so that the properties from Theorem 5.24 hold. The pair (X, T) is then called a *topological space*.

Thus a metric space (X, d) is a special case of a topological space in which the sets in T are chosen according to Definition 5.18.

Since without a metric we do not have balls available, the concept of a neighborhood must then be defined as follows: A set $U \subset X$ is called a neighborhood of a point $x \in X$ if there exists an open set $V \in T$ with $x \in V$ and $V \subset U$. In the case of metric spaces this is equivalent to Definition 5.16(ii).

The Hausdorff separation axiom from Theorem 5.17 does not necessarily hold for general topological spaces. If it does hold, the space (X, T) is called a *Hausdorff space*. This explains why the separation property was called an "axiom": for general topological spaces it is not automatically satisfied but must be required separately.

For metric spaces, however, the Hausdorff separation axiom is always satisfied, as proven in Theorem 5.17. Hence metric spaces are always Hausdorff spaces.

Chapter 6

Limits and Continuity

Last update: July 21, 2026

In this chapter we want to generalize the concept of convergent sequences to general metric spaces and in particular to $\mathbb{R}^n$.

6.1 Limits

Definition 6.1 Let (X,d) be a metric space and $(x_k)_{k\in\mathbb{N}}$ a sequence of points in X. The sequence is called convergent to the limit $a \in X$ if the following holds:

For every $\varepsilon > 0$ there exists an $N(\varepsilon) \in \mathbb{N}$ such that $d(x_k, a) < \varepsilon$ for all $k \geq N(\varepsilon)$. (6.1)

In this case we write, as usual, $\lim_{k\to\infty} x_k = a$ or briefly $x_k \to a$ as $k \to \infty$. □

Remark 6.2 The condition (6.1) in the definition of convergence can be expressed in several equivalent ways:

(i) Since $d(x,y) \geq 0$ always holds, condition (6.1) simply requires that the (real) sequence $(d(x_k,a))_{k\in\mathbb{N}}$ is a null sequence. Thus it is equivalent to

$$\lim_{k\to\infty} d(x_k, a) = 0. \tag{6.2}$$

Since $d(x_k, a)$ is a real number for each k, the limit in (6.2) is the real limit defined in [7, Definition 8.3]. If the metric is generated by a norm via $d(x,y) = \|x-y\|$, this can equivalently be written as

$$\lim_{k\to\infty} \|x_k - a\| = 0. \tag{6.3}$$

In the special case $X = \mathbb{R}^n$ it does not matter which norm $\|\cdot\|_p$ from Definition 5.6 we use, because Theorem 5.14 implies that the convergence $\lim_{k\to\infty} \|x_k - a\|_p = 0$ holds for one of these norms if and only if it holds for all of them.

Question 6.1 *Suppose that for a sequence in $\mathbb{R}^3$ for all $\varepsilon > 0$ you know $N(\varepsilon)$ such that $\|x_k - a\|_\infty < \varepsilon$ holds for all $k \geq N(\varepsilon)$. How do you have to choose $\widetilde{N}(\varepsilon)$ such that $\|x_k - a\|_1 < \varepsilon$ holds for all $k \geq \widetilde{N}(\varepsilon)$?*

(ii) Using the equivalence

$$d(x_k, a) < \varepsilon \quad \Leftrightarrow \quad x_k \in B_\varepsilon(a),$$

equation (6.1) can also be written as

$$\text{For every } \varepsilon > 0 \text{ there exists an } N(\varepsilon) \in \mathbb{N} \text{ such that } x_k \in B_\varepsilon(a) \text{ for all } k \geq N(\varepsilon). \tag{6.4}$$

(iii) Using the fact that every ε-ball $B_\varepsilon(x)$ is a neighborhood of x and that every neighborhood U of x contains an ε-ball, it follows from (6.4) that (6.1) can equivalently be written as

$$\begin{array}{l}\text{For every neighborhood } U \text{ of } a \text{ there exists an } N(U) \in \mathbb{N} \\ \text{such that } x_k \in U \text{ for all } k \geq N(U).\end{array} \tag{6.5}$$

□

In the special case $X = \mathbb{R}^n$ the points x_k are vectors of the form

$$x_k = \begin{pmatrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{kn} \end{pmatrix}.$$

In this case convergence can easily be checked using the components of x_k, as the following theorem shows.

Theorem 6.3 Let $(x_k)_{k\in\mathbb{N}}$ be a sequence of points in $\mathbb{R}^n$. Then the sequence converges with respect to any p-norm from Definition 5.6 to the limit

$$a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix},$$

if and only if each of the component sequences x_{ki}, $i = 1, \ldots, n$ converges with

$$\lim_{k\to\infty} x_{ki} = a_i.$$

Proof: By Remark 6.2(i) we have

$$\lim_{k\to\infty} x_k = a \quad \Leftrightarrow \quad \lim_{k\to\infty} \|x_k - a\|_p = 0 \quad \Leftrightarrow \quad \lim_{k\to\infty} \|x_k - a\|_\infty = 0.$$

From the definition of the ∞-norm we immediately obtain the equivalence

$$\lim_{k\to\infty} \|x_k - a\|_\infty = 0 \quad \Leftrightarrow \quad \lim_{k\to\infty} |x_{ki} - a_i| = 0 \text{ for all } i = 1, \ldots, n,$$

which proves the claim. □

From [7, Corollary 8.15] we obtain the following property of closed intervals $[a, b]$: If $(x_k)_{k\in\mathbb{N}}$ is a convergent real sequence with $x_k \in [a, b]$ for all $k \in \mathbb{N}$, then $\lim_{k\to\infty} x_k \in [a, b]$.

The following theorem shows that this property can be generalized to arbitrary closed sets and is in fact equivalent to the definition of closedness.

Theorem 6.4 Let (X,d) be a metric space and $A \subset X$. Then A is closed if and only if for every convergent sequence $(x_k)_{k\in\mathbb{N}}$ with $x_k \in A$ for all $k \in \mathbb{N}$ we have $\lim_{k\to\infty} x_k \in A$.

Proof: "$\Rightarrow$": Let A be closed and let $a := \lim_{k\to\infty} x_k$. We prove by contradiction that $a \in A$. Assume $a \notin A$. Since $X \setminus A$ is open, $X \setminus A$ is a neighborhood of a. By the definition of convergence it follows that $x_k \in X \setminus A$ for all sufficiently large k. This contradicts $x_k \in A$ for all $k \in \mathbb{N}$.

"$\Leftarrow$": Assume that the limit of every convergent sequence $x_k \in A$ again lies in A. We prove that A is closed by showing that $X \setminus A$ is open. Let $x \in X \setminus A$ be an arbitrary point. We must prove that there exists an $\varepsilon > 0$ such that $B_\varepsilon(x) \subset X \setminus A$. Suppose this were not the case. Then for every $k \in \mathbb{N} \setminus \{0\}$ and $\varepsilon = 1/k$ there exists an $x_k \in B_\varepsilon(x) \cap A$. Thus $d(x_k, x) < \varepsilon = 1/k$, and therefore x_k is a convergent sequence in A with $x_k \to x \notin A$ as $k \to \infty$. This contradicts the fact that the limit of x_k must lie in A. $\square$

Just like the concept of convergent sequences, we can also generalize the notion of a Cauchy sequence.

Definition 6.5 Let (X,d) be a metric space and $(x_k)_{k\in\mathbb{N}}$ a sequence of points in X. Then $(x_k)_{k\in\mathbb{N}}$ is called a *Cauchy sequence* if for every $\varepsilon > 0$ there exists a $C(\varepsilon) \in \mathbb{N}$ such that

$$d(x_k, x_m) < \varepsilon$$

for all $k, m \geq C(\varepsilon)$. $\square$

Exactly as in $\mathbb{R}$, one proves that every convergent sequence is a Cauchy sequence. However, the converse does not necessarily hold in general metric spaces. Therefore we need the following definition.

Definition 6.6 A metric space (X,d) is called *complete* if every Cauchy sequence converges (to some $a \in X$).

If in addition $X = V$ is a vector space and the metric d is defined by a norm $\|\cdot\|$, then $(V, \|\cdot\|)$ is called a *complete normed vector space* or a *Banach space*. $\square$

Question 6.2 *Consider a metric space (X,d), where X is an arbitrary non-empty set and d is the metric with $d(x,y) = 1$ whenever $x \neq y$, cf. (5.7). Which sequences are convergent in this space? Which sequences are Cauchy sequences? Is the space complete?*

In [7, Section 8.F] we saw that the property that every Cauchy sequence converges is equivalent to the completeness axiom in $\mathbb{R}$. Definition 6.6 is therefore nothing but a generalization of the completeness axiom to metric spaces. In particular, it follows that the metric space $(\mathbb{R}, d)$ with $d(x,y) = |x-y|$ is complete. The following theorem shows that the same holds for $\mathbb{R}^n$.

Theorem 6.7 $(\mathbb{R}^n, \|\cdot\|_p)$ is a Banach space for every p-norm from Definition 5.6, and therefore also a complete metric space with the associated metric.

Proof: Completely analogously to the proof of Theorem 6.3, one proves that a sequence $(x_k)_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ is a Cauchy sequence if and only if each of its component sequences $(x_{ki})_{k\in\mathbb{N}}$ is a Cauchy sequence. Since every Cauchy sequence in $\mathbb{R}$ converges, each component sequence converges, and by Theorem 6.3 the sequence x_k itself also converges. □

Another example of a complete metric space is given by the following theorem.

Theorem 6.8 The normed space $(C([a,b],\mathbb{R}), \|\cdot\|_\infty)$ (cf. Example 5.15) is a Banach space, and therefore also a complete metric space with the associated metric.

Proof: Let f_k be a Cauchy sequence of functions in $C([a,b],\mathbb{R})$. Then for every $x \in [a,b]$ we have

$$|f_k(x) - f_m(x)| \le d(f_k, f_m) = \|f_k - f_m\|_\infty,$$

which shows that the real sequence $f_k(x)$ is also a Cauchy sequence. Since $\mathbb{R}$ is complete, $f_k(x)$ converges to a limit $f(x)$. We now prove that the f_k converge uniformly to the function f defined in this way.

For a given $\varepsilon > 0$ there exists a $C(\varepsilon) \in \mathbb{N}$ such that $\|f_k - f_m\|_\infty < \varepsilon$ for all $m, k \ge C(\varepsilon)$, which implies

$$|f_k(x) - f_m(x)| < \varepsilon$$

for all $x \in [a,b]$ and all $m, k \ge C(\varepsilon)$. Hence

$$|f_k(x) - f(x)| \le |f_k(x) - f_m(x)| + |f_m(x) - f(x)| < \varepsilon + |f_m(x) - f(x)|$$

for all $x \in [a,b]$.

For each fixed $x \in [a,b]$ we have $|f_m(x) - f(x)| \to 0$, hence there exists an $m(x) \in \mathbb{N}$ such that $|f_{m(x)}(x) - f(x)| < \varepsilon$, and therefore

$$|f_k(x) - f(x)| < 2\varepsilon$$

for all $k \ge C(\varepsilon)$. Since none of the expressions in this inequality, and also $C(\varepsilon)$, depend on $m(x)$, this inequality holds for all $x \in [a,b]$, and thus

$$\sup_{x\in[a,b]} |f_k(x) - f(x)| \le 2\varepsilon$$

for all $k \ge C(\varepsilon)$, which proves the uniform convergence of f_k to f.

Hence f is continuous by Theorem 4.6 and therefore $f \in C([a,b],\mathbb{R})$. The Cauchy sequence f_k therefore converges, and consequently the space $C([a,b],\mathbb{R})$ is complete. □

6.2 Continuity

Both the sequence criterion for continuity for real functions from [7, Definition 14.6] and the ε-δ criterion from [7, Theorem 14.16] can be directly transferred to general metric spaces using the previous results. Here we consider functions $f : X \to Y$ for metric spaces (X, d_X) and (Y, d_Y) (we add the index to the metrics to emphasize that they are not necessarily the same; we will explicitly use the metrics only in the ε-δ criterion). The functions from $\mathbb{R}$ to $\mathbb{R}$ (or from $D \subset \mathbb{R}$ to $\mathbb{R}$) considered in [7, Section 14.A] are special cases of these general functions. Note that in this general framework, we do not need to consider separate domains $D \subset X$, because any subset of a metric space is again a metric space, and therefore we can always take X as the domain D of interest.

Definition 6.9 Let (X, d_X) and (Y, d_Y) be metric spaces and $f : X \to Y$ a function. Then f is called *continuous* at a point $x \in X$ if for every sequence $x_k \to x$ we have

$$\lim_{k\to\infty} f(x_k) = f(x).$$

The function f is called *continuous on* X if it is continuous at every point $x \in X$. □

Example 6.10 From the reverse triangle inequality it follows in a normed vector space $(V, \|\cdot\|)$ that

$$|\,\|x_k\| - \|x\|\,| \le \|x_k - x\| \to 0$$

if $x_k \to x$, because by Remark 6.2(i), $x_k \to x$ is equivalent to $\|x_k - x\| \to 0$. Thus the norm is a continuous mapping from V to $\mathbb{R}$. Note, however, that continuity generally holds only for the norm used in the definition of convergence and not necessarily for other norms on the same vector space. □

Just as in the real case, we now make some statements about the continuity of compositions of continuous functions.

Theorem 6.11 Let (X, d_X), (Y, d_Y), (Z, d_Z) be metric spaces, and $f : Y \to Z$ and $g : X \to Y$ be functions. If g is continuous at $x \in X$ and f is continuous at $y = g(x) \in Y$, then the composition $f \circ g : X \to Z$ is continuous at x.

Proof: Let $x_k \to x$. From the continuity of g we get $g(x_k) \to g(x)$, and from the continuity of f we get $f(g(x_k)) \to f(g(x))$ as $k \to \infty$. This proves the claim. □

As mentioned at the beginning of Chapter 5, a mapping to $\mathbb{R}^m$ is written as a vector of component mappings $f_1, \dots, f_m$ to $\mathbb{R}$. For the continuity of such a mapping, we have:

Theorem 6.12 Let

$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{pmatrix}$$

with $f_i : X \to \mathbb{R}$, $i = 1, \dots, m$. Then for every p-norm from Definition 5.6, f is continuous at a point $x \in X$ if and only if all component functions f_i are continuous at x.

Proof: Follows immediately from the definition of continuity and Theorem 6.3. □

The following theorem establishes the continuity of some functions from $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$. The notation $\mathbb{R} \times \mathbb{R}$ is just another way of writing $\mathbb{R}^2$, emphasizing that the function arguments are not interpreted as vectors in $\mathbb{R}^2$ (although one could do so if desired).

Theorem 6.13 The following mappings are continuous:

$$\begin{array}{rlll}
\text{(i)} & \text{add:} & \mathbb{R}\times\mathbb{R}\to\mathbb{R}, & (x,y) \mapsto x+y \\
\text{(ii)} & \text{mult:} & \mathbb{R}\times\mathbb{R}\to\mathbb{R}, & (x,y) \mapsto xy \\
\text{(iii)} & \text{quot:} & \mathbb{R}\times\mathbb{R}\setminus\{0\}\to\mathbb{R}, & (x,y) \mapsto x/y.
\end{array}$$

Proof: Let (x_k, y_k) be a sequence converging to a point $(x, y) \in \mathbb{R} \times \mathbb{R}$ (with (x_k, y_k) and (x, y) in $\mathbb{R} \times \mathbb{R} \setminus \{0\}$ in case (iii)). By Theorem 6.3 we have $x_k \to x$ and $y_k \to y$. The required convergences $x_k + y_k \to x + y$, $x_k y_k \to xy$, and $x_k/y_k \to x/y$ then follow from [7, Theorems 8.9, 8.10, and 8.13]. □

Corollary 6.14 Let (X, d) be a metric space and $f, g : X \to \mathbb{R}$ continuous functions. Then $f + g$ and fg are continuous. If additionally $g(x) \neq 0$ for all $x \in X$, then the function f/g is also continuous.

Proof: Follows immediately from Theorems 6.13 and 6.11, since the functions can be written as $f + g = \text{add} \circ (f, g)$, $fg = \text{mult} \circ (f, g)$, and $f/g = \text{quot} \circ (f, g)$. □

Example 6.15 We consider some examples of functions from $\mathbb{R}^n$ to $\mathbb{R}$, where we write $x \in \mathbb{R}^n$ as usual as $x = (x_1, x_2, \dots, x_n)^T$.

(i) For each $i = 1, \dots, n$, the coordinate function $x \mapsto x_i$ is continuous, because if $x_k \to a \in \mathbb{R}^n$, Theorem 6.3 implies that $x_{ki} \to a_i$.

(ii) A monomial function of degree r is a function $g : \mathbb{R}^n \to \mathbb{R}$ of the form

$$g(x) = x_1^{k_1} x_2^{k_2} \cdot \ldots \cdot x_n^{k_n}$$

with $k_1, \dots, k_n \in \mathbb{N}$ and $k_1 + k_2 + \ldots + k_n = r$. Hence g is a product of coordinate functions from (i), and thus continuous by Corollary 6.14.

(iii) A polynomial of degree $\leq r$ in $\mathbb{R}^n$ is a function of the form

$$f(x) = \sum_{\substack{k_1,k_2,\dots,k_n=0 \\ k_1+k_2+\ldots+k_n\leq r}}^{r} c_{k_1k_2\dots k_n} x_1^{k_1} x_2^{k_2} \cdot \ldots \cdot x_n^{k_n}.$$

Each summand is the product of a constant (and thus continuous) function and a function from (ii), and hence continuous by Corollary 6.14. By the same corollary, the entire function f is continuous. □

Question 6.3 *For $n = 2$, list all combinations of $x_1^{k_1} x_2^{k_2}$ that can appear in a polynomial of degree ≤ 3.*

Now we turn to the second criterion for continuity known from $\mathbb{R}$, the ε-δ criterion.

Theorem 6.16 Let (X, d_X) and (Y, d_Y) be metric spaces and $f : X \to Y$ a function. Then f is continuous at a point $a \in X$ if and only if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $x \in X$,

$$d_X(x, a) < \delta \;\Rightarrow\; d_Y(f(x), f(a)) < \varepsilon.$$

Proof: The proof proceeds completely analogously to the corresponding proof in $\mathbb{R}$, cf. [7, Theorem 14.16], replacing the absolute values there with the appropriate metrics.

First, we show that the ε-δ criterion implies the sequence definition of continuity 6.9. Let $(x_k)_{k\in\mathbb{N}}$ be an arbitrary sequence in X with $x_k \to a$. We need to show that $f(x_k)$ converges to $f(a)$.

Let $\varepsilon > 0$ be given and let $\delta > 0$ come from the criterion. Since $x_k \to a$, there exists an $N_a(\delta) \in \mathbb{N}$ such that $d_X(x_k, a) < \delta$ for all $k \geq N_a(\delta)$, and hence

$$d_Y(f(x_k), f(a)) < \varepsilon.$$

Thus the convergence criterion from Remark 6.2(i) is satisfied with $N(\varepsilon) = N_a(\delta)$.

Conversely, assume that f is continuous at a in the sequential sense but that the ε-δ criterion does not hold. Then there exists an $\varepsilon > 0$ such that for every $\delta > 0$ there is an $x_\delta \in X$ with $d_X(x_\delta, a) < \delta$ and $d_Y(f(x_\delta), f(a)) \geq \varepsilon$.

Now set $\delta = 1/k$ and choose $x_k = x_\delta$. Then $d_X(x_k, a) < 1/k$, so $x_k \to a$ by Remark 6.2(i). However,

$$d_Y(f(x_k), f(a)) \geq \varepsilon$$

for all $k \geq 1$, so $d_Y(f(x_k), f(a)) \not\to 0$, i.e., $f(x_k) \not\to f(a)$, which contradicts the continuity of f at a. Therefore, the ε-δ criterion must hold. □

Example 6.17 Let (X, d) be a metric space and $x_0 \in X$ a fixed point. Consider the mapping $f : X \to \mathbb{R}$ given by

$$f(x) = d(x, x_0).$$

Then, by the reverse triangle inequality (5.6), for any $x, y \in X$ we have

$$|f(x) - f(y)| = |d(x, x_0) - d(y, x_0)| \leq d(x, y).$$

For $a = y$, this shows that the ε-δ criterion is satisfied with $\delta = \varepsilon$ (using the usual absolute-value metric on $\mathbb{R}$). Hence f is continuous at every $a \in X$. □

Since the ε-δ criterion works exactly as in $\mathbb{R}$, it can be applied in the same way. In particular, it follows that the proof of Theorem 4.6 remains valid in arbitrary metric spaces, i.e., from the uniform convergence

$$\lim_{k\to\infty} \sup_{x\in X} d_Y(f_k(x), f(x)) = 0$$

of a sequence of continuous functions $f_k : X \to Y$ to a limit function $f : X \to Y$, the limit function f is continuous, just as in the real case.

6.3 Linear Functions

We now briefly consider a special class of functions that will play an important role, in particular in the definition of the derivative.

Definition 6.18 Let V, W be vector spaces over $\mathbb{R}$. A function $f : V \to W$ is called *linear* if for all $v, \tilde{v} \in V$ and all $\lambda \in \mathbb{R}$ the equations

$$f(v + \tilde{v}) = f(v) + f(\tilde{v}) \quad \text{and} \quad f(\lambda v) = \lambda f(v)$$

hold. For linear functions, the notation fv is often used instead of $f(v)$. □

Example 6.19 (i) If $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$, any linear function can be represented by a matrix once a basis is fixed (here we use, as usual, the standard basis (5.2)), and its action is given by matrix-vector multiplication:

Consider the standard basis vectors e_i from (5.2). Any vector $x = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$ can be written as

$$x = x_1 e_1 + x_2 e_2 + \dots + x_n e_n.$$

By linearity of f,

$$f(x) = x_1 f(e_1) + x_2 f(e_2) + \dots + x_n f(e_n).$$

Hence, f is uniquely determined by the n vectors $f(e_1), f(e_2), \dots, f(e_n) \in \mathbb{R}^m$. Writing these as columns in a matrix $A = (f(e_1), f(e_2), \dots, f(e_n)) \in \mathbb{R}^{m \times n}$, we have

$$Ax = x_1 f(e_1) + x_2 f(e_2) + \dots + x_n f(e_n) = f(x).$$

Thus every linear map $f : \mathbb{R}^n \to \mathbb{R}^m$ is of the form

$$f(x) = Ax$$

for the matrix $A \in \mathbb{R}^{m \times n}$ constructed in this way.

Question 6.4 *What is the shape of the matrix A when f maps from $\mathbb{R}^n$ to $\mathbb{R}$?*

(ii) Consider the space $C^\infty(\mathbb{R}, \mathbb{R})$ of infinitely often (continuously) differentiable real functions. For any $f \in C^\infty(\mathbb{R}, \mathbb{R})$, its derivative $f' = \frac{d}{dx} f$ is again a function in $C^\infty(\mathbb{R}, \mathbb{R})$. Hence the derivative $\frac{d}{dx}$ is a function from $C^\infty(\mathbb{R}, \mathbb{R})$ to $C^\infty(\mathbb{R}, \mathbb{R})$. Since $(f + g)' = f' + g'$ and $(\lambda f)' = \lambda f'$, the derivative is indeed a linear function.

(iii) The Riemann integral, by Theorem 3.3, is a linear function from the set of Riemann-integrable functions to $\mathbb{R}$. □

The following theorem gives a criterion for the continuity of a linear map.

Theorem 6.20 Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed vector spaces and let $f : V \to W$ be a linear map. Then f is continuous if and only if there exists a constant $C \in \mathbb{R}$ such that for all $x \in V$,

$$\|f(x)\|_W \le C \|x\|_V.$$

A linear map satisfying this condition is called *bounded*.

Proof: "$\Rightarrow$": From the second condition for linear functions it follows that $f(0) = 0$ for any linear map.

Assume f is continuous. Then it is, in particular, continuous at $x = 0$. Hence there exists, for $\varepsilon = 1$, a $\delta > 0$ such that $\|f(x)\|_W < 1$ for all $x \in V$ with $\|x\|_V < \delta$. Set $C := 2/\delta$. For any $x \in V$, set $\lambda := 1/(C\|x\|_V)$ and $y = \lambda x$. Then $\|y\|_V = \lambda\|x\|_V = 1/C < \delta$, so $\|f(y)\|_W < 1$. Since

$$f(y) = f(\lambda x) = \lambda f(x) = \frac{1}{C\|x\|_V} f(x),$$

we get

$$\|f(x)\|_W = \Big\|C\|x\|_V f(y)\Big\|_W = C\|x\|_V\|f(y)\|_W < C\|x\|_V.$$

"$\Leftarrow$": If there exists $C > 0$ with $\|f(x)\|_W \leq C\|x\|_V$ for all $x \in V$, then by linearity,

$$\|f(x) - f(y)\|_W = \|f(x - y)\|_W \leq C\|x - y\|_V.$$

Hence the ε-δ criterion is satisfied with $\delta = \varepsilon/C$. Therefore, f is continuous. □

Example 6.21 (i) For $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$, let $f(x) = Ax$ for a matrix $A \in \mathbb{R}^{m\times n}$. We prove the boundedness condition from Theorem 6.20 for the Euclidean norm. Let the entries of the matrix be denoted by a_{ij}. Then, for the entries of the vector $y = Ax \in \mathbb{R}^m$, we have

$$y_i = \sum_{k=1}^{n} a_{ik}x_k.$$

Let $a_{i\cdot}$ denote the i-th row of A. Then we can write this as the inner product $\langle a_{i\cdot}^T, x\rangle$. By the Cauchy-Schwarz inequality and Theorem 5.14,

$$|y_i| = |\langle a_{i\cdot}^T, x\rangle| \leq \|a_{i\cdot}^T\|_2\|x\|_2 \leq \sqrt{n}\|a_{i\cdot}^T\|_\infty\|x\|_2 \leq \sqrt{n}\alpha\|x\|_2,$$

with $\alpha = \max_{i,k}|a_{ik}|$. Hence, using Theorem 5.14 again,

$$\|Ax\|_2 = \|y\|_2 \leq \sqrt{m}\|y\|_\infty \leq \sqrt{nm}\alpha\|x\|_2.$$

Therefore, $f(x) = Ax$ is bounded with $C = \sqrt{nm}\alpha$, and thus every linear map $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous.[1]

(ii) Consider the set of continuous functions $C([a, b], \mathbb{R})$ with the ∞-norm, see Example 5.15. Then, for the Riemann (and also for the Lebesgue) integral,

$$\left|\int_a^b f(x)\,dx\right| \leq \int_a^b |f(x)|\,dx \leq \int_a^b \|f\|_\infty\,dx = (b - a)\|f\|_\infty.$$

Hence, the integral map is bounded with $C = b-a$, and therefore continuous on the normed vector space $(C([a, b], \mathbb{R}), \|\cdot\|_\infty)$.

(iii) Consider the set of infinitely differentiable functions $C^\infty([0, 1], \mathbb{R})$ with the ∞-norm. We show that the derivative $\frac{d}{dx}$ is not continuous. We prove this by showing it is not bounded, which by Theorem 6.20 implies it cannot be continuous.

[1] Just as with convergence, continuity is independent of the choice of norm in $\mathbb{R}^n$ and $\mathbb{R}^m$, see Theorem 5.14.

We need to show that for every $C > 0$, there exists $f \in C^\infty([0,1], \mathbb{R})$ with

$$\left\| \frac{d}{dx} f \right\|_\infty > C \|f\|_\infty.$$

Let $C > 0$ be arbitrary. For $n \in \mathbb{N}$ with $n > C$, consider $f(x) = x^n$. Then $\|f\|_\infty = 1$ and $\frac{d}{dx} f(1) = n$, so

$$\left\| \frac{d}{dx} f \right\|_\infty = n = n\|f\|_\infty > C\|f\|_\infty.$$

□

Question 6.5 *Find a nonlinear function $f : \mathbb{R} \to \mathbb{R}$ that shows that Theorem 6.20 does not hold if f is not linear.*

The set of linear maps $f : V \to W$ is itself a vector space, since for two linear maps f and g, both $f + g$ and λf are again linear. It is natural to ask whether we can define a norm on this space. This is indeed possible if we restrict to the set of bounded linear maps.

Definition 6.22 For a bounded linear map $f : V \to W$, we define the norm

$$\|f\| := \sup\{\|f(x)\|_W \mid x \in V \text{ with } \|x\|_V \leq 1\}.$$

□

This is indeed a norm (satisfying the conditions in Definition 5.4), although we do not verify it in detail here. Note that $\|f\|$ depends on the choice of norms $\|\cdot\|_V$ and $\|\cdot\|_W$, even if this is not explicit in the notation (one could write $\|f\|_{V,W}$, but this is cumbersome).

From the definition, it immediately follows that $\|f\| \leq C$ for the C from the boundedness condition in Theorem 6.20. Moreover, for all $x \in V$ with $x \neq 0$,

$$\|f(x)\|_W = \left\| f\left(\frac{\|x\|_V}{\|x\|_V} x\right) \right\|_W = \|x\|_V \left\| f\left(\frac{1}{\|x\|_V} x\right) \right\|_W \leq \|x\|_V \|f\|,$$

since $\|(1/\|x\|_V)x\|_V = 1$. For $x = 0$, the inequality also holds since $f(0) = 0$. Hence, the inequality holds for all $x \in V$.

In the case $f : \mathbb{R}^n \to \mathbb{R}^m$, $f(x) = Ax$, we also write $\|A\|$ instead of $\|f\|$. With the Euclidean norm, from Example 6.21(i), we have

$$\|A\| \leq \sqrt{nm} \max_{i,k} |a_{ik}|.$$

6.4 Alternative Continuity Criteria

We conclude this section with three alternative descriptions of continuity that do not rely on sequence convergence (as in Definition 6.9) nor on a metric (as in the ε-δ criterion in Theorem 6.16).

Theorem 6.23 Let (X, d_X) and (Y, d_Y) be metric spaces, and $f : X \to Y$. Then:

(i) f is continuous at $a \in X$ if and only if for every neighborhood V of $f(a)$ there exists a neighborhood U of a with $f(U) \subset V$.

(ii) f is continuous on all of X if and only if the *preimage*

$$f^{-1}(V) := \{x \in X \mid f(x) \in V\}$$

of every open set $V \subset Y$ is itself open.

(iii) f is continuous on all of X if and only if the preimage

$$f^{-1}(V) := \{x \in X \mid f(x) \in V\}$$

of every closed set $V \subset Y$ is itself closed.

Proof: (i) Every δ-ball $B_\delta(a)$ is a neighborhood of a, and every neighborhood of $f(a)$ contains an ε-ball $B_\varepsilon(f(a))$. Hence (i) is equivalent to the ε-δ criterion.

(ii) Suppose f is continuous and $V \subset Y$ is open. We show $f^{-1}(V)$ is open: for any $x \in f^{-1}(V)$, there exists $\varepsilon > 0$ with $B_\varepsilon(x) \subset f^{-1}(V)$. Let $x \in f^{-1}(V)$. Since V is open and $f(x) \in V$, V is a neighborhood of $f(x)$. By (i), there exists a neighborhood $U \subset X$ of x with $f(U) \subset V$. Then U contains a ball $B_\varepsilon(x)$, and $f(B_\varepsilon(x)) \subset f(U) \subset V$. Thus $B_\varepsilon(x) \subset f^{-1}(V)$, so $f^{-1}(V)$ is open.

Conversely, suppose preimages of open sets are open. Let $a \in X$. Any neighborhood V of $f(a)$ contains an ε-ball $B_\varepsilon(f(a))$. Since $B_\varepsilon(f(a))$ is open, $U = f^{-1}(B_\varepsilon(f(a)))$ is open with $a \in U$, and $f(U) \subset V$. Hence (i) is satisfied, and f is continuous.

(iii) To show (iii) is equivalent to (ii), note that for any $A \subset Y$,

$$f^{-1}(Y \setminus A) = \{x \in X \mid f(x) \notin A\} = X \setminus f^{-1}(A).$$

If (iii) holds and $V \subset Y$ is open, then $f^{-1}(V) = X \setminus f^{-1}(Y \setminus V)$. Since $Y \setminus V$ is closed, $f^{-1}(Y \setminus V)$ is closed, so $f^{-1}(V)$ is open. The converse follows similarly. □

An important application of parts (ii) and (iii) is the following example.

Example 6.24 Let (X, d) be a metric space and $f : X \to \mathbb{R}$ continuous. Then for any $c \in \mathbb{R}$, the set

$$O := \{x \in X \mid f(x) < c\}$$

is open, and

$$A := \{x \in X \mid f(x) \leq c\}$$

is closed.

Proof: We have $O = f^{-1}((-\infty, c))$ and $A = f^{-1}((-\infty, c])$. The claim follows from Theorem 6.23 (ii) and (iii), since $(-\infty, c)$ is open and $(-\infty, c]$ is closed in $\mathbb{R}$. □

Question 6.6 *What can you say about the set $\{x \in X \mid f(x) = c\}$ being open or closed?*

Chapter 7

Compactness

Last update: July 21, 2026

In Chapter 5 we have already seen how open and closed intervals can be generalized to metric spaces (and therefore in particular to $\mathbb{R}^n$). In this chapter we extend this generalization to compact intervals $[a, b]$, whose extension gives rise to the *compact sets*.

7.1 Definition and Examples

Compact sets can be defined in several equivalent ways. We choose one of them for our definition and formulate the others later in corresponding theorems.

Definition 7.1 Let (X, d) be a metric space and $A \subset X$.

(i) A family of open sets $U_\alpha \subset X$, $\alpha \in \mathcal{I}$, is called an *open cover* of A if

$$A \subset \bigcup_{\alpha \in \mathcal{I}} U_\alpha$$

holds.

(ii) The set A is called *compact* if for every open cover U_α, $\alpha \in \mathcal{I}$, there exists a finite set $\widetilde{\mathcal{I}} \subset \mathcal{I}$ such that

$$A \subset \bigcup_{\alpha \in \widetilde{\mathcal{I}}} U_\alpha.$$

The set $\{U_\alpha \,|\, \alpha \in \widetilde{\mathcal{I}}\}$ is then called a *finite subcover*. □

Question 7.1 *Find an open cover for $A = \mathbb{N}$.*

Example 7.2 We consider the concept just defined for subsets of $\mathbb{R}$ with $d(x, y) = |x - y|$.

(i) We show by contradiction that every compact interval $I = [a, b] \subset \mathbb{R}$ satisfies the above condition. For this purpose we assume that there exists such an interval I and an open cover for which no finite subcover exists. We can then divide the interval I into two equally large closed subintervals I_1 and I_1', for at least one of which (without loss of generality I_1) there can also be no finite subcover (otherwise the two finite subcovers of I_1 and I_1' together would yield a finite subcover of I). We proceed iteratively with this subinterval and subdivide it again.

In this way we obtain a nested sequence of intervals I_k with $|I_k| \leq (1/2)^k |I|$ for which no finite subcover exists. By the completeness axiom there exists an $x \in \mathbb{R}$ with $x \in I_k \subset I$ for all $k \in \mathbb{N}$, and since the U_α form a cover of I, there must exist an $U_{\alpha'}$ with $x \in U_{\alpha'}$. Since $U_{\alpha'}$ is open, it contains a ball around x, i.e., an open interval $(x-\varepsilon, x+\varepsilon) \subset U_{\alpha'}$. If we now choose k so large that $(1/2)^k|I| < \varepsilon$, then from $x \in I_k$ it follows that $I_k \subset U_{\alpha'}$. Hence $\{U_{\alpha'}\}$ is a finite subcover of I_k, which contradicts the conclusion that none of the intervals I_k possesses a finite subcover.

(ii) Open or half-open intervals in $\mathbb{R}$ are not compact. Consider for example the interval $A = [a, b)$ with $a, b \in \mathbb{R}$. The sets $U_\alpha = (a-1, b-1/\alpha)$, $\alpha \in \mathcal{I} = \mathbb{N} \setminus \{0\}$, form a cover of A, because every point $x \in A$ satisfies $x \geq a$ and $x < b$ and lies (as can be verified) in the set U_α for every $\alpha \geq 2/(b-x)$. However, there is no finite subcover, since in every finite subset $\widetilde{\mathcal{I}} \subset \mathcal{I}$ there exists a maximal α^*. Consequently all points in the interval $[b-1/\alpha^*, b)$ are not contained in the subcover.

(iii) Unbounded subsets A of $\mathbb{R}$ are not compact. Indeed, consider the cover $U_\alpha = (\alpha - 1, \alpha+1)$ for $\alpha \in \mathcal{I} = \mathbb{Z}$ (which covers all of $\mathbb{R}$ and therefore also every set $A \subset \mathbb{R}$). Any finite subcover covers only a bounded set and therefore cannot cover the whole set A. □

In the following theorems we consider some more general compact sets.

Theorem 7.3 Every set in $\mathbb{R}^n$ of the form

$$Q = [a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_n, b_n]$$

with $a_i \leq b_i$ (Q is called a *cuboid*) is compact.

Proof: The proof proceeds analogously to Example 7.2(i): We assume that there exists a cover U_α, $\alpha \in \mathcal{I}$, for which no finite subcover exists, set $Q_0 := Q$, and construct — instead of the subintervals I_k in Example 7.2(i) — subrectangles Q_k by halving the individual intervals. At each step we choose Q_k among the 2^n subrectangles arising from subdividing Q_{k-1} in such a way that again no finite subcover exists for it. For the sequence of rectangles Q_k constructed in this way, the edge lengths are halved at every step, which implies that for any two points $x, y \in Q_k$ the inequality $\|x-y\|_\infty \leq (1/2)^k C$ holds, where C is the maximal edge length of Q.

If we now choose an arbitrary point x_k from each rectangle Q_k, then $\|x_k - x_m\|_\infty \leq (1/2)^k C$ holds for all $k \in \mathbb{N}$ and all $m \geq k$ (since every point x_m is also contained in Q_k). Hence the sequence x_k is a Cauchy sequence and therefore converges to some $x \in \mathbb{R}^n$. For this limit we obtain (by letting m tend to infinity in the above inequality) the estimate $\|x_k - x\| \leq (1/2)^k C$. Since every rectangle Q_k is closed, it follows from Theorem 6.4 that $x \in Q_k$ for all k, and in particular $x \in Q$. Thus there exists some U_α with $x \in U_\alpha$, and because U_α is open there exists a ball $B_\varepsilon(x) \subset U_\alpha$ (we again take the ball with respect to the ∞-norm). Choosing k so large that $(1/2)^k C < \varepsilon$ yields $Q_k \subset B_\varepsilon(x) \subset U_\alpha$, which means that $\{U_\alpha\}$ is a finite subcover of Q_k. But by construction no Q_k admits such a finite subcover, which gives a contradiction. □

Theorem 7.4 Let (X, d) be a metric space and $(x_k)_{k\in\mathbb{N}}$ a convergent sequence in X with limit x. Then the set

$$A := \{x_k \mid k \in \mathbb{N}\} \cup \{x\}$$

is compact.

Proof: Let U_α be an open cover of A. Then there exist α_k, $k \in \mathbb{N}$, and α^* with $x_k \in U_{\alpha_k}$ and $x \in U_{\alpha^*}$. Since U_{α^*} is open, Remark 6.2(iii) implies the existence of $N(U_{\alpha^*}) \in \mathbb{N}$ such that $x_k \in U_{\alpha^*}$ holds for all $k \geq N(U_{\alpha^*})$. Thus

$$U_{\alpha_1}, U_{\alpha_2}, \ldots, U_{\alpha_{N(U_{\alpha^*})-1}}, U_{\alpha^*}$$

is a finite subcover. □

7.2 Properties of Compact Sets

Definition 7.5 (i) A set A in a metric space (X, d) is called *bounded* if there exist $x \in X$ and $r > 0$ such that $A \subset B_r(x)$.

(ii) A sequence $(a_k)_{k\in\mathbb{N}}$ in a metric space (X, d) is called *bounded* if the set $A = \{a_k \mid k \in \mathbb{N}\}$ is bounded. □

Theorem 7.6 Every compact set A in a metric space (X, d) is bounded and closed.

Proof: Boundedness: Let $a \in A$ be arbitrary. Then the sets $U_\alpha = B_\alpha(a)$ with $\alpha \in \mathcal{I} = \mathbb{N} \setminus \{0\}$ form an open cover of A, since every point $x \in A$ is certainly contained in the ball $B_{2d(x,a)}(a)$. Since A is compact, there exists a finite subcover U_α, $\alpha \in \widetilde{\mathcal{I}}$. As $\widetilde{\mathcal{I}}$ is finite, it contains a maximal $\alpha \in \mathbb{N}$, and it follows that $A \subset B_\alpha(a)$, which proves boundedness.

Closedness: We must show that $X \setminus A$ is open. For this purpose, for every $x \in X \setminus A$ we must prove that there exists a ball $B_\varepsilon(x) \subset X \setminus A$. For $\alpha \in \mathcal{I} = \mathbb{N} \setminus \{0\}$ we define

$$U_\alpha := \{y \in X \mid d(x, y) > 1/\alpha\}.$$

Each of these sets U_α is open (cf. Example 5.19(v)), and their union contains all points of X except for x. Thus the U_α form a cover of A, and there exists a finite subcover. If we choose the maximal $\alpha \in \mathbb{N}$ in this subcover, then $A \subset U_\alpha$, and therefore

$$B_{1/\alpha}(x) \subset X \setminus U_\alpha \subset X \setminus A.$$

Thus we have found an open ball around x in $X \setminus A$ with $\varepsilon = 1/\alpha$. □

Theorem 7.7 Let (X, d) be a metric space, $K \subset X$ compact, and $A \subset K$ closed. Then A is compact.

Proof: Let U_α be an open cover of A. If we add the open set $X \setminus A$ to this cover, we obtain an open cover of K for which a finite subcover exists. This is then also a finite subcover of A. If we remove (if present) the set $X \setminus A$ from the finite subcover, the remaining sets still form a finite subcover of A, since $X \setminus A$ contains no points of A. Hence the remaining sets form a finite subcover of A with respect to the original cover U_α. □

The converse of Theorem 7.6 unfortunately does not hold in general metric spaces. However, it does hold in the important special case $X = \mathbb{R}^n$, as shown by the following theorem.

Theorem 7.8 (Heine–Borel) A subset $A \subset \mathbb{R}^n$ is compact if and only if it is bounded and closed.

Proof: "$\Rightarrow$": This holds in every metric space by Theorem 7.6, and therefore also in $\mathbb{R}^n$.

"$\Leftarrow$": If A is bounded, then there exists a ball $B_r(x)$ (without loss of generality with respect to the ∞-norm) such that $A \subset B_r(x)$. Hence A is also contained in the rectangle

$$Q = [x_1 - r, x_1 + r] \times [x_2 - r, x_2 + r] \times \ldots \times [x_n - r, x_n + r].$$

Since this set is compact by Theorem 7.3, the statement follows from Theorem 7.7. □

Question 7.2 *Find a compact set in $\mathbb{R}$ that is not an interval.*

7.3 Compactness and Limits

In the following we consider several properties of compact sets in connection with limits. We begin with a corollary of Theorem 7.4 and Theorem 7.6, which generalizes [7, Theorem 8.7] to metric spaces.

Corollary 7.9 Every convergent sequence in a metric space is bounded.

Proof: This follows immediately from Theorem 7.4 and Theorem 7.6. □

The following theorem generalizes the Bolzano–Weierstrass Theorem [7, Theorem 8.42] to metric spaces.

Theorem 7.10 Let (X, d) be a complete metric space, $A \subset X$ compact, and $(x_k)_{k\in\mathbb{N}}$ a sequence in A. Then there exists a subsequence that converges to a limit $a \in A$.

Proof: Assume that there is no convergent subsequence. Then every point $x \in A$ possesses an open neighborhood U_x that contains only finitely many elements of the sequence. Indeed, if this were not the case, there would exist a point such that infinitely many elements of the sequence lie in every neighborhood $B_{1/p}(x)$, $p \geq 1$. Define inductively a subsequence with $k_0 := 0$ and $k_p \geq k_{p-1} + 1$ minimal such that $x_{k_p} \in B_{1/p}(x)$. Then $x_{k_p} \in B_{1/p_0}(x)$ for all $p \geq p_0$, and the subsequence x_{k_p} would converge to x by Remark 6.2(ii).

The open neighborhoods U_x now form a cover of A, and therefore a finite subcover exists. However, since each set in the finite subcover contains only finitely many sequence elements, only finitely many different elements of the sequence lie in A. This means that the sequence becomes constant. It thus converges and has a convergent subsequence, which is a contradiction. □

As already seen in Theorem 7.6, the statement of the preceding theorem can be strengthened in the special case $X = \mathbb{R}^n$.

Corollary 7.11 Every bounded sequence in $\mathbb{R}^n$ possesses a convergent subsequence.

Proof: Analogously to the proof of Theorem 7.8, one shows that the sequence is contained in a cuboid Q, which is compact by Theorem 7.3. The claim then follows from Theorem 7.10. □

7.4 Compactness and Continuity

The results in this section show that compact sets play a role for continuous functions on metric spaces (and consequently also in $\mathbb{R}^n$) similar to that played by closed intervals for real-valued functions. To show this, we first need a preparatory theorem.

Theorem 7.12 Let (X, d_X) and (Y, d_Y) be metric spaces and $f : X \to Y$ continuous. Then the following holds: if $K \subset X$ is compact, then $f(K) \subset Y$ is also compact.

Proof: Let U_α, $\alpha \in \mathcal{I}$, be an open cover of $f(K)$. By Theorem 6.23, the sets $V_\alpha := f^{-1}(U_\alpha)$ are open. Moreover, they cover K, and therefore there exists a finite subcover V_α, $\alpha \in \widetilde{\mathcal{I}}$, of K. Hence the sets $U_\alpha = f(V_\alpha)$, $\alpha \in \widetilde{\mathcal{I}}$, also cover $f(K)$ and therefore form a finite subcover. □

The following theorem generalizes [7, Theorem 14.14] to general metric spaces.

Theorem 7.13 Let (X, d) be a metric space and $K \subset X$ compact. Then every continuous function $f : X \to \mathbb{R}$ is bounded on K, i.e. there exists $M \in \mathbb{R}$ such that $|f(x)| \le M$ for all $x \in K$, and it attains its maximum and minimum, i.e. there exist $p, q \in K$ such that

$$f(p) = \sup\{f(x) \mid x \in K\} \quad \text{and} \quad f(q) = \inf\{f(x) \mid x \in K\}.$$

Proof: By Theorem 7.12, $f(K)$ is compact in $\mathbb{R}$, and therefore bounded and closed. The boundedness of f follows immediately. For the existence of p, let $p_k \in K$ be a sequence such that

$$\lim_{k\to\infty} f(p_k) = \sup\{f(x) \mid x \in K\}.$$

By Theorem 7.10, there exists a convergent subsequence $p_{k_j} \to p \in K$. The sequence $f(p_{k_j})$ is then a subsequence of the sequence $f(p_k)$ and, as a subsequence of a convergent sequence, converges to the same limit. Since f is continuous, it follows that

$$f(p) = \lim_{j\to\infty} f(p_{k_j}) = \sup\{f(x) \mid x \in K\}.$$

The statement for the infimum follows analogously. □

Question 7.3 *Why does the sequence p_k used in the previous proof exist?*

As a further generalization we consider [7, Theorem 14.20], which states that every continuous function on bounded and closed intervals is uniformly continuous. For this we need the following definition.

Definition 7.14 Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \to Y$ is called *uniformly continuous* on a set $A \subset X$ if the following holds:

For every $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $x, x' \in A$,

$$d_X(x, x') < \delta \;\Rightarrow\; d_Y(f(x), f(x')) < \varepsilon.$$

□

Theorem 7.15 Let (X, d_X) and (Y, d_Y) be metric spaces and let $K \subset X$ be compact. Then every continuous function $f : K \to Y$ is also uniformly continuous on K.

Proof: Let $\varepsilon > 0$ be given. By the ε–δ criterion of continuity (applied with $\varepsilon/2$ instead of ε), for every $a \in K$ there exists $\delta_a > 0$ such that

$$d_X(x, a) < \delta_a \;\Rightarrow\; d_Y(f(x), f(a)) < \varepsilon/2.$$

Now consider the balls

$$B_{\delta_a/2}(a), \quad a \in K,$$

which clearly cover the set K. Hence there exists a finite subcover

$$B_{\delta_{a_1}/2}(a_1), \ldots, B_{\delta_{a_m}/2}(a_m). \tag{7.1}$$

We now prove that the condition of uniform continuity holds for

$$\delta := \min\{\delta_{a_1}/2, \ldots, \delta_{a_m}/2\}.$$

Let $x, x' \in K$ be arbitrary points with $d_X(x, x') < \delta$. Since the balls in (7.1) cover K, the point x lies in one of these balls, say $B_{\delta_{a_k}/2}(a_k)$. Thus $d_X(x, a_k) < \delta_{a_k}/2$ and therefore $d_Y(f(x), f(a_k)) < \varepsilon/2$.

From $d_X(x, x') < \delta$ we obtain

$$d_X(x', a_k) \le d_X(x', x) + d_X(x, a_k) < \delta + \delta_{a_k}/2 \le \delta_{a_k},$$

and therefore $d_Y(f(x'), f(a_k)) < \varepsilon/2$.

Hence

$$d_Y(f(x), f(x')) \le d_Y(f(x), f(a_k)) + d_Y(f(a_k), f(x')) < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

□

7.5 Equivalence of Norms in $\mathbb{R}^n$

A consequence of Theorem 5.14 is that for any two p-norms $\|\cdot\|_p$, $\|\cdot\|_q$ there exist constants $\alpha > 0$ and $\beta > 0$ such that for all $x \in \mathbb{R}^n$ the inequality

$$\alpha\|x\|_p \leq \|x\|_q \leq \beta\|x\|_p$$

holds. Norms with this property are called *equivalent* , because they generate the same topology (that is, the same open and closed sets) and the same notions of convergence.

The following theorem shows (using several results from this chapter in the proof) that this statement holds not only for the p-norms, but in fact for all possible norms[1] in $\mathbb{R}^n$.

Theorem 7.16 Let $\|\cdot\|$ and $|||\cdot|||$ be two arbitrary norms in $\mathbb{R}^n$. Then there exist constants $\alpha > 0$ and $\beta > 0$ such that for all $x \in \mathbb{R}^n$ the inequality

$$\alpha\|x\| \leq |||x||| \leq \beta\|x\|$$

holds.

Proof: It suffices to prove the statement for an arbitrary norm $\|\cdot\|$ and the norm $\|\cdot\|_1$, because if for two arbitrary norms the inequalities

$$\alpha\|x\| \leq \|x\|_1 \leq \beta\|x\| \qquad \text{and} \qquad \tilde{\alpha}|||x||| \leq \|x\|_1 \leq \tilde{\beta}|||x|||$$

hold for all $x \in \mathbb{R}^n$, then it follows that

$$\alpha\|x\| \leq \|x\|_1 \leq \tilde{\beta}|||x||| \Rightarrow \frac{\alpha}{\tilde{\beta}}\|x\| \leq |||x|||$$

and

$$\tilde{\alpha}|||x||| \leq \tilde{\beta}\|x\|_1 \leq \beta\|x\| \Rightarrow |||x||| \leq \frac{\beta}{\tilde{\alpha}}\|x\|.$$

From this the claim follows. Hence we prove the existence of α and β such that

$$\alpha\|x\| \leq \|x\|_1 \leq \beta\|x\|$$

for all $x \in \mathbb{R}^n$. It suffices to show the inequality for all $x \neq 0$, since for $x = 0$ it holds trivially because $\|x\| = \|x\|_1 = 0$.

To prove the first inequality we proceed similarly to the proof of Theorem 5.14(i). For $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ we define

$$y_i := x_i e_i,$$

where e_i is taken from (5.2). Define the values $c_i := \|e_i\|$. From the norm property we obtain

$$\|y_i\| = \|x_i e_i\| = |x_i|\,\|e_i\| = c_i|x_i|.$$

From the definition of the 1-norm it follows that

$$\|x\|_1 = \sum_{i=1}^{n} \|y_i\|_1 \quad \text{and} \quad c_i\|y_i\|_1 = c_i|x_i| = \|y_i\|.$$

[1] There are quite many of these. For example, one can verify that for every symmetric matrix $Q \in \mathbb{R}^{n\times n}$ satisfying $x^TQx > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$, the expression $\|x\|_Q := \sqrt{x^TQx}$ defines a norm.

Using the triangle inequality for the norm $\|\cdot\|$ we obtain

$$\|x\| \leq \|y_1\| + \|y_2\| + \ldots + \|y_n\| = c_1\|y_1\|_1 + c_2\|y_2\|_1 + \ldots + c_n\|y_n\|_1 \leq \left(\max_{i=1,\ldots,n} c_i\right) \|x\|_1.$$

Thus the first inequality follows with

$$\alpha = \frac{1}{\max_{i=1,\ldots,n} c_i}.$$

To prove the second inequality we set

$$\gamma := \inf\{\|x\| \,:\, \|x\|_1 = 1\}.$$

If $\gamma > 0$, then the second inequality follows with $\beta = 1/\gamma$. Indeed, for $x \neq 0$ and $\lambda := 1/\|x\|_1 > 0$ we have

$$\|\lambda x\|_1 = \lambda\|x\|_1 = \frac{\|x\|_1}{\|x\|_1} = 1,$$

and therefore

$$\|x\| = \frac{\lambda}{\lambda}\|x\| = \frac{1}{\lambda}\|\lambda x\| \geq \frac{1}{\lambda}\gamma = \gamma\|x\|_1.$$

Thus it remains to prove $\gamma > 0$.

Let x_k be a sequence of vectors in $\mathbb{R}^n$ with $\|x_k\|_1 = 1$ and $\gamma = \lim_{k\to\infty}\|x_k\|$. Since $\|\cdot\|_1$ is continuous, the set

$$\{x \in \mathbb{R}^n \mid \|x\|_1 \leq 1\}$$

is closed by Example 6.24. Moreover it is bounded and therefore compact by Theorem 7.8. Consequently, by Theorem 7.10 there exists a convergent subsequence x_{k_j} of x_k. Hence there exists $x^* \in \mathbb{R}^n$ with

$$\|x_{k_j} - x^*\|_1 \to 0 \quad \text{as } j \to \infty,$$

and again by the continuity of $\|\cdot\|_1$ it follows that

$$\|x^*\|_1 = \lim_{j\to\infty}\|x_{k_j}\|_1 = 1.$$

Thus $x^* \neq 0$ and therefore $\|x^*\| \neq 0$.

From the reverse triangle inequality and the first inequality already proved we obtain

$$\big|\, \|x^*\| - \|x_{k_j}\| \,\big| \leq \|x^* - x_{k_j}\| \leq \|x^* - x_{k_j}\|_1/\alpha \to 0$$

as $j \to \infty$, and therefore

$$\gamma = \lim_{j\to\infty}\|x_{k_j}\| = \|x^*\| > 0,$$

which proves the claim. $\square$

Chapter 8

Differentiability in $\mathbb{R}^n$

Last update: July 21, 2026

In this chapter we return to vector-valued functions $f : D \to \mathbb{R}^m$, $D \subset \mathbb{R}^n$, with

$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{pmatrix},$$

with (component) functions $f_i : D \to \mathbb{R}$, $i = 1, \dots, m$. We want to define when such a function is differentiable, what kind of object the derivative is, and how it can be computed.

For real-valued functions $f : D \to \mathbb{R}$ with $D \subset \mathbb{R}$ we introduced two equivalent definitions of differentiability at a point $x \in D$:

(i) The existence of the limit of the difference quotient

$$a = \lim_{\substack{h \to 0 \\ h \neq 0, x+h \in D}} \frac{f(x+h) - f(x)}{h} \in \mathbb{R}$$

(ii) The existence of a number $a \in \mathbb{R}$ and a function $\varphi : \mathbb{R} \to \mathbb{R}$ with

$$\lim_{h \to 0} \frac{\varphi(h)}{h} = 0,$$

such that

$$f(y) = f(x) + a(y - x) + \varphi(y - x) \tag{8.1}$$

holds for all $y \in D$.

If one of these two conditions[1] is satisfied, then the other one is also satisfied, and the numbers a in the two conditions coincide; this a is precisely the value of the derivative at x, which we usually denote by $f'(x)$ or—more rarely—by $\frac{d}{dx} f(x)$.

For differentiability in $\mathbb{R}^n$ we will initially use the second definition. We will see later why this formulation is more suitable.

[1] Looking more closely at the proof of the equivalence of the two conditions [7, Theorem 16.4], one sees that it suffices if condition (ii) holds for all $y \in (x - \varepsilon, x + \varepsilon) \cap D$ for an arbitrary $\varepsilon > 0$. Hence φ only needs to be defined on the interval $(-\varepsilon, \varepsilon)$.

8.1 Differentiability

We begin by formally deriving a "vector-valued" version of (8.1). Writing first $y = x + z$, we can rewrite (8.1) as

$$f(x+z) = f(x) + az + \varphi(z). \tag{8.2}$$

If we now move to the vector-valued case, then $x, z \in \mathbb{R}^n$ and $f(x), f(x+z) \in \mathbb{R}^m$. For (8.2) to make sense in terms of dimensions, we must therefore have $az \in \mathbb{R}^m$ and $\varphi(z) \in \mathbb{R}^m$. The second condition is easy to satisfy if we define φ as a mapping from $\mathbb{R}^n$ (or a subset of it) to $\mathbb{R}^m$. But how should we generalize the product az so that the result can be added to the other terms from $\mathbb{R}^m$?

If a were still a real number, then az (in the sense of scalar multiplication of a vector) would again lie in $\mathbb{R}^n$—but not, like the other terms in the equation, in $\mathbb{R}^m$. If one replaced a by a vector, one could interpret the multiplication as the scalar product of two vectors. However, the result of that would be a real number, which again would not fit.

If we want to keep the multiplication—which we do—the only possibility is to replace the number a by a matrix $A \in \mathbb{R}^{m\times n}$. With this choice we have $Az \in \mathbb{R}^m$ for all $z \in \mathbb{R}^n$.

Since every matrix corresponds uniquely (after choosing a basis) to a linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ according to Example 6.19(i), we can also view A as a linear mapping. In the following we do not distinguish in the notation between a linear mapping and the associated matrix.

Formally, the extension of (ii) is then defined as follows.

Definition 8.1 Let $f : D \to \mathbb{R}^m$ with $D \subset \mathbb{R}^n$ be given.

(i) The function f is called *(totally) differentiable* (or *Fréchet differentiable*) at a point $x \in D$ if there exist a linear mapping $A : \mathbb{R}^n \to \mathbb{R}^m$, a neighborhood $U \subset D$ of x, and a function $\varphi : V \to \mathbb{R}^m$ defined on

$$V := \{z \in \mathbb{R}^n \,|\, x + z \in U\}$$

such that

$$\lim_{\substack{z\to 0\\ z\in V\setminus\{0\}}} \frac{\varphi(z)}{\|z\|} = 0$$

and

$$f(x+z) = f(x) + Az + \varphi(z)$$

holds for all $z \in V$. In this case we call $Df(x) := A$ the *(total or Fréchet) derivative* of f at x. The associated matrix in $\mathbb{R}^{m\times n}$, which we also denote by $Df(x)$, is called the *Jacobian matrix*.

(ii) The function f is called *(totally) differentiable* (or *Fréchet differentiable*) on D if it is differentiable in the sense of (i) for all $x \in D$. The derivative is then a function $Df : D \to \mathbb{R}^{m\times n}$, i.e. to every point $x \in D$ there corresponds a Jacobian matrix $Df(x) \in \mathbb{R}^{m\times n}$. □

The limit $\lim_{z\to 0, z\in V}$ in (i) is to be understood in the sense of limits for functions, i.e. for every sequence $z_k \in V \setminus \{0\}$ with $z_k \to 0$ as $k \to \infty$ we must have $\varphi(z_k)/\|z_k\| \to 0$.

As in the real case, this condition on φ is also written as

$$\varphi(z) = o(\|z\|)$$

using the *Landau symbol* $o(\cdot)$.

Note that in part (i) of this definition we implicitly required that x lies in the interior of the domain D, since the definition can only be applied if a neighborhood U of x exists in D. Differentiability on D according to (ii) therefore implicitly assumes that the set D is open. This avoids problems with boundary points. These are even more complicated here than in $\mathbb{R}$, since there are not only "left" and "right" directions (and the corresponding derivatives), but many more directions. In the following the domain will therefore always be assumed to be open.

Question 8.1 *For which n, m is the total derivative $Df(x)$ of a function $f : D \to \mathbb{R}^m$, $D \subset \mathbb{R}^n$, a value in $\mathbb{R}$, and what is that function in this case? For which n, m is it a row or column vector?*

Analogously to the real case one proves the following theorem.

Theorem 8.2 If a function $f : D \to \mathbb{R}^m$, with $D \subset \mathbb{R}^n$ open, is differentiable at a point $x \in D$, then f is also continuous at x.

Proof: We must show that $\lim_{z\to 0} f(x+z) = f(x)$. Using the notation from Definition 8.1 we have $\lim_{z\to 0} Az = 0$ (because linear mappings from $\mathbb{R}^n$ to $\mathbb{R}^m$ are continuous according to Example 6.21) and

$$\lim_{z\to 0} \|\varphi(z)\| = \lim_{z\to 0} \frac{\|\varphi(z)\|}{\|z\|}\|z\| = \left(\lim_{z\to 0} \frac{\|\varphi(z)\|}{\|z\|}\right)\left(\lim_{z\to 0} \|z\|\right) = 0 \cdot 0 = 0.$$

Hence the claim follows from

$$\lim_{z\to 0} f(x+z) = \lim_{z\to 0} \big(f(x) + Az + \varphi(z)\big) = f(x) + \lim_{z\to 0} Az + \lim_{z\to 0} \varphi(z) = f(x).$$

$\square$

To compute the derivative $Df(x)$ we must determine the individual entries of this matrix. How this can be done in general will be seen in the following section. For a special case, however, we can compute $Df(x)$ directly using Definition 8.1.

Example 8.3 Let $f(x) = Ax$ be a linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ with a matrix $A \in \mathbb{R}^{m\times n}$. Then $Df(x) = A$ for all $x \in \mathbb{R}^n$, because by linearity we have

$$f(x+z) = A(x+z) = Ax + Az = f(x) + Az,$$

so that the equation from Definition 8.1 is satisfied with $\varphi \equiv 0$. The derivative of a linear mapping is therefore simply the matrix of the mapping itself. $\square$

Question 8.2 *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be an affine mapping, i.e. $f(x) = Ax + b$ with $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. What is $Df(x)$? Does it depend on b?*

The question now arises how one can compute the Jacobian matrix A. For this purpose, partial derivatives are useful.

8.2 Partial Derivatives

If we consider the component functions $f_i(x)$ of a vector-valued function, we can write them explicitly as

$$f_i(x_1, x_2, \ldots, x_n).$$

Here the x_j are not regarded as components of a vector but as n real variables. This is merely a different notation; the function itself of course does not change.

We now choose an arbitrary index $j \in \{1, 2, \ldots, n\}$ and consider only x_j as the function argument. This means that we fix values $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n$, each from $\mathbb{R}$, and consider the mapping

$$g_{ij} : x_j \mapsto f_i(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n) \tag{8.3}$$

from $\mathbb{R}$ to $\mathbb{R}$ (or from a suitably chosen domain $D \subset \mathbb{R}$ to $\mathbb{R}$). Since this is a real-valued function in the usual sense, we can apply the standard notion of differentiability on $\mathbb{R}$, which leads to the following definition.

Definition 8.4 Let $f : D \to \mathbb{R}^m$ with $D \subset \mathbb{R}^n$ be given.

(i) If the function $g_{ij} : \mathbb{R} \to \mathbb{R}$ defined according to (8.3) for fixed $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n$ is defined and (in the usual sense on $\mathbb{R}$) differentiable at a point $x_j \in \mathbb{R}$, then

$$\frac{\partial f_i}{\partial x_j}(x) := g'_{ij}(x_j)$$

is called the j-th *partial derivative*[2] of the i-th component function at

$$x = (x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n)^T \in \mathbb{R}^n.$$

(ii) If at a point $x \in D$ all partial derivatives $\frac{\partial f_i}{\partial x_j}(x)$, $i = 1, \ldots, m$, $j = 1, \ldots, n$, exist, then f is called *partially differentiable* at x.

(iii) If f is partially differentiable at all $x \in D$, then f is called *partially differentiable* on D. In this case the partial derivatives $\frac{\partial f_i}{\partial x_j}$ are functions from $\mathbb{R}^n$ to $\mathbb{R}$. □

[2] This can also be written as

$$\frac{\partial}{\partial x_j} f_i(x).$$

Example 8.5 We consider the functions from the introductory section of Chapter 5.

(i) $f : \mathbb{R}^2 \to \mathbb{R}$ with $f(x) = x_1^2 + 2x_2^2$. This function has only one component function $f_1 = f$. We have

$$g_{11} : x_1 \mapsto x_1^2 + 2x_2^2 \quad \text{and} \quad g_{12} : x_2 \mapsto x_1^2 + 2x_2^2$$

and therefore

$$\frac{\partial f_1}{\partial x_1}(x) = 2x_1 \quad \text{and} \quad \frac{\partial f_1}{\partial x_2}(x) = 4x_2.$$

(ii) $f : \mathbb{R} \to \mathbb{R}^2$ with

$$f(x) = \begin{pmatrix} \sin x \\ \cos x \end{pmatrix}.$$

This function has two component functions $f_1(x) = \sin x$ and $f_2(x) = \cos x$. The argument is one-dimensional, so $x = (x_1)$, and we have $g_{11}(x_1) = \sin x_1$ and $g_{21}(x_1) = \cos x_1$. Hence

$$\frac{\partial f_1}{\partial x_1}(x) = \cos x \quad \text{and} \quad \frac{\partial f_2}{\partial x_1}(x) = -\sin x.$$

(iii) $f : \mathbb{R}^3 \to \mathbb{R}^2$ with

$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \begin{pmatrix} 2(x_1x_2 + x_1x_3 + x_2x_3) \\ x_1x_2x_3 \end{pmatrix}.$$

Here we have

$$\frac{\partial f_1}{\partial x_1}(x) = 2x_2 + 2x_3, \quad \frac{\partial f_1}{\partial x_2}(x) = 2x_1 + 2x_3, \quad \frac{\partial f_1}{\partial x_3}(x) = 2x_1 + 2x_2$$

and

$$\frac{\partial f_2}{\partial x_1}(x) = x_2x_3, \quad \frac{\partial f_2}{\partial x_2}(x) = x_1x_3, \quad \frac{\partial f_2}{\partial x_3}(x) = x_1x_2.$$

□

Remark 8.6 Note: the x_j in the "denominator" of $\frac{\partial f_i}{\partial x_j}(x)$ is merely a symbolic placeholder indicating with respect to which variable the derivative is taken. The $x = (x_1, \ldots, x_n)^T$ in the argument, however, indicates the point at which the derivative is evaluated. Although the same symbols x_i are used here, they have very different meanings: one can assign numbers or vectors to the argument x (for example writing $\frac{\partial f_2}{\partial x_2}((5,7,9)^T)$), but not to the symbol x_j in the partial derivative. □

The relationship between partial and total differentiability is clarified by the following theorem.

Theorem 8.7 Let $f : D \to \mathbb{R}^m$ with $D \subset \mathbb{R}^n$ and $x \in D$ be given. Then:

(i) If f is totally differentiable at x, then f is also partially differentiable at x and we have

$$Df(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \cdots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}. \tag{8.4}$$

(ii) If f is partially differentiable in a neighborhood $N \subset D$ of x and all partial derivatives are continuous at the point x, then f is also totally differentiable at x (which again implies (8.4)).

Proof: (i) Suppose f is totally differentiable at x. We write the entries of the matrix $A = Df(x)$ as a_{ij} and denote the i-th row of A by $a_{i\cdot}$.

Applying the definition of total differentiability with $z = he_j$ yields that for all sufficiently small $h \in \mathbb{R}$ with $h \neq 0$ the equation

$$f(x + he_j) = f(x) + Ahe_j + \varphi(he_j)$$

holds with $\varphi(he_j)/h \to 0$ as $h \to 0$.

Considering the i-th row of this vector-valued equation, we obtain

$$f_i(x + he_j) = f_i(x) + a_{i\cdot}he_j + \varphi_i(he_j).$$

From the definition of e_j it follows that

$$a_{i\cdot}he_j = ha_{ij} \quad \text{and} \quad f_i(x + he_j) = g_{ij}(x_j + h)$$

(with g_{ij} from (8.3)). Hence the equation can also be written, with $\tilde{\varphi}_i(h) := \varphi_i(he_j)$, as

$$g_{ij}(x_j + h) = g_{ij}(x_j) + a_{ij}h + \tilde{\varphi}_i(h).$$

Since $\tilde{\varphi}_i(h)/h \to 0$ as $h \to 0$, it follows from [7, Theorem 16.4] that g_{ij} is differentiable with $g'_{ij}(x_j) = a_{ij}$. Thus the partial derivative $\partial f_i/\partial x_j$ exists at x and

$$\frac{\partial f_i}{\partial x_j}(x) = g'_{ij}(x_j) = a_{ij}.$$

(ii) We verify Definition 8.1(i) using the ∞-norm[3].

Since N is a neighborhood of x, there exists $\varepsilon > 0$ such that $U := B_\varepsilon(x) \subset N$. Let $z \in V := B_\varepsilon(0)$ be arbitrary. Using the components z_k of z, define

$$v^{(j)} := x + \sum_{k=1}^{j} z_k e_k, \quad \text{for } j = 0, \dots, n.$$

Because U is a ball and $x + z \in U$, it follows that $v^{(j)} \in U$ for all $j = 0, \dots, n$. From the definition we obtain $v^{(0)} = x$ and $v^{(n)} = x + z$, and

$$f_i(v^{(j)}) - f_i(v^{(j-1)}) = g_{ij}(x_j + z_j) - g_{ij}(x_j).$$

By the mean value theorem of differential calculus in $\mathbb{R}$, there exists for all i, j some $\xi_{ij} \in (0, z_j)$ such that

$$f_i(v^{(j)}) - f_i(v^{(j-1)}) = g_{ij}(x_j + z_j) - g_{ij}(x_j) = g'_{ij}(x_j + \xi_{ij})z_j = \frac{\partial f_i}{\partial x_j}(y^{(ij)})z_j$$

[3] Since all norms on $\mathbb{R}^n$ are equivalent, we can choose any norm for the proof.

with $y^{(ij)} = v^{(j-1)} + \xi_{ij} e_j$.

This implies

$$f_i(x+z) - f_i(x) = \sum_{j=1}^{n} \left(f_i(v^{(j)}) - f_i(v^{(j-1)}) \right) = \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(y^{(ij)}) z_j.$$

Setting $a_{ij} = \frac{\partial f_i}{\partial x_j}(x)$ gives

$$f(x+z) - f(x) = Az + \varphi(z)$$

with $\varphi(z) = (\varphi_1(z), \ldots, \varphi_m(z))^T$ and

$$\varphi_i(z) = \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(y^{(ij)}) z_j - a_{i\cdot} z = \sum_{j=1}^{n} \left(\frac{\partial f_i}{\partial x_j}(y^{(ij)}) - \frac{\partial f_i}{\partial x_j}(x) \right) z_j.$$

Total differentiability now follows if we can show that

$$\frac{\varphi(z)}{\|z\|_\infty} \to 0$$

as $\|z\|_\infty \to 0$ for all $i = 1, \ldots, n$. It suffices to show that $|\varphi_i(z)|/\|z\|_\infty \to 0$ for all i.

We have

$$|\varphi_i(z)| \leq \sum_{j=1}^{n} \left| \frac{\partial f_i}{\partial x_j}(y^{(ij)}) - \frac{\partial f_i}{\partial x_j}(x) \right| |z_j| \leq \sum_{j=1}^{n} \left| \frac{\partial f_i}{\partial x_j}(y^{(ij)}) - \frac{\partial f_i}{\partial x_j}(x) \right| \|z\|_\infty.$$

Hence

$$\frac{|\varphi_i(z)|}{\|z\|_\infty} \leq \sum_{j=1}^{n} \left| \frac{\partial f_i}{\partial x_j}(y^{(ij)}) - \frac{\partial f_i}{\partial x_j}(x) \right|.$$

As $\|z\|_\infty \to 0$, all components of z converge to 0, which implies $v^{(j)} \to x$, $\xi_{ij} \to 0$, and therefore $y^{(ij)} \to x$. Since the partial derivatives are assumed to be continuous, it follows that

$$\left| \frac{\partial f_i}{\partial x_j}(y^{(ij)}) - \frac{\partial f_i}{\partial x_j}(x) \right| \to 0$$

for all $i = 1, \ldots, m$, $j = 1, \ldots, n$, and hence the desired convergence holds. □

Question 8.3 *The partial derivatives of f at x only capture the behaviour of f along the n coordinate directions. Why does total differentiability of f at x nevertheless allow one to approximate f near x in every direction $z \in \mathbb{R}^n$? Why is the continuity of the partial derivatives important?*

Example 8.8 Since the partial derivatives of all functions from Example 8.5 are continuous, the condition of Theorem 8.7(i) is satisfied and we obtain

(i) $Df(x) = \begin{pmatrix} 2x_1 & 4x_2 \end{pmatrix}$

(ii) $Df(x) = \begin{pmatrix} \cos x_1 \\ -\sin x_1 \end{pmatrix}$

(iii) $Df(x) = \begin{pmatrix} 2x_2 + 2x_3 & 2x_1 + 2x_3 & 2x_1 + 2x_2 \\ x_2 x_3 & x_1 x_3 & x_1 x_2 \end{pmatrix}.$

□

If one compares conditions (i) and (ii) in Theorem 8.7, one observes that (ii) requires more than (i), namely the continuity of the partial derivatives at x. Thus Theorem 8.7 can be summarized by the implications

continuously partially differentiable ⇒ (totally) differentiable ⇒ partially differentiable.

Note that the partial derivatives are continuous if and only if the total derivative is continuous (by Theorem 8.7). This is the reason for the following definition.

Definition 8.9 A function $f : D \to \mathbb{R}^m$, $D \subset \mathbb{R}^n$ open, that is partially differentiable on all of D and whose partial derivatives are continuous is called *continuously differentiable* . The set of such functions is denoted by $C^1(D, \mathbb{R}^m)$. □

In addition to the total derivative and the partial derivatives, there are also various "mixed forms" that we want to briefly consider at the end of this section. For this purpose we consider two cases for functions from $C^1(D, \mathbb{R}^m)$.

In the first case we consider the Jacobian matrices $Df_i(x)$ of the components f_i of a function $f = (f_1, \dots, f_m)^T$. Since f_i maps into $\mathbb{R}$, the matrices $Df_i(x)$ are single-row matrices:

$$Df_i(x) = \left(\frac{\partial f_i}{\partial x_1}(x), \dots, \frac{\partial f_i}{\partial x_n}(x) \right).$$

For the Jacobian matrix $Df(x)$ of the full function we then have

$$Df(x) = \begin{pmatrix} Df_1(x) \\ \vdots \\ Df_m(x) \end{pmatrix}.$$

The second case we consider concerns functions $f : \mathbb{R}^s \to \mathbb{R}^m$ whose vector-valued argument splits into two (or more) vectors. For example, in the case of two vectors we write the function in the form $f(y, z)$, with $y \in \mathbb{R}^p$ and $z \in \mathbb{R}^q$ and $p + q = s$.

An example of this is the Euclidean inner product

$$f(y, z) = \langle y, z \rangle = \sum_{i=1}^{n} y_i z_i,$$

which defines a mapping from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}$ (here $p = q = n$ and $s = 2n$).

To differentiate such functions, one possibility is to combine the individual vectors in the argument into a single vector

$$x = \begin{pmatrix} y \\ z \end{pmatrix} \in \mathbb{R}^s,$$

write the function in the form $f(x)$, and define the derivative in the usual way.

The inner product can then be written as

$$f(x) = \sum_{i=1}^{n} x_i x_{n+i},$$

and the partial derivatives can be computed as

$$\frac{\partial f}{\partial x_j}(x) = \begin{cases} x_{j+n}, & j = 1, \dots, n, \\ x_{j-n}, & j = n+1, \dots, 2n. \end{cases}$$

From this we obtain the Jacobian matrix

$$Df(x) = \left(x_{n+1} \ x_{n+2} \ \dots \ x_{2n} \ x_1 \ x_2 \ \dots \ x_n \right).$$

Alternatively, one can keep one argument fixed and consider the functions

$$y \mapsto f(y,z) \qquad \text{and} \qquad z \mapsto f(y,z),$$

which are then mappings from $\mathbb{R}^p$ and $\mathbb{R}^q$ to $\mathbb{R}^m$, respectively, and differentiate them separately.

The resulting derivatives are denoted by

$$\frac{d}{dy} f(y,z) \qquad \text{and} \qquad \frac{d}{dz} f(y,z),$$

or alternatively by

$$D_y f(y,z) \qquad \text{and} \qquad D_z f(y,z).$$

These are matrices in $\mathbb{R}^{m \times p}$ and $\mathbb{R}^{m \times q}$, respectively.

For the example of the inner product we obtain

$$D_y \langle y, z \rangle = (z_1, \dots, z_n) = z^T \qquad \text{and} \qquad D_z \langle y, z \rangle = (y_1, \dots, y_n) = y^T.$$

If we compare the partial derivatives in $D_y f(y,z)$ and $D_z f(y,z)$ with those in the "full" Jacobian matrix $Df(x)$, we obtain with

$$x = \begin{pmatrix} y \\ z \end{pmatrix}$$

the identity

$$Df(x) = (D_y f(y,z),\ D_z f(y,z)). \tag{8.5}$$

Thus it does not matter which approach one uses; in the end—when assembled correctly—the same result is obtained.

In the "extreme case" one can decompose a vector-valued argument x into its n individual entries $x_1, \dots, x_n \in \mathbb{R}$ and obtain

$$Df(x) = (D_{x_1} f(x),\ \dots,\ D_{x_n} f(x))$$

with

$$D_{x_j} f(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_j}(x) \\ \vdots \\ \frac{\partial f_m}{\partial x_j}(x) \end{pmatrix}.$$

8.3 The Directional Derivative

Fix $x \in \mathbb{R}^n$ and $v \in \mathbb{R}^n$ with $v \neq 0$. The mapping $t \mapsto x + tv$, $t \in \mathbb{R}$, defines a straight line in $\mathbb{R}^n$. Now consider a function $f : \mathbb{R}^n \to \mathbb{R}^m$. Then

$$h : t \mapsto f(x + tv) \tag{8.6}$$

defines a function from $\mathbb{R}$ to $\mathbb{R}^m$ which gives precisely the values of the function f along the line $x + tv$, $t \in \mathbb{R}$, and for which $h(0) = f(x)$ holds.

Example 8.10 Consider the function

$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \begin{pmatrix} 2(x_1x_2 + x_1x_3 + x_2x_3) \\ x_1x_2x_3 \end{pmatrix}.$$

Let $x = (1, 1, 1)^T$ and $v = (1, 1, 0)$. Then we obtain

$$h(t) = \begin{pmatrix} 2((1+t)(1+t) + (1+t) + (1+t)1) \\ (1+t)(1+t) \end{pmatrix}.$$

Thus the function describes how the surface area and the volume of a cuboid change when we keep the height fixed at $x_3 = 1$ and choose the length and width as $x_1 = x_2 = 1 + t$. □

The derivative $Dh(0) \in \mathbb{R}^m$ of h at $t = 0$ is a vector whose i-th component gives the slope of the component function f_i at the point x in the direction of the vector v. This vector is called the *directional derivative* of f at the point x in the direction v. The following theorem shows how the directional derivative can be computed easily using the Jacobian matrix $Df(x)$.

Question 8.4 *What happens with the directional derivative $Dh(0)$ if v is replaced by $2v$ or by $-v$?*

Theorem 8.11 Let $f : D \to \mathbb{R}^m$ with $D \subset \mathbb{R}^n$ open, differentiable at $x \in D$. Let $v \in \mathbb{R}^n$ with $v \neq 0$ be given. Then the function h from (8.6) is differentiable at $t = 0$ and

$$Dh(0) = Df(x)v.$$

Proof: From the differentiability of f we obtain

$$f(x + tv) = f(x) + Df(x)tv + \varphi(tv).$$

Hence

$$h(t) = h(0) + (Df(x)v)t + \psi(t)$$

with $\psi(t) := \varphi(tv)$. Since

$$\frac{\psi(t)}{t} = \frac{\varphi(tv)}{t} = \frac{\|v\|}{\|v\|}\frac{\varphi(tv)}{t} = \|v\|\frac{\varphi(tv)}{\|tv\|} \to 0$$

for $t \to 0$, it follows that h is differentiable at 0 and $Dh(0) = Df(x)v$. ∎

In the following we will always write directional derivatives in the form $Df(x)v$. The auxiliary function h used here is merely a technical construction for the definition, although it will appear again in some arguments later on.

Note that the magnitude of the directional derivative depends on the length of the vector v. If one wants to compare directional derivatives in different directions v and w, one should therefore use vectors with $\|v\| = \|w\|$. Usually it is convenient to choose $\|v\| = \|w\| = 1$.

Example 8.12 (i) For the data from Example 8.10 we have

$$Df(x)v = \begin{pmatrix} 2x_2+2x_3 & 2x_1+2x_3 & 2x_1+2x_2 \\ x_2x_3 & x_1x_3 & x_1x_2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2x_1+2x_2+4x_3 \\ (x_1+x_2)x_3 \end{pmatrix} = \begin{pmatrix} 8 \\ 2 \end{pmatrix}.$$

Thus the slope of the surface area in the direction of the vector v is 8, and the slope of the volume is 2.

(ii) If we set $v = e_j$ for the standard basis vector e_j, we obtain the directional derivative

$$Df(x)e_j = \begin{pmatrix} \frac{\partial f_1}{\partial x_j}(x) \\ \vdots \\ \frac{\partial f_m}{\partial x_j}(x) \end{pmatrix} = D_{x_j}f(x),$$

which is precisely the vector of partial derivatives with respect to x_j. □

With the help of the directional derivative one can also explain why the (total) derivative in $\mathbb{R}^n$ is not defined via the difference quotient. To see this, consider the special case of a function $f : \mathbb{R}^n \to \mathbb{R}$. A formal generalization of this definition would be

$$\lim_{\substack{z\to 0\\ z\neq 0,\, x+z\in D}} \frac{f(x+z)-f(x)}{\|z\|},$$

where z is now a vector in $\mathbb{R}^n$. The limit $z \to 0$ is understood as usual as the limit over all possible sequences $z^{(k)} \to 0$, and requires that the limits for all such sequences coincide.

For every vector $v \in \mathbb{R}^n$ with $v \neq 0$ and every real sequence $t_k \to 0$, the sequence $z^{(k)} := t_kv$ is such a sequence. For this choice one can rewrite the difference quotient using the function h from (8.6), and one sees that the limit of the difference quotient is precisely $Df(x)v$. Depending on how one chooses v (or more generally the sequence $z^{(k)}$), one obtains different limits. This is the reason why this expression cannot be used to define the (total) derivative. However, it can of course be used to define partial derivatives or (componentwise) directional derivatives.

8.4 The Gradient

In this section we consider the special case of real-valued functions, i.e., functions $f : D \to \mathbb{R}$ with $D \subset \mathbb{R}^n$. For such functions, the derivative $Df(x)$ is a one-row matrix, which we can also interpret as a row vector. By transposing this vector, we obtain a column vector, called the gradient.

Definition 8.13 For a function $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$, differentiable at $x \in D$, the column vector[4]

$$\nabla f(x) := Df(x)^T$$

is called the *gradient* of f at x. The symbol "∇" is also called "nabla". □

From this definition (together with the definition of the Euclidean inner product), the directional derivative can be written using the gradient as

$$Df(x)v = \langle \nabla f(x),\, v\rangle.$$

The geometric interpretation of the gradient is that it points in $\mathbb{R}^n$ in the direction of the steepest increase of the function f. Formally, this is expressed in the following theorem.[5]

Theorem 8.14 Let $D \subset \mathbb{R}^n$ be open and $f : D \to \mathbb{R}$ differentiable at $x \in D$ with $Df(x) \neq 0$. Then, for $v^* = \nabla f(x)/\|\nabla f(x)\|_2$ and all $v \in \mathbb{R}^n$ with $\|v\|_2 = 1$, the inequality

$$Df(x)v^* \geq Df(x)v$$

holds.

Proof: Using $Df(x)v = \langle \nabla f(x),\, v\rangle$ and the Cauchy-Schwarz inequality, we obtain

$$Df(x)v = \langle \nabla f(x), v\rangle \leq \|\nabla f(x)\|_2 \|v\|_2 = \|\nabla f(x)\|_2.$$

On the other hand, for v^*:

$$Df(x)v^* = \langle \nabla f(x), v^*\rangle = \frac{1}{\|\nabla f(x)\|_2}\langle \nabla f(x), \nabla f(x)\rangle = \|\nabla f(x)\|_2.$$

□

Gradient ascent and gradient descent

The definition of the gradient implies that f increases as fast as possible in the direction of $\nabla f(x)$. Consequently, it also decreases as fast as possible in the direction of $-\nabla f(x)$. This gives rise to the so-called *gradient ascent* and *gradient descent* algorithms. In both algorithms one picks an initial point $x^{(0)} \in D$ and a step size $\alpha > 0$. In gradient ascent one then computes

$$x^{(k+1)} := x^{(k)} + \alpha \nabla f(x^{(k)})$$

[4] Some books define the gradient as a row vector.

[5] While we have so far used an arbitrary norm in $\mathbb{R}^n$, the following theorem is specific to the 2-norm, since this norm naturally corresponds to the inner product via $\langle x, x\rangle = \|x\|_2^2$.

and in gradient descent one computes

$$x^{(k+1)} := x^{(k)} - \alpha \nabla f(x^{(k)}).$$

In the following, we consider the gradient descent algorithm. Under appropriate conditions and for small enough α, the fact that f decreases in the direction of $-\nabla f(x)$, it follows that

$$f(x^{(k+1)}) \leq f(x^{(k)})$$

and in many cases $f(x^{(k)})$ converges to a minimum of f as $k \to \infty$ and $x^{(k)}$ converges to the corresponding minimizer.

Example 8.15 Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x) = (x_1 - 1)^2 + (x_2 - 2)^2.$$

Clearly, $f(x) \geq 0$ for all $x \in \mathbb{R}^2$ and $f(x) = 0$ if and only if $x = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. We perform three steps of the gradient descent algorithm for $x^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ with $\alpha = 1/4$ and $f(x^{(0)}) = 5$. To this end, observe that

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \end{pmatrix} = \begin{pmatrix} 2(x_1 - 1) \\ 2(x_2 - 2) \end{pmatrix}.$$

Hence the steps of the gradient descent algorithm become

$$x^{(1)} = x^{(0)} - \alpha \nabla f(x^{(0)}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \frac{1}{4} \begin{pmatrix} 2(0-1) \\ 2(0-2) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix} \quad \rightsquigarrow \quad f(x^{(1)}) = \frac{5}{4}$$

$$x^{(2)} = x^{(1)} - \alpha \nabla f(x^{(1)}) = \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix} - \frac{1}{4} \begin{pmatrix} 2(\frac{1}{2}-1) \\ 2(1-2) \end{pmatrix} = \begin{pmatrix} \frac{3}{4} \\ \frac{3}{2} \end{pmatrix} \quad \rightsquigarrow \quad f(x^{(2)}) = \frac{5}{16}$$

$$x^{(3)} = x^{(2)} - \alpha \nabla f(x^{(2)}) = \begin{pmatrix} \frac{3}{4} \\ \frac{3}{2} \end{pmatrix} - \frac{1}{4} \begin{pmatrix} 2(\frac{3}{4}-1) \\ 2(\frac{3}{2}-2) \end{pmatrix} = \begin{pmatrix} \frac{7}{8} \\ \frac{7}{4} \end{pmatrix} \quad \rightsquigarrow \quad f(x^{(3)}) = \frac{5}{64}$$

Continuing these computations one sees that $f(x^{(k)}) = \frac{5}{4^k}$ converges to 0 and $x^{(k)}$ converges to $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$. □

Question 8.5 *What happens in the example if you use larger values of α? Try values up to $\alpha = 3$.*

The gradients descent algorithm is very important for the training of neural networks, however, here the gradient is taken with respect to to the parameters of the network.

Recall from (5.3)–(5.3) that the components f_i of the map f represented by a neural network are given by

$$y_i^{(1)} = \sigma\left(\sum_{j=1}^{n} w_{i,j}^{(0)} x_j + b_i^{(0)}\right)$$
$$y_i^{(l+1)} = \sigma\left(\sum_{j=1}^{n_l} w_{i,j}^{(l)} y_j^{(l)} + b_i^{(l)}\right)$$
$$f_i(x) = \sum_{j=1}^{n_L} w_{i,j}^{(L)} y_j^{(L)} + b_i^{(L)}.$$

Let us write all parameters $w_{i,j}^{(l)}$ and $b_i^{(l)}$ into a large vector $\theta \in \mathbb{R}^M$ and let us write the function realized by the neural network as $f(x, \theta)$. Now suppose we have a lot of data pairs $(\tilde{x}_p, \tilde{y}_p)$, $p = 1, \ldots, N$, (so-called *training data*) and we would like to find a parameter vector $\theta^\star \in \mathbb{R}^M$ such that the network function satisfies

$$f(\tilde{x}_p, \theta^*) \approx \tilde{y}_p$$

as good as possible. Then we can define the so-called loss function

$$\ell(\theta) = \frac{1}{N} \sum_{p=1}^{N} \|f(\tilde{x}_p, \theta) - \tilde{y}_p\|_2^2$$

and try to find $\theta^\star$ minimizing this loss function. For this, we can use the gradient descent algorithm

$$\theta^{(k+1)} := \theta^{(k)} - \alpha \nabla \ell(\theta^{(k)}). \tag{8.7}$$

While in practice several modifications are used to make this algorithm faster and more reliable, the gradient descent idea is still the key ingredient of most neural network training algorithms until today.

8.5 Derivative Rules in $\mathbb{R}^n$

The rules for derivatives in $\mathbb{R}^n$ follow as direct generalizations of the rules in $\mathbb{R}$. We state them here informally without proof. Assume open domains $D \subset \mathbb{R}^n$ and differentiability where indicated.

Scalar multiplication: $f : D \to \mathbb{R}^m$, $\lambda \in \mathbb{R}$:

$$D(\lambda f)(x) = \lambda Df(x)$$

Sum rule: $f, g : D \to \mathbb{R}^m$:

$$D(f + g)(x) = Df(x) + Dg(x)$$

Product rule: $f, g : D \to \mathbb{R}$ (note: to multiply, the functions must be real-valued):

$$D(fg)(x) = f(x)Dg(x) + g(x)Df(x)$$

(The multiplications on the right-hand side are understood as "scalar times vector".)

Question 8.6 *Why can't we write the product rule in the familiar form $D(fg)(x) = Df(x)g(x) + f(x)Dg(x)$ from the last semester?*

Quotient rule: $f, g : D \to \mathbb{R}$, $g(x) \neq 0$:

$$D\left(\frac{f}{g}\right)(x) = \frac{g(x)Df(x) - f(x)Dg(x)}{(g(x))^2}$$

Chain rule: $f : D \to \mathbb{R}^m$, $g : E \to \mathbb{R}^n$, $E \subset \mathbb{R}^p$, $g(E) \subset D$:

$$D(f \circ g)(x) = Df(g(x))Dg(x).$$

In the last case it is useful to check the dimensions: $Df(y) \in \mathbb{R}^{m\times n}$ and $Dg(x) \in \mathbb{R}^{n\times p}$, so the product is well-defined and lies in $\mathbb{R}^{m\times p}$, consistent with $f \circ g : E \to \mathbb{R}^m$.

We previously considered functions of the form $f(y, z)$ with $f : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^m$. If another function $g : \mathbb{R}^p \to \mathbb{R}^q$ is given, we can form the "partial composition"

$$h(y) = f(y, g(y)).$$

To apply the chain rule to this form, we write $x = \begin{pmatrix} y \\ z \end{pmatrix}$ and consider f as $f(x)$, for which $Df(x) = (D_y f(y, z),\ D_z f(y, z))$. Then the composition can be written as $f \circ G(y)$ with

$$G(y) = \begin{pmatrix} y \\ g(y) \end{pmatrix}.$$

Its derivative is

$$DG(y) = \begin{pmatrix} \mathrm{Id} \\ Dg(y) \end{pmatrix},$$

where Id is the $p \times p$ identity matrix. The chain rule then yields

$$\begin{aligned} Dh(y) = Df(G(y))DG(y) &= \Big(D_y f(y, g(y)),\ D_z f(y, g(y))\Big) \begin{pmatrix} \mathrm{Id} \\ Dg(y) \end{pmatrix} \\ &= D_y f(y, g(y)) + D_z f(y, g(y))Dg(y). \end{aligned} \tag{8.8}$$

Example 8.16 We compute the derivative of the squared 2-norm $y \mapsto \|y\|_2^2$. This can be written using the inner product $f(y, z) = \langle y, z\rangle$ and $g(y) = y = \mathrm{Id}y$ as $\|y\|_2^2 = \langle y, g(y)\rangle = f(y, g(y))$. Since $D_y f(y, z) = z^T$ and $D_z f(y, z) = y^T$, as well as $Dg(y) = \mathrm{Id}$, we get

$$D(\|y\|_2^2) = D_y f(y, g(y)) + D_z f(y, g(y))Dg(y) = g(y)^T + y^T \mathrm{Id} = y^T + y^T = 2y^T.$$

□

Example 8.17 When we want to apply the gradient descent algorithm (8.7) for training the neural network, we need to compute $\nabla \ell(\theta) = D\ell(\theta)^T$. We explain how to do this for a neural network with only one layer $L = 1$ and one-dimensional input x and output $f(x)$. In this case, we can write $f = f_1$ in one formula

$$f(x,\theta) = \sum_{j=1}^{n_1} w_{1,j}^{(1)} \sigma\left(w_{j,1}^{(0)} x + b_j^{(0)}\right) + b^{(1)}$$

and ℓ becomes

$$\ell(\theta) = \frac{1}{N} \sum_{p=1}^{N} (f(\tilde{x}_p,\theta) - \tilde{y}_p)^2.$$

The sum rule and the chain rule give us

$$D\ell(\theta) = \frac{1}{N} \sum_{p=1}^{N} 2(f(\tilde{x}_p,\theta) - \tilde{y}_p) D_\theta f(\tilde{x}_p,\theta),$$

where the entries of $D_\theta f(\tilde{x}_p,\theta)$ are of the form

$$\begin{aligned}
\frac{\partial f}{\partial w_{1,j}^{(1)}}(\tilde{x}_p,\theta) &= \sigma\left(w_{j,1}^{(0)} \tilde{x}_p + b_j^{(0)}\right), \\
\frac{\partial f}{\partial w_{j,1}^{(0)}}(\tilde{x}_p,\theta) &= w_{1,j}^{(1)} \sigma'\left(w_{j,1}^{(0)} \tilde{x}_p + b_j^{(0)}\right) \tilde{x}_p, \\
\frac{\partial f}{\partial b_j^{(0)}}(\tilde{x}_p,\theta) &= w_{1,j}^{(1)} \sigma'\left(w_{j,1}^{(0)} \tilde{x}_p + b_j^{(0)}\right), \\
\frac{\partial f}{\partial b^{(1)}}(\tilde{x}_p,\theta) &= 1.
\end{aligned}$$

If there is more than one layer, then one can apply similar formulas to compute the derivatives of the values $y_j^{(l)}$ recursively using the chain rule. Using that during this computation the arguments of σ and σ' are the same as when evaluating $f(\tilde{x}_p,\theta)$, this leads to a very efficient way of computing $\nabla\ell$: One first evaluates f and stores all arguments of the activation functions σ during the evaluation. Then one recursively evaluates the formulas needed for computing the derivatives. This method is called back-propagation and is one reason for the success of neural networks in machine learning, because for other classes of functions the evaluation of the gradients requires much more complicated computations. □

8.6 Higher Partial Derivatives

Question 8.7 *Does the partial derivative $\frac{\partial f_i}{\partial x_j}(x)$ depend on x_k if $k \neq i$ and $k \neq j$?*

A partial derivative $\frac{\partial f_i}{\partial x_j}$ is itself a function from $\mathbb{R}^n$ to $\mathbb{R}$. As such, it can—just as for functions in $\mathbb{R}$—be differentiated again if it is itself differentiable. Formally, this is made precise by the following definition.

Definition 8.18 Let $f : D \to \mathbb{R}^m$, $D \subset \mathbb{R}^n$ open. For given indices i and $j_1, j_2, \ldots \in \{1, \ldots, n\}$ and $k \geq 2$, the *higher partial derivatives of order k*

$$\frac{\partial^k f_i}{\partial x_{j_1} \ldots \partial x_{j_k}}$$

are defined for all $x \in D$ inductively as

$$\frac{\partial^k f_i}{\partial x_{j_1} \ldots \partial x_{j_k}}(x) := \frac{\partial}{\partial x_{j_k}} \frac{\partial^{k-1} f_i}{\partial x_{j_1} \ldots \partial x_{j_{k-1}}}(x),$$

assuming that the indicated partial derivative on the right-hand side exists, i.e., that the function

$$x \mapsto \frac{\partial^{k-1} f_i}{\partial x_{j_1} \ldots \partial x_{j_{k-1}}}(x)$$

(from D to $\mathbb{R}$) is partially differentiable.

If all higher partial derivatives exist and are continuous for all $x \in D$ and all $k = 1, \ldots, p$ for some $p \in \mathbb{N}$, the function f is called p-times continuously differentiable. The set of these functions is denoted $C^p(D, \mathbb{R}^m)$. □

Note that for $k = 1$ we recover the usual first-order partial derivatives, which is why the definition starts at $k = 2$. In non-inductive form, higher partial derivatives can also be written as

$$\frac{\partial^k f_i}{\partial x_{j_1} \partial x_{j_2} \ldots \partial x_{j_k}}(x) = \frac{\partial}{\partial x_{j_k}} \frac{\partial}{\partial x_{j_{k-1}}} \cdots \frac{\partial}{\partial x_{j_2}} \frac{\partial}{\partial x_{j_1}} f_i(x).$$

The following theorem shows that the order of differentiation does not matter.

Theorem 8.19 Let π be a permutation of $\{1, \ldots, k\}$, i.e., a bijection of this set onto itself. Then for any function $f \in C^k(D, \mathbb{R}^m)$, $D \subset \mathbb{R}^n$ open, and all $x \in D$,

$$\frac{\partial^k f_i}{\partial x_{j_1} \ldots \partial x_{j_k}}(x) = \frac{\partial^k f_i}{\partial x_{j_{\pi(1)}} \ldots \partial x_{j_{\pi(k)}}}(x).$$

Proof: First, consider the case $k = 2$. Without loss of generality, let $n = 2$, $j_1 = 1$, $j_2 = 2$, and $x = 0$, and write $f(x_1, x_2)$ instead of $f_i(x)$. We need to show

$$\frac{\partial}{\partial x_2} \frac{\partial}{\partial x_1} f(0,0) = \frac{\partial}{\partial x_1} \frac{\partial}{\partial x_2} f(0,0).$$

Choose $\delta > 0$ such that $U = \{(x_1, x_2) \in \mathbb{R}^2 \mid |x_1| < \delta, |x_2| < \delta\} \subset D$. Define

$$F(x_1) := f(x_1, x_2) - f(x_1, 0),$$

which is differentiable with

$$F'(x_1) = \frac{\partial f}{\partial x_1}(x_1, x_2) - \frac{\partial f}{\partial x_1}(x_1, 0).$$

By the Mean Value Theorem, there exists $\xi_1 \in (-|x_1|, |x_1|)$ with

$$F(x_1) - F(0) = F'(\xi_1)x_1.$$

Now consider the real function

$$x_2 \mapsto \frac{\partial f}{\partial x_1}(\xi_1, x_2),$$

which is differentiable with derivative $\frac{\partial^2 f}{\partial x_2 \partial x_1}$. Applying the Mean Value Theorem again gives $\xi_2 \in (-|x_2|, |x_2|)$ such that

$$\frac{\partial f}{\partial x_1}(\xi_1, x_2) - \frac{\partial f}{\partial x_1}(\xi_1, 0) = \frac{\partial^2 f}{\partial x_2 \partial x_1}(\xi_1, \xi_2)x_2.$$

Combining these, we obtain

$$f(x_1, x_2) - f(x_1, 0) - f(0, x_2) + f(0, 0) = \frac{\partial^2 f}{\partial x_2 \partial x_1}(\xi_1, \xi_2)x_1x_2.$$

Repeating the same argument with $\tilde{F}(x_2) := f(x_1, x_2) - f(0, x_2)$ gives

$$f(x_1, x_2) - f(0, x_2) - f(x_1, 0) + f(0, 0) = \frac{\partial^2 f}{\partial x_1 \partial x_2}(\tilde{\xi}_1, \tilde{\xi}_2)x_1x_2.$$

Since the left-hand sides coincide, taking $x_1 \to 0$ and $x_2 \to 0$ and using continuity of the partial derivatives yields the equality.

For $k > 2$, any permutation can be written as a composition of adjacent transpositions, and the result follows inductively by the same argument. ∎

Example 8.20 Consider $f : \mathbb{R}^2 \to \mathbb{R}$ with $f(x) = \cos(x_1x_2)$. Then

$$\begin{aligned}
\frac{\partial^3 f}{\partial x_2 \partial x_1 \partial x_1} &= \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2}\cos(x_1x_2) \\
&= \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_1}(-\sin(x_1x_2)x_1) \\
&= \frac{\partial}{\partial x_1}(-\cos(x_1x_2)x_1x_2 - \sin(x_1x_2)) \\
&= \sin(x_1x_2)x_1x_2^2 - 2\cos(x_1x_2)x_2,
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^3 f}{\partial x_1 \partial x_1 \partial x_2} &= \frac{\partial}{\partial x_2}\frac{\partial}{\partial x_1}\frac{\partial}{\partial x_1}\cos(x_1x_2) \\
&= \frac{\partial}{\partial x_2}\frac{\partial}{\partial x_1}(-\sin(x_1x_2)x_2) \\
&= \frac{\partial}{\partial x_2}(-\cos(x_1x_2)x_2^2) \\
&= \sin(x_1x_2)x_1x_2^2 - 2\cos(x_1x_2)x_2.
\end{aligned}$$

As expected, the derivatives coincide. □

Remark 8.21 In general, for $k \geq 2$ partial derivatives cannot be arranged into a matrix, because they are determined by more than two indices (i and $j_1, \ldots, j_k$). An exception is the case $m = 1$ and $k = 2$, i.e., second derivatives of a real-valued function $f : D \to \mathbb{R}$. In this case, one defines the *Hessian matrix*

$$H_f(x) := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix}.$$

By Theorem 8.19, H_f is symmetric for $f \in C^2(D, \mathbb{R})$. □

Chapter 9

Applications of Differential Calculus in $\mathbb{R}^n$

Last update: July 21, 2026

9.1 Taylor Polynomials

In [7, Section 17.B] we have seen that for sufficiently often differentiable real functions $f : \mathbb{R} \to \mathbb{R}$, the Taylor polynomial

$$P(x) = \sum_{j=0}^{k} \frac{f^{(j)}(x_0)}{j!}(x - x_0)^j$$

satisfies the equation

$$f(x) = P(x) + \frac{(x - x_0)^{k+1}}{(k+1)!} f^{(k+1)}(\xi).$$

In this section we want to generalize this statement to functions on $\mathbb{R}^n$. We restrict ourselves to real-valued functions $f : D \to \mathbb{R}$ with $D \subset \mathbb{R}^n$. For functions with values in $\mathbb{R}^m$, a Taylor polynomial P_i can be constructed for each component function f_i, $i = 1, \ldots, m$, which we will not carry out explicitly here.

We have already encountered polynomials $P : \mathbb{R}^n \to \mathbb{R}$. Such a polynomial of degree k is given by

$$P(x) = \sum_{\substack{k_1, k_2, \ldots, k_n = 0 \\ k_1 + k_2 + \ldots + k_n \le k}}^{k} c_{k_1 k_2 \ldots k_n} x_1^{k_1} x_2^{k_2} \cdot \ldots \cdot x_n^{k_n}.$$

In order to write this expression—and several other expressions that we will need in the following—in a shorter form, we introduce the following multi-indices.

Definition 9.1 (i) A multi-index of length n is a row vector

$$\alpha = (\alpha_1, \ldots, \alpha_n)$$

with $\alpha_i \in \mathbb{N}$.

(ii) For a multi-index α we define

$$|\alpha| := \alpha_1 + \ldots + \alpha_n \quad \text{and} \quad \alpha! := \alpha_1! \cdot \ldots \cdot \alpha_n!.$$

(iii) For a vector $x \in \mathbb{R}^n$ we define

$$x^\alpha := x_1^{\alpha_1} \cdot \ldots \cdot x_n^{\alpha_n}.$$

(iv) For a multi-index α and a function $f \in C^{|\alpha|}(D, \mathbb{R})$ with $D \subset \mathbb{R}^n$ open we define

$$D^\alpha f := \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}},$$

with

$$\partial x_j^{\alpha_j} := \underbrace{\partial x_j \ldots \partial x_j}_{\alpha_j \text{ times}}.$$

□

A polynomial $P : \mathbb{R}^n \to \mathbb{R}$ of degree k can then be written compactly as

$$P(x) = \sum_{|\alpha| \le k} c_\alpha x^\alpha.$$

The sum is to be understood as the sum over all multi-indices of length n with $|\alpha| \le k$.

Question 9.1 *What are the coefficients c_α for the polynomial $P : \mathbb{R}^3 \to \mathbb{R}$ given by*

$$P(x) = x_1 + x_2 + x_1 x_2 + x_3^2 + x_1 x_2^2 + x_1 x_2 x_3?$$

To derive the Taylor polynomial we need the following lemma.

Lemma 9.2 Let $f \in C^k(D, \mathbb{R})$ with $D \subset \mathbb{R}^n$ open. Let $x \in D$ and $z \in \mathbb{R}^n$ be such that $x + tz \in D$ holds for all $t \in [0, 1]$. Then the function

$$g : [0, 1] \to \mathbb{R}, \quad g : t \mapsto f(x + tz)$$

lies in $C^k([0, 1], \mathbb{R})$ with

$$g^{(k)}(t) = \sum_{|\alpha| = k} \frac{k!}{\alpha!} D^\alpha f(x + tz) z^\alpha.$$

Proof (sketch): Using the chain rule, one first proves by induction over k the formula

$$g^{(k)}(t) = \sum_{i_1, \ldots, i_k = 1}^{n} \frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}} (x + tz) z_{i_1} \cdots z_{i_k} \tag{9.1}$$

(which we do not carry out in detail here; see Forster II, Chapter 7).

After that, in each summand in (9.1) and for each $m = 1, \ldots, n$, let α_m denote the number of indices j for which $i_j = m$. From Theorem 8.19 it then follows for $\alpha = (\alpha_1, \ldots, \alpha_n)$ that

$$\frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(x + tz) z_{i_1} \cdots z_{i_k} = D^\alpha f(x + tz) z^\alpha.$$

In the sum in (9.1) there are exactly $k!/\alpha!$ index tuples $i_1, \ldots, i_k$ in which the number m appears exactly α_m times (there are $k!$ arrangements in total, but for each $\alpha_m!$ only the number m is permuted with itself, which produces identical tuples for each m). Thus the expression $D^\alpha f(x + tz) z^\alpha$ appears exactly $k!/\alpha!$ times in the sum, which implies

$$\sum_{i_1, \ldots, i_k = 1}^{n} \frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(x + tz) z_{i_1} \cdots z_{i_k} = \sum_{|\alpha| = k} \frac{k!}{\alpha!} D^\alpha f(x + tz) z^\alpha$$

and thus proves the claim. ∎

Definition 9.3 For $f \in C^k(D, \mathbb{R}^m)$, $D \subset \mathbb{R}^n$ open and $x_0 \in D$, we define the Taylor polynomial of degree k of f at x_0 as

$$P(x) := \sum_{|\alpha| \leq k} \frac{D^\alpha f(x_0)}{\alpha!} (x - x_0)^\alpha.$$

□

Remark 9.4 By explicitly computing the partial derivatives contained in D^α, one verifies the identities

$$\sum_{|\alpha| = 0} \frac{D^\alpha f(x_0)}{\alpha!} (x - x_0)^\alpha = f(x_0), \quad \sum_{|\alpha| = 1} \frac{D^\alpha f(x_0)}{\alpha!} (x - x_0)^\alpha = Df(x_0)(x - x_0)$$

and

$$\sum_{|\alpha| = 2} \frac{D^\alpha f(x_0)}{\alpha!} (x - x_0)^\alpha = \frac{1}{2} (x - x_0)^T H_f(x_0)(x - x_0)$$

with the Hessian matrix $H_f(x)$ from Remark 8.21. Thus, for example, the Taylor polynomial of degree $k = 2$ can be written in the form

$$P(x) = f(x_0) + Df(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^T H_f(x_0)(x - x_0).$$

□

The following theorem states the Taylor formula in $\mathbb{R}^n$, i.e., the relationship between $P(x)$ and $f(x)$.

Theorem 9.5 Let $f \in C^{k+1}(D, \mathbb{R})$ with $D \subset \mathbb{R}^n$ open. Let $x_0, x \in D$ be such that $x_0 + tz \in D$ holds for $z := x - x_0$ and all $t \in [0, 1]$.

Then there exists a $\theta \in [0, 1]$ such that

$$f(x) = P(x) + \sum_{|\alpha| = k+1} \frac{D^\alpha f(x_0 + \theta z)}{\alpha!} (x - x_0)^\alpha.$$

Proof: We consider the function $g(t) := f(x_0 + tz)$. By Lemma 9.2 this is a function in $C^{k+1}([0,1],\mathbb{R})$, which implies that the real Taylor polynomial of g at $t_0 = 0$

$$P_g(t) := \sum_{j=0}^{k} \frac{g^{(j)}(0)}{j!} t^j$$

satisfies the equation

$$g(1) = P_g(1) + \frac{g^{(k+1)}(\theta)}{(k+1)!} \tag{9.2}$$

for some $\theta \in [0,1]$ (see [7, Theorem 17.9]).

From Lemma 9.2 we obtain

$$\frac{g^{(j)}(0)}{j!} = \sum_{|\alpha|=j} \frac{D^\alpha f(x_0)}{\alpha!} z^\alpha,$$

which implies that $P_g(1) = P(x)$. Again from Lemma 9.2 we obtain

$$\frac{g^{(k+1)}(\theta)}{(k+1)!} = \sum_{|\alpha|=k+1} \frac{D^\alpha f(x_0+\theta z)}{\alpha!} z^\alpha,$$

and therefore the claim follows from equation (9.2). ∎

Remark 9.6 Since $f \in C^{k+1}(D,\mathbb{R})$, the function

$$x \mapsto \sum_{|\alpha|=k+1} |D^\alpha f(x)|$$

is continuous. Choose $\delta > 0$ such that $\overline{B}_\delta(x_0) \subset D$. Then, by Theorem 7.13, the function attains its maximum M on the compact set $\overline{B}_\delta(x_0)$.

Moreover, for each component of $z = x - x_0$ we have the inequality $|z_i| \le \|x - x_0\|_\infty$, and therefore

$$|(x-x_0)^\alpha| = |z^\alpha| = |z_1^{\alpha_1}|\cdots|z_n^{\alpha_n}| \le \|x-x_0\|_\infty^{\alpha_1}\cdots\|x-x_0\|_\infty^{\alpha_n} = \|x-x_0\|_\infty^{|\alpha|}.$$

Question 9.2 *Find a vector $z \in \mathbb{R}^n$ with $z \neq 0$ and a multi-index α with $|z|^\alpha = \|z\|_\infty^{|\alpha|}$.*

Since for all $x \in \overline{B}_\delta(x_0)$ and $\theta \in [0,1]$ we also have $x_0 + \theta z \in \overline{B}_\delta(x_0)$, it follows from the triangle inequality that

$$\left| \sum_{|\alpha|=k+1} \frac{D^\alpha f(x_0+\theta z)}{\alpha!}(x-x_0)^\alpha \right| \le \sum_{|\alpha|=k+1} \left| \frac{D^\alpha f(x_0+\theta z)}{\alpha!} \right| \|x-x_0\|_\infty^{|\alpha|} \le M\|x-x_0\|_\infty^{k+1}.$$

Analogously to the real case (see [7, Remark 17.11]), we therefore obtain for all $x \in \overline{B}_\delta(x_0)$

$$f(x) - P(x) = O(\|x-x_0\|_\infty^{k+1}).$$

□

9.2 The Mean Value Theorem of Differential Calculus in $\mathbb{R}^n$

For real functions, the mean value theorem states that for a function that is continuous on $[a, b]$ and differentiable on (a, b) there exists a $\xi \in (a, b)$ such that

$$f(b) - f(a) = f'(\xi)(b - a).$$

For the generalization to $\mathbb{R}^n$ we rewrite this statement slightly. Let $x = a$ and $z = b - a$. Then there exists a value $\theta \in (0, 1)$ such that

$$f(x + z) - f(x) = f'(x + \theta z)z. \tag{9.3}$$

Now consider $f \in C^1(D, \mathbb{R})$ with $D \subset \mathbb{R}^n$ open. Then the Taylor polynomial for $k = 0$ is simply the constant function $P(x) \equiv f(x_0)$. Under the assumptions of Theorem 9.5 with $k = 0$ and $t = 1$ we obtain

$$f(x_0 + z) - f(x_0) = \sum_{|\alpha|=1} \frac{D^\alpha f(x_0 + \theta z)}{\alpha!} z^\alpha = Df(x_0 + \theta z)z, \tag{9.4}$$

which is precisely the generalization of (9.3). Thus, the mean value theorem is nothing other than a special case of the Taylor formula.

However, a difficulty arises when we consider functions $f : D \to \mathbb{R}^m$. Although equation (9.4) holds for each component function $f_i : D \to \mathbb{R}$, the value of θ depends on i, so that in general we cannot find a point $x_0 + \theta z$ for which (9.4) holds for the entire vector-valued function.

To prove a mean value theorem that also holds for vector-valued functions, we therefore proceed somewhat differently. We again consider the real case. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is continuously differentiable. Then from the fundamental theorem of calculus and the substitution rule we obtain

$$f(x + z) - f(x) = \int_x^{x+z} f'(y)\, dy = \int_0^1 f'(x + tz)\, dt\, z.$$

Comparing this with (9.3) shows that this can be understood as a "mean value theorem in integral form": the derivative $f'(x + \theta z)$ is replaced by the integral over this derivative.

Question 9.3 *Comparing the last formula with* (9.3) *shows the identity*

$$\int_0^1 f'(x + tz)\, dt = f'(x + \theta z)$$

for a $\theta \in (0, 1)$*. Where have we seen a similar identity before?*

In this integral form, the theorem can now also be formulated and proved in $\mathbb{R}^n$. For this we first need the definition of the integral for matrix-valued functions

$$t \mapsto A(t) = \begin{pmatrix} a_{11}(t) & \cdots & a_{1n}(t) \\ \vdots & \ddots & \vdots \\ a_{n1}(t) & \cdots & a_{nn}(t) \end{pmatrix}.$$

Analogously to (5.9), we define this componentwise by

$$\int_a^b A(t)dt = \begin{pmatrix} \int_a^b a_{11}(t)dt & \cdots & \int_a^b a_{1n}(t)dt \\ \vdots & \ddots & \vdots \\ \int_a^b a_{n1}(t)dt & \cdots & \int_a^b a_{nn}(t)dt \end{pmatrix},$$

where we of course assume that the (real) component functions are integrable. In the following theorem these functions are continuous, which implies Riemann integrability.

Theorem 9.7 Let $D \subset \mathbb{R}^n$ be open and $f \in C^1(D, \mathbb{R}^m)$. Let $x \in D$ and $z \in \mathbb{R}^n$ be chosen such that $x + tz$ lies in D for all $t \in [0, 1]$. Then

$$f(x+z) - f(x) = \left(\int_0^1 Df(x+tz)dt\right) z.$$

Proof: Set $g_i(t) := f_i(x + tz)$. By the chain rule we have

$$g_i'(t) = Df_i(x+tz)z = \left(\frac{\partial f_i}{\partial x_1}(x+tz) \ldots \frac{\partial f_i}{\partial x_n}(x+tz)\right) z = \sum_{j=1}^n \frac{\partial f_i}{\partial x_j}(x+tz)z_j.$$

From the fundamental theorem of calculus and the linearity of the integral we obtain

$$g_i(1) - g_i(0) = \int_0^1 g_i'(t)dt = \int_0^1 Df_i(x+tz)dt\, z.$$

Thus we obtain

$$f(x+z)-f(x) = \begin{pmatrix} g_1(1) - g_1(0) \\ \vdots \\ g_m(1) - g_m(0) \end{pmatrix} = \begin{pmatrix} \int_0^1 Df_1(x+tz)dt\, z \\ \vdots \\ \int_0^1 Df_m(x+tz)dt\, z \end{pmatrix} = \left(\int_0^1 Df(x+tz)dt\right) z,$$

which proves the claim. $\square$

9.3 Extrema

We now return to functions $f : D \to \mathbb{R}$, i.e., to functions with values in $\mathbb{R}$. For such functions the question of extrema arises just as in the real case, that is, maxima and minima of f. As in the real case, we want to determine them with the help of derivatives, and likewise we must define *local* extrema.

Definition 9.8 Let $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$. We say that f has a *local minimum* at a point $x \in D$ if there exists $\varepsilon > 0$ such that

$$f(x) \leq f(y) \quad \text{for all } y \in B_\varepsilon(x) \cap D.$$

The minimum is called *strict* if the above inequality holds with $<$ for $y \neq x$. The point x is called a *(strict) minimizer* of f.

Analogously we define the (strict) local maximum with $\geq$ and $>$, respectively. $\square$

Example 9.9 (i) Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ with $f(x) = \|x\|_2^2$. This function has a strict local minimum at $x = 0$, since $f(x) = 0$ there, while for all $x \neq 0$ we have $f(x) > 0$. In fact, this is even a global minimum.

(ii) Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(x) = x^T Ax + b^T x + c$ for a matrix $A \in \mathbb{R}^{n\times n}$, a vector $b \in \mathbb{R}^n$, and a real number $c \in \mathbb{R}$. For $A = \mathrm{Id}$, $b = 0$, and $c = 0$ this is exactly the function in (i). The question now is whether this function also possesses minimizers (or maximizers) for other choices of A, b, and c. □

The following theorem gives a necessary condition for the existence of an extremum if D is an open set.

Theorem 9.10 Let $f \in C^1(D, \mathbb{R})$ with $D \subset \mathbb{R}^n$ open. If f has a local minimum or maximum at $x \in D$, then

$$Df(x) = 0.$$

Proof: Let $B_\varepsilon(x)$ be the ball from Definition 9.8. Then the function $g(t) := f(x + tv)$ is defined, continuously differentiable, and has a local minimum or maximum at $t = 0$ for every $v \in \mathbb{R}^n$ with $\|v\| \leq \varepsilon$ and every $t \in (-1, 1)$. From [7, Theorem 17.2] and Theorem 8.11 we obtain

$$0 = g'(0) = Df(x)v.$$

Since this holds for all $v \in \mathbb{R}^n$ with $\|v\| \leq \varepsilon$, it follows that $Df(x) = 0$. □

Definition 9.11 A point $x \in D$ with $Df(x) = 0$ is called a *critical point.* □

Example 9.12 We consider the functions from Example 9.9.

(i) For $f(x) = \|x\|_2^2$ we have

$$Df(x) = 2(x_1, \dots, x_n).$$

Thus $Df(x) = 0$ holds exactly when $x = 0$. Hence the point $x = 0$ is the only critical point, i.e., the only candidate for a local minimum or maximum.

(ii) If A is a symmetric matrix, then for $f(x) = x^T Ax + b^T x + c$ we have

$$Df(x) = (2Ax + b)^T,$$

which can be computed similarly to Example 8.16. Thus the condition $Df(x) = 0$ is equivalent to

$$2Ax + b = 0.$$

If we assume that A is invertible, then this equation has the unique solution

$$x = -\frac{1}{2}A^{-1}b.$$

Thus, for symmetric and invertible A, this x is the only critical point. □

To ensure that a point x with $Df(x) = 0$ is indeed a local minimum or maximum (and to distinguish minima from maxima), we need the second derivatives. In the real case, a point x with $f'(x) = 0$ is, according to [7, Theorem 17.7], a local minimum (respectively maximum) if $f''(x) > 0$ (respectively $f''(x) < 0$).

Here we can—since the functions considered take values in $\mathbb{R}$—write the second derivatives in the form of the Hessian matrix $H_f(x)$ from Remark 8.21. The properties of this matrix that we need are given in the following definition.

Definition 9.13 Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then we call

(i) *A positive definite* if $x^T Ax > 0$ for all $x \in \mathbb{R}^n$ with $x \neq 0$,

(ii) *A negative definite* if $-A$ is positive definite.

□

Positive (and therefore also negative) definiteness of a matrix A can be checked in various ways. We state two criteria here (without proofs):

(a) A symmetric matrix is positive definite if and only if all eigenvalues λ of A satisfy the inequality $\lambda > 0$.

(b) A symmetric matrix is positive definite if and only if its leading principal minors, i.e., the determinants of the matrices

$$(a_{ij})_{i,j=1,\dots,k}$$

for all $k = 1, \dots, n$, are positive.

Question 9.4 *How do these criteria translate into conditions for A being negative definite?*

Example 9.14 (i) Consider the matrix

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}.$$

We verify the two criteria.

(a) The characteristic polynomial of the matrix is

$$\chi_A(\lambda) = \lambda^2 - 5\lambda + 5,$$

from which the eigenvalues are obtained by computing the roots:

$$\lambda_{1,2} = \frac{5}{2} \pm \sqrt{\left(\frac{5}{2}\right)^2 - 5} = \frac{5}{2} \pm \sqrt{\frac{5}{4}} = \frac{5 \pm \sqrt{5}}{2}.$$

Since both of these values are positive (because $\sqrt{5} < 5$), the matrix is positive definite.

(b) The principal minors of the matrix are

$$\det(3) = 3 \quad \text{and} \quad \det\begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix} = 3 \cdot 2 - 1 \cdot 1 = 5.$$

Thus the matrix is positive definite. □

The following theorem now gives necessary conditions for the Hessian matrix for the existence of a local minimum or maximum.

Theorem 9.15 Let $f \in C^2(D, \mathbb{R})$ with $D \subset \mathbb{R}^n$ open and let $x \in D$ be a critical point, i.e., $Df(x) = 0$.

If the Hessian matrix $H_f(x)$ is positive definite (respectively negative definite), then $f(x)$ is a strict local minimum (respectively maximum).

Proof: We consider the positive definite case and first prove the following auxiliary result: There exists a $c > 0$ such that

$$z^T H_f(x) z \geq 2c \quad \text{for all } z \in \mathbb{S}^{n-1} := \{z \in \mathbb{R}^n \mid \|z\|_2 = 1\}. \tag{9.5}$$

Proof of (9.5): The mapping

$$z \mapsto z^T H_f(x) z$$

is continuous and therefore attains a minimum on the compact set $\mathbb{S}^{n-1}$ at some $z_0 \in \mathbb{S}^{n-1}$.

Question 9.5 *Which results from the previous chapters tell us that $\mathbb{S}^{n-1}$ is a compact set?*

Setting

$$c := z_0^T H_f(x) z_0 / 2,$$

we have $c > 0$, since $H_f(x)$ is positive definite, and thus (9.5) follows.

As a second auxiliary claim we show that there exists an $\varepsilon > 0$ with $B_\varepsilon(x) \subset D$ such that

$$z^T H_f(y) z > c \quad \text{for all } z \in \mathbb{S}^{n-1} \text{ and all } y \in B_\varepsilon(x) \tag{9.6}$$

with $c > 0$ from (9.5).

Proof of (9.6): Since D is open, we have $B_\varepsilon(x) \subset D$ for all sufficiently small $\varepsilon > 0$, which we consider in the following. Because the mapping $x \mapsto H_f(x)$ is continuous by assumption, the mapping

$$(y, z) \mapsto z^T H_f(y) z$$

is also continuous.

Assume now that no $\varepsilon > 0$ satisfying (9.6) exists. Then there would exist sequences $y_n \to x$ and $z_n \in \mathbb{S}^{n-1}$ with

$$z_n^T H_f(y_n) z_n \leq c.$$

Since $\mathbb{S}^{n-1}$ is compact, the sequence z_n has a convergent subsequence $z_{n_j} \to z \in \mathbb{S}^{n-1}$. By continuity of H_f and (9.5) we obtain

$$2c \leq z^T H_f(x) z = \lim_{j \to \infty} z_{n_j}^T H_f(y_{n_j}) z_{n_j} \leq c,$$

which is a contradiction because $c > 0$. Hence (9.6) follows.

For arbitrary vectors $z \in \mathbb{R}^n$ with $z \neq 0$ and all $y \in B_\varepsilon(x)$ we obtain with $\tilde{z} := z/\|z\|_2 \in \mathbb{S}^{n-1}$

$$z^T H_f(y) z = (\|z\|_2 \tilde{z})^T H_f(y) (\|z\|_2 \tilde{z}) = \tilde{z}^T H_f(y) \tilde{z} \|z\|_2^2 \geq c \|z\|_2^2.$$

To prove the local minimum we now consider the Taylor formula for $k = 1$ with x in place of x_0 and $y \in B_\varepsilon(x)$ in place of x. Using the formulas from Remark 9.4 and $Df(x) = 0$ we can write the formula as

$$f(y) = f(x) + \frac{1}{2} z^T H_f(x + \theta z) z \tag{9.7}$$

with $z = y - x$ and some $\theta \in [0, 1]$.

Since from $y \in B_\varepsilon(x)$ and $\theta \in [0, 1]$ it follows that $x + \theta z \in B_\varepsilon(x)$, we obtain from (9.6) and (9.7)

$$f(y) = f(x) + \frac{1}{2} z^T H_f(x + \theta z) z > f(x) + \frac{c}{2} \|y - x\|_2^2 > f(x)$$

for all $y \in B_\varepsilon(x)$ with $y \neq x$. Thus $f(x)$ is a strict local minimum. □

Example 9.16 We consider the functions from Examples 9.9 and 9.12.

(i) For $f(x) = \|x\|_2^2$ one computes

$$H_f(x) = 2\mathrm{Id}$$

for all $x \in \mathbb{R}^n$, in particular for $x = 0$. This matrix is positive definite since

$$z^T 2\mathrm{Id} z = 2\|z\|_2^2 > 0$$

for all $z \in \mathbb{R}^n$ with $z \neq 0$. Thus $f(x)$ has a strict local minimum at $x = 0$.

(ii) For $f(x) = x^T A x + b^T x + c$ the Hessian matrix is

$$H_f(x) = 2A$$

for all $x \in \mathbb{R}^n$, in particular for $x = -\frac{1}{2} A^{-1} b$. Thus this point x is a strict minimizer of f if A is positive definite and a strict maximizer if A is negative definite.

As a concrete numerical example consider

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad c = 5. \tag{9.8}$$

Then we obtain a minimizer (note that A is positive definite by Example 9.14) at

$$x = -\frac{1}{2} A^{-1} b = \begin{pmatrix} -\frac{1}{10} \\ -\frac{1}{5} \end{pmatrix}.$$

The value of f at this point is $f(x) = 97/20$. Figure 9.1 shows the graph of this function and illustrates that the minimum indeed occurs at the stated point.

As another concrete numerical example consider

$$A = \begin{pmatrix} -3 & 1 \\ 1 & 2 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad c = 5. \tag{9.9}$$

This matrix has the eigenvalues $\lambda_{1/2} = -\frac{1}{2} \pm \frac{1}{2}\sqrt{29}$. Since one eigenvalue is positive and the other negative, the matrix is neither positive nor negative definite. Hence we cannot

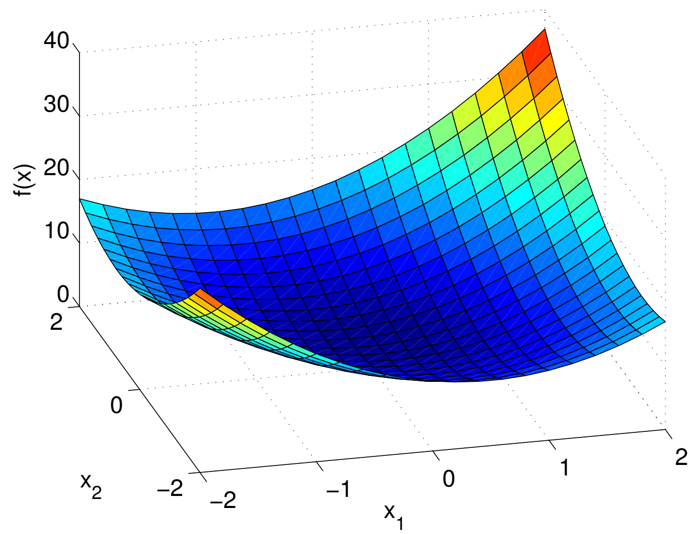

Figure 9.1: Graph of the function $f(x) = x^T Ax + b^T x + c$ with A, b, c from (9.8)

conclude the existence of either a minimum or a maximum, and indeed the function has neither. Instead, the point

$$x = -\frac{1}{2}A^{-1}b = \begin{pmatrix} \frac{1}{14} \\ -\frac{2}{7} \end{pmatrix}$$

—the only point with $Df(x) = 0$—is a so-called saddle point (the function decreases in one direction and increases in another), as shown in Figure 9.2.

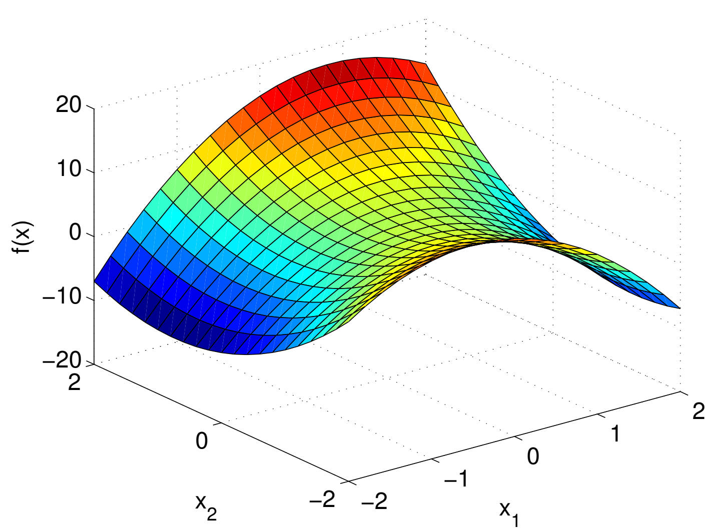

Figure 9.2: Graph of the function $f(x) = x^T Ax + b^T x + c$ with A, b, c from (9.9)

$\square$

Chapter 10

Implicit and Inverse Functions

Last update: July 21, 2026

10.1 Implicit Functions

Implicitly defined functions are functions that are not given by an explicit formula (as almost all functions we have encountered so far), but instead by an implicit equation.

Some examples will first illustrate this concept.

Example 10.1 (i) Circle: The unit circle (or the 1-sphere $\mathbb{S}^1$) in $\mathbb{R}^2$ is the set of all points $x = (x_1, x_2)^T$ with $\|x\|_2^2 = 1$. This can also be written as the set

$$\{x \in \mathbb{R}^2 \,|\, f(x) = 0\}$$

with

$$f(x) = \|x\|_2^2 - 1 = x_1^2 + x_2^2 - 1.$$

We can now try to represent the points on the circle (or at least a subset of them) as the graph of a function $g : D \to \mathbb{R}$ with $D \subset \mathbb{R}$. The function g is then defined by the equation

$$f(x_1, g(x_1)) = 0.$$

The circle (or at least part of it) can then be represented as the graph of this function, i.e., as the set

$$\{(x_1, x_2)^T \in \mathbb{R}^2 \,|\, x_2 = g(x_1),\ x_1 \in D\}.$$

If one draws the circle in a coordinate system, one sees that it is not possible to represent the entire circle in this form, since the circle contains both the point $(0, 1)^T$ and the point $(0, -1)^T$, while the function g can take only one of the values $g(0) = 1$ or $g(0) = -1$.

However, the representation succeeds if we consider not one but two functions, namely $g, \tilde{g} : D \to \mathbb{R}$ with $D = [-1, 1]$ and

$$g(x_1) = \sqrt{1 - x_1^2} \quad \text{and} \quad \tilde{g}(x_1) = -\sqrt{1 - x_1^2}.$$

These satisfy the desired condition

$$f(x_1, g(x_1)) = x_1^2 + g(x_1)^2 - 1 = x_1^2 + 1 - x_1^2 - 1 = 0$$

and

$$f(x_1, \tilde{g}(x_1)) = x_1^2 + \tilde{g}(x_1)^2 - 1 = x_1^2 + 1 - x_1^2 - 1 = 0.$$

Geometrically, the graph of g represents the upper semicircle and the graph of $\tilde{g}$ represents the lower semicircle.

Translated to this example, the question we are going to address later in this chapter is: Given a point $x \in \mathbb{S}^1$, i.e., with $f(x_1, x_2) = 0$, can we find an open interval I around x_1 and a function $g : I \to \mathbb{R}$ with $f(x_1, g(x_1)) = 0$ for all $x_1 \in I$.

Question 10.1 *For which points $x \in \mathbb{S}^1$ is this possible? For which points is it not possible?*

(ii) Sphere: Analogously to the circle, the sphere (or the 2-sphere $\mathbb{S}^2$) in $\mathbb{R}^3$ can be written as

$$\{x \in \mathbb{R}^3 \mid f(x) = 0\}$$

with

$$f(x) = \|x\|_2^2 - 1 = x_1^2 + x_2^2 + x_3^2 - 1.$$

A function whose graph describes the sphere (or part of it) must now have the form $g : \mathbb{R}^2 \to \mathbb{R}$: to each point $(x_1, x_2)^T$ in the plane we assign the third coordinate x_3 of a point $x = (x_1, x_2, x_3)^T$ on the sphere. Such a function is characterized by the implicit equation

$$f(x_1, x_2, g(x_1, x_2)) = 0.$$

(Whether such a function exists and how many of them are needed to cover the entire sphere can be considered by the reader.)

(iii) Quadratic equations: Consider quadratic equations of the form

$$x^2 + px + q = 0.$$

If we now seek a function $g : \mathbb{R}^2 \to \mathbb{R}$ that assigns to each pair p, q a solution $x = g(p, q)$ of the equation, then g is again implicitly defined by

$$f(g(p, q), p, q) = 0$$

with

$$f(x, p, q) = x^2 + px + q.$$

(Functions g satisfying this condition should be familiar from school.)

(iv) Nonlinear systems of equations: For a given $x \in \mathbb{R}$ we seek $y = (y_1, y_2)^T \in \mathbb{R}^2$ such that the system

$$\begin{aligned} x^3 + y_1^3 + y_2^3 - 7 &= 0 \\ xy_1 + y_1y_2 + y_2x + 2 &= 0 \end{aligned}$$

is satisfied.

If we want to know whether this system can be solved for y (i.e., whether we can give a formula for y depending on x), this is equivalent to asking whether there exists a function $g : \mathbb{R} \to \mathbb{R}^2$ such that $f(x, g(x)) = 0$ with

$$f(x,y) = \begin{pmatrix} f_1(x,y) \\ f_2(x,y) \end{pmatrix} := \begin{pmatrix} x^3 + y_1^3 + y_2^3 - 7 \\ xy_1 + y_1y_2 + y_2x + 2 \end{pmatrix}.$$

□

In this section we want to investigate two questions:

- When does an implicitly defined function g exist?
- If such a function g exists, what can we say about its differentiability?

We will begin with the second question.

Remark 10.2 In the following we use the notation $f(x, y)$ for the defining function and $y = g(x)$ for the implicitly defined function. Here $x \in \mathbb{R}^k$, $y \in \mathbb{R}^m$, and

$$f : D_1 \times D_2 \to \mathbb{R}^m$$

with $D_1 \subset \mathbb{R}^k$, $D_2 \subset \mathbb{R}^m$, while

$$g : D_1 \to \mathbb{R}^m.$$

□

The examples above take the following form in this notation:

(i) $k = m = 1$, $x = x_1$, $y = x_2$, $f(x, y) = x^2 + y^2 - 1$

(ii) $k = 2$, $m = 1$, $x = (x_1, x_2)^T$, $y = x_3$, $f(x, y) = x_1^2 + x_2^2 + y^2 - 1$

(iii) $k = 2$, $m = 1$, $x = (p, q)^T$, $y = x$, $f(x, y) = y^2 + x_1y + x_2$

(iv) $k = 1$, $m = 2$ and x, y, f as given in the example.

Question 10.2 *Why does the example*

$$k = 1,\ m = 2,\ x = x_1,\ y = (y_1, y_2)^T,\ f(x, y) = x_1^2 + y_1/y_2$$

not fit the standard form?

Note that the dimension of y is always equal to the dimension of the codomain of f. The reason for this choice will become clear in the following theorem. This theorem makes a statement about the differentiability of the function g. For this we recall the definitions of the derivatives

$$D_x f(x, y) \in \mathbb{R}^{m\times k} \quad \text{and} \quad D_y f(x, y) \in \mathbb{R}^{m\times m}$$

from Section 8.2.

Theorem 10.3 Let the functions f, g be given as in Remark 10.2. Let $a \in D_1$ and let $U_1 \subset D_1$ be an open neighborhood of a such that g is continuous on U_1, $g(U_1) \subset D_2$, and for all $x \in U_1$ the equation

$$f(x, g(x)) = 0$$

holds. Furthermore, assume that f is (totally) differentiable at the point (a, b) with $b = g(a)$ and that the Jacobian matrix $D_y f(a, b) \in \mathbb{R}^{m \times m}$ is invertible.

Then g is differentiable at the point a and

$$Dg(a) = -(D_y f(a, b))^{-1} D_x f(a, b).$$

Proof: Without loss of generality we may assume $a = 0$ and $b = g(a) = 0$.[1] For brevity we write

$$A := D_x f(0, 0), \qquad B := D_y f(0, 0).$$

From the differentiability of f it follows, using (8.5), that on sufficiently small neighborhoods $\widetilde{U}_1 \subset \mathbb{R}^k$ of $0 \in \mathbb{R}^k$ and $\widetilde{U}_2 \subset \mathbb{R}^m$ of $0 \in \mathbb{R}^m$ there exists a function $\varphi : \widetilde{U}_1 \times \widetilde{U}_2 \to \mathbb{R}^m$ such that

$$f(x, y) = (A,\, B) \begin{pmatrix} x \\ y \end{pmatrix} + \varphi(x, y) = Ax + By + \varphi(x, y)$$

and

$$\frac{\varphi(x, y)}{\left\| \begin{pmatrix} x \\ y \end{pmatrix} \right\|} \to 0$$

for

$$\begin{pmatrix} x \\ y \end{pmatrix} \to 0.$$

Since all norms in $\mathbb{R}^n$ are equivalent, we may choose $\| \cdot \|$ here to be the 1-norm, for which

$$\left\| \begin{pmatrix} x \\ y \end{pmatrix} \right\|_1 = \|x\|_1 + \|y\|_1.$$

Thus

$$\frac{\varphi(x, y)}{\|x\|_1 + \|y\|_1} \to 0.$$

Since $g(0) = 0$ and g is continuous at $a = 0$, it follows that

$$\begin{pmatrix} x \\ g(x) \end{pmatrix} \to 0 \quad \text{for } x \to 0,$$

and hence also

$$\frac{\varphi(x, g(x))}{\|x\|_1 + \|g(x)\|_1} \to 0$$

[1] Otherwise we consider $f(x + a, y + b)$ and $g(x + a) - b$ instead of f and g. Note that this does not change the derivatives of f and g.

for $x \to 0$. In particular, for all $c > 0$ it follows that

$$\|\varphi(x, g(x))\|_1 \le c(\|x\|_1 + \|g(x)\|_1) \tag{10.1}$$

for all x with $\|x\|_1$ sufficiently small.

Because

$$0 = f(x, g(x)) = Ax + Bg(x) + \varphi(x, g(x))$$

and because B is assumed to be invertible, we obtain

$$g(x) = -B^{-1}Ax - B^{-1}\varphi(x, g(x)).$$

From this we obtain

$$g(x) = -B^{-1}Ax + \tilde{\varphi}(x)$$

with

$$\tilde{\varphi}(x) = -B^{-1}\varphi(x, g(x)),$$

from which, using (10.1), we obtain in particular

$$\begin{aligned}\|g(x)\|_1 &\le \|B^{-1}A\|_1\|x\|_1 + \|B^{-1}\|_1\|\varphi(x, g(x))\|_1 \\ &\le \|B^{-1}A\|_1\|x\|_1 + \|B^{-1}\|_1 c\|x\|_1 + \|B^{-1}\|_1 c\|g(x)\|_1\end{aligned}$$

for all x with $\|x\|_1$ sufficiently small.

If we now choose

$$c = \frac{1}{2\|B^{-1}\|_1},$$

it follows that

$$\|g(x)\|_1 \le \underbrace{2(\|B^{-1}A\|_1 + \|B^{-1}\|_1 c)}_{=:C}\|x\|_1,$$

and hence

$$\|x\|_1 + \|g(x)\|_1 \le (1 + C)\|x\|_1.$$

To prove that g is differentiable with

$$Dg(0) = B^{-1}A$$

we must now show that

$$\tilde{\varphi}(x)/\|x\|_1 \to 0$$

for $x \to 0$, or equivalently

$$\|\tilde{\varphi}(x)\|_1/\|x\|_1 \to 0.$$

Since

$$\frac{\|\tilde{\varphi}(x)\|_1}{\|x\|_1} \le (1 + C)\frac{\|\tilde{\varphi}(x)\|_1}{\|x\|_1 + \|g(x)\|_1} \le (1 + C)\|B^{-1}\|_1\frac{\|\varphi(x, g(x))\|_1}{\|x\|_1 + \|g(x)\|_1} \to 0$$

for $x \to 0$, this condition is satisfied, which proves the claim. $\square$

The theorem therefore tells us that the implicitly defined function g—under the stated assumptions—is differentiable if it is merely continuous, and it allows us to compute its derivative.

Example 10.4 (i) Consider Example 10.1(i), i.e., $f(x,y) = x^2 + y^2 - 1$. Here

$$D_x f(x,y) = 2x \quad \text{and} \quad D_y f(x,y) = 2y.$$

This "1×1 matrix" is invertible exactly when $b = g(a) \neq 0$, and at these points we obtain

$$Dg(a) = -(D_y f(a, g(a)))^{-1} D_x f(a, g(a)) = -\frac{2a}{2g(a)} = -\frac{a}{g(a)}.$$

For $g(x) = \sqrt{1-x^2}$ we obtain

$$Dg(x) = -\frac{x}{\sqrt{1-x^2}},$$

and for $\tilde{g}(x) = -\sqrt{1-x^2}$ we obtain

$$D\tilde{g}(x) = \frac{x}{\sqrt{1-x^2}}.$$

Of course, we could also have computed this directly using the chain rule. However, it should be noted that here we did not actually need to know the explicit formula for $g(x)$. To compute the derivative $Dg(x)$ at $x = a$, it would have been sufficient to know the values a and $b = g(a)$.

The points $a = 0$, where Theorem 10.3 cannot be applied, correspond here precisely to the points $b = -1$ and $b = 1$. At these points the functions g and $\tilde{g}$ are indeed not differentiable.

Question 10.3 *Where are the exceptional points with $a = 0$ located on the sphere?*

Note, however, that the theorem does not make any statement about non-differentiability. It may therefore happen that $D_y f(x,y)$ is not invertible but the function g is nevertheless differentiable at x. An example is the constant function $f \equiv 0$. In this case, for example, the function $g \equiv 0$ satisfies the condition $f(x, g(x)) = 0$. This function is differentiable everywhere, but since $D_y f \equiv 0$, the theorem cannot be applied to any pair (x,y).

(ii) We again consider the same example and show that the continuity assumption in Theorem 10.3 cannot be omitted; that is, continuity of g does not automatically follow from $f(x, g(x)) = 0$, even if f is continuous.

Consider the function

$$g(x) = \begin{cases} \sqrt{1-x^2}, & x \in [-1, 0], \\ -\sqrt{1-x^2}, & x \in (0, 1]. \end{cases}$$

At $(a,b) = (0,1)$ all assumptions of the theorem are satisfied except for the continuity of g. Since g is not continuous at this point, it cannot be differentiable there. □

Remark 10.5 If f satisfies the conditions of Theorem 10.3 and is additionally continuously differentiable in a neighborhood of (a,b), then g is continuously differentiable in a neighborhood of a.

Differentiability follows because $D_y f(a)$ (like any matrix) is invertible if and only if its determinant is nonzero. Since the determinant is a continuous mapping from $\mathbb{R}^{n \times n}$ to $\mathbb{R}$ (because it is a polynomial in the coefficients of the matrix), if it is nonzero at (a,b) then by the ε-δ criterion it remains nonzero in a neighborhood of (a, b). Consequently, Theorem 10.3 can be applied in a whole neighborhood of a, and therefore g is differentiable there.

That the derivative of g is moreover continuous follows from the formula for Dg and the fact that matrix inversion is also a continuous mapping. This follows from the fact that the entries of A^{-1} can be computed using Cramer's rule from the system $Ax = e_i$, and Cramer's rule in turn is based on the continuous determinant and the (also continuous) real division. □

10.2 Banach's Fixed Point Theorem

We now turn to the first question posed at the beginning, namely the question of the existence of an implicitly defined function. To prove the corresponding theorem we use a famous result from analysis, the Banach fixed point theorem. Recall (cf. Definition 6.6) that a Banach space is a complete normed vector space, that is, a vector space on which a norm $\|\cdot\|$ is defined such that with respect to the associated metric $d(x, y) = \|x - y\|$ every Cauchy sequence converges.

Theorem 10.6 (Banach's Fixed Point Theorem) Let V be a Banach space, $K \subset V$ be closed, and $T : K \to K$ a mapping for which there exists a $\kappa < 1$ such that

$$\|T(x) - T(y)\| \leq \kappa \|x - y\| \tag{10.2}$$

for all $x, y \in K$ (such a mapping T is called a *contraction* on K). Then there exists a unique fixed point $x^* \in K$ of T, i.e., a unique x^* such that $T(x^*) = x^*$.

Proof: Choose an arbitrary $x_0 \in K$ and consider the sequence defined inductively for $m = 0, 1, 2, \ldots$ by

$$x_{m+1} = T(x_m) \in K.$$

From the contraction inequality (10.2) it follows inductively that

$$\|T(x_m) - T(x_{m+1})\| \leq \kappa \|T(x_{m-1}) - T(x_m)\| \leq \ldots \leq \kappa^m \|x_0 - T(x_0)\|.$$

Hence for all $n \geq m \geq 1$ we obtain

$$\begin{aligned}
\|x_m - x_n\| &= \|T(x_{m-1}) - T(x_{n-1})\| \\
&\leq \sum_{l=m-1}^{n-2} \|T(x_l) - T(x_{l+1})\| \\
&\leq \sum_{l=m-1}^{n-2} \kappa^l \|x_0 - T(x_0)\| \\
&\leq \kappa^{m-1} \sum_{l=0}^{\infty} \kappa^l \|x_0 - T(x_0)\| \\
&= \frac{\kappa^{m-1}}{1 - \kappa} \|x_0 - T(x_0)\|.
\end{aligned}$$

Since $\kappa^{m-1} \to 0$ as $m \to \infty$, the sequence x_m is a Cauchy sequence. Because V is a Banach space, the sequence converges to a limit $x^* \in V$, which by the closedness of K also lies in K.

Because T is continuous (which follows from the contraction inequality (10.2)), we obtain

$$T(x^*) = T(\lim_{m\to\infty} x_m) = \lim_{m\to\infty} T(x_m) = \lim_{m\to\infty} x_{m+1} = x^*.$$

Thus x^* is a fixed point.

To prove uniqueness, let x^* and x^{**} be two fixed points of T in K. Then

$$\|x^* - x^{**}\| = \|T(x^*) - T(x^{**})\| \leq \kappa \|x^* - x^{**}\|.$$

Since $\kappa < 1$, this can only hold if $\|x^* - x^{**}\| = 0$, hence $x^* = x^{**}$. □

Question 10.4 *Would the inequality $\|x^* - x^{**}\| \leq \kappa\|x^* - x^{**}\|$ be possible if $\|x^* - x^{**}\| < 0$? If yes, why don't we need to consider this in the proof?*

10.3 The Implicit Function Theorem

With the help of Banach's fixed point theorem we can now answer the question of the existence of an implicit function with the following theorem.

Theorem 10.7 (Implicit Function Theorem) Let the function f from Remark 10.2 be continuously differentiable on $U_1 \times U_2$, where $U_1 \subset D_1$ and $U_2 \subset D_2$ are open sets. Let $(a, b) \in U_1 \times U_2$ satisfy

$$f(a, b) = 0$$

and assume that the Jacobian matrix $D_y f(a, b) \in \mathbb{R}^{m\times m}$ is invertible.

Then there exist open neighborhoods $V_1 \subset U_1$ of a and $V_2 \subset U_2$ of b and a continuously differentiable mapping $g : V_1 \to \overline{V_2}$ such that

$$f(x, g(x)) = 0 \quad \text{for all } x \in V_1.$$

Moreover, for every $x \in V_1$ the point $y = g(x)$ is the unique point in V_2 for which $f(x, y) = 0$ holds.

Question 10.5 *Why doesn't Example 10.1(i), in which we found* two *implicitly defined functions g_1 and g_2, contradict the uniqueness statement in Theorem 10.7?*

Proof: To prove the theorem it suffices to establish the existence of a continuous function g with the stated properties, because the continuous differentiability then follows from Theorem 10.3 and Remark 10.5. For this g the derivative formula from Theorem 10.3 also holds.

Step 1 (preparatory definitions): As in the proof of Theorem 10.3 we may assume without loss of generality that $a = 0$ and $b = g(a) = 0$, and let $B = D_y f(0, 0)$. On $U_1 \times U_2$ we define the mapping

$$G(x, y) := y - B^{-1} f(x, y).$$

From this definition we obtain the equivalence

$$f(x, y) = 0 \quad \Leftrightarrow \quad G(x, y) = y. \tag{10.3}$$

For the derivative of G with respect to y we have

$$D_y G(x, y) = \mathrm{Id} - B^{-1} D_y f(x, y),$$

and in particular

$$D_y G(0, 0) = \mathrm{Id} - B^{-1} B = 0.$$

Since $D_y G$ is continuous, there exist neighborhoods $W_1 \subset U_1$ of 0 and $W_2 \subset U_2$ of 0 such that

$$\|D_y G(x, y)\| \le \frac{1}{2} \quad \text{for all } (x, y) \in W_1 \times W_2. \tag{10.4}$$

We now choose $\varepsilon_2 > 0$ such that $\overline{V_2} \subset W_2$ holds for

$$V_2 := B_{\varepsilon_2}(0),$$

take any $\delta > 0$ with $\delta < \varepsilon_2/2$, and choose $\varepsilon_1 > 0$ such that for all $x \in V_1 := B_{\varepsilon_1}(0)$ the inequality

$$\sup_{x \in V_1} \|G(x, 0)\| \le \delta \tag{10.5}$$

holds (such an ε_1 exists because G is continuous with $G(0, 0) = 0$).

Finally we define K as the set of continuous functions from V_1 to $\overline{V_2}$, i.e.

$$K = C(V_1, \overline{V_2}).$$

Note that K is a closed subset of the Banach space of continuous functions $C(V_1, \mathbb{R}^m)$ with the norm[2]

$$\|g\|_\infty = \sup_{x \in V_1} \|g(x)\|.$$

We will apply the Banach fixed point theorem on the space $V = C(V_1, \mathbb{R}^m)$.

Step 2 (application of the Banach fixed point theorem):

Define a mapping

$$T : C(V_1, \mathbb{R}^m) \to C(V_1, \mathbb{R}^m)$$

by

$$T(g) := G(\cdot, g(\cdot)).$$

[2] Note that the norm $\|g\|_\infty$ on the left is a norm on the vector space of functions while the norm $\|g(x)\|$ on the right is a norm on $\mathbb{R}^m$.

This means that the value of the function $T(g)$ at the point x is

$$T(g)(x) = G(x, g(x)).$$

We now show that T is a contraction on K.

Let $g_1, g_2 \in K$ and let $x \in V_1$ be arbitrary. From the mean value theorem and Lemma 5.11 (note that this lemma holds only for the 2-norm, which we therefore use here) and with the abbreviations $y_1 = g_1(x)$, $y_2 = g_2(x)$ we obtain

$$\begin{aligned}
\|T(g_2)(x) - T(g_1)(x)\| &= \|G(x, g_2(x)) - G(x, g_1(x))\| \\
&= \|G(x, y_2) - G(x, y_1)\| \\
&= \left\| \int_0^1 D_y G(x, y_1 + t(y_2 - y_1))(y_2 - y_1)\, dt \right\| \\
&\le \int_0^1 \|D_y G(x, y_1 + t(y_2 - y_1))(y_2 - y_1)\|\, dt \\
&\le \int_0^1 \|D_y G(x, y_1 + t(y_2 - y_1))\| \|y_2 - y_1\|\, dt \\
&\le \int_0^1 \frac{1}{2}\|y_2 - y_1\|\, dt = \frac{1}{2}\|y_2 - y_1\| = \frac{1}{2}\|g_2(x) - g_1(x)\|.
\end{aligned}$$

Since this inequality holds for all $x \in V_1$, taking the supremum gives

$$\|T(g_2) - T(g_1)\|_\infty \le \frac{1}{2}\|g_2 - g_1\|_\infty.$$

Thus (10.2) holds with $\kappa = 1/2$. Moreover, taking $g_1 \equiv 0$ and using that all $g \in K$ map into $V_2 = B_{\varepsilon_2}(0)$ gives

$$\|T(g_2) - T(g_1)\|_\infty = \frac{1}{2}\|g_2\|_\infty \le \frac{\varepsilon_2}{2}.$$

Furthermore

$$\|T(g_1)\|_\infty = \|T(0)\|_\infty \le \|G(\cdot, 0)\|_\infty \le \delta.$$

Using the reverse triangle inequality we obtain

$$\|T(g_2)\|_\infty < \frac{\varepsilon_2}{2} + \|T(g_1)\|_\infty \le \frac{\varepsilon_2}{2} + \delta < \varepsilon_2.$$

Hence $T(g_2) \in K = C(V_1, \overline{V_2})$, so T maps K into itself and all assumptions of Banach's fixed point theorem are satisfied. Consequently, there exists a unique $g \in K$ with

$$G(x, g(x)) = g(x),$$

which, by (10.3), is exactly the desired implicit function. ∎

Example 10.8 We consider the functions from Example 10.1 and use the set

$$M := \{(x,y) \in \mathbb{R}^k \times \mathbb{R}^m \mid f(x,y) = 0\}.$$

(i) $k = m = 1$, $x = x_1$, $y = x_2$, $f(x,y) = x^2 + y^2 - 1$

We have $D_y f(x,y) = 2y$. This "1×1 matrix" is invertible for all $y \neq 0$, which implies that the implicit function g exists for all $(x,y) \in M$ with $y \neq 0$. These correspond precisely to the points with $x \in (-1,1)$. Hence there exists (at least) one continuously differentiable implicit function g on $x = (-1,1)$ (in fact there are exactly two of them).

(ii) $k = 2$, $m = 1$, $x = (x_1, x_2)^T$, $y = x_3$, $f(x,y) = x_1^2 + x_2^2 + y^2 - 1$

Again we have $D_y f(x,y) = 2y$, so the implicit function exists for all $(x,y) \in M$ with $y \neq 0$. These correspond here to points $x \in \mathbb{R}^2$ with $x_1^2 + x_2^2 < 1$.

(iii) $k = 2$, $m = 1$, $x = (p,q)^T$, $y = x$, $f(x,y) = y^2 + x_1 y + x_2$

Here we have $D_y f(x,y) = 2y + x_1$, so g exists for all $(x,y) \in M$ with $2y \neq -x_1$. The set of all $x = (x_1, x_2)$ for which this holds is, however, not so easy to determine in this case.

(iv) $k = 1$, $m = 2$ and

$$f(x,y) = \begin{pmatrix} f_1(x,y) \\ f_2(x,y) \end{pmatrix} = \begin{pmatrix} x^3 + y_1^3 + y_2^3 - 7 \\ xy_1 + y_1 y_2 + y_2 x + 2 \end{pmatrix}.$$

Here

$$D_y f(x,y) = \begin{pmatrix} 3y_1^2 & 3y_2^2 \\ x + y_2 & y_1 + x \end{pmatrix}.$$

If

$$\det(D_y f(x,y)) = 3(y_1^3 + y_1^2 x - y_2^3 - y_2^2 x) \neq 0,$$

then this matrix is invertible. This is the case, for example, at $x = 2$ and $y = (-1,0)^T$, so the equation can be uniquely solved in a neighborhood of $x = 2$.

At this point we have

$$D_y f(x,y) = \begin{pmatrix} 3 & 0 \\ 2 & 1 \end{pmatrix} \quad \Rightarrow \quad (D_y f(x,y))^{-1} = \frac{1}{3} \begin{pmatrix} 1 & 0 \\ -2 & 3 \end{pmatrix}$$

as well as

$$D_x f(x,y) = \begin{pmatrix} 3x^2 \\ y_1 + y_2 \end{pmatrix} = \begin{pmatrix} 12 \\ -1 \end{pmatrix}.$$

Hence

$$Dg(x) = -\frac{1}{3} \begin{pmatrix} 1 & 0 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} 12 \\ -1 \end{pmatrix} = \begin{pmatrix} -4 \\ 9 \end{pmatrix}.$$

□

Let us briefly summarize what the implicit function theorem states and what it does not state.

The theorem says (under the given conditions)

- that the equation can be uniquely solved in a neighborhood $V_1 \times V_2$ of a given solution point (x, y),
- that the solutions can be written as a differentiable function $g : V_1 \to V_2$, and it provides the form of its derivative.

The theorem does *not* say

- how to find a formula for the implicit function g,
- how large the neighborhood V_1 is on which g can be defined,
- how to find a point $(x, y) = (a, b)$ at which the theorem can be applied.

At first this may sound somewhat disappointing, since for practical computations these last questions are often the most important. Nevertheless, the theorem is very important for many applications because it guarantees that a solution exists at all and that it has a "nice" structure, namely that it is the graph of a differentiable function. This forms the basis for many further methods (for example in numerical mathematics) that can then be used to actually compute the solutions.

We conclude this section with an application of the theorem that shows that—even if g itself is unknown—one can still derive interesting information from the implicit function theorem.

Consider functions $f : \mathbb{R}^2 \to \mathbb{R}$. Such functions can always be visualized graphically as a surface over $\mathbb{R}^2$. As on a map, points where f takes the same value can be drawn in $\mathbb{R}^2$ as contour lines. Figure 10.1 illustrates this for the function $f(x) = x_1^2 + 2x_2^2$ from Example 5.2(i).

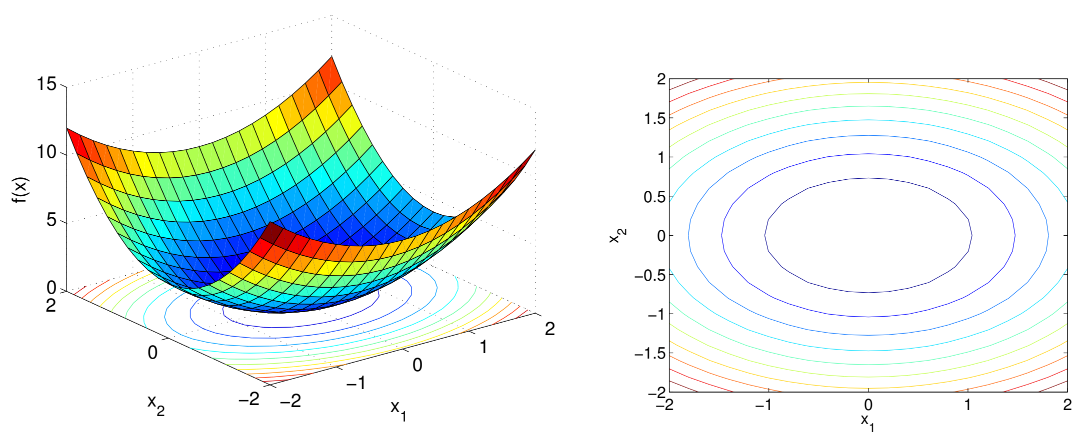

Figure 10.1: Representation of the function $f(x) = x_1^2 + 2x_2^2$. Left: graph and contour lines. Right: contour lines only

Formally, the contour line corresponding to a value $c \in \mathbb{R}$ is defined as

$$H(c) := \{x \in \mathbb{R}^2 \mid f(x) = c\}.$$

In Figure 10.1, all contour lines appear locally to consist of graphs of differentiable functions. This is not a coincidence.

Consider the function

$$F(x_1, x_2) := f(x_1, x_2) - c$$

and a point $x = (x_1, x_2)^T \in H(c)$ at which the derivative $DF(x) = Df(x)$ (or equivalently the gradient $\nabla f(x) = Df(x)^T$) is nonzero.

Then either $D_{x_1} f(x) \neq 0$ or $D_{x_2} f(x) \neq 0$, and we can apply Theorem 10.7 to F either with $x = x_2$ and $y = x_1$ or with $x = x_1$ and $y = x_2$.

Thus we obtain either a differentiable function $x_1 = g(x_2)$ with

$$f(g(x_2), x_2) = c$$

or a differentiable function $x_2 = g(x_1)$ with

$$f(x_1, g(x_1)) = c.$$

Question 10.6 *What does it mean geometrically for two vectors $x, y \in \mathbb{R}^n$ if their scalar product satisfies $\langle x, y \rangle = 0$?*

In both cases the contour line is therefore given locally by a differentiable function.

Besides the interpretation that the gradient always points in the "direction of steepest ascent" of a function, the implicit function theorem provides another geometric interpretation of the gradient.

Since g is differentiable and the function $x_1 \mapsto f(x_1, g(x_1))$ is constant, we obtain

$$\begin{aligned} 0 &= D(f(x_1, g(x_1))) \\ &= D_{x_1} f(x_1, g(x_1)) + D_{x_2} f(x_1, g(x_1))\, g'(x_1) \\ &= \left\langle \nabla f(x_1, g(x_1)), \begin{pmatrix} 1 \\ g'(x_1) \end{pmatrix} \right\rangle . \end{aligned}$$

The second vector in this scalar product is precisely the tangent direction to the contour line at the point x.

Since the scalar product of two vectors is zero exactly when the vectors are orthogonal, it follows that the gradient is always orthogonal to the tangent of the contour line, and hence orthogonal to the contour line itself[3].

10.4 Inverse Functions

Another application of the theorem on implicit functions is the proof of the existence of inverse functions. For real-valued functions $f : D \to \mathbb{R}$, $D \subset \mathbb{R}$, we learned in [7, Section

[3]One says that a vector is orthogonal (or perpendicular) to a graph at a point $(x, g(x))$ if the function g is differentiable at x and the vector is orthogonal to the tangent line at that point.

15.C] a criterion under which an inverse function f^{-1} exists, and in [7, Theorem 16.10] a formula for its derivative was derived. We now want to generalize this to functions $f : D \to \mathbb{R}^n$, $D \subset \mathbb{R}^n$. Note that here we consider functions whose domain and codomain have the same dimension n.

Question 10.7 *Why do we make this restriction? Hint: Think of linear functions $f(x) = Ax$.*

A criterion for the existence of an inverse function in $\mathbb{R}$ is the strict monotonicity of a function. A sufficient condition for strict monotonicity is, according to [7, Theorem 17.6], the inequality $f'(x) > 0$ or $f'(x) < 0$ for all x in an interval $[a, b]$.

Suppose now that at a point $a \in \mathbb{R}$ we know that $f'(a) \neq 0$, and moreover that f' is continuous at a. Then it follows that $f'(x) > 0$ (or $f'(x) < 0$) for all x in a neighborhood V of a. This implies monotonicity on V and thus the existence of an inverse function.

We now want to generalize this criterion to $\mathbb{R}^n$. However, the question arises what the appropriate generalization of $f'(a) \neq 0$ in $\mathbb{R}^n$ is. The natural conjecture $Df(a) \neq 0$ is unfortunately false, as shown for example by $f(x) = (0, x_2)^T$ in $\mathbb{R}^2$ with

$$Df(x) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \neq 0.$$

If an inverse function $g = f^{-1}$ existed for this function, then we would have

$$g(f(x)) = x,$$

and therefore $x_1 = g_1(f(x)) = g_1(0, x_2)$. But this cannot be true, because the function $g_1(f(x)) = g_1(0, x_2)$ does not "know" the value x_1, and therefore cannot return it as a function value.

The correct condition for the existence of an inverse function in a neighborhood of a point $a \in \mathbb{R}^n$ is not $Df(a) \neq 0$, but rather the invertibility of $Df(a)$, as the following theorem shows.

Theorem 10.9 (Existence of the Inverse Function) Let $D \subset \mathbb{R}^n$ be open and $f : D \to \mathbb{R}^n$. Let $a \in D$ be such that f is continuously differentiable in a neighborhood of a and $Df(a) \in \mathbb{R}^{n \times n}$ is invertible.

Then there exist open neighborhoods $V \subset D$ of a and $V' \subset \mathbb{R}^n$ of $f(a)$ with $f(V) \subset V'$ and a continuously differentiable function $f^{-1} : V' \to \mathbb{R}^n$ with $f^{-1}(V') \subset V$ such that

$$f^{-1} \circ f(x) = x \quad \text{for all } x \in V \qquad \text{and} \qquad f \circ f^{-1}(x) = x \quad \text{for all } x \in V'.$$

Moreover, for $b := f(a)$

$$D(f^{-1})(b) = (Df(a))^{-1},$$

where $(Df(a))^{-1}$ denotes the inverse of the Jacobian matrix $Df(a)$.

Proof: Consider the mapping $F : \mathbb{R}^n \times D \to \mathbb{R}^n$ given by

$$F(x, y) := x - f(y).$$

We have $F(b, a) = b - f(a) = 0$, and since $D_yF(x, y) = -Df(y)$, the derivative D_yF is invertible at $y = a$. Therefore Theorem 10.7 can be applied to F, and there exist neighborhoods $V_1 \subset \mathbb{R}^n$ of b and $V_2 \subset D$ of a, as well as a continuously differentiable function $g : V_1 \to V_2$ such that

$$F(x, g(x)) = 0$$

for all $x \in V_1$. If necessary, we replace V_2 by $g(V_1)$ so that $g(V_1) = V_2$ holds. Note that the modified V_2 is still a neighborhood of a.

We now claim that $f^{-1} = g$ is the desired function. First we have

$$f \circ g(x) = x - F(x, g(x)) = x$$

for all $x \in V_1$, which is the second desired property with $V' = V_1$.

For the derivative of g we obtain

$$Dg(b) = -(D_yF(b, a))^{-1}D_xF(b, a) = -(-Df(a))^{-1}I = (Df(a))^{-1},$$

which is exactly the desired formula.

Moreover,

$$g \circ f(x) = g \circ (f \circ g) \circ f(x) = (g \circ f) \circ (g \circ f)(x)$$

for all $x \in V_2$. Note that $(f \circ g)$ can be inserted because $g(V_1) = V_2$, and therefore by the second property already proven $f(V_2) = f \circ g(V_1) = V_1$. Thus for all points $y \in g \circ f(V_2)$ we obtain $g \circ f(y) = y$, i.e. the first desired property. It remains to find a neighborhood V of b such that $V \subset g \circ f(V_2)$.

To do this we first use $f \circ g(V') = V'$, which together with $g(V') \subset V_2$ implies in particular $f(V_2) \supset V'$. Next we use that $g : V' \to V_2$ itself again satisfies the assumptions of the theorem. Hence, by the same argument, there exist neighborhoods $\widetilde{V}_1 \subset V'$ and $\widetilde{V}_2 \subset V_2$ of b and a with

$$g(V') \supset g(\widetilde{V}_1) \supset \widetilde{V}_2.$$

This implies

$$g \circ f(V_2) \supset g(V') \supset \widetilde{V}_2.$$

The set $V = \widetilde{V}_2$ therefore has the desired property. $\square$

Example 10.10 In [7, Section 15.F] we saw that every complex number $z \in \mathbb{C}$ can be written as

$$z = re^{i\varphi} = r(\cos\varphi + i \sin\varphi)$$

with polar coordinates $r \geq 0$ and $\varphi \in \mathbb{R}$. As a consequence, every vector $x = (x_1, x_2)^T \in \mathbb{R}^2$ can be written as

$$x = (r \cos\varphi, r \sin\varphi)^T =: f(r, \varphi).$$

This f is therefore a mapping $f : D \to \mathbb{R}^2$ with $D = (0, \infty) \times \mathbb{R} \subset \mathbb{R}^2$ (we thus regard (r, φ) as a vector in $\mathbb{R}^2$[4]). We now investigate whether this mapping has an inverse.

Since

$$Df(r,\varphi) = \begin{pmatrix} \frac{\partial f_1}{\partial r} & \frac{\partial f_1}{\partial \varphi} \\ \frac{\partial f_2}{\partial r} & \frac{\partial f_2}{\partial \varphi} \end{pmatrix}(r,\varphi) = \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix},$$

we have

$$\det Df(r,\varphi) = r(\cos^2\varphi + \sin^2\varphi) = r > 0$$

for all $(r, \varphi) \in D$ (which is why we removed the points with $r = 0$ from D). Thus f has an inverse function f^{-1} on a suitable neighborhood V for all points in D.

For this inverse we obtain with $x = f(r, \varphi)$

$$D(f^{-1})(x) = \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix}^{-1} = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\dfrac{\sin\varphi}{r} & \dfrac{\cos\varphi}{r} \end{pmatrix}.$$

Using the relations $r = \|x\|_2$, $x_1 = r\cos\varphi$ and $x_2 = r\sin\varphi$, this derivative can be written purely in terms of x without using r and φ:

$$D(f^{-1})(x) = \begin{pmatrix} \dfrac{x_1}{\|x\|_2} & \dfrac{x_2}{\|x\|_2} \\ -\dfrac{x_2}{\|x\|_2^2} & \dfrac{x_1}{\|x\|_2^2} \end{pmatrix}.$$

Note that f can actually only be "inverted" on proper subsets $V \subset D$ with $V \neq D$. If an inverse function existed on the entire $f(D)$, then for all $(r, \varphi) \in D$ we would in particular have

$$f^{-1}(f(r,\varphi)) = (r,\varphi) \quad \text{and} \quad f^{-1}(f(r,\varphi+2\pi)) = (r,\varphi+2\pi). \tag{10.6}$$

However, due to the periodicity of sine and cosine,

$$f(r,\varphi) = f(r,\varphi+2\pi),$$

which implies

$$f^{-1}(f(r,\varphi)) = f^{-1}(f(r,\varphi+2\pi)),$$

contradicting (10.6). □

The inverse function theorem "inherits" all the advantages and disadvantages of the implicit function theorem. In particular, it provides a (local) existence statement and a formula for the derivative of the inverse function, but no explicit formula for the inverse function itself and no information about the size of the neighborhood V on which f^{-1} is defined.

[4]We could also define f on the whole $\mathbb{R}^2$, but we do not need negative r to represent all points $x \in \mathbb{R}^2$. Why we also remove the points with $r = 0$ will become clear shortly.

10.5 Extrema with Constraints

As a final application of the implicit function theorem, we revisit the calculation of extrema from Section 9.3, but this time with an additional difficulty: we want to solve the minimization problem only for those points for which $g(x) = 0$ holds. Formally, this leads to the following extended definition of a minimum.

Definition 10.11 Let $f : D \to \mathbb{R}$ with $D \subset \mathbb{R}^n$ and $g : D \to \mathbb{R}^m$. We say that f has a *local minimum under the constraint* $g(x) = 0$ at a point $x \in D$ if $g(x) = 0$ and there exists $\varepsilon > 0$ such that

$$f(x) \leq f(y) \quad \text{for all } y \in B_\varepsilon(x) \cap D \text{ with } g(y) = 0.$$

The minimum is called *strict* if the above inequality holds with $<$ for $y \neq x$. The point x is called a *(strict) minimizer of* f *under the constraint* $g(x) = 0$.

Analogously, we define the (strict) local maximum under a constraint using $\geq$ and $>$, respectively. □

Question 10.8 *Can a function* $f : \mathbb{R}^n \to \mathbb{R}$ *that has no local minima or maxima (like the saddle point function from Example 9.9(ii) with* A, b, c *from (9.9)) have local minima or maxima under a constraint?*

The following theorem provides necessary optimality conditions analogous to Theorem 9.10.

Theorem 10.12 Let $f \in C^1(D, \mathbb{R})$ with $D \subset \mathbb{R}^n$ open and $g \in C^1(D, \mathbb{R}^m)$, $m \leq n$. If f has a local minimum or maximum under the constraint $g(x) = 0$ at $x^* \in D$ and $Dg(x^*)$ has full rank, then there exists a vector $\lambda \in \mathbb{R}^m$ such that

$$Df(x^*) = \lambda^T Dg(x^*). \tag{10.7}$$

The vector λ is called a *Lagrange multiplier*.

Proof: The matrix $Dg(x^*)$ has full rank and is a matrix with n columns and m rows. Since $m < n$, there exist m linearly independent columns. By renumbering the variables x_i, these columns can be moved; thus we may assume that the last m columns are linearly independent. Note that this only permutes the columns of Df and Dg, so equation (10.7) remains valid under such a renumbering.

We now split the vector x in the form

$$x = \begin{pmatrix} z \\ y \end{pmatrix}$$

with $z \in \mathbb{R}^{n-m}$ and $y \in \mathbb{R}^m$, and write g in the form $g(z, y)$. Then g satisfies the assumptions of the implicit function theorem. Writing the local minimum similarly as $x^* = (z^*, y^*)$, we can find neighborhoods V_1 of z^* and V_2 of y^* and a continuously differentiable function $h : V_1 \to V_2$ such that $g(z, h(z)) = 0$.

From the definition of a local minimum under constraints it follows that z^* is a local minimum of the continuously differentiable function

$$c(z) = f(z, h(z)).$$

Theorem 9.10, equation (8.8), and Theorem 10.3 yield

$$\begin{aligned} 0 = Dc(z^*) &= D_z f(z^*, h(z^*)) + D_y f(z^*, h(z^*)) Dh(z^*) \\ &= D_z f(z^*, h(z^*)) - D_y f(z^*, h(z^*))(D_y g(x^*))^{-1} D_z g(x^*). \end{aligned}$$

Now set

$$\lambda = \left[D_y f(z^*, h(z^*))(D_y g(x^*))^{-1}\right]^T.$$

Then

$$\begin{aligned} D_z f(z^*, h(z^*)) - \lambda^T D_z g(x^*) &= D_z f(z^*, h(z^*)) - D_y f(z^*, h(z^*))(D_y g(x^*))^{-1} D_z g(x^*) \\ &= Dc(x^*) = 0, \end{aligned}$$

and

$$\begin{aligned} D_y f(z^*, h(z^*)) - \lambda^T D_y g(x^*) &= D_y f(z^*, h(z^*)) - D_y f(z^*, h(z^*))(D_y g(x^*))^{-1} D_y g(x^*) \\ &= D_y f(z^*, h(z^*)) - D_y f(z^*, h(z^*)) = 0. \end{aligned}$$

Combining these relations gives

$$Df(x^*) = (D_y f(z^*, y^*), D_z f(z^*, y^*)) = (\lambda^T D_z g(x^*), \lambda^T D_y g(x^*)) = \lambda^T Dg(x^*).$$

□

Remark 10.13 Note that equation (10.7) does not guarantee that $g(x^*) = 0$ holds. Therefore, when computing candidates for extrema under constraints, the condition $g(x^*) = 0$ must be added as an additional equation to determine x^* and λ. □

Question 10.9 *How many scalar equations and how many scalar unknowns do you get when you consider the equation (10.7) together with $g(x^*) = 0$?*

Example 10.14 We consider the last numerical example from Example 9.16, namely $f(x) = x^T Ax + b^T x + c$ with

$$A = \begin{pmatrix} -3 & 1 \\ 1 & 2 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad c = 5.$$

Now we add the constraint $g(x) = 0$ with

$$g(x) = x_1^2 + x_2^2 - 1.$$

Thus we seek the extrema of f on the unit circle, i.e. on the set of all x with $\|x\|_2 = 1$. Here the dimensions are $n = 2$ and $m = 1$, and we have

$$Df(x) = (2Ax + b)^T, \qquad Dg(x) = (2x_1, 2x_2).$$

The condition that Dg has full rank is therefore satisfied for $x \neq 0$. Thus equation (10.7) together with $g(x) = 0$ gives

$$\begin{aligned} -6x_1 + 2x_2 + 1 &= \lambda\, 2x_1, \\ 2x_1 + 4x_2 + 1 &= \lambda\, 2x_2. \end{aligned}$$

Solving this linear system in terms of λ yields

$$x_1 = \frac{1}{2}\frac{\lambda - 1}{\lambda^2 + \lambda - 7}, \qquad x_2 = \frac{1}{2}\frac{\lambda + 4}{\lambda^2 + \lambda - 7}.$$

Substituting into the remaining equation $g(x) = 0$, i.e. $x_1^2 + x_2^2 - 1 = 0$, we obtain

$$\frac{1}{4}\frac{2\lambda^2 + 6\lambda + 17}{(\lambda^2 + \lambda - 7)^2} - 1 = 0,$$

which has as solutions the zeros of the polynomial

$$P(\lambda) = 4\lambda^4 + 8\lambda^3 - 54\lambda^2 - 62\lambda + 179.$$

Computing the roots (with MAPLE) gives the approximate values

$$\lambda_1 = 2.779408674, \quad \lambda_2 = 1.605041266, \quad \lambda_3 = -2.793395794, \quad \lambda_4 = -3.591054146,$$

with the corresponding extrema

$$\begin{pmatrix} 0.2538732892 \\ 0.9672374900 \end{pmatrix}, \quad \begin{pmatrix} -0.1073224411 \\ -0.9942242700 \end{pmatrix}, \quad \begin{pmatrix} 0.9529537485 \\ -0.3031157474 \end{pmatrix}, \quad \begin{pmatrix} -0.9960563225 \\ 0.08872321925 \end{pmatrix}.$$

Figure 9.2 suggests why there are exactly four extrema with this sign structure. □

Chapter 11

The Riemann Integral in $\mathbb{R}^n$

Last update: July 21, 2026

In Section 9.2 we already saw that the integration of functions $f : \mathbb{R} \to \mathbb{R}^m$ can simply be defined componentwise. For vector-valued functions this can be done exactly in the same way as we did there for matrix-valued functions, namely

$$\int_a^b f(x)\,dx = \begin{pmatrix} \int_a^b f_1(x)\,dx \\ \vdots \\ \int_a^b f_m(x)\,dx \end{pmatrix}.$$

A more complicated question is how to integrate functions $f : D \to \mathbb{R}^m$ for $D \subset \mathbb{R}^n$, that is, when the domain D is multidimensional. To clarify this question it suffices to consider the case $m = 1$, since the extension to $m \geq 2$ can again be carried out componentwise.

In this chapter we first consider the Riemann integral. We can define it analogously to the case $D \subset \mathbb{R}$ once we know how to define step functions in $\mathbb{R}^n$ and how to integrate them.

11.1 Step Functions

Step functions in $\mathbb{R}^n$ are again functions that are piecewise constant. In $\mathbb{R}^n$, however, the regions on which the functions are constant are no longer intervals. Instead, one uses rectangular boxes (rectangles) according to the following definition.

Definition 11.1 (i) A set $Q \subset \mathbb{R}^n$ is called a *rectangle* if it is of the form

$$Q = [a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_n, b_n]$$

for numbers $a_i, b_i \in \mathbb{R}$ with $a_i < b_i$ for $i = 1, \ldots, n$. The *interior* of the rectangle is given by

$$\operatorname{int} Q = (a_1, b_1) \times (a_2, b_2) \times \ldots \times (a_n, b_n).$$

(ii) The *volume* of a rectangle is defined as

$$\operatorname{Vol}(Q) := \prod_{i=1}^{n} (b_i - a_i) = (b_1 - a_1) \cdot \ldots \cdot (b_n - a_n).$$

(iii) A family $\mathcal{Q} = \{Q_1, \ldots, Q_q\}$ of $q \in \mathbb{N}$ rectangles is called a *rectangle decomposition* of a set $A \subset \mathbb{R}^n$ if the following holds:

(a) $A = \bigcup_{k=1,\ldots,q} Q_k$

(b) For all $k, j = 1, \ldots, q$ with $k \neq j$ we have $\operatorname{int} Q_k \cap \operatorname{int} Q_j = \emptyset$ (that is, the sets $\operatorname{int} Q_k$ are *pairwise disjoint*).

For a set $A \subset \mathbb{R}^n$ with rectangle decomposition $\mathcal{Q}$ we define

$$\operatorname{Vol}(A) := \sum_{k=1}^{q} \operatorname{Vol}(Q_k). \tag{11.1}$$

(iv) For a rectangle $Q_k = [a_1, b_1] \times \ldots \times [a_n, b_n]$ of the decomposition $\mathcal{Q}$ of A from (iii), we define the *half-open rectangle*

$$Q_k^* = [a_1, b_1) \times [a_2, b_2) \times \ldots \times [a_n, b_n).$$

The union of all these rectangles is denoted by

$$A^* := \bigcup_{k=1,\ldots,q} Q_k^*.$$

□

Question 11.1 *Can you describe the shape of $A \backslash A^*$ in general—or in some special cases?*

Note that the set A in (iii) must always be bounded, since otherwise it cannot be decomposed into finitely many bounded rectangles. It also follows immediately that the boundary of A must consist piecewise of boundaries of rectangles.

Definition (iv) may appear somewhat technical at first glance, but it describes something quite simple: Q_k^* is just Q_k without all boundaries with the larger coordinates (that is, without the right, upper, rear, etc. boundary). The set A^* is simply the set A without the boundaries with the larger coordinates. Note that for every rectangle decomposition of A we obtain the same set A^*.

The reason for this definition is that it allows us to find, for every point $x \in A^*$, a unique rectangle Q_k with $x \in Q_k^*$. This property will significantly simplify several proofs later.

Definition 11.2 (i) A bounded function $f : A \to \mathbb{R}$ is called a *step function* on a set $A \subset \mathbb{R}^n$ if there exists a rectangle decomposition $\mathcal{Q}$ of A and real numbers $c_k \in \mathbb{R}$, $k = 1, \ldots, q$, such that

$$f|_{Q_k^*} \equiv c_k$$

holds for every rectangle $Q_k \in \mathcal{Q}$.

(ii) A bounded function $f : \mathbb{R}^n \to \mathbb{R}$ is called a *step function* on $\mathbb{R}^n$ if there exists a set $A \subset \mathbb{R}^n$ such that $f|_A$ is a step function on A and $f|_{\mathbb{R}^n \backslash A} \equiv 0$.

The sets of step functions defined in this way are denoted by $T(A)$ and $T(\mathbb{R}^n)$. □

Question 11.2 *Does a step function always have exactly q different values? Can it have less—or even more?*

This generalizes Definition 1.1 of step functions in $\mathbb{R}$. The condition that f is constant on Q_k^* (and not only on $\operatorname{int} Q_k$) was not needed in $\mathbb{R}$. In Lemma 11.4 we will see why it is helpful in $\mathbb{R}^n$.

Just as in $\mathbb{R}$, we can further subdivide the decomposition of a step function into smaller rectangles without losing the step function property.

For this reason, as in $\mathbb{R}$, for any two step functions f and g on the same set A we may assume without loss of generality that they are defined on the same rectangle decomposition by refining the original decompositions if necessary so that both functions are constant on the rectangles of the new subdivision. Such a refinement always exists because for any two rectangles Q_1 and Q_2 the sets $Q_1 \cap Q_2$ and $\overline{Q_1 \setminus Q_2}$ can again be represented by rectangles. For this reason, sums and products of step functions in $\mathbb{R}^n$ are again step functions.

Question 11.3 *Consider $Q_1 = [0,3] \times [0,2] \times [0,2]$ and $Q_2 = [2,4] \times [1,3] \times [1,4]$. What is the minimal number of rectangles one needs for a rectangle decomposition of $\overline{Q_1 \setminus Q_2}$?*

The integral of a step function can now be defined exactly analogously to the real-valued case.

Definition 11.3 For a step function $f \in T(A)$ (or $f \in T(\mathbb{R}^n)$ with defining set A) and a rectangle decomposition $\mathcal{Q} = \{Q_1, \ldots, Q_q\}$ of A with $f|_{\operatorname{int} Q_k} \equiv c_k$, we define the integral as

$$\int_A f(x)\,dx := \sum_{k=1}^{q} \operatorname{Vol}(Q_k)\, c_k.$$

□

The intuitive interpretation of this integral is that we compute the volume of the set under the (piecewise constant) graph, with the usual negative sign if f takes negative values.

As in $\mathbb{R}$, one proves that the value of the integral does not depend on the choice of the decomposition $\mathcal{Q}$. Likewise, monotonicity and linearity hold as in $\mathbb{R}$, i.e. Theorem 1.4 remains valid.

Already in differential calculus we used the fact that vector-valued function arguments $x \in \mathbb{R}^n$ can be decomposed into smaller subvectors $y \in \mathbb{R}^p$ and $z \in \mathbb{R}^s$ with $p + s = n$, and instead of writing $f(x)$ we can also write $f(y,z)$. The following lemma shows several properties of step functions under such a representation.

Lemma 11.4 Let $f \in T(A)$ with $A \subset \mathbb{R}^n$ and $A = Y \times Z$ for $Y \subset \mathbb{R}^p$ and $Z \subset \mathbb{R}^s$ with $p + s = n$. Then the following statements hold.

(i) We have $A^* = Y^* \times Z^*$ and the function

$$y \mapsto g(y) := f(y,z)$$

is, for every fixed $z \in Z^*$, a step function on Y.

(ii) The function

$$z \mapsto \int_Y f(y,z)\,dy$$

is a step function on[1] Z.

(iii) We have[2]

$$\int_A f(x)\,dx = \int_Z \int_Y f(y,z)\,dy\,dz.$$

Proof:

(i) Each rectangle $Q_k \in \mathcal{Q}$, $k = 1, \ldots, q$, can be written in the form

$$Q_k = Q_{k,Y} \times Q_{k,Z},$$

where $Q_{k,Y}$ is a p-dimensional rectangle and $Q_{k,Z}$ is an s-dimensional rectangle. With this decomposition we obtain $Q_k^* = Q_{k,Y}^* \times Q_{k,Z}^*$, which in particular implies the identity $A^* = Y^* \times Z^*$. We define $\mathcal{Q}_Y := \{Q_{1,Y}, \ldots, Q_{q,Y}\}$ and $\mathcal{Q}_Z := \{Q_{1,Z}, \ldots, Q_{q,Z}\}$

For $z \in Z^*$ we define the set

$$\mathcal{Q}_Y(z) := \{\, Q_{k,Y} \in \mathcal{Q}_Y \mid k = 1, \ldots, q,\ z \in Q_{k,Z}^* \,\}.$$

This defines a rectangle decomposition of Y (the number of rectangles in $\mathcal{Q}_Y(z)$ is typically smaller than q, but we keep the original numbering). The corresponding sets $Q_{k,Y}^*$ then form a decomposition of Y^*.

For all $Q_{k,Y} \in \mathcal{Q}_Y(z)$ and all $y \in Q_{k,Y}^*$ we have

$$x = \begin{pmatrix} y \\ z \end{pmatrix} \in Q_k^*.$$

Hence

$$g(y) = f(y,z) = f(x) = c_k,$$

so g is a step function.

(ii) For each $\tilde{z} \in Z^*$ we define the family of rectangles

$$\mathcal{Q}_Z(\tilde{z}) := \{\, Q_{k,Z} \in \mathcal{Q}_Z \mid \tilde{z} \in Q_{k,Z}^* \,\} \tag{11.2}$$

and the rectangle

$$Q_Z(\tilde{z}) := \bigcap_{Q_{k,Z} \in \mathcal{Q}_Z(\tilde{z})} Q_{k,Z}.$$

[1] According to (i) the integral is defined for all $z \in Z^*$. For $z \in Z \setminus Z^*$ we may assign arbitrary values to the function, since these values play no role in the definition of step functions.

[2] The following "nested" integral is to be understood as

$$\int_Z \int_Y f(y,z)\,dy\,dz = \int_Z \left(\int_Y f(y,z)\,dy \right) dz.$$

Statement (iii) is also called the "Fubini theorem for step functions".

Since only finitely many rectangles $Q_{k,Z}$ exist, there are also only finitely many different rectangles $Q_Z(\tilde{z})$ even though there are infinitely many $\tilde{z} \in Z^*$. Their interiors are disjoint by construction through intersections. Hence these rectangles form a rectangle decomposition of Z, which we denote by

$$\widetilde{\mathcal{Q}}_Z = \{\, \widetilde{Q}_{j,Z} \mid j = 1, \dots, \tilde{q} \,\}.$$

If $\widetilde{Q}_{j,Z} \in \widetilde{\mathcal{Q}}_Z$, then $\widetilde{Q}_{j,Z} = Q_Z(\tilde{z})$ for some $\tilde{z} \in Z^*$. From the construction of $\mathcal{Q}_Z(\tilde{z})$ it follows that

$$Q_Z(z) = Q_Z(\tilde{z}) = \widetilde{Q}_{j,Z} \quad \text{for all } z \in \widetilde{Q}^*_{j,Z},$$

and therefore

$$f(y,z) = f(y,\tilde{z}) \quad \text{for all } y \in Y^*,\ z \in \widetilde{Q}^*_{j,Z}.$$

Hence

$$z \mapsto \int_Y f(y,z)\,dy$$

is constant in z on $\widetilde{Q}^*_{j,Z}$ and therefore a step function with respect to the decomposition $\widetilde{\mathcal{Q}}_Z$.

(iii) Consider a rectangle $\widetilde{Q}_{j,Z} \in \widetilde{\mathcal{Q}}_Z$. Then for each $\tilde{z} \in \widetilde{Q}_{j,Z}$ we obtain the same family of rectangles $\mathcal{Q}_Z(\tilde{z})$ in (11.2), which we denote by $\mathcal{Q}_Z(j)$ since it depends only on the index j.

For all $z \in \widetilde{Q}^*_{j,Z}$ we therefore have

$$\int_Y f(y,z)\,dy = \sum_{\substack{k=1 \\ Q_{k,Z} \in \mathcal{Q}_Z(j)}}^{q} \mathrm{Vol}(Q_{k,Y})\, c_k.$$

Each rectangle $Q_{k,Z}$ is contained in at least one, possibly several, of the families $\mathcal{Q}_Z(j)$. In any case, $Q_{k,Z}$ is exactly the union of all corresponding rectangles $\widetilde{Q}_{j,Z}$. Therefore

$$\mathrm{Vol}(Q_{k,Z}) = \sum_{\substack{j=1 \\ Q_{k,Z} \in \mathcal{Q}_Z(j)}}^{\tilde{q}} \mathrm{Vol}(\widetilde{Q}_{j,Z}).$$

Thus we obtain

$$
\begin{aligned}
\int_Z \int_Y f(y,z)\,dy\,dz &= \sum_{j=1}^{\tilde{q}} \operatorname{Vol}(\widetilde{Q}_{j,Z}) \sum_{\substack{k=1\\ Q_{k,Z}\in\mathcal{Q}_Z(j)}}^{q} \operatorname{Vol}(Q_{k,Y})\,c_k \\
&= \sum_{k=1}^{q} \underbrace{\sum_{\substack{j=1\\ Q_{k,Z}\in\mathcal{Q}_Z(j)}}^{\tilde{q}} \operatorname{Vol}(\widetilde{Q}_{j,Z})}_{=\operatorname{Vol}(Q_{k,Z})} \operatorname{Vol}(Q_{k,Y})\,c_k \\
&= \sum_{k=1}^{q} \underbrace{\operatorname{Vol}(Q_{k,Z})\operatorname{Vol}(Q_{k,Y})}_{=\operatorname{Vol}(Q_k)}\,c_k \\
&= \sum_{k=1}^{q} \operatorname{Vol}(Q_k) c_k = \int_A f(x)\,dx.
\end{aligned}
$$

$\square$

Remark 11.5 (i) Since the roles of y and z in Lemma 11.4 can simply be interchanged, the statement holds analogously for fixed $y \in Y^*$ for the functions $z \mapsto f(y,z)$ and the corresponding integral. Hence

$$\int_Y \int_Z f(y,z)\,dz\,dy = \int_A f(x)\,dx = \int_Z \int_Y f(y,z)\,dy\,dz.$$

Thus, the order of integration can be interchanged arbitrarily.

(ii) Repeated application of the theorem yields that for a set A which can be decomposed in the form $A = Y_1 \times Y_2 \times \ldots \times Y_N$ and the corresponding vector decomposition $x^T = (y_1^T, y_2^T, \ldots, y_N^T)$, we obtain the equation

$$\int_A f(x)\,dx = \int_{Y_1} \int_{Y_2} \cdots \int_{Y_N} f(y_1, y_2, \ldots, y_N)\,dy_N \ldots dy_2\,dy_1.$$

If A is a rectangle of the form $[a_1, b_1] \times \ldots \times [a_n, b_n]$, this yields

$$\int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_n}^{b_n} f(x_1, x_2, \ldots, x_n)\,dx_n \ldots dx_2\,dx_1,$$

where the x_i are now one-dimensional variables and the integrals are therefore integrals on $\mathbb{R}$. $\square$

11.2 Definition of the Riemann Integral

The Riemann integral can now be defined completely analogously to the real-valued case.

Definition 11.6 Let $f : D \to \mathbb{R}$ be a function, $D \subset \mathbb{R}^n$ open, and let $A \subset D$ be a set for which a rectangle decomposition exists. We then define the *upper integral*

$$\int_A^* f(x)\,dx := \inf \left\{ \int_A \varphi(x)\,dx \;\middle|\; \varphi \in T(A),\ \varphi \geq f \right\}$$

and the *lower integral*

$$\int_{*A} f(x)\,dx := \sup \left\{ \int_A \varphi(x)\,dx \;\middle|\; \varphi \in T(A),\ \varphi \leq f \right\},$$

where the integrals of φ are understood in the sense of Definition 11.3. □

Definition 11.7 We call a function $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$ open, *Riemann integrable* on a rectangle-decomposable set $A \subset D$ if

$$\int_A^* f(x)\,dx = \int_{*A} f(x)\,dx.$$

In this case we define the Riemann integral of f by

$$\int_A f(x)\,dx := \int_A^* f(x)\,dx.$$

□

Exactly as in $\mathbb{R}$ one proves that the integral is a monotone and linear functional, and just as in $\mathbb{R}$ the following theorem follows from the definition and the properties of the supremum and infimum.

Theorem 11.8 A function $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$ open, is Riemann integrable on a rectangle-decomposable set $A \subset D$ if and only if for every $\varepsilon > 0$ there exist step functions $\varphi, \psi \in T(A)$ with $\varphi \leq f \leq \psi$ and

$$\int_A \psi(x)\,dx - \int_A \varphi(x)\,dx \leq \varepsilon.$$

Also in $\mathbb{R}^n$ one can show for various classes of functions that they are Riemann integrable. Clearly, step functions are always Riemann integrable. The following theorem shows that this is also true for all continuous functions.

Theorem 11.9 Let $A \subset \mathbb{R}^n$ be a rectangle-decomposable set. Then every continuous function $f : A \to \mathbb{R}$ is Riemann integrable on A.

Proof: We show that the assumption of Theorem 11.8 is satisfied. Let $\varepsilon > 0$ be arbitrary. Since A is bounded and closed as a finite union of rectangles, A is compact. Hence, by Theorem 7.15, f is uniformly continuous.

Set $\varepsilon' := \varepsilon / \operatorname{Vol}(A)$ and choose $\delta > 0$ as in Definition 7.14 of uniform continuity, using the ∞-norm. Now choose a rectangle decomposition $\mathcal{Q} = \{Q_1, \ldots, Q_q\}$ with rectangles of edge length $< \delta$ (which can always be constructed by sufficiently frequent subdivision of the original rectangle decomposition).

Then for any two points x, x' in a rectangle $Q_k \in \mathcal{Q}$ we have $\|x - x'\|_\infty < \delta$ and therefore $|f(x) - f(x')| < \varepsilon'$. Note that (11.1) also holds for the new rectangle decomposition since the total volume does not change under subdivision.

Define

$$c_k = \min\{f(x) \mid x \in Q_k\}, \qquad c_k' = \max\{f(x) \mid x \in Q_k\}.$$

Then $c_k' - c_k < \varepsilon'$.

Consider the step functions

$$\varphi|_{Q_k^*} \equiv c_k, \qquad \psi|_{Q_k^*} \equiv c_k'.$$

Then $\varphi \le f \le \psi$ and

$$\begin{aligned}
\int_A \psi(x)\,dx - \int_A \varphi(x)\,dx &= \int_A (\psi(x) - \varphi(x))\,dx \\
&= \sum_{k=1}^{q} \operatorname{Vol}(Q_k)(c_k' - c_k) \\
&< \sum_{k=1}^{q} \operatorname{Vol}(Q_k)\varepsilon' \\
&= \operatorname{Vol}(A)\varepsilon' = \varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, the claim follows from Theorem 11.8. $\square$

As in $\mathbb{R}$ (cf. Section 1.3), one can use the construction of the step functions in this proof to compute the integral. However, already in the next section we will learn a method that is much more efficient.

Remark 11.10 For a function $f : \mathbb{R}^n \to \mathbb{R}$ we define

$$\operatorname{supp}(f) := \overline{\{x \in \mathbb{R}^n \mid f(x) \neq 0\}}$$

to be the *support* of the function.

For a continuous function with compact support and for all rectangles $A, \widetilde{A} \subset \mathbb{R}^n$ with $\operatorname{supp}(f) \subset A$ and $\operatorname{supp}(f) \subset \widetilde{A}$ we then have

$$\int_A f(x)\,dx = \int_{\widetilde{A}} f(x)\,dx,$$

since f (and therefore also the integral) is zero outside $\operatorname{supp}(f)$. We therefore also write this integral briefly as

$$\int_{\mathbb{R}^n} f(x)\,dx.$$

$\square$

Question 11.4 *What is the support of the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by*

$$f(x_1, x_2) = \begin{cases} \sqrt{1 - x_1^2 - x_2^2}, & \text{if } \sqrt{1 - x_1^2 - x_2^2} \in \mathbb{R} \\ 0, & \text{if } \sqrt{1 - x_1^2 - x_2^2} \in \mathbb{C} \setminus \mathbb{R} \end{cases} \quad ?$$

What is the geometric meaning of its integral over $\mathbb{R}^2$?

11.3 Fubini's Theorem for Continuous Functions

So far we have defined the integral, but we do not yet have a proper method for computing it easily. The following theorem, together with the subsequent remark, provides the basis for this.

Theorem 11.11 (Fubini's Theorem for Continuous Functions)
Let $A \subset \mathbb{R}^n$ be a rectangle-decomposable set that can be written in the form $A = Y \times Z$ for two rectangle-decomposable sets $Y \subset \mathbb{R}^p$ and $Z \subset \mathbb{R}^s$ with $p + s = n$. Then for every continuous function $f : A \to \mathbb{R}$ the following equality holds:

$$\int_A f(x)\,dx = \int_Z \int_Y f(y, z)\,dy\,dz.$$

In particular, the two functions defined on the right-hand side are Riemann integrable.

Proof: Choose $\varepsilon > 0$ arbitrarily and the functions $\varphi \le f \le \psi$ as in the proof of Theorem 11.9, where we now define[3]

$$\varepsilon' := \min\left\{\frac{\varepsilon}{\mathrm{Vol}(A)}, \frac{\varepsilon}{\mathrm{Vol}(Y)}\right\}.$$

By Lemma 11.4(i), the functions $y \mapsto \varphi(y, z)$ and $y \mapsto \psi(y, z)$ are step functions. By monotonicity of the integral we also have for all $z \in Z$

$$\int_Y \psi(y, z)\,dy - \int_Y \varphi(y, z)\,dy < \mathrm{Vol}(Y)\varepsilon' \le \varepsilon.$$

Hence $y \mapsto f(y, z)$ is Riemann integrable by Theorem 11.8, with

$$\tilde{\varphi}(z) := \int_Y \varphi(y, z)\,dy \le \int_Y f(y, z)\,dy \le \int_Y \psi(y, z)\,dy =: \tilde{\psi}(z).$$

By Lemma 11.4(ii), the functions $\tilde{\varphi}$ and $\tilde{\psi}$ are again step functions. For these we have by Lemma 11.4(iii)

$$\int_Z \tilde{\psi}(z)\,dz - \int_Z \tilde{\varphi}(z)\,dz = \int_A \psi(x) - \varphi(x)\,dx < \mathrm{Vol}(A)\varepsilon' \le \varepsilon.$$

[3]Note that $\mathrm{Vol}(Y) > \mathrm{Vol}(A)$ may occur. For example, consider $A = [0, 1] \times [0, 1/2]$, $Y = [0, 1]$ and $Z = [0, 1/2]$.

Hence $z \mapsto \int_Y f(y,z)\,dy$ is Riemann integrable by Theorem 11.8. From monotonicity of the integral and Lemma 11.4(iii) it then follows that

$$\left|\int_A f(x)\,dx - \int_Z \int_Y f(y,z)\,dy\,dz\right| \leq \int_A \psi(x)\,dx - \underbrace{\int_Z \int_Y \varphi(y,z)\,dy\,dz}_{=\int_A \varphi(x)\,dx} < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, the claimed equality follows. □

Analogous to Remark 11.5, the following observations hold.

Remark 11.12 (i) Since the roles of y and z in Theorem 11.11 can simply be interchanged, the statement holds analogously for fixed $y \in Y$ for the functions $z \mapsto f(y,z)$ and the corresponding integral. Hence

$$\int_Y \int_Z f(y,z)\,dz\,dy = \int_A f(x)\,dx = \int_Z \int_Y f(y,z)\,dy\,dz.$$

Thus we may interchange the order of integration arbitrarily.

(ii) Repeated application of the theorem yields that for a set A that can be decomposed in the form

$$A = Y_1 \times Y_2 \times \ldots \times Y_N$$

and the corresponding vector decomposition $x^T = (y_1^T, y_2^T, \ldots, y_N^T)$, we obtain

$$\int_A f(x)\,dx = \int_{Y_1} \int_{Y_2} \cdots \int_{Y_N} f(y_1, y_2, \ldots, y_N)\,dy_N \ldots dy_2\,dy_1.$$

If A is a rectangle of the form $[a_1, b_1] \times \ldots \times [a_n, b_n]$, this yields

$$\int_A f(x)\,dx = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_n}^{b_n} f(x_1, x_2, \ldots, x_n)\,dx_n \ldots dx_2\,dx_1,$$

where the x_i are now one-dimensional variables and the integrals are therefore real Riemann integrals.

□

Part (ii) of this remark now provides a simple way to reduce integration in $\mathbb{R}^n$ over rectangles A to integration in $\mathbb{R}$. We consider the following examples.

Example 11.13 (i) $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x) = x_1^2 + x_2^2$, $A = [-1,1] \times [-1,1]$

We obtain

$$\begin{aligned}
\int_A f(x)\,dx &= \int_{-1}^{1} \int_{-1}^{1} (x_1^2 + x_2^2)\,dx_2\,dx_1 \\
&= \int_{-1}^{1} \left(x_1^2 x_2 + \frac{x_2^3}{3} \Big|_{-1}^{1} \right) dx_1 \\
&= \int_{-1}^{1} \left(2x_1^2 + \frac{2}{3} \right) dx_1 \\
&= \frac{2x_1^3}{3} + \frac{2x_1}{3} \Big|_{-1}^{1} = \frac{4}{3} + \frac{4}{3} = \frac{8}{3}.
\end{aligned}$$

(ii) **(Volume of the three-dimensional unit sphere)**

The surface of the upper hemisphere with radius 1 in $\mathbb{R}^3$ is described by the function

$$f(x_1, x_2) = \sqrt{1 - x_1^2 - x_2^2}$$

for all $(x_1, x_2)^T \in \mathbb{R}^2$ with $x_1^2 + x_2^2 \leq 1$.

If we define $f(x_1, x_2) = 0$ for all other x_1, x_2, then f is a continuous function with compact support

$$\operatorname{supp}(f) = \{x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1\},$$

hence Riemann integrable, and the volume of the hemisphere is exactly the volume under the graph.

For the integration we may use any rectangle A with $\operatorname{supp}(f) \subset A$; here we choose $A = [-1, 1] \times [-1, 1]$.

For every fixed $x_1 \in \mathbb{R}$ we have $f(x_1, x_2) = \sqrt{1 - x_1^2 - x_2^2}$ if $x_1^2 + x_2^2 \leq 1$ (i.e. $|x_2| \leq \sqrt{1 - x_1^2}$) and $f(x_1, x_2) = 0$ otherwise. Hence

$$\begin{aligned}
\int_{-1}^{1} f(x_1, x_2)\, dx_2 &= \int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} \sqrt{1 - x_1^2 - x_2^2}\, dx_2 \\
&= (1 - x_1^2) \int_{-1}^{1} \sqrt{1 - x_2^2}\, dx_2 \\
&= (1 - x_1^2) \frac{\pi}{2}.
\end{aligned}$$

Here we first performed a substitution $\varphi(x_2) = \sqrt{1 - x_1^2}\, x_2$ and then used the result from Example 2.11(v).

Thus

$$\begin{aligned}
\int_A f(x)\, dx &= \int_{-1}^{1} \int_{-1}^{1} f(x_1, x_2)\, dx_2\, dx_1 \\
&= \int_{-1}^{1} (1 - x_1^2) \frac{\pi}{2}\, dx_1 \\
&= \frac{\pi}{2} \int_{-1}^{1} (1 - x_1^2)\, dx_1 \\
&= \frac{\pi}{2} \left(x_1 - \frac{x_1^3}{3} \right) \Big|_{-1}^{1} = \frac{2}{3}\pi.
\end{aligned}$$

Hence the volume of the whole sphere is $\frac{4}{3}\pi$. □

Remark 11.14 Definition 11.7 so far only allows integration over domains A that can be decomposed into rectangles. Fubini's theorem now provides the possibility to define integrals also on the following more general sets defined inductively.

Let $Y_1 \subset \mathbb{R}^{n_1}$. For each $y_1 \in Y_1$ let a set $Y_2(y_1) \subset \mathbb{R}^{n_2}$ be given. For each pair (y_1, y_2) with $y_1 \in Y_1$ and $y_2 \in Y_2(y_1)$ let a set $Y_3(y_1, y_2) \subset \mathbb{R}^{n_3}$ be given, and so on.

For N such consecutively defined sets and $n = n_1 + \ldots + n_N$ we define

$$A := \{(y_1, \ldots, y_N) \in \mathbb{R}^n \mid y_k \in Y_k(y_1, \ldots, y_{k-1}) \text{ for all } k = 1, \ldots, N\}.$$

For $x^T = (y_1^T, \ldots, y_N^T)$ and $f : A \to \mathbb{R}$ we then define the integral by

$$\int_A f(x)\,dx = \int_{Y_1} \int_{Y_2(y_1)} \cdots \int_{Y_N(y_1,\ldots,y_{N-1})} f(y_1, y_2, \ldots, y_N)\,dy_N \ldots dy_2\,dy_1.$$

If A can be decomposed into rectangles, this coincides with the integral from Definition 11.7 by Fubini's theorem. However, in this way we can represent many further sets (for example circles in $\mathbb{R}^2$ using $n_1 = n_2 = 1$, $Y_1 = [-1, 1]$, $Y_2(y_1) = [-\sqrt{1 - y_1^2}, \sqrt{1 - y_1^2}]$) and integrate over them.

Similarly to the proof of Theorem 11.9, one can show that this integral exists, for example, when A is compact and f is continuous.

Integration over even more general sets is possible with the integral definition of Lebesgue. However, this is not part of this lecture. □

Question 11.5 *Can you write the set $\{x \in \mathbb{R}^2 \mid \|x\|_1 \leq 1\}$ in the form of Remark 11.14?*

11.4 Integrals over Partial Arguments

In Fubini's theorem, integrals appear that are taken only with respect to part of the function arguments. In this section we consider functions that are defined by such integrals.

For this purpose we consider functions of the form

$$f : A \times D \to \mathbb{R}, \quad (x, y) \mapsto f(x, y) \tag{11.3}$$

with $D \subset \mathbb{R}^n$. The set $A \subset \mathbb{R}^m$ is assumed to be rectangle-composable (but the statements could be extended to the sets from Remark 11.14).

Our goal is to derive statements about the continuity and differentiability of the function

$$g : D \to \mathbb{R}, \quad g(y) := \int_A f(x, y)\,dx. \tag{11.4}$$

For both properties we first prove an auxiliary lemma and then the corresponding theorem.

Lemma 11.15 Let f from (11.3) be continuous and let $(y_k)_{k\in\mathbb{N}}$ be a convergent sequence in D with $y^* := \lim_{k\to\infty} y_k \in D$. Then the functions

$$F_k : A \to \mathbb{R}, \quad F_k(x) := f(x, y_k)$$

converge uniformly to the limit function

$$F(x) := f(x, y^*).$$

Proof: By Theorem 7.4 the set

$$M := \{y_k \mid k \in \mathbb{N}\} \cup \{y^*\}$$

is compact. Since A is also compact, the set $A \times M$ is compact as well.

Let $\varepsilon > 0$ be given. We must find $K(\varepsilon)$ such that $|F_k(x) - F(x)| < \varepsilon$ holds for all $k \geq K(\varepsilon)$ and all $x \in A$.

Since f is continuous, it is uniformly continuous on $A \times M$. Thus there exists $\delta > 0$ such that

$$\|(x, y) - (x', y')\|_\infty < \delta \Rightarrow |f(x, y) - f(x', y')| < \varepsilon.$$

Choose $K(\varepsilon)$ such that $\|y_k - y^*\|_\infty < \delta$ for all $k \geq K(\varepsilon)$. Then

$$\|(x, y_k) - (x, y^*)\|_\infty = \|y_k - y^*\|_\infty < \delta,$$

and therefore for all $k \geq K(\varepsilon)$ and all $x \in A$

$$|F_k(x) - F(x)| = |f(x, y_k) - f(x, y^*)| < \varepsilon.$$

Thus the desired uniform convergence holds. □

Theorem 11.16 Let f from (11.3) be continuous. Then the function g from (11.4) is also continuous.

Proof: Let $y \in D$ be arbitrary and let $y_k \to y$ be a sequence in D. With the notation from Lemma 11.15 we have

$$g(y_k) = \int_A F_k(x)\, dx, \qquad g(y) = \int_A F(x)\, dx.$$

By Theorem 4.13, which also holds in $\mathbb{R}^n$ and can be proved exactly as in $\mathbb{R}$, and which can be applied because of the uniform convergence established in Lemma 11.15, we may interchange limit and integration. Hence

$$\lim_{k\to\infty} g(y_k) = \lim_{k\to\infty} \int_A F_k(x)\, dx = \int_A \lim_{k\to\infty} F_k(x)\, dx = \int_A F(x)\, dx = g(y).$$

Thus g is continuous at y. □

We now investigate the differentiability of g defined in (11.4). Again we first formulate an auxiliary statement.

Lemma 11.17 Let f be as in (11.3). For some $j \in \{1, \dots, n\}$ assume that f is continuously partially differentiable with respect to y_j on D. Let $y \in D$ and let $(h_k)_{k\in\mathbb{N}}$ be a real sequence with $h_k \to 0$, $h_k \neq 0$, and

$$y_k := y + h_k e_j \in D$$

for all $k \in \mathbb{N}$, where e_j denotes the j-th unit vector in $\mathbb{R}^n$.

Then the sequence of functions

$$F_k : A \to \mathbb{R}, \quad F_k(x) := \frac{f(x, y_k) - f(x, y)}{h_k}$$

converges uniformly to

$$F : A \to \mathbb{R}, \quad F(x) := \frac{\partial f}{\partial y_j}(x, y).$$

Proof: Write

$$D_j f(x,y) := \frac{\partial f}{\partial y_j}(x,y).$$

Let $\varepsilon > 0$ be arbitrary. Since D is open, there exists $\eta > 0$ with $\overline{B}_\eta(y) \subset D$. Because $D_j f$ is continuous on $A \times D$, it is uniformly continuous on the compact set $A \times \overline{B}_\eta(y)$.

Hence there exists $\delta > 0$ such that for all $x, x' \in A$ and $y' \in \overline{B}_\eta(y)$

$$\|(x,y) - (x',y')\|_\infty < \delta \Rightarrow |D_j f(x,y) - D_j f(x',y')| < \varepsilon.$$

Without loss of generality we may assume $\delta \le \eta$.

Applying the mean value theorem in $\mathbb{R}$ to the function $h \mapsto f(x, y + he_j)$, we obtain for each k a number $\theta_k \in [0,1]$ such that

$$F_k(x) = D_j f(x, y + \theta_k h_k e_j).$$

Choose $K(\varepsilon)$ such that $|h_k| < \delta$ for all $k \ge K(\varepsilon)$. Then

$$\|(y + \theta_k h_k e_j) - y\|_\infty \le |h_k| < \delta,$$

hence $y + \theta_k h_k e_j \in \overline{B}_\eta(y)$.

Therefore for every $x \in A$

$$|F(x) - F_k(x)| = |D_j f(x,y) - D_j f(x, y + \theta_k h_k e_j)| < \varepsilon.$$

Thus the convergence is uniform. □

Theorem 11.18 Let f be as in (11.3). For some $j \in \{1, \ldots, n\}$ assume that f is continuously partially differentiable with respect to y_j on D.

Then the function g from (11.4) is continuously partially differentiable with respect to y_j, and

$$\frac{\partial}{\partial y_j} g(y) = \frac{\partial}{\partial y_j} \int_A f(x,y)\,dx = \int_A \frac{\partial}{\partial y_j} f(x,y)\,dx.$$

Thus we may interchange differentiation and integration, i.e. we may *differentiate under the integral.*

Proof: Choose F_k, F, h_k, and y_k as in Lemma 11.17. Because of the uniform convergence $F_k \to F$, Theorem 4.13 implies

$$\lim_{k\to\infty} \int_A F_k(x)\,dx = \int_A \lim_{k\to\infty} F_k(x)\,dx = \int_A F(x)\,dx = \int_A \frac{\partial}{\partial y_j} f(x,y)\,dx.$$

On the other hand

$$\begin{aligned}\int_A F_k(x)\,dx &= \int_A \frac{f(x,y_k) - f(x,y)}{h_k}\,dx \\ &= \frac{1}{h_k}\left(\int_A f(x,y_k)\,dx - \int_A f(x,y)\,dx\right) \\ &= \frac{g(y + h_k e_j) - g(y)}{h_k}.\end{aligned}$$

Hence for every sequence $h_k \to 0$ with $h_k \neq 0$ the limit

$$\lim_{k\to\infty} \frac{g(y+h_k e_j) - g(y)}{h_k}$$

exists. This is exactly the derivative of the function

$$g_j(h) := g(y + he_j)$$

at $h = 0$, which is the definition of the partial derivative $\partial g/\partial y_j(y)$. Thus the stated formula holds. $\square$

Remark 11.19 Applying Theorem 11.18 to the components of the derivative $D_y g$ yields

$$D_y \int_A f(x,y)\,dx = \int_A D_y f(x,y)\,dx,$$

provided the assumptions of Theorem 11.18 hold for all $j = 1, \ldots, n$, which is the case exactly when $D_y f(x,y)$ exists and is continuous for all $(x,y) \in A \times D$. $\square$

Example 11.20 We compute the (real) integral

$$\int_0^b x^2 \cos x\,dx.$$

This can be done using two integrations by parts, but one can also apply Theorem 11.18 cleverly.

Consider the function

$$g(y) := \int_0^b \cos(xy)\,dx$$

for $y \in D = (1/2, 2)$.

Since $\sin(xy)/y$ is an antiderivative of $\cos(xy)$ with respect to x, we obtain

$$g(y) = \int_0^b \cos(xy)\,dx = \left.\frac{\sin(xy)}{y}\right|_0^b = \frac{\sin(by)}{y}.$$

By Theorem 11.18 with $A = [0,b]$ we have

$$\frac{\partial}{\partial y} g(y) = \int_0^b \frac{\partial}{\partial y}\cos(xy)\,dx = -\int_0^b x\sin(xy)\,dx.$$

Since the integrand is again continuously differentiable in y, we apply the theorem once more:

$$\frac{\partial^2}{\partial y^2} g(y) = -\int_0^b \frac{\partial}{\partial y}\big(x\sin(xy)\big)\,dx = -\int_0^b x^2\cos(xy)\,dx.$$

Hence

$$\int_0^b x^2\cos x\,dx = -g''(1).$$

Because $g(y) = \sin(by)/y$, we compute

$$g'(y) = -\frac{\sin(by)}{y^2} + \frac{b\cos(by)}{y},$$

and

$$g''(y) = \frac{2\sin(by)}{y^3} - \frac{2b\cos(by)}{y^2} - \frac{b^2\sin(by)}{y}.$$

Therefore

$$\int_0^b x^2 \cos x \, dx = -2\sin b + 2b\cos b + b^2 \sin b.$$

$\square$

Question 11.6 *For the function $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = \frac{\exp(x)}{\sin(x)\cos(x)}$, which formula is easier to evaluate:*

$$\frac{\partial}{\partial y} \int_{[-1,1]} f(x,y)\, dx \quad \text{or} \quad \int_{[-1,1]} \frac{\partial}{\partial y} f(x,y)\, dx?$$

11.5 The Transformation Formula

One of the most important rules for computing integrals in $\mathbb{R}$ is the substitution formula. Under the assumptions of Theorem 2.10, the formula

$$\int_a^b f(\varphi(x))\varphi'(x)\, dx = \int_{\varphi(a)}^{\varphi(b)} f(y)\, dy$$

holds.

We now state—due to time constraints without giving a complete proof—the corresponding formula for $\mathbb{R}^n$, which is also called the substitution formula but more often the *transformation formula*.

Theorem 11.21 (Transformation Theorem) Let $U \subset \mathbb{R}^n$ be an open set and let $\Phi \in C^1(U, \mathbb{R}^n)$ be an invertible function. Let $A \subset U$ be such that $f : \Phi(U) \to \mathbb{R}$ is integrable over $\Phi(A)$. Then $f \circ \Phi \,|\det(D\Phi)|$ is integrable over A and

$$\int_A f(\Phi(x))\, |\det(D\Phi(x))|\, dx = \int_{\Phi(A)} f(y)\, dy.$$

The version of this theorem stated here appears in the book *Analysis 3* by O. Forster [4, §13] and is stated for the Lebesgue integral. Other books give different formulations with slightly different assumptions (for example, instead of invertibility one may assume injectivity and $\det(D\Phi(x)) \neq 0$ for all $x \in U$).

The proof in the general case is rather complicated. To at least explain the idea, we consider the case $n = 2$, $U = A = \mathbb{R}^2$, f a step function from $T(\mathbb{R}^2)$, and $\Phi(x) = Bx$ for a matrix $B \in \mathbb{R}^{2\times 2}$ with $\det(B) \neq 0$.

In this case Φ is a linear mapping and

$$D\Phi(x) = B \quad \text{for all } x \in \mathbb{R}^2.$$

The function $f \circ \Phi$ is then constant with value c_k on $\operatorname{int} P_k$ for the sets

$$P_k := \{x \in \mathbb{R}^2 \mid Bx \in Q_k\},$$

where the rectangles

$$Q_k = [a_{k1}, b_{k1}] \times [a_{k2}, b_{k2}]$$

belong to the rectangle decomposition $\mathcal{Q}$ associated with f.

Hence

$$\int_{\mathbb{R}^2} f \circ \Phi(x)\, dx = \sum_{k=1}^{q} \operatorname{Vol}(P_k)\, c_k.$$

Since $\det(B) \neq 0$, the matrix B is invertible and we obtain

$$P_k = B^{-1} Q_k.$$

This set P_k is a parallelogram with vertices

$$p_{ij} = B^{-1} x_{ij},$$

where the x_{ij}, $i, j = 1, 2$, are the vertices of Q_k.

This quadrilateral is a parallelogram and therefore has area

$$\begin{aligned}
\left|\det\left((p_{21} - p_{11},\, p_{12} - p_{11})\right)\right| &= \left|\det\left(B^{-1}(x_{21} - x_{11},\, x_{12} - x_{11})\right)\right| \\
&= |\det(B^{-1})|\, \left|\det\left((x_{21} - x_{11},\, x_{12} - x_{11})\right)\right| \\
&= \frac{1}{|\det(B)|} \operatorname{Vol}(Q_k).
\end{aligned}$$

Thus

$$\int_{\mathbb{R}^2} f \circ \Phi(x)\, |\det(D\Phi(x))|\, dx = \sum_{k=1}^{q} \operatorname{Vol}(P_k)\, |\det(B)|\, c_k = \sum_{k=1}^{q} \operatorname{Vol}(Q_k)\, c_k = \int_{\mathbb{R}^2} f(y)\, dy.$$

This proof can be carried out in $\mathbb{R}^n$ in a very similar way (using boxes Q_k and parallelotopes P_k). For general integrands f, one can approximate them by step functions.

The most difficult part is the treatment of nonlinear functions Φ. In this case the function Φ^{-1} must be approximated on sufficiently small boxes by its derivative. One can then proceed similarly to the above argument, but one must prove that the summed approximation errors remain sufficiently small (see Forster [4], §2 and §13).

Alternatively, the theorem can be derived by induction using Fubini's theorem and reducing the statement to substitution in $\mathbb{R}$ (Heuser II [9], Section 205).

Example 11.22 The transformation from polar coordinates to Cartesian (i.e. the usual) coordinates is given by the mapping

$$\Phi : (0, \infty) \times [0, 2\pi) \to \mathbb{R}^2, \qquad \Phi(r, \varphi) := \begin{pmatrix} r\cos\varphi \\ r\sin\varphi \end{pmatrix}.$$

For the function

$$f(x) = \|x\|_2 = \sqrt{x_1^2 + x_2^2}$$

we obtain for example

$$f \circ \Phi(r, \varphi) = \sqrt{r^2\cos^2\varphi + r^2\sin^2\varphi} = r.$$

On $(0, \infty) \times [0, 2\pi)$ we have

$$D\Phi(r, \varphi) = \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix}$$

and therefore

$$\det(D\Phi(r, \varphi)) = r\cos^2\varphi + r\sin^2\varphi = r.$$

Suppose we want to compute the integral of f over the disk

$$K := \{x \in \mathbb{R}^2 : \|x\| \le 1\}.$$

Except for the point 0, which does not contribute to the value of the integral, we can represent this disk as $\Phi(A)$ with

$$A = (0, 1] \times [0, 2\pi).$$

Using the transformation formula and Fubini's theorem we obtain

$$\int_K f(x)\,dx = \int_A f(\Phi(r, \varphi))|\det(D\Phi(r, \varphi))|\,d(r, \varphi) = \int_A r^2\,d(r, \varphi).$$

Hence

$$\int_0^{2\pi}\int_0^1 r^2\,dr\,d\varphi = \frac{2\pi}{3}.$$

$\square$

Chapter 12

An Introduction to Vector Analysis

In this final chapter of this lecture we give an introduction into the main concepts and some of the main results from vector analysis, a branch of analysis that combines analysis and geometry. Unlike the previous chapters, we will mostly describe and illustrate the results by examples, but not provide full proofs, because they would exceed the time frame for this lecture.

12.1 Submanifolds of $\mathbb{R}^n$

Submanifolds of $\mathbb{R}^n$ generalize the idea of subspaces to objects that are not necessarily flat but can be curved. Formally, they consist of surface pieces, which are images of differentiable maps from lower dimensional spaces into $\mathbb{R}^n$. Here and in what follows we use the convention that d is a natural number with $1 \leq d < n$ and k is either a natural number ≥ 1 or ∞.

Definition 12.1 Let $W \subset \mathbb{R}^d$ be open, $\Phi \in C^k(W; \mathbb{R}^n)$ with

(i) Φ is injective,

(ii) for all $y \in W$, $\operatorname{rank}(D\Phi(y)) = d$, i.e., the rank of the Jacobian is maximal,

(iii) $\Phi^{-1} : \Phi(W) \to \mathbb{R}^d$ is continuous.

Then the set $\Phi(W)$ is called a d-dimensional C^k *surface piece* with *chart* Φ. □

Definition 12.2 A set $M \subset \mathbb{R}^n$ is called a d-dimensional C^k submanifold of $\mathbb{R}^n$ if for all $p \in M$ there exists an open set $U \subset \mathbb{R}^n$ with $p \in U$, such that $U \cap M$ is a d-dimensional C^k surface piece.

A family of charts Φ_i, $i \in I$, that satisfy the conditions of Definition 12.1 with sets W_i, are called an *atlas* of M, if $M = \bigcup_{i \in I} \Phi_i(W_i)$. □

Example 12.3 $\mathbb{S}^1 = \{x \in \mathbb{R}^2 \,|\, \|x\|_2 = 1\}$ is a 1-dimensional C^∞ submanifold of $\mathbb{R}^2$. We give two different atlases that show that this is true:

Atlas 1 consists of the maps Φ_1, Φ_2, Φ_3, $\Phi_4 : (-1,1) \to \mathbb{R}^2$ given by

$$\Phi_1(y) = \begin{pmatrix} y \\ \sqrt{1-y^2} \end{pmatrix}, \quad \Phi_2(y) = \begin{pmatrix} y \\ -\sqrt{1-y^2} \end{pmatrix}$$

$$\Phi_3(y) = \begin{pmatrix} \sqrt{1-y^2} \\ y \end{pmatrix}, \quad \Phi_4(y) = \begin{pmatrix} -\sqrt{1-y^2} \\ y \end{pmatrix}.$$

Atlas 2 is given by the maps

$$\Phi_1 : (-\pi, \pi) \to \mathbb{R}^2, \;\; y \mapsto \begin{pmatrix} \cos y \\ \sin x \end{pmatrix}$$

and

$$\Phi_2 : (0, 2\pi) \to \mathbb{R}^2, \;\; y \mapsto \begin{pmatrix} \cos y \\ \sin x \end{pmatrix}.$$

$\square$

Question 12.1 *Can you find an atlas for $\mathbb{S}^1$ that consists of only one map Φ and one open set $W \subset \mathbb{R}$?*

Example 12.4 This example describes some common classes of submanifolds.

(i) Graphs: Let $W \subset \mathbb{R}^d$ open, $f \in C^k(W, \mathbb{R}^{n-d})$. Then

$$M := \left\{ \begin{pmatrix} y \\ f(y) \end{pmatrix} \middle| \, y \in W \right\}$$

is a d-dimensional C^k submanifold of $\mathbb{R}^n$. It's atlas conists of the single map $\Phi : W \to \mathbb{R}^n$, $\Phi(y) = \begin{pmatrix} y \\ f(y) \end{pmatrix}$: Φ is injective, because $y_1 \neq y_2$ implies $\begin{pmatrix} y_1 \\ f(y_1) \end{pmatrix} \neq \begin{pmatrix} y_2 \\ f(y_2) \end{pmatrix}$ regardless of what f is, $D\Phi(y) = \begin{pmatrix} \mathrm{Id}_{\mathbb{R}^d} \\ Df(y) \end{pmatrix}$ has rank d since the first d rows are linearly independent and $\Phi^{-1}\left(\begin{pmatrix} y \\ f(y) \end{pmatrix}\right) = y$ is easily seen to be continuous.

(ii) Sets of roots or zeros: Let $D \subset \mathbb{R}^n$ open, $F \in C^k(D, \mathbb{R}^{n-d})$ with $\mathrm{rank}(DF(x)) = n - d$ for all $x \in M := \{x \in D \,|\, F(x) = 0\}$. Given a $p \in M$, similar to the proof of Theorem 10.12, we can assume that the last $n - d$ columns are linearly independent and split x into y and z. Then according to the Implicit Function Theorem 10.7, in a neighborhood U of p the points on M are uniquely described as the graph $(y, g(y))$ for $g \in C^k(W, \mathbb{R}^{n-d})$ (note that y now plays the role of x in the implicit function theorem). As the derivative of g has full rank in y corresponding to p, according to (a) this shows that $\Phi(y) = \begin{pmatrix} y \\ g(y) \end{pmatrix}$ is a chart and $M \cap U$ is a d-dimensional C^k surface piece. Hence, M is a d-dimensional C^k submanifold of $\mathbb{R}^n$.

(iii) Curves: This special case appears if $d = 1$, i.e., the manifold is one-dimensional. If $\gamma \in C^k([a,b], \mathbb{R}^n)$ is injective with $D\gamma(y) \neq 0$ for all $y \in [a,b]$, then

$$M := \gamma((a,b))$$

is a one-dimensional C^k submanifold of $\mathbb{R}^n$. □

Definition 12.5 For an d-dimensional submanifold M of $\mathbb{R}^n$ and a point $p \in M$ with chart Φ in a neighborhood of p, we define the *tangent space* of M at p as

$$T_pM := \operatorname{span}\left(\left\{\frac{\partial \Phi}{\partial y_1}(\Phi^{-1}(p)), \ldots, \frac{\partial \Phi}{\partial y_d}(\Phi^{-1}(p))\right\}\right).$$

Its orthogonal complement $N_pM := (T_pM)^\perp$ is called the *normal space* of M at p. □

Example 12.6 We consider the sphere with the second atlas and $p = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. This point is covered by the first chart and it holds that $p = \Phi_1(0)$. We get that

$$T_pM := \operatorname{span}\left(\left\{\frac{\partial \Phi_1}{\partial y}(0)\right\}\right) = \operatorname{span}\begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

When drawing the affine space $\{p\} + T_pM$, one sees that it is tangent to M in p. As this is always the case for the tangent space, this is the reason for its name. □

Question 12.2 *Draw the 1d unit sphere* $\mathbb{S}^1$ *in* $\mathbb{R}^2$ *and visualize the tangent space and the normal space in three different points* p *using the corresponding affine spaces* $\{p\} + T_pM$ *and* $\{p\} + N_pM$*.*

Remark 12.7 (i) T_pM and N_pM are subspaces of $\mathbb{R}^n$ with $\dim T_pM = d$ and $\dim N_pM = n - d$.

(ii) T_pM and N_pM are independent of the chart that is used for its computation (the reason will be explained after this remark).

(iii) $\{p\} + T_pM$ is tangential to M in p (cf. Example 12.6)

(iv) For $M = \{x \in D \,|\, F(x) = 0\}$ as in Example 12.4(ii), it holds that

$$N_pM = \operatorname{span}(\{\nabla F_1(p), \ldots, \nabla F_{n-d}(p)\}), \quad T_pM = \ker(DF(p)).$$

□

The fact that usually more that one chart is needed to describe a manifold implies that the images of two charts may overlap, see also Example 12.3. Let $\Phi_1 : W_1 \to \mathbb{R}^n$ and $\Phi_2 : W_2 \to \mathbb{R}^n$ be two charts with

$$V := \Phi_1(W_1) \cap \Phi_2(W_2) \neq \emptyset.$$

Then the maps

$$\begin{aligned}\Psi_{12} :&\Phi_1^{-1}(V) \to \Phi_2^{-1}(V), \quad y \to \Phi_2^{-1}(\Phi_1(y))\\ \Psi_{21} :&\Phi_2^{-1}(V) \to \Phi_1^{-1}(V), \quad y \to \Phi_1^{-1}(\Phi_2(y))\end{aligned}$$

map the point $y_1 \in W_1$ with $\Phi_1(y_1) = p$ to the point $y_2 \in W_2$ with $\Phi_2(y_2) = p$, and vice versa. These two maps are bijective, hence invertible, with $\Psi_{12}^{-1} = \Psi_{21}$ and both Ψ_{12} and Ψ_{21} are C^k maps. These maps are called the *changes of charts*. Using these maps one can show that the tangent space is indeed independent of the choice of the chart: From the equality $\Phi_1(\Psi_{21}(y)) = \Phi_2(y)$ it follows using the chain rule that

$$D\Phi_2(y) = \frac{d}{dy}\Phi_2(y) = \frac{d}{dy}(\Phi_1(\Psi_{21}(y))) = D\Phi_1(\Psi_{21}(y))D\Psi_{21}(y)$$

and the fact that $D\Psi_{21}(y)$ is invertible, it follows that $D\Phi_2(y)$ and $D\Phi_1(\Psi_{21}(y))$span the same space.

Question 12.3 *Visualize the set where $\Phi_1((-1,1))$ and $\Phi_3((-1-1))$ from Atlas 1 in Example 12.3 overlap. What is Ψ_{13} in this example?*

12.2 Integration on Submanifolds

We now define integrals of functions $f : M \to \mathbb{R}$, which are defined on a submanifold M. This is done in the following definition.

Definition 12.8 Let M be a d-dimensional C^1 submanifold of $\mathbb{R}^n$, $f : M \to \mathbb{R}$ a map, and $\Phi : W \to \mathbb{R}^n$ a chart of M.

(i) The function $g_\Phi : W \to [0,\infty)$, $g_\Phi(y) := \det\left(D\Phi(y)^T D\Phi(y)\right)$ is called the *Gramian determinant* of Φ.

(ii) If $M = \Phi(W)$ (i.e., the atlas has only one chart) and $f : M \to \mathbb{R}$, we define the *integral of f over the set M by*

$$\int_M f(x)dS(x) := \int_W f(\Phi(y))\sqrt{g_\Phi(y)}dy,$$

provided that the integral on the right hand side exists.

(iii) If the atlas has more than one chart, we divide M into disjoint sets that are contained in the image of a single chart each, and sum the integrals according to (ii) over these sets.

□

Sometimes the argument "(x)" is dropped, i.e., one writes $\int_M f dS$ instead of $\int_M f(x)dS(x)$.

Example 12.9 Computation of Volumes: We can compute the volume of the manifold as

$$\mathrm{Vol}_d(M) = \int_M 1 dS(x).$$

Here a volume for $d = 1$ is a length and a volume for $d = 2$ is an area. A concrete example is the length of the circle with radius 1. According to Example 12.3 the half circle M_h is covered by the chart Φ_1 from Atlas 2 as $\Phi_1([-\pi/2, \pi/2])$. We have that

$$g_{\phi_1}(y) = \det\left((-\sin x, \cos x)\begin{pmatrix} -\sin y \\ \cos y \end{pmatrix}\right) = \det(\sin^2 y + \cos^2 y) = \det(1) = 1.$$

This implies that

$$\mathrm{Vol}_1(M_h) = \int_{M_h} 1 dS(x) = \int_{-\pi/2}^{\pi/2} 1 \cdot 1 dy = \int_{-\pi/2}^{\pi/2} 1 dy = \pi,$$

which implies that the length of the whole circle is 2π. □

Remark 12.10 (i) The value of the integral does not depend on the chosen chart. This can be proved using the change of charts Ψ_{ij} and the Transformation Theorem 11.21.

(ii) Just as the determinant appearing in the Transformation Theorem 11.21, the Gramian determinant g_Φ can be interpreted as an area transformation factor: Consider the case that $d = 2$ and $\Phi(y) = Ay$, $A = (v_1, v_2)$, $v, w \in \mathbb{R}^n$. Then a rectangle Q in $\mathbb{R}^2$ becomes a 2-dimensional parallelogram $P = \Phi(Q)$ in $\mathbb{R}^n$ and $f_\phi(y) = \det(D\Phi(y)^T D\Phi(y)) = \det(A^T A)$ is the squared area of this parallelogram.

(iii) In case $d = 1$, on writes $\int_M f(x) ds(x)$ instead of $\int_M f(x) dS(x)$. □

Example 12.11 We present some special cases of the Gramian determinant:

(i) If $d = 1$: $\sqrt{g_\Phi(y)} = \|D\Phi(y)\|_2$ (the Euclidean norm of the $n \times 1$-matrix $D\Phi(y)$ interpreted as a vector in $\mathbb{R}^n$)

(ii) If $d = 2, n = 3$: Using the cross-product for two vectors $v, w \in \mathbb{R}^3$, which is defined by

$$v \times w = \begin{pmatrix} v_2 w_3 - v_3 w_2 \\ v_3 w_1 - v_1 w_3 \\ v_1 w_2 - v_2 w_1 \end{pmatrix}$$

one obtains that

$$\sqrt{g_\Phi(y)} = \left\| \frac{\partial \Phi}{\partial y_1} \times \frac{\partial \Phi}{\partial y_2} \right\|_2.$$

Question 12.4 *Find out what $v \times w$ means geometrically. You may remember this from physics lessons at school, otherwise wikipedia or a large language model may yield the answer.*

(iii) If $d = 2, n = 3$ and Φ is a graph of the form $\Phi(y) = \begin{pmatrix} y \\ h(y) \end{pmatrix}$ with $h : W \to \mathbb{R}$, $W \subset \mathbb{R}^2$, we obtain from (ii) that

$$\sqrt{g_\Phi(y)} = \left\| \begin{pmatrix} 1 \\ 0 \\ \frac{\partial h}{\partial y_1} \end{pmatrix} \times \begin{pmatrix} 0 \\ 1 \\ \frac{\partial h}{\partial y_2} \end{pmatrix} \right\|_2 = \left\| \begin{pmatrix} -\frac{\partial h}{\partial y_1} \\ -\frac{\partial h}{\partial y_2} \\ 1 \end{pmatrix} \right\|_2 = \sqrt{1 + \|\nabla h(y)\|_2^2}.$$

□

12.3 Oriented Integrals on Submanifolds

In this section we briefly explain two concepts of integrals for functions with codomains $\subset \mathbb{R}^n$ on submanifolds in $\mathbb{R}^n$. We will look at the case $d = 1$ and $n \geq 2$ arbitrary, which leads to the concept of *line* or *path integral*, and the case $d = 2$ and $n = 3$, which we will interpret as the flow through a surface. We start with the first case.

Definition 12.12 Let $a < b$ be real numbers, $\gamma \in C^1([a,b], \mathbb{R}^n)$ and $F \in C(D, \mathbb{R}^n)$ with $\gamma((a,b)) \subset D \subset \mathbb{R}^n$. Then we call

$$\int_\gamma F(x) \cdot d\vec{s}(x) := \int_a^b \langle F(\gamma(y)), D\gamma(y)\rangle dy$$

the *line integral* (or *path integral*) of F along γ. □

Remark 12.13 (i) The reason for the term *oriented* is the following. For the integral for $f : \mathbb{R} \to \mathbb{R}$ we have the relation

$$\int_a^b f(x)dx = -\int_b^a f(x)dx,$$

i.e., it does make a difference whether we integrate forward or backward. This is not the case for the one-dimensional surface integral from Definition 12.8. Consider this integral for $d = 1$ and a manifold M that has an atlas with a single chart γ on $W = (0,1)$. Then

$$g_\Phi(y) = \det D\gamma(y)^T D\gamma(y) = \|D\gamma(y)\|_2^2. \tag{12.1}$$

This yields

$$\int_M f(x)ds(x) = \int_0^1 f(\Phi(y))\sqrt{g_\gamma(y)}dy = \int_0^1 f(\gamma(y))\|D\gamma(y)\|_2 dy.$$

If we parametrize M in the opposite direction, i.e., if we choose $\tilde{\gamma}(y) = \gamma(1-y)$, then we have $D\tilde{\gamma}(y) = -D\gamma(1-y)$ and we obtain

$$\begin{aligned}\int_M f(x)ds(x) &= \int_0^1 f(\Phi(y))\sqrt{g_{\tilde{\gamma}}(y)}dy = \int_0^1 f(\tilde{\gamma}(y))\|D\tilde{\gamma}(y)\|_2 dy \\ &= \int_0^1 f(\gamma(1-y))\|D\gamma(1-y)\|_2 dy = \int_0^1 f(\gamma(y))\|D\gamma(y)\|_2 dy,\end{aligned}$$

i.e., exactly the same value. Hence, the definition of the surface integral does not take into account the orientation, i.e., the direction in which we run through the curve.

Question 12.5 *How does the observation that was just made relate to Remark 12.10(i)?*

The oriented line integral, in contrast, satisfies

$$\begin{aligned}&\int_{\tilde{\gamma}} F(x) \cdot d\vec{s}(x) = \int_a^b \langle F(\tilde{\gamma}(y)), D\tilde{\gamma}(y)\rangle dy = \int_a^b \langle F(\gamma(1-y)), -D\gamma(1-y)\rangle dy \\ &= -\int_a^b \langle F(\gamma(1-y)), D\gamma(1-y)\rangle dy = -\int_a^b \langle F(\gamma(y)), D\gamma(y)\rangle dy = -\int_\gamma F(x) \cdot d\vec{s}(x).\end{aligned}$$

Hence, the change of orientation of γ changes the sign of the integral.

(ii) Apart from the orientation, however, the value of the line integral does not depend on the choice of γ. Using the chain rule, one computes that for any $\tilde{\gamma} \in C^1([a,b],\mathbb{R}^n)$ with $\tilde{\gamma}((a,b)) = \gamma((a,b))$, $\tilde{\gamma}(a) = \gamma(a)$, and $\tilde{\gamma}(b) = \gamma(b)$ the value of the line integral remains the same.

(iii) A physical interpretation of the line integral is obtained if F is a force field (e.g., the gravitation field). Then $\int_\gamma F(y) \cdot d\vec{s}(y)$ describes the amount of energy that is lost when moving along the curve γ.

(iv) The set $\gamma((a,b))$ need not be a submanifold, because neither injectivity of γ nor $D\gamma(y) \neq 0$ is needed for Definition 12.12.

(v) If, however, $M = \gamma((a,b))$ is a submanifold, then there is a relation between the line integral from Definition 12.12 and the surface integral from Definition 12.8. It holds that

$$\int_\gamma F(x) \cdot d\vec{s}(x) = \int_M \langle F(x), \tau(x)\rangle ds(x),$$

where

$$\tau(x) = \frac{D\gamma(\gamma^{-1}(x))}{\|D\gamma(\gamma^{-1}(x))\|_2} \quad \text{for } x \in M = \gamma((a,b))$$

is called the unit tangent field of γ.

This identity follows from

$$\begin{aligned}\int_\gamma F(x) \cdot d\vec{s}(x) &= \int_a^b \langle F(\gamma(y)), D\gamma(y)\rangle dy = \int_a^b \langle F(\gamma(y)), \tau(\gamma(y))\rangle \|D\gamma(y)\|_2\rangle dy \\ &= \int_a^b \langle F(\gamma(y)), \tau(\gamma(y))\rangle \sqrt{g_\Phi(y)} dy = \int_M \langle F(x), \tau(x)\rangle ds(x),\end{aligned}$$

where we used Definition 12.12 in the first equality, the definition of τ and the fact that $D\gamma(y) \neq 0$ in the second equality, the identity (12.1) in the third equality and Definition 12.8 in the last equality. □

Now we look at the case $d = 2$ and $n = 3$, i.e., M is a two-dimensional submanifold of $\mathbb{R}^3$. Here we use the following definition.

Definition 12.14 (i) A 2-dimensional C^1 submanifold of $\mathbb{R}^3$ is called *orientable*, if there is a continuous map $\nu : M \to \mathbb{R}^3$ such that $\|\nu(p)\|_2 = 1$ and $\nu(p) \in N_pM$ for all $p \in M$. The map ν is then called a *unit normal field* and defines the orientation on M.

(ii) A chart $\Phi : W \to \mathbb{R}^3$ of a surface piece of $\Phi(W) \subset M$ is called *compatible* with the orientation defined by ν if

$$\nu(\Phi(y)) = \frac{\frac{\partial\Phi}{\partial y_1}(y) \times \frac{\partial\Phi}{\partial y_2}(y)}{\left\|\frac{\partial\Phi}{\partial y_1}(y) \times \frac{\partial\Phi}{\partial y_2}(y)\right\|_2}$$

for all $y \in W$.

Question 12.6 *How does this definition fit the geometric interpretation of the cross product, cf. Question 12.4?*

(iii) If $M = \Phi(W)$ (i.e., the atlas has only one chart) and Φ satisfies (ii), then for $F \in C(M, \mathbb{R}^3)$ we define

$$\int_M F(y) \cdot d\vec{S}(y) := \int_W \left\langle F(\Phi(y)), \frac{\partial \Phi}{\partial y_1}(y) \times \frac{\partial \Phi}{\partial y_2}(y) \right\rangle dy.$$

(iii) If the atlas has more than one chart, then as in Definition 12.8 we divide M into disjoint sets that are contained in the image of a single chart each, and sum the integrals according to (iii) over these sets. □

Remark 12.15 (i) The physical interpretation of this integral is as follows: If F is a velocity field, then $\int_{\Phi(W)} F(y) \cdot d\vec{S}(y)$ is the volume flow (=volume/time) through $\Phi(W)$.

(ii) If the chart is not compatible with ν, then it must be compatible with $-\nu$. It can thus still be used to compute the integral, but the sign changes.

(iii) The value of the integral does not depend on Φ except for the sign.

(iv) The oriented integral relates to the surface integral from Definition 12.8 as follows:

$$\int_M F(y) \cdot d\vec{S}(y) = \int_M \langle F(y), \nu(y) \rangle dS(y).$$

This holds since

$$\nu(\Phi(y)) = \frac{\frac{\partial \Phi}{\partial y_1}(y) \times \frac{\partial \Phi}{\partial y_2}(y)}{\left\| \frac{\partial \Phi}{\partial y_1}(y) \times \frac{\partial \Phi}{\partial y_2}(y) \right\|_2} \quad \text{and} \quad \sqrt{g_\Phi(y)} = \left\| \frac{\partial \Phi}{\partial y_1}(y) \times \frac{\partial \Phi}{\partial y_2}(y) \right\|_2$$

because of Example 12.11(ii).

(v) An example for a non-orientable 2d submanifold in $\mathbb{R}^3$ is the Möbius strip (`https://en.wikipedia.org/wiki/M%C3%B6bius_strip`). □

12.4 The Classical Integral Theorems

The classical integral theorems generalize the fundamental theorem of calculus, see Theorem 2.8, which states that

$$\int_a^b f'(x) dx = f(b) - f(a).$$

It thus relates the derivative of f in the interval $[a, b]$ with the values of f at the boundary points a and b of the interval.

We now generalize this to submanifolds M with boundary ∂M. The general form of the Theorem of Stokes then states in an abstract notation that

$$\int_M d\omega = \int_{\partial M} \omega.$$

Here ω is a so-called differential form and $d\omega$ its derivative. Instead of explaining what this object is and giving a rigorous definition of the integrals in this identity, in the following we will state and explain four special cases of the theorem, where Theorems 12.17–12.19 are known as the "classical integral theorems". Throughout this section we assume that the integrals used in the theorems exist, without stating suitable sufficient conditions (e.g., continuity, C^1) of the integrands that are needed for this.

The first theorem concerns the line integral from Definition 12.12, where $d = 1$ and $f : \mathbb{R}^n \to \mathbb{R}$ and $n \geq 2$. Here the relation to the fundamental theorem of calculus is easy to see and the proof is relatively straightforward using this fundamental theorem.

Theorem 12.16 For the line integral from Definition 12.12 it holds that

$$\int_\gamma \nabla f(x) \cdot d\vec{s}(x) = f(\gamma(b)) - f(\gamma(a)).$$

The second theorem deals with the case that $d = 2$ and $n = 2$.

Theorem 12.17 (Green's Theorem) Let $F \in C^1(U, \mathbb{R}^2)$ for an open set $U \subset \mathbb{R}^2$ and $D \subset \mathbb{R}^2$ be an open and bounded set with $\overline{D} \subset U$, whose boundary ∂D is a one-dimensional C^1 submanifold of $\mathbb{R}^2$. Then

$$\int_D \left(\frac{\partial F_2}{\partial x_1}(x) - \frac{\partial F_1}{\partial x_2}(x) \right) dx = \int_{\partial D} F(x) \cdot d\vec{s}(x).$$

Here the orientation of ∂D is chosen such that D lies left of ∂D in the direction of traversing the curves that form the atlas for ∂D.

Our third theorem is the classical Theorem of Stokes, which generalizes Green's Theorem to $d = 2$ and $n = 3$.

Theorem 12.18 (Classical Stokes Theorem) Let $F : M \to \mathbb{R}^3$, where M is a two-dimensional submanifold, whose boundary ∂M is a one-dimensional submanifold, both of $\mathbb{R}^3$. Then

$$\int_M \operatorname{curl}(F)(x) \cdot d\vec{S}(x) = \int_{\partial M} F(x) \cdot d\vec{s}(x).$$

Here

$$\operatorname{curl}(F) = \nabla \times F = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \frac{\partial}{\partial x_3} \end{pmatrix} \times \begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix} = \begin{pmatrix} \frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3} \\ \frac{\partial F_1}{\partial x_3} - \frac{\partial F_3}{\partial x_1} \\ \frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2} \end{pmatrix}.$$

Here the orientation is chosen as in Theorem 12.17.

We end this section on classical integration theorems with the classical Theorem of Gauss, which considers $d = 3$ and $n = 3$.

Theorem 12.19 (Classical Gauss Theorem) Let $F \in C^1(U, \mathbb{R}^3)$ for an open set $U \subset \mathbb{R}^3$ and $D \subset \mathbb{R}^3$ be an open and bounded set with $\overline{D} \subset U$, whose boundary ∂D is a 2-dimensional C^1 submanifold of $\mathbb{R}^2$. Then

$$\int_D \operatorname{div}(F)(x)dx = \int_{\partial D} F(x) \cdot d\vec{S}(x).$$

Here

$$\operatorname{div}(F) = \frac{\partial F_1}{\partial x_1} + \frac{\partial F_2}{\partial x_2} + \frac{\partial F_3}{\partial x_3}$$

and the orientation of ∂D is defined by the outer normal field w.r.t. D.

Question 12.7 *In light of Remark 12.15(i), can you give a physical meaning of the integral on the right-hand side of the theorem?*

12.5 The General Theorem of Gauss and Consequences

It seems plausible that the classical Theorem of Gauss 12.19 can be generalized to dimensions $d = n \neq 3$. This is indeed possible and is known as the General Theorem of Gauss.

Theorem 12.20 (General Gauss Theorem) Let $n \geq 2$, $F \in C^1(U, \mathbb{R}^n)$ for an open set $U \subset \mathbb{R}^n$ and $D \subset \mathbb{R}^n$ an open and bounded set with $\overline{D} \subset U$, whose boundary ∂D is an $(n-1)$-dimensional C^1 submanifold of $\mathbb{R}^n$. Then

$$\int_D \operatorname{div}(F)(x)dx = \int_{\partial D} \langle F(x), \nu(x) \rangle dS(x).$$

Here

$$\operatorname{div}(F) = \sum_{i=1}^{n} \frac{\partial F_n}{\partial x_n}$$

and ν is the outer normal field w.r.t. D.

Question 12.8 *Verify that Green's Theorem is the special case of the Theorem of Gauss for $n = 2$.*

Hint: *Use the formula from Remark 12.13(v) and convince yourself that* $\begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix} = \begin{pmatrix} -\nu_2 \\ \nu_1 \end{pmatrix}$.

Remark 12.21 In this remark we discuss some consequences from Gauss' General Theorem. Throughout this remark, we let D, U, and ν as in Theorem 12.20.

(i) For $f \in C^1(U, \mathbb{R})$ we obtain

$$\int_D \frac{\partial f}{\partial x_j}(x)dx = \int_{\partial D} f(x)\nu_j(x)dS(x).$$

To see this, apply Theorem 12.20 to $F(x) = f(x)e_j$, with e_j being the j-th canonical basis vector.

(ii) For $f, g \in C^1(U, \mathbb{R})$ we obtain

$$\int_D \frac{\partial f}{\partial x_j}(x)g(x)dx = \int_{\partial D} f(x)g(x)\nu_j(x)dS(x) - \int_D f(x)\frac{\partial g}{\partial x_j}(x)dx.$$

This follows by applying (i) to fg and using the product rule.

(iii) For $f \in C^1(U, \mathbb{R})$ and the Laplace operator

$$\Delta f = \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2} = \operatorname{div}(\nabla f)$$

it holds that

$$\int_D \Delta f(x)dx = \int_{\partial D} \langle \nabla f(x), \nu(x)\rangle dS(x).$$

This formula is obtained by applying Theorem 12.20 to $F = \nabla f$.

(iv) For $f, g \in C^1(U, \mathbb{R})$ the identity

$$\int_D f(x)\Delta g(x)dx = \int_{\partial D} \langle f(x)\nabla g(x), \nu(x)\rangle \, dS(x) - \int_D \langle \nabla f(x), \nabla g(x)\rangle dx$$

holds. This is seen by applying Theorem 12.20 to $F = f\nabla g$ and using that $\operatorname{div}(f\nabla g) = \langle \nabla f, \nabla g\rangle + f\Delta g$.

(v) Adding (iv) for $f\Delta g$ and for $-g\Delta f$ yields

$$\int_D (f(x)\Delta g(x) - g(x)\Delta f(x))dx = \int_{\partial D} \langle f(x)\nabla g(x) - g(x)\nabla f(x), \nu(x)\rangle \, dS(x).$$

(vi) Suppose we want to find a C^2 map $\Phi : \overline{D} \to \mathbb{R}$ with $\Phi|_{\partial D} = g$ for a given function g, such that

$$\int_D \|\nabla \Phi(x)\|_2^2 dx$$

becomes minimal. This means that if we slightly change Φ to $\widetilde{\Phi}_\varepsilon = \Phi + \varepsilon h$, where $\varepsilon \in \mathbb{R}$ and $h : \overline{D} \to \mathbb{R}$ is arbitrary with $h|_{\partial D} = 0$, then

$$\int_D \|\nabla \widetilde{\Phi}_\varepsilon(x)\|_2^2 dx \geq \int_D \|\nabla \Phi(x)\|_2^2 dx.$$

This implies that the map $f : \mathbb{R} \to \mathbb{R}$, $f(\varepsilon) = \int_D \|\nabla \widetilde{\Phi}_\varepsilon(x)\|_2^2 dx$ has a minimum in $\varepsilon = 0$, and thus

$$\begin{aligned}
0 = f'(x) &= \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \int_D \|\nabla \widetilde{\Phi}_\varepsilon(x)\|_2^2 dx \\
&= \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \int_D \langle \nabla\Phi(x) + \varepsilon\nabla h(x), \nabla\Phi(x) + \varepsilon\nabla h(x)\rangle \, dx \\
&= \int_D \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \langle \nabla\Phi(x) + \varepsilon\nabla h(x), \nabla\Phi(x) + \varepsilon\nabla h(x)\rangle \, dx \\
&= 2\int_D \langle \nabla\Phi(x), \nabla h(x)\rangle dx \\
&= 2\left[\int_{\partial D} \langle h(x)\nabla\Phi(x), \nu(x)\rangle \, dS(x) - \int_D \Delta\Phi(x)h(x)dx\right],
\end{aligned}$$

where we used (iv) in the last step. Now since $h(x) = 0$ for all $x \in \partial D$, the left integral in the last line equals 0, and it follows that

$$\int_D \Delta\Phi(x)h(x)dx = 0.$$

Now since h is arbitrary, this means that $\Delta\Phi(x) = 0$ must hold for all $x \in D$. This is because if $\Delta\Phi(x) \neq 0$ for some x, then because of continuity of $\Delta\Phi$ there is a neighborhood $B_\delta(x)$ such that $\Delta\Phi(y) \neq 0$ for all $y \in B_\delta(x)$ and if we choose h positive on this ball and 0 outside, then $\int_D \Delta\Phi(x)h(x)dx \neq 0$ would follow. We can thus conclude that

$$\Delta\Phi(x) = 0$$

holds for all $x \in D$. This equation is called the *Euler-Lagrange Equation* and can be used to compute the minimizing Φ. $\square$

Bibliography

[1] H. Amman and J. Escher, *Analysis I*, Birkhäuser Verlag, 3. Auflage, 2006.

[2] C. Blatter, *Analysis I*, Springer Verlag, 2. Auflage, 1977.

[3] C. Blatter, *Analysis II*, Springer Verlag, 3. Auflage, 1992.

[4] O. Forster, *Analysis 3*, Vieweg Verlag, Braunschweig, 1. Auflage, 1981.

[5] O. Forster, *Analysis 2*, Vieweg Verlag, Braunschweig, 1. Auflage, 1984.

[6] O. Forster, *Analysis 1*, Vieweg + Teubner Verlag, 10. Auflage, 2011.

[7] L. Grüne and M. Stoll, *Foundations of Higher Mathematics.* Lecture Notes, University of Bayreuth, 2026. Available from `https://num.math.uni-bayreuth.de/de/team/lars-gruene/skripten/`.

[8] H. Heuser, *Lehrbuch der Analysis, Teil 1*, Teubner, Stuttgart, 3. Auflage, 1984.

[9] H. Heuser, *Lehrbuch der Analysis, Teil 2*, Teubner, Stuttgart, 6. Auflage, 1991.

[10] K. Königsberger, *Analysis 1*, Springer Verlag, 6. Auflage, 2003.

[11] K. Königsberger, *Analysis 2*, Springer Verlag, 3. Auflage, 2003.

[12] W. Rudin, *Analysis*, Oldenbourg Verlag, 4. Auflage, 2009.